Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Journal of Functional Analysis 248 (2007) 1–26

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

Supremum concentration inequality and modulus of continuity for sub-nth chaos processes

Frederi G. Viens [a,b,*,1], Andrew B. Vizcarra [b,1]

[a] *Department of Statistics, Purdue University, 150 N. University St., West Lafayette, IN 47907-2067, USA*
[b] *Department of Mathematics, Purdue University, 150 N. University St., West Lafayette, IN 47907-2067, USA*

Received 22 February 2006; accepted 25 March 2007

Available online 7 May 2007

Communicated by Paul Malliavin

Abstract

This article provides a detailed analysis of the behavior of suprema and moduli of continuity for a large class of random fields which generalize Gaussian processes, sub-Gaussian processes, and random fields that are in the nth chaos of a Wiener process. An upper bound of Dudley type on the tail of the random field's supremum is derived using a generic chaining argument; it implies similar results for the expected supremum, and for the field's modulus of continuity. We also utilize a sharp and convenient condition using iterated Malliavin derivatives, to arrive at similar conclusions for suprema, via a different proof, which does not require full knowledge of the covariance structure.

Keywords: Stochastic analysis; Malliavin derivative; Wiener chaos; Sub-Gaussian process; Concentration; Suprema of processes; Dudley–Fernique theorem; Borell–Sudakov inequality

1. Introduction

The regularity properties of random processes have long been studied, going as far back as Kolmogorov's celebrated chaining argument and criterion (see [12, Theorem I.2.1]), and in the 70s and 80s the sharp work of Fernique, Talagrand, and others (see [6,13]) for the Gaussian case. These latter works drew upon ideas of R. Dudley, who in 1967 (see [4]) gave a sufficient condition

* Corresponding author.
E-mail addresses: viens@purdue.edu (F.G. Viens), avizcarr@purdue.edu (A.B. Vizcarra).
1 Supported by NSF grant 0606615.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2007.03.019

for the boundedness of Gaussian processes based on the so called ε-entropy integral. All these results, of upper bound type, are largely also valid in the case of sub-Gaussian processes (see [8]). The question remains of how many such results are true for other classes of processes. While the majorizing measure conditions of Talagrand show that supremum estimates can be achieved for processes with tails decaying no slower than exponentially (in the nomenclature of [8], these are processes with increments in the Orlicz space relative to convex Young functions of the type $\exp(z^q) - 1$, with $q \geqslant 1$), the authors of this note provided in [17] an extension of Dudley's entropy upper bound to all $q = 2/n$ with n an integer.

They presented their work by defining a new class of processes with the so-called *sub-nth chaos* property. These sub-nth chaos processes are an extension of sub-Gaussian processes. Such a process (random field) X on an arbitrary index set I is essentially required to have increments $X(t) - X(s)$ with tails that decay no slower than $\exp(-c|z|^{2/n}/\delta(s,t))$ where δ is some pseudo-metric on I. This definition was motivated by results including: (i) the fundamental observation [17, Lemma 3.3], that if the Malliavin derivative of a random variable X is almost surely bounded, then X must be sub-Gaussian; and (ii) the discovery in that same article of conditions on the nth Malliavin derivatives extending this observation to the sub-nth chaos case.

Also in [17], using a new concise Malliavin-derivative-based proof, a Borell–Sudakov[2]-type concentration inequality for such processes was proved, which shows that the supremum of a sub-nth chaos process is again a sub-nth chaos random variable with a well-controlled scale. While the proof in [17] completely generalized the standard Borell–Sudakov inequality to the sub-Gaussian and sub-2nd-chaos cases, it ran into inefficiencies in the case of higher order chaos.

In this article we correct these inefficiencies by combining a new use of iterated Malliavin derivatives, via a fractional exponential Poincaré lemma, with the innovative relation between Malliavin derivatives and suprema of processes discovered in [17]. This technique provides a new usage of the versatile Malliavin calculus, whose many applications are discussed in [9]; we summarize the Malliavin derivative's properties that we use herein, making our treatment essentially self-contained.

This article also contains an alternate way of establishing concentration inequalities for suprema of sub-nth chaos processes. Inspired by Ledoux and Talagrand's generic chaining arguments (see [14]), we use chaining to prove a Dudley entropy upper bound which holds in its usual expectation form, as well as in a tail form, for suprema of sub-nth chaos processes. These bounds allow us to prove, without the use of Malliavin derivatives, that concentration for these suprema is of sub-nth-chaos type, with respect to a scale which we estimate precisely, generalizing, up to a universal constant, the classical concentration of Borell–Sudakov-type for Gaussian processes.

The Malliavin-derivative-based and chaining-argument-based proofs of sub-nth-chaos supremum concentration use hypotheses that are evidently morally close, but it is not yet possible to say how close. Our fractional exponential Poincaré inequality appears to indicate that Malliavin-derivative conditions are slightly stronger. They do allow a simpler, much more accessible proof. We also believe that in many Brownian-based applications, checking that an nth Malliavin derivative is almost surely bounded may be significantly easier than checking the moment conditions for the sub-nth chaos property, particularly for solutions of stochastic equations; this is due to the need to use iterations of Itô's formula for the latter, which is algebraically usually more difficult

[2] Borell–Sudakov concentration inequalities are also known as Borell–Sudakov–Tsirel'son inequalities. Because of the appearance in the work of R. Dudley of these results' basic elements for scalar random variables, the inequalities should perhaps be called Dudley–Borell–Sudakov–Tsirel'son. For conciseness, we will continue to use the appellation Borell–Sudakov.

than calculating iterated Malliavin derivatives. In specific problems, it should be easy to identify which condition is most convenient.

As an application of our estimates, we choose a very basic and general one, which does not, in principle, rely on Malliavin derivative calculations: we derive almost-sure moduli of continuity for sub-nth chaos processes, with results that generalize those stated for general one-parameter Gaussian processes such as in [15]. The fractional exponential Poincaré inequality then allows one to use Malliavin derivative estimates to check that the sub-nth chaos conditions are satisfied; in this sense, Malliavin derivatives can be crucial in establishing path regularity.

Throughout this paper, $d\bar{s}$ denotes the Lebesgue measure on $(\mathbf{R}_+)^k$ for any integer k. Summarizing, our main results can be stated (partially) as follows.

- Assume there exists a positive integer n and a non-random pseudo-metric δ on an index set I such that, for a separable random field X on I, either of the following two conditions hold for all pairs $(x, y) \in I^2$:

  $$\mathbf{E}\left[\exp\left(\left(\frac{X}{\delta(x,y)}\right)^{2/n}\right)\right] \leqslant 2,$$

  or, almost surely,

  $$\left|D_{\cdot}^{(n)}\big(X(x)-X(y)\big)\right|_{L^2(dr)}^2 := \int_{(\mathbf{R}_+)^n} \left|D_{\bar{s}}^{(n)}\big(X(x)-X(y)\big)\right|^2 d\bar{s} \leqslant \delta^2(x,y),$$

  where $D^{(n)}$ is the nth iterated Malliavin derivative with respect to a Brownian motion defined on the same probability space as X. Then,

  $$\mu := \mathbf{E}\Big[\sup_{t\in I} X_t\Big] \leqslant C_n \int_0^\infty \big(\log N_\delta(\varepsilon)\big)^{n/2}\, d\varepsilon$$

  and

  $$\mathbf{P}\Big[\Big|\sup_{t\in I} X_t - \mu\Big| > u\Big] \leqslant 2\exp\left(-\frac{1}{2}\left(\frac{u}{C_n D}\right)^{2/n}\right),$$

  where C_n is a constant depending only on n, $N_\delta(\varepsilon)$ is the smallest number of δ-balls of radius ε needed to cover I, and D is the diameter of I under δ.
- Moreover, if I is an interval (or a smooth one-dimensional manifold) and $\delta(x,y) \leqslant d(|x-y|)$ for some increasing function d, then the function

  $$f_d(h) := d(h)\big(\log(1/h)\big)^{n/2}$$

  is, up to a constant depending on n only, almost surely a uniform modulus of continuity for X on I.

The structure of this article is the following. Section 2 gives the definitions of the sub-nth chaos property, and introduces the context of supremum concentration inequalities. Section 3 gives these inequalities using the generic chaining method. Section 4 uses the iterated Malliavin derivatives method. Section 5 finishes the article with the example of almost-sure moduli of continuity.

2. Sub-nth chaos processes

2.1. Definitions

We first recall the definition of sub-nth chaos random variables and processes as an extension of sub-Gaussian processes to heavier-tailed objects.

Definition 2.1. Let n be a positive integer. A centered random variable X is said to have the sub-nth-Gaussian chaos property (or is a sub-nth chaos r.v., or is a sub-Gaussian chaos r.v. of order n, etc.) relative to the scale M if

$$\mathbf{E}\left[\exp\left(\left(\frac{X}{M}\right)^{2/n}\right)\right] \leqslant 2. \tag{1}$$

When $n=1$, such an X is sub-Gaussian relative to the scale $\sqrt{5}M$, see [17].

Remark 2.2. Note that the following statement implies the sub-nth-Gaussian chaos property (1) and is also implied by it, with different universal multiplicative constant c in each implication: for all $u>0$

$$\mathbf{P}\big[|X|>u\big] \leqslant 2\exp\left(-\frac{u^{2/n}}{cM^{2/n}}\right). \tag{2}$$

Specifically, (1) implies (2) with $c=1$, and (2) implies (1) with $c=1/3$.

Definition 2.3. Let δ be a pseudo-metric on a set I. A centered random field X on I is said to be a sub-nth-Gaussian chaos field with respect to δ if for any $s,t\in I$, the random variable $X(t)-X(s)$ has the sub-nth-Gaussian chaos property relative to the scale $\delta(s,t)$.

Definition 2.4. Let δ and X be as in the previous definition. We use the notation $N_\delta(\varepsilon)$, and say that $N_\delta(\varepsilon)$ is a metric entropy for X, if $N_\delta(\varepsilon)$ is the smallest number of balls of radius ε in the pseudo-metric δ needed to cover I.

2.2. Overview of supremum concentration inequalities

In the next two sections, we prove a tail estimate version of the Dudley upper bound for sub-nth chaos processes, use it to derive estimates on the location and concentration of a sub-nth chaos process's supremum, and obtain similar concentration results via Malliavin derivative conditions.

For reference and comparison, we recall the statements of standard Gaussian estimates of supremum location and concentration. Let Z be a separable centered Gaussian field on an index set I that is almost-surely bounded, with canonical metric δ defined by $\delta^2(x,y)=\mathbf{E}[(Z(x)-Z(y))^2]$. Let $Y=\sup_I Z$. The so-called Dudley upper bound is $\mathbf{E}[Y]\leqslant K\int_0^\infty\sqrt{\log N_\delta(\varepsilon)}\,d\varepsilon$ where K is a universal constant (see [4]) and $N_\delta(\varepsilon)$ is defined as in the previous section. A corresponding lower bound, due to Fernique (see [5]) with a smaller universal constant, is known to hold for homogeneous (shift invariant in law) Gaussian processes on groups. The often-called Borell–Sudakov (or simply Borell) inequality, which may be improperly named as it first appears at the scalar process level in Dudley's work, is $\mathbf{P}[|Y - \mathbf{E}Y| > u] \leqslant 2\exp(-u^2/2\sigma^2)$ where $\sigma^2 := \max_{x\in I} Var[Z(x)]$. This result says precisely that the supremum of a Gaussian field is sub-Gaussian with respect to the scale σ.

These Gaussian results (except for Fernique's lower bound) were generalized in [17] to sub-Gaussian fields, and, under some additional assumption on Malliavin derivatives, to sub-2nd-chaos fields. In the next section, by using a general chaining argument, we show that such generalizations do not need stronger assumptions, and work in all chaoses. We also prove a Malliavin-derivative-based concentration result, using a non-chaining argument, in Section 4. Both concentration results show, analogously to the Gaussian case, that the random field X's supremum is a sub-nth chaos random variable, relative to scales which consistently generalize the σ in the Gaussian case above. This means that concentration results for suprema of sub-nth chaos fields can be obtained using two separate sets of hypotheses: one using standard sub-nth chaos estimates on X as defined in the previous section, and one using Malliavin-derivative boundedness conditions.

It may be said that the concentration results via Malliavin derivatives are more powerful, because they do not require that the process have the sub-nth chaos property: they only require that a process's supremum have a finite expectation, and that each one-dimensional distribution of the process be a sub-nth chaos random variable. This very weak assumption on the process's covariance structure was not known to be sufficient for Borell–Sudakov concentration until it was noticed in [17]. This article proves that it is sufficient for sub-nth-chaoses as well.

At the same time, at the random variable level, because of Proposition 4.2 (consequence of a fractional exponential Poincaré inequality proved in Appendix A), the Malliavin derivative conditions seem stronger than mere sub-nth-chaos conditions, although this is not yet clear that the gap between the two is of any significance. It is clear, though, that depending on the situation, one or the other set of hypotheses may be most advantageous.

3. Generic chaining argument method

Our first result is a tail estimate similar to the Dudley inequality, which generalizes the latter to sub-nth-chaos processes.

Theorem 3.1. *For each fixed positive integer n, there exist universal constants C_n and C_n', depending only on n, such that if X defined on I is a separable sub-nth-Gaussian chaos field with respect to the pseudo-metric δ, then for each $t_0 \in I$ and $s \geqslant 0$,*

$$\mathbf{P}\Big[\sup_{t\in I} |X_t - X_{t_0}| > C_n M + s C_n' D\Big] \leqslant 2e^{-s^{2/n}/2},$$

where $M = \int_0^\infty (\log N_\delta(\varepsilon))^{n/2}\, d\varepsilon$ and $D = Diam_\delta(I)$ is the diameter of I under δ.

In addition, if D and M are finite, then X is almost-surely bounded.

The above constants can be taken to be equal to

$$C_n = 2^{2n+1}(q+1)/\big(1-q^{-1}\big) f_n(q),$$
$$C_n' = \frac{1}{2}(q+1)/\big(1-q^{-1}\big),$$

where q is any fixed value > 1 and $f_n(q) = \sum_{\ell=1}^{\infty} q^{-\ell+1}\ell^{n/2}$. One may optimize q to obtain a value of C'_n that does not depend on n: with $q = 1+\sqrt{2}$, $C'_n = \sqrt{2} + 3/2$. Depending on the respective values of M and D, it may be preferable to minimize C_n instead, in which case C'_n will depend on n, or to perform some other optimization which takes s into consideration.

One may remove the assumption that X is separable, by changing the definition of the probability in Theorem 3.1: generically define

$$\mathbf{P}\Big[\sup_{t\in I}|X_t| > C\Big] = \sup\Big\{\mathbf{P}\Big[\sup_{t\in F}|X_t| > C\Big]\colon\ F\subset I;\ F \text{ finite}\Big\}.$$

The same remark holds for other theorems below which require separability, replacing expectations of suprema by suprema of expectations of suprema over finite sets. We will not comment further on this point.

Proof of Theorem 3.1. Since X is separable, by the monotone convergence theorem, we may and do assume that I is finite. Without loss of generality, we can assume that $X_{t_0} = 0$; if this is not true, consider the random field $Y_t = X_t - X_{t_0}$.

Let $q > 1$ be fixed and let ℓ_0 be the largest integer ℓ in $\mathbb{Z}$ such that $N_\delta(q^{-\ell}) = 1$. For every $\ell \geqslant \ell_0$, we consider a family of cardinality $N_\ell := N_\delta(q^{-\ell})$ of balls of radius $q^{-\ell}$ covering I. One may therefore construct a partition $\mathcal{A}_\ell$ of I of cardinality N_ℓ on the basis of this covering with sets of diameter less than $2q^{-\ell}$. Denote by I_ℓ the collection of points which are centers of each ball A of $\mathcal{A}_\ell$. For each t in I, denote by $A_\ell(t)$ the element of $\mathcal{A}_\ell$ that contains t. For every t and every ℓ, let then $s_\ell(t)$ be the element of I_ℓ such that $t \in A_\ell(s_\ell(t))$. Note that $\delta(t, s_\ell(t)) \leqslant q^{-\ell}$ for every t and $\ell \geqslant \ell_0$. Also note that

$$\delta\big(s_\ell(t), s_{\ell-1}(t)\big) \leqslant q^{-\ell} + q^{-\ell+1} = (q+1)q^{-\ell}. \tag{3}$$

Hence, by the second inequality in [17, Lemma 4.6], the series $\sum_{\ell>\ell_0}(X_{s_\ell(t)} - X_{s_{\ell-1}(t)})$ converges in $L^1(\Omega)$, and also $s_\ell(t)$ converges to t in $L^1(\Omega)$ as $\ell \to \infty$. By the telescoping property of the above sum, we thus get that almost surely for every t,

$$X_t = X_{t_0} + \sum_{\ell>\ell_0}(X_{s_\ell(t)} - X_{s_{\ell-1}(t)}) = \sum_{\ell>\ell_0}(X_{s_\ell(t)} - X_{s_{\ell-1}(t)}), \tag{4}$$

where $s_{\ell_0}(t) := t_0$. This decomposition is the basis of the so-called chaining argument.

Let $\{c_\ell\}_{\ell>\ell_0}$ be a sequence of positive numbers which will be chosen later. Note that if $\forall t \in I$ and $\forall \ell > \ell_0$, $|X_{s_\ell(t)} - X_{s_{\ell-1}(t)}| \leqslant c_\ell$ then from (4), $\forall t \in I$, $|X_t| \leqslant \sum_{\ell>\ell_0} c_\ell$. Then

$$\begin{aligned}
\mathbf{P}\Big(\sup_{t\in I}|X_t| > \sum_{\ell>\ell_0} c_\ell\Big) &= \mathbf{P}\Big(\exists t\in I\colon\ |X_t| > \sum_{\ell>\ell_0} c_\ell\Big)\\
&\leqslant \mathbf{P}\big(\exists t\in I, \exists \ell>\ell_0\colon\ |X_{s_\ell(t)} - X_{s_{\ell-1}(t)}| > c_\ell\big)\\
&\leqslant \mathbf{P}\big(\exists \ell>\ell_0, \exists t\in I_\ell, \exists t'\in I_{\ell-1}\colon\ |X_t - X_{t'}| > c_\ell\big)\\
&\leqslant \sum_{\ell>\ell_0} 2N_\ell^2 \exp\bigg(-\bigg(\frac{c_\ell}{\delta(t,t')}\bigg)^{2/n}\bigg) \qquad (5)
\end{aligned}$$

$$\leqslant 2\sum_{\ell>\ell_0} N_\ell^2 \exp\biggl(-\biggl(\frac{c_\ell}{(q+1)q^{-\ell}}\biggr)^{2/n}\biggr) \tag{6}$$

$$= 2\sum_{\ell>\ell_0} \exp\bigl(2\log N_\ell - b_\ell^{2/n}\bigr), \tag{7}$$

where in line (5) we used (2), and in line (6) we used (3), and in line (7) $c_\ell := (q+1)q^{-\ell}b_\ell$.

Set

$$b_\ell = 8^{n/2}\bigl[(\log N_\ell)^{n/2} + (\ell-\ell_0)^{n/2}\bigr] + s. \tag{8}$$

Since for any $u, v, \alpha \geqslant 0$, $(u+v)^\alpha$ exceeds $(u^\alpha + v^\alpha)/2$, we have that

$$\begin{aligned} b_\ell^{2/n} &\geqslant 4\bigl[(\log N_\ell)^{n/2} + (\ell-\ell_0)^{n/2}\bigr]^{2/n} + s^{2/n}/2 \\ &\geqslant 2\bigl[\log N_\ell + (\ell-\ell_0)\bigr] + s^{2/n}/2. \end{aligned} \tag{9}$$

Thus,

$$\begin{aligned} 2\sum_{\ell>\ell_0} \exp\bigl(2\log N_\ell - b_\ell^{2/n}\bigr) &\leqslant 2\sum_{\ell>\ell_0} \exp\bigl(2\log N_\ell - 2\log N_\ell - 2(\ell-\ell_0) - s^{2/n}/2\bigr) \\ &\leqslant 2e^{-s^{2/n}/2}\sum_{\ell>\ell_0} \exp\bigl(-2(\ell-\ell_0)\bigr) \\ &\leqslant 2e^{-s^{2/n}/2}. \end{aligned}$$

To finish the proof, we note that $N_\ell \geqslant 2$ for $\ell > \ell_0$ and that the series $f_n(q) := \sum_{\ell>0} q^{-\ell+1}\ell^{n/2}$ converges, hence

$$\begin{aligned} \sum_{\ell>\ell_0} c_\ell &= (q+1)8^{n/2}\sum_{\ell>\ell_0} q^{-\ell}\bigl[(\log N_\ell)^{n/2} + (\ell-\ell_0)^{n/2}\bigr] + s(q+1)\sum_{\ell>\ell_0} q^{-\ell} \\ &= (q+1)8^{n/2}\biggl(\sum_{\ell>\ell_0} q^{-\ell}(\log N_\ell)^{n/2} + q^{-\ell_0-1}f_n(q)\biggr) + s(q+1)\sum_{\ell>\ell_0} q^{-\ell} \\ &= (q+1)8^{n/2}q^{-\ell_0-1}\biggl(\sum_{\ell>\ell_0} q^{-\ell+\ell_0+1}(\log N_\ell)^{n/2} + f_n(q)\biggr) + s\frac{q+1}{1-q^{-1}}q^{-\ell_0-1}. \end{aligned}$$

Now with $A := \sum_{\ell>\ell_0} q^{-\ell+\ell_0+1}(\log N_\ell)^{n/2} \geqslant (\frac{1}{2})^{n/2}$ and $B := f_n(q) \geqslant 1$, we use the relation $A + B \leqslant 2\cdot 2^{n/2}AB$, obtaining

$$\begin{aligned} \sum_{\ell>\ell_0} c_\ell &\leqslant 2(q+1)8^{n/2}2^{n/2}f_n(q)\sum_{\ell>\ell_0} q^{-\ell}(\log N_\ell)^{n/2} + s\frac{q+1}{1-q^{-1}}q^{-\ell_0-1} \\ &\leqslant 2\frac{q+1}{1-q^{-1}}4^n f_n(q)\sum_{\ell>\ell_0}\int_{q^{-\ell-1}}^{q^{-\ell}} \bigl(\log N_\delta(\varepsilon)\bigr)^{n/2}\,d\varepsilon + s\frac{q+1}{1-q^{-1}}q^{-\ell_0-1} \\ &\leqslant C_n M + 2C_n' s q^{-\ell_0-1}, \end{aligned}$$

where $C_n = 2^{2n+1}(q+1)/(1-q^{-1})f_n(q)$, and $C'_n = \frac{1}{2}(q+1)/(1-q^{-1})$, and we used the definition of M and the fact that $N_\delta(\varepsilon)$ is decreasing in ε.

We now notice that by definition of ℓ_0, there exists a point $t^* = s_{\ell_0}(t) \in I$ such that

$$q^{-\ell_0-1} \leqslant \max\{\delta(t, t^*): t \in I\} \leqslant q^{-\ell_0}.$$

Therefore, we can bound $q^{-\ell_0-1}$ above by $\frac{D}{2}$ where $D = Diam_\delta(I)$, the diameter of I under δ, and conclude that

$$\mathbf{P}\Big(\sup_{t\in I}|X_t - X_{t_0}| > C_n M + sC'_n D\Big) \leqslant \mathbf{P}\Big(\sup_{t\in I}|X_t - X_{t_0}| > \sum_{\ell>\ell_0} c_\ell\Big) \leqslant 2e^{-s^{2/n}/2},$$

which ends the proof of the theorem, except for the statement of its last sentence.

To prove that last statement, now assume that X is not almost surely bounded, so that in particular I is infinite. Thus we have

$$\mathbf{P}[\forall N \in \mathbf{N}, \exists t \in I: |X_t| > N] = p > 0.$$

This implies

$$\mathbf{P}\Big[\forall s > 0, \sup_t |X_t| > C_n M + sC'_n D\Big] \geqslant p$$

which contradicts the result of the previous paragraph unless D is infinite or M is infinite. □

The tail estimate of the previous theorem shows that the supremum of a sub-nth chaos process is a sub-nth chaos random variable, although the scale parameter in this property is not quite clear, since there is no explicit mention of the mean of $\sup X$. As a first step to make this scale clearer, we notice the following, which also allows a sharpening of the last statement of Theorem 3.1.

Corollary 3.2. *With n, X, I, δ, and M as in Theorem* 3.1*, there exist a universal constant C_n depending only on n, such that for each $t_0 \in I$ and $s \geqslant 1$,*

$$\mathbf{P}\Big(\sup_{t\in I}|X_t - X_{t_0}| > sC_n M\Big) \leqslant 2e^{-s^{2/n}/2}.$$

In particular, if M is finite, then X is almost-surely bounded.

The constant C_n is the same as in Theorem 3.1: it can be taken as $\min_{q>1} 2^{2n+1}(q+1)/(1-q^{-1})f_n(q)$.

Proof. We reclaim the calculations in the proof of Theorem 3.1. The definition of c_ℓ remains the same in form, but instead of (8), we use

$$b_\ell = 8^{n/2}s\big[(\log N_\ell)^{n/2} + (\ell - \ell_0)^{n/2}\big].$$

Making use of the inequality $2AB \geqslant A + B$ for $A, B \geqslant 1$, we still obtain (9), and thus still also

$$\mathbf{P}\Big(\sup_{t\in I}|X_t| > \sum_{\ell>\ell_0} c_\ell\Big) \leqslant 2e^{-s^{2/n}/2}.$$

The calculation method for $\sum_{\ell>\ell_0} c_\ell$ is again identical, with the factor s appearing multiplicatively instead of additively:

$$\sum_{\ell>\ell_0} c_\ell \leqslant 2s\frac{q+1}{1-q^{-1}}4^n f_n(q)M = sC_nM,$$

where $C_n = 2^{2n+1}(q+1)/(1-q^{-1})f_n(q)$. $\square$

An upper bound on the expected supremum, whose form is most reminiscent of what is usually called "Dudley's" inequality, was already proved in [17], using a distinct chaining argument from what we do above, but follows now trivially from the previous corollary.

Corollary 3.3. *Let n, X, I, C_n be as in Theorem* 3.1 *and assume X is centered. Then*

$$\mathbf{E}\sup_{t\in I} X_t \leqslant C_n'' \int_0^\infty \big(\log N_\delta(\varepsilon)\big)^{n/2}\,d\varepsilon,$$

where $C_n'' = C_n(1 + 2\int_1^\infty e^{-s^{2/n}/2}\,ds)$ depends only on n.

Proof. Using the same notation as in Theorem 3.1, and the fact that X is centered,

$$\begin{aligned}
\mathbf{E}\sup_{t\in I} X_t &= \mathbf{E}\sup_{t\in I}(X_t - X_{t_0}) \\
&\leqslant \mathbf{E}\sup_{t\in I}|X_t - X_{t_0}| \\
&= \int_0^\infty \mathbf{P}\Big(\sup_{t\in I}|X_t - X_{t_0}| > u\Big)\,du \\
&\leqslant C_nM + \int_{C_nM}^\infty \mathbf{P}\Big(\sup_{t\in I}|X_t - X_{t_0}| > u\Big)\,du \\
&= C_nM + C_nM\int_1^\infty \mathbf{P}\Big(\sup_{t\in I}|X_t - X_{t_0}| > sC_nM\Big)\,ds \\
&\leqslant C_nM + 2C_nM\int_1^\infty e^{-s^{2/n}/2}\,ds \\
&= C_n''M. \qquad \square
\end{aligned}$$

We are now in a position to state and prove a full concentration inequality of Dudley–Borell–Sudakov type, for $\sup X$. Throughout this article, we assume that the expected maximum in the statements of the Borell–Sudakov-type inequality (Theorems 3.4, 4.1, Corollary 3.5) is finite.

Theorem 3.4. *Let n, X, I, C_n, C_n', M, and D be as in Theorem* 3.1, *and C_n'' be as in Corollary* 3.3. *Let $\mu = \mathbf{E}\sup_{t\in I} X_t$. Then for all $u \geqslant 2(C_n + C_n'')M$,*

$$\mathbf{P}\Big[\Big|\sup_{t\in I} X_t - \mu\Big| > u\Big] \leqslant 2\exp\left(-\frac{1}{2}\left(\frac{u}{2C_n' D}\right)^{2/n}\right).$$

Note that the statement of the theorem can also be made to hold for all $u \in (\varepsilon, 2(C_n + C_n'')M]$ for any fixed $\varepsilon > 0$, as long as one is willing to change the constant C_n' above, allowing it to depend also on ε. It is preferable to use the form given above, however, in order to be able to use the sharper constant D in the tail estimate. Thus, up to a constant depending only on n, D appears as a scale with respect to which $\sup_{t\in I} X_t - \mu$ is a sub-nth chaos random variable.

Proof. By Theorem 3.1, for any $u \geqslant \mu$, defining s via $C_n M + sC_n' D = u - \mu$, we have

$$\begin{aligned}\mathbf{P}\Big[\Big|\sup_{t\in I} X_t - \mu\Big| > u\Big] &\leqslant \mathbf{P}\Big[\sup_{t\in I} |X_t| + \mu > u\Big]\\ &= \mathbf{P}\Big[\sup_{t\in I} |X_t| > C_n M + sC_n' D\Big]\\ &\leqslant 2e^{-s^{2/n}/2}.\end{aligned}$$

Now define $r = u/(KD)$ where K will be chosen below. If we impose $s \geqslant r$, since $rKD - \mu = C_n M + sC_n' D$, it follows that $rKD \geqslant C_n M + rC_n' D + \mu$. Thus choosing $K = 2C_n'$ we get

$$r \geqslant \frac{C_n M + \mu}{(K - C_n')D} = \frac{C_n M + \mu}{C_n' D}.$$

Hence, translating this into a condition on u, we have that if

$$u \geqslant \frac{C_n M + \mu}{C_n' D} KD = 2(C_n M + \mu)$$

so that indeed $s \geqslant r$. We immediately get

$$\mathbf{P}\Big[\Big|\sup_{t\in I} X_t - \mu\Big| > u\Big] \leqslant 2e^{-s^{2/n}/2} \leqslant 2e^{-r^{2/n}/2} = 2\exp\left(-\frac{1}{2}\left(\frac{u}{2C_n' D}\right)^{2/n}\right).$$

Since by Corollary 3.3, $\mu \leqslant C_n'' M$, the lower bound on u above holds as soon as $u \geqslant 2(C_n + C_n'')M$, which ends the proof of the theorem. □

If instead we use the result from Corollary 3.2, following a similar argument above with $sC_n M = u - \mu$, $r = u/(KM)$, $K = 2C_n$, we also get the following.

Corollary 3.5. *Let n, X, I, C_n, and M be as in Corollary* 3.2, *and C''_n be as in Corollary* 3.3. *Let $\mu = \mathbf{E}\sup_{t\in I} X_t$. Then for all $u \geqslant 2C''_n M$,*

$$\mathbf{P}\Big[\Big|\sup_{t\in I} X_t - \mu\Big| > u\Big] \leqslant 2\exp\biggl(-\frac{1}{2}\biggl(\frac{u}{2C_n M}\biggr)^{2/n}\biggr).$$

In this result, M seems to be a sub-nth chaos scale. The reader might wonder which of Theorem 3.4 and Corollary 3.5 is sharpest. The answer is trivial if it is possible to find a relation between D and M. Although this is not clear in a general sub-nth chaos case, we may expect that typically D should be smaller than M up to a constant, the reason being that D has to do with expectations of increments, with a supremum outside the expectation, whereas M is an upper bound on an expectation with a supremum inside. This is very clear in the centered Gaussian case, as the following calculation shows. In calculating D, by writing $X_t - X_s = (X_t - X_{t_0}) - (X_s - X_{t_0})$ we can assume there exists t_0 such that $X_{t_0} = 0$. Thence

$$\begin{aligned} D^2 &= \sup_{s,t\in I} \mathbf{E}\big[(X_t - X_s)^2\big] \leqslant \sup_{s,t\in I} \mathbf{E}\big[2X_t^2 + 2X_s^2\big] = 4\sup_{t\in I}\mathbf{E}\big[X_t^2\big] \\ &= 2\pi\Big(\sup_{t\in I}\mathbf{E}\big[|X_t|\big]\Big)^2 \leqslant 2\pi\Big(\mathbf{E}\Big[\sup_{t\in I}|X_t|\Big]\Big)^2 \\ &\leqslant 2\pi\Big(\mathbf{E}|X_{t_0}| + 2\mathbf{E}\Big[\sup_{t\in I} X_t\Big]\Big)^2 = 8\pi\mathbf{E}^2\Big[\sup_{t\in I} X_t\Big] \leqslant 8\pi\big(C''_1 M\big)^2, \end{aligned}$$

by Corollary 3.3, so that $D \leqslant \sqrt{8\pi}\, C''_1 M$.

4. Iterated Malliavin derivatives method

The use of boundedness of (iterated) Malliavin derivatives as conditions for the nth chaos property was first introduced in [17].

For purposes of comparison, we cite the basic result in the sub-Gaussian scale. Let X be an almost-surely bounded and centered random field on an index set I, and assume that for every $x \in I$, $X(x)$ is a member of the Malliavin–Sobolev space $\mathbf{D}^{1,2}$ (see definition below). Assume for each $x \in I$ there exists a non-random value $\sigma(x)$ such that

$$\big|D_{\cdot}X(x)\big|^2_{L^2(dr)} := \int_0^\infty \big|D_r X(x)\big|^2\, dr \leqslant \sigma^2(x).$$

Then, as in the classical Dudley–Borell–Sudakov concentration inequality, with $\sigma^2 = \sup_{x\in I}\sigma^2(x)$, we have

$$\mathbf{P}\Big[\Big|\sup_{x\in I} X(x) - \mathbf{E}\sup_{x\in I} X(x)\Big| > u\Big] \leqslant 2\exp\biggl(-\frac{u^2}{2\sigma^2}\biggr). \tag{10}$$

The idea here is that a random variable with a Malliavin derivative that is almost-surely bounded in $L^2(dr)$ by a non-random constant has the sub-Gaussian property with respect to the scale of that non-random constant, and at the process level, we get sub-Gaussian concentration of a process's supremum relative to the smallest possible scale, namely the max of the process's one-dimensional distributions' scales. In this section, we generalize the above result to the sub-nth chaos level, using n-thly iterated Malliavin derivatives. This idea was already used on [17], although the results therein were not optimally stated or proved, an issue we correct here. The following review of Malliavin derivatives' properties will be useful to the reader.

Let W be a Brownian motion in the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbf{P})$, where $\mathcal{F}$ is the sigma-field generated by the process W. For any centered Gaussian random variable $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, we have that $X = \int_0^\infty f(s)\,dW(s)$ for some non-random $f \in L^2(\mathbf{R}_+)$. The Malliavin derivative of X at time $s \in \mathbf{R}_+$ is defined a.e. as $D_s X := f(s)$. For any function Φ on $\mathbf{R}$ that is continuously differentiable a.e. and such that $Y := \Phi(X) \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, the Malliavin derivative of Y is defined via the chain rule as $D_s Y = \Phi'(X) D_s X = \Phi'(X) f(s)$, $\mathbf{P} \times dr$-a.e. This chain rule also applies to functions of a Gaussian vector in $\mathcal{F}$ in the obvious way. The Malliavin derivative operator is extended as a closed operator to the set $\mathbf{D}^{1,2}$ of random variables whose Malliavin derivatives are in $L^2(\Omega \times \mathbf{R}_+)$. See details of the extension in [11]. The nth Malliavin derivative operator $D^{(n)}$ is defined by iterating D n-fold, and thus depends on n time parameters, while still acting on a single r.v. Thus $\mathbf{D}^{n,2}$ is the set of all random variables X in $\mathcal{F}$ such that $D^{(n)} X \in L^2(\Omega \times (\mathbf{R}_+)^n)$. Lastly, a word on chaos expansions (see [11] or [17] for details). Every random variable in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ can be expanded as $X = \mathbf{E}[X] + \sum_{k=1}^\infty I_k(f_k)$ where I_k is the iterated Wiener–Itô integral with respect to W, and f_k is a symmetric function in $L^2((\mathbf{R}_+)^k)$. Elementary work with Hermite polynomials shows that $D_s I_k(f_k) = k I_{k-1}(f_k(\cdot, x))$. In particular, the Malliavin derivative of a non-random quantity is 0, the Malliavin derivative of a Gaussian r.v. is non-random, the Malliavin derivative of a 2nd-chaos random variable is Gaussian, etc., and the nth Malliavin derivative of $I_n(f_n)$ is $n! f_n$, which implies that any condition on the nth Malliavin derivative of X ignores its terms in the chaoses of orders $0, 1, \ldots, n-1$.

We have the following supremum concentration inequality.

Theorem 4.1. *Assume a separable almost-surely bounded random field X on an index set I is such that for all $x \in I$, $X(x) \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ and for some integer n, for all $x \in I$, $D^{(n)} X(x)$ is almost-surely bounded in $L^2(dr)$ by a non-random constant. More specifically, define*

$$\sigma^2(x) := \operatorname{ess\,sup}\left\{ \int_{(\mathbf{R}_+)^n} \left| D^{(n)}_{r_1,\ldots,r_n} X(x) \right|^2 dr_1 \ldots dr_n \right\} \tag{11}$$

and assume $\sigma := \sup_{x \in I} \sigma(x) < \infty$. Also assume that pairs of values of X are a.s. distinct. Then there is a constant C_n depending only on n such that for any $\varepsilon > 0$, for u large enough,

$$\mathbf{P}\left[\left|\sup_{x\in I} X(x) - \mathbf{E}\sup_{x\in I} X(x)\right| > u\right] \leqslant 2(1+\varepsilon)\exp\left(-\frac{1}{2(1+\varepsilon)}\left(\frac{u}{C_n \sigma}\right)^{2/n}\right). \tag{12}$$

In fact, $\mathbf{P}$-almost surely,

$$\int_{(\mathbf{R}_+)^n} \left| D^{(n)}_{r_1,\ldots,r_n} \sup_{x\in I} X(x) \right|^2 dr_1 \ldots dr_n \leqslant \sigma^2. \tag{13}$$

In order to prove this theorem, we need the following consequence of a fractional exponential Poincaré inequality, whose proof is relegated to Appendix A.

Proposition 4.2. *Let Y be a centered random variable in $L^2(\Omega)$ of the form $Y = \sum_{k=n}^{\infty} I_k(f_k)$, satisfying almost surely*

$$\int\limits_{(\mathbf{R}_+)^n} d\bar{s}\left|D_{\bar{s}}^{(n)} Y\right|^2 \leqslant M^2$$

for some non-random value M. Then there exists a constant C_n depending only on n such that Y is a sub-nth chaos r.v. relative to the scale $C_n M$, i.e.

$$\mathbf{E}\left[\exp\left(\left|\frac{Y}{C_n M}\right|^{2/n}\right)\right] \leqslant 2.$$

The constant C_n can be taken as $C_n = (\log K(n)/\log 2)^{n/2}$ where $K(n)$ is defined and estimated in (A.4), (A.5).

Proof of Theorem 4.1.

Step 1 (*Setup*). We can assume without loss of generality that our index set is finite: $I = I_N = \{1, 2, \ldots, N\}$. Indeed, by separability, $\sup_{x\in I} X(x)$ can be evaluated by replacing I by a countable subset. Recall that $\mu := \mathbf{E}\sup_{x\in I} X(x)$ is finite, and thus the random variable $\sup_{x\in I} X(x) - \mu$ is well defined, and is the almost sure limit of the same r.v. with I replaced by a sequence I_N of increasing sets of size N converging to I; the dominated convergence theorem yields the conclusion of the theorem since its estimates do not depend on N.

Step 2 (*Proof of (13)*). Denote $X_m = X(m)$ and define $S_m = \max\{X_1, X_2, \ldots, X_m\}$, so that $S_{m+1} = \max\{X_m, S_m\}$. In order to prove that $\max_I X \in \mathbf{D}^{n,2}$, an approximation technique can be used (see the proof of [17, Theorem 3.6]): one shows that $\mathbf{1}_{X_{m+1}>Sm}$ can be approximated in $\mathbf{D}^{1,2}$ by a smooth function of $X_{m+1} - S_m$ whose Malliavin derivative tends to 0 for almost every (ω, s) in $L^2(\Omega) \times \mathcal{H}$ because $X_{m+1} - S_m \neq 0$ a.s. In particular, $D.\mathbf{1}_{X_{m+1}>Sm} = 0$ in $L^2(\Omega) \times \mathcal{H}$, and for any $k \leqslant n$, the kth-order Malliavin derivative of $\mathbf{1}_{X_{m+1}>Sm}$ is 0 in $L^2(\Omega \times (R_+)^k)$ as well.

This justifies the following computation, where equalities hold in $L^2(\Omega \times (R_+)^n)$:

$$\begin{aligned}
D^{(n)}_{s_n,\ldots,s_2,s_1} S_{m+1} &= D^{(n-1)}_{s_n,\ldots,s_2}(D_{s_1} X_{m+1}\mathbf{1}_{X_{m+1}>Sm} + D_{s_1} S_m \mathbf{1}_{X_{m+1}<Sm}) \\
&= D^{(n-2)}_{s_n,\ldots,s_3}\big([D_{s_2} D_{s_1} X_{m+1}]\mathbf{1}_{X_{m+1}>Sm} + [D_{s_2} D_{s_1} S_m]\mathbf{1}_{X_{m+1}<Sm}\big) \\
&\;\;\vdots \\
&= \big[D^{(n)}_{s_n,\ldots,s_2,s_1} X_{m+1}\big]\mathbf{1}_{X_{m+1}>Sm} + \big[D^{(n)}_{s_n,\ldots,s_2,s_1} S_m\big]\mathbf{1}_{X_{m+1}<Sm}.
\end{aligned} \tag{14}$$

Now let $\sigma_m^{*2} = \max\{\sigma^2(1); \ldots; \sigma^2(m)\}$. Proceeding by induction, assume that $\|D^{(n)}_{\cdot} S_m\|_2^2 \leqslant \sigma_m^{*2}$ almost surely, which is satisfied for $m = 1$ by our hypothesis on $D^{(n)}_{\cdot} X(1)$ since it is assumed to be bounded in $L^2((\mathbf{R}_+)^n)$ by $\sigma(1) = \sigma_1^*$. Our hypothesis and equality (14) implies that almost surely

$$\begin{aligned}
&\int\limits_{(\mathbf{R}_+)^n} \left|D^{(n)}_{s_n,\dots,s_2,s_1} S_{m+1}\right|^2 ds_n \cdots ds_2\, ds_1 \\
&\quad= \mathbf{1}_{X_{m+1}>Sm} \int\limits_{(\mathbf{R}_+)^n} \left|D^{(n)}_{s_n,\dots,s_2,s_1} X_{m+1}\right|^2 ds_n \cdots ds_2 ds_1 \\
&\qquad+ \mathbf{1}_{X_{m+1}<Sm} \int\limits_{(\mathbf{R}_+)^n} \left|D^{(n)}_{s_n,\dots,s_2,s_1} S_m\right|^2 ds_n \cdots ds_2\, ds_1 \\
&\quad\leqslant \sigma^2(m+1)\mathbf{1}_{X_{m+1}>Sm} + \sigma^{*2}_m \mathbf{1}_{X_{m+1}<Sm} \\
&\quad\leqslant \sigma^{*2}_{m+1},
\end{aligned}$$

induction implies (13) when $m = N$.

Step 3 (*Translation into a concentration inequality*). We can write the chaos decomposition of $\sup_{x\in I} X(x)$ as $\sum_{k=0}^{\infty} I_k(f_k)$. Since the nth Malliavin derivative kills off the first n terms, we direct our attention to the random variable

$$Y = \sum_{k=n}^{\infty} I_k(f_k).$$

Statement (13) along with Proposition 4.2, with $M = \sigma$, and Chebyshev's inequality, implies the conclusion (12) of the theorem, with $\varepsilon = 0$ but with $\sup_{x\in I} X(x) - \mathbf{E}\sup_{x\in I} X(x) = \sum_{k=1}^{\infty} I_k(f_k)$ replaced by Y. The technique used in [17] to recuperate all the terms in the chaos expansion of $\sup_{x\in I} X(x)$, at the cost of adding an $\varepsilon > 0$ in the statement (12), can now be invoked, to conclude the proof of the theorem (see the proof of Corollary 4.14, and especially the result of Lemma 4.15 on the tails of purely kth chaos random variables, in [17], which is a consequence of results sometimes attributed to Ch. Borell, found for instance in [7]; see also [2]; all details are omitted). □

One may wish to compare the Borell–Sudakov inequalities from Theorem 3.4, and from Theorem 4.1, since they provide sub-nth chaos concentrations with respect to two distinct scales D and σ.

Before discussing the differences between these scales, let us note that the hypotheses of the two theorems are close, but not actually comparable. Theorem 4.1 requires less than Theorem 3.4 in terms of joint distribution, since Theorem 3.4 needs X to be a sub-nth chaos process, while Theorem 4.1 only needs each random variable $X(x)$ to have the sub-nth chaos property. On the other hand, Theorem 3.4 needs only the basic sub-nth chaos property as in our original definitions, while Theorem 4.1 requires a bit more, since condition (11) implies this property via Proposition 4.2, whose converse is not known.

Having said this, let us now show that D and σ are not necessarily comparable, and explain in what cases they are. In order to make any meaningful comparisons, it is necessary to believe that while the boundedness of the nth Malliavin derivative in Proposition 4.2 is perhaps not necessary for the sub-nth chaos property, it is still presumably extremely close. We will not argue in favor or against this belief here, but assuming one has it, one immediately concludes that if X is a sub-nth chaos field on I, the "nearly" best choice for its scale metric δ is

$$\delta(s,t) = \operatorname{ess\,sup}\left\{ \int\limits_{(\mathbf{R}_+)^n} \left| D^{(n)}_{r_1,\ldots,r_n}\big(X(t) - X(s)\big)\right|^2 dr_1 \ldots dr_n \right\}^{1/2}$$
$$= \operatorname{ess\,sup} \left| D^{(n)}\big(X(t) - X(s)\big)\right|_{L^2(\mathbf{R}^n_+)}.$$

We can first prove $D \leqslant 2\sigma$ as follows:

$$\begin{aligned}
D &= \sup_{s,t\in I} \delta(s,t) \\
&\leqslant \sup_{s,t\in I} \operatorname{ess\,sup}\left(\left| D^{(n)}\big(X(t)\big)\right|_{L^2(\mathbf{R}^n_+)} + \left| D^{(n)}\big(X(s)\big)\right|_{L^2(\mathbf{R}^n_+)}\right) \\
&= 2 \sup_{t\in I} \operatorname{ess\,sup} \left| D^{(n)}\big(X(t)\big)\right|_{L^2(\mathbf{R}^n_+)} \\
&= 2\sigma.
\end{aligned}$$

An opposite inequality, cannot hold in general, but we can easily prove $\sigma \leqslant D$ assuming for instance that for some t_0, $X(t_0) = 0$. Indeed, we can then write

$$\begin{aligned}
\sigma &= \sup_{t\in I} \operatorname{ess\,sup} \left| D^{(n)}\big(X(t) - X(t_0)\big)\right|_{L^2(\mathbf{R}^n_+)} \\
&\leqslant \sup_{s,t\in I} \operatorname{ess\,sup} \left| D^{(n)}\big(X(t) - X(s)\big)\right|_{L^2(\mathbf{R}^n_+)} \\
&= \sup_{s,t\in I} \delta(s,t) = D.
\end{aligned}$$

To finish this discussion, let us give an example where σ can be much larger than D. Consider the last example for X, and define $Y(t) = X(t) + Z$ where Z is an nth-chaos random variable independent of X, with an nth chaos norm $|D^{(n)}Z|_{L^2(\mathbf{R}^n_+)} = z$, a given non-random constant. Since X and Z are assumed to be independent, their Malliavin derivatives are supported on disjoint subsets of $\mathbf{R}^n_+$. Thus we have for Y that

$$\begin{aligned}
\sigma_Y^2 &= \sup_{t\in I} \operatorname{ess\,sup} \left| D^{(n)}\big(X(t)\big) + D^{(n)} Z\right|^2_{L^2(\mathbf{R}^n_+)} \\
&= \sup_{t\in I} \operatorname{ess\,sup} \left| D^{(n)}\big(X(t)\big)\right|^2_{L^2(\mathbf{R}^n_+)} + \left| D^{(n)} Z\right|^2_{L^2(\mathbf{R}^n_+)} \\
&= \sigma_X^2 + z^2.
\end{aligned}$$

On the other hand, we obviously have that X and Y share the same diameter D since they share the same scale metric δ due to having the same increments. Thus by choosing z arbitrarily large, D can be made arbitrarily small compared to σ_Y. Hence Theorem 4.1 can sometimes be much less sharp than Theorem 3.4. To avoid such a situation, one may redefine σ in Theorem 4.1 by considering only the oscillation function of X. We will not delve deeper into this issue except to say that no statement of the Borel–Sudakov inequality we have seen in the literature has noted that σ may not be sharp; even in the Gaussian case ($n = 1$), where $\sigma^2 = \max_{t\in I} Var[X(t)]$, this lack of sharpness may occur as described above.

5. Modulus of continuity

Definition 5.1. Let f be a continuous increasing function in $\mathbf{R}_+$ such that $\lim_{0+} f = 0$. Let $\{Y(t)\colon t \in I\}$ be a random field on an index set I endowed with a metric ρ. We say that f is an almost sure uniform modulus of continuity for Y on (I, ρ) if there exists an almost-surely positive random variable α_0 such that

$$\alpha < \alpha_0 \quad \Rightarrow \quad \sup_{s,t \in I\colon \rho(s,t) < \alpha} \bigl|Y(t) - Y(s)\bigr| \leqslant f(\alpha).$$

Corollary 3.3 has an immediate consequence for continuity (see [1]). Consider the random field $Y_{u,v} = X_v - X_u$. Let δ be the canonical metric for X on T defined as $\delta(u, v) = (\mathbf{E}(X(v) - X(u))^2)^{1/2}$ and let δ_Y be the canonical metric for Y on $T \times T$. Then

$$\begin{aligned} \delta_Y\bigl((u, v), (u', v')\bigr) &= \bigl[\mathbf{E}\bigl((X_v - X_u) - (X_{v'} - X_{u'})\bigr)^2\bigr]^{1/2} \\ &\leqslant 2 \max\bigl(\delta(u, v), \delta\bigl(u', v'\bigr)\bigr). \end{aligned}$$

This implies that $N_{\delta_Y}(\varepsilon) \leqslant N_\delta(\varepsilon/2)$. So, Corollary 3.3 implies

$$\begin{aligned} \mathbf{E}\Bigl[\sup_{(u,v)\in T\times T,\ \delta(u,v)\leqslant\eta} |X_v - X_u|\Bigr] &= \mathbf{E}\Bigl[\sup_{(u,v)\in T\times T,\ \delta(u,v)\leqslant\eta} |Y_{u,v}|\Bigr] \\ &\leqslant C''_n \int_0^{2\eta} \bigl(\log N_\delta(\varepsilon/2)\bigr)^{n/2}\, d\varepsilon \\ &= 2C''_n \int_0^{\eta} \bigl(\log N_\delta(\varepsilon)\bigr)^{n/2}\, d\varepsilon. \end{aligned}$$

Thus, we have proved the following result.

Proposition 5.2. *Let X be as in Corollary* 3.3. *Then*

$$\mathbf{E}\Bigl[\sup_{\delta(u,v)\leqslant\eta} |X_v - X_u|\Bigr] \leqslant K_n \int_0^{\eta} \bigl(\log N_\delta(\varepsilon)\bigr)^{n/2}\, d\varepsilon,$$

where $K_n = 2C''_n$ with C''_n the constant depending only on n defined in Corollary 3.3.

A well-known phenomenon in the theory of Gaussian regularity is that a modulus of continuity of X relative to δ is also given almost surely as the right-hand side of the previous inequality. The same effect is shown here for sub-nth chaos processes, based on a chaining argument similar to that used to prove Theorem 3.1. In our attempt to derive the next result as a mere consequence of Theorem 3.1, we found that the number of modifications needing to be made to the proof of the latter in order to get the next theorem to work, warranted a whole new proof. It is given below.

Theorem 5.3. *Let X be as in Theorem* 3.1. *There exists a random number $\eta_0 > 0$ almost surely, such that*

$$\sup_{\delta(u,v)\leqslant\eta} |X_v - X_u| \leqslant k_n \int_0^\eta \bigl(\log N_\delta(\varepsilon)\bigr)^{n/2}\, d\varepsilon + 4k_n g(\eta),$$

for all $\eta < \eta_0$, where k_n can be taken as $2^{n+1}(q+1)$ with $q > 1$, where

$$g(\eta) := \eta \log^{n/2}\bigl(\log(1/\eta)\bigr).$$

Before proving this theorem, we discuss its scope and a corollary in Euclidean space.

Remark 5.4. There seems to be a slight inefficiency in the above result, due to the presence of the term $g(\eta) = \eta \log^{n/2} \log(1/\eta)$: the theorem cannot differentiate between sub-nth chaos random fields which are a.s Lipschitz continuous relative to δ, and those who have $g(\eta)$ as an a.s. δ-modulus of continuity.

Therefore this theorem, as stated, does not provide a sharp result for trivial examples such as $X(t) = \sum_{i=1}^M f_i(t) G_i$ where $(G_i)_{i=1}^M$ is a finite-dimensional sub-nth chaos random vector, and the f_i's are smooth vector fields. We leave it to the reader to check that the proof below can be modified to allow for a sharp result in this case, but of course the result is trivial in that situation.

There may be other less trivial examples of nearly Lipschitz random fields for which the theorem is not sharp, but we suspect these are always pathological, i.e. are highly inhomogeneous. For the vast majority of random fields, including all those encountered in the literature, the entropy integral in the above theorem will dominate $g(\eta)$. The next result shows that in typical examples in Euclidean space, the extra term is not needed.

The corollary below on moduli of continuity can also be established using the so-called Garsia–Rodemich–Rumsey real variable lemma, however, we prefer the current presentation based fully on probabilistic chaining arguments, to avoid seeking techniques outside of probability theory.

Corollary 5.5. *Let X be a separable sub-nth chaos field relative to the pseudo-metric δ on $E \times E$, where E is a subset of the d-dimensional Euclidean space, with Euclidean norm denoted by $|\cdot|$. Assume there is an increasing univariate function on $\mathbf{R}_+$, also denoted by δ, such that the right-hand derivative of δ at* 0 *exists* (*possibly equal to* $+\infty$), *satisfying $\delta(s,t) \leqslant \delta(|s-t|)$ and the condition*

$$\lim_{r\to 0^+} \delta(r)\left(\log\frac{1}{r}\right)^{n/2} = 0. \tag{15}$$

Then, up to a non-random constant c, $f_\delta(r) := \delta(r)(\log r^{-1})^{n/2}$ is an almost-sure uniform modulus of continuity for X on any compact subset of E.

The constant c can be taken as any constant exceeding $4k_n d^{n/2}$ where k_n is as in Theorem 5.3, *and therefore does not depend on the distribution of X.*

Proof of Theorem 5.3. Following the same construction as in the proof of Theorem 3.1, we get that almost surely for every $u, v \in T$ and $\ell' > \ell_0$,

$$X_v - X_u = X_{s_{\ell'}(v)} - X_{s_{\ell'}(u)} + \sum_{\ell>\ell'} \big(X_{s_\ell(v)} - X_{s_{\ell-1}(v)}\big) - \sum_{\ell>\ell'} \big(X_{s_\ell(u)} - X_{s_{\ell-1}(u)}\big). \tag{16}$$

From the first inequality in [17, Lemma 4.6], we have that for $c > 0$,

$$\mathbf{P}\big(|X_v - X_u| > c\delta(u, v)\big) \leqslant 2\exp\big(-c^{2/n}\big).$$

Applying the above inequality, we have

$$\begin{aligned}&\mathbf{P}\big(\exists u, v \in T\colon |X_{s_\ell(v)} - X_{s_\ell(u)}| > \delta\big(s_\ell(u), s_\ell(v)\big)\big(\log\big(\ell^2 N_\ell^2\big)\big)^{n/2}\big)\\&\quad \leqslant 2N_\ell^2 \exp\big(-\log\big(\ell^2 N_\ell^2\big)\big) \leqslant 2\ell^{-2},\end{aligned}$$

where $N_\ell = N_\delta(q^{-\ell})$. Since this is a summable series, by the Borel–Cantelli lemma, there exists a random almost-surely finite integer $L_{0,1} > \ell_0$ such that with probability one,

$$\ell \geqslant L_{0,1} \quad \Rightarrow \quad |X_{s_\ell(v)} - X_{s_\ell(u)}| \leqslant \delta\big(s_\ell(u), s_\ell(v)\big)\big(\log\big(\ell^2 N_\ell^2\big)\big)^{n/2}$$

for all $u, v \in T$. Similarly, we also get that

$$\ell \geqslant L_{0,2} \quad \Rightarrow \quad |X_{s_\ell(u)} - X_{s_{\ell-1}(u)}| \leqslant \delta\big(s_\ell(u), s_{\ell-1}(u)\big)\big(\log\big(\ell^2 N_\ell N_{\ell-1}\big)\big)^{n/2}$$

for all $u \in T$, for a possibly larger random integer $L_{0,2}$, so that we denote $L_0 = \max(L_{0,1}, L_{0,2})$.

Putting these into (16) in which we replace ℓ' by some $L \geqslant L_0$, we get that

$$\begin{aligned}|X_v - X_u| &\leqslant \delta\big(s_L(u), s_L(v)\big)\big(\log\big(L^2 N_L^2\big)\big)^{n/2} + \sum_{\ell>L} \delta\big(s_\ell(u), s_{\ell-1}(u)\big)\big(\log\big(\ell^2 N_\ell N_{\ell-1}\big)\big)^{n/2}\\&\quad + \sum_{\ell>L} \delta\big(s_\ell(v), s_{\ell-1}(v)\big)\big(\log\big(\ell^2 N_\ell N_{\ell-1}\big)\big)^{n/2}.\end{aligned}$$

Now define $\eta_0 := q^{-L_0}$ and for any $\eta \leqslant \eta_0$, let L be the unique integer such that $q^{-L-1} < \eta \leqslant q^{-L}$. Hence for any pair (u, v) such that $\delta(u, v) \leqslant \eta$,

$$\begin{aligned}\delta\big(s_L(u), s_L(v)\big) &\leqslant \delta\big(s_L(u), u\big) + \delta(u, v) + \delta\big(v, s_L(v)\big)\\&\leqslant q^{-L} + \eta + q^{-L}\\&\leqslant 3q^{-L}.\end{aligned}$$

Using this, (3), and the fact that $N_{\ell-1} \leqslant N_\ell$ for any $\ell > \ell_0$, we get that for the above pair (u, v),

$$\begin{aligned}|X_v - X_u| &\leqslant 3q^{-L}\bigl(\log\bigl(L^2 N_L^2\bigr)\bigr)^{n/2} + 2\sum_{\ell>L}(q+1)q^{-\ell}\bigl(\log\bigl(\ell^2 N_\ell N_{\ell-1}\bigr)\bigr)^{n/2}\\ &\leqslant 2(q+1)\sum_{\ell\geqslant L} q^{-\ell}\bigl(\log\bigl(\ell^2 N_\ell^2\bigr)\bigr)^{n/2}\\ &\leqslant k_n\bigl[F(L)+I(L)\bigr],\end{aligned} \tag{17}$$

where

$$k_n := 2^{n+1}(q+1),$$

and

$$F(L) := \sum_{\ell\geqslant L} q^{-\ell}\log^{n/2}\ell,$$

and also

$$I(L) := \sum_{\ell\geqslant L} q^{-\ell}(\log N_\ell)^{n/2}.$$

Note that for $\varepsilon \in (q^{-\ell-1}, q^{-\ell}]$, $N_\ell \leqslant N_\delta(\varepsilon)$ and $\ell \leqslant \log(1/\varepsilon)/\log q$. Therefore, approximating the series $I(L)$ and $F(L)$ by Riemann integrals, we get

$$I(L) \leqslant \int_0^\eta \log^{n/2}\bigl(N_\delta(\varepsilon)\bigr)\,d\varepsilon, \tag{18}$$

and

$$F(L) \leqslant \int_0^\eta \log^{n/2}\biggl(\frac{\log(1/\varepsilon)}{\log q}\biggr)\,d\varepsilon \leqslant 2\int_0^\eta \log^{n/2}\bigl(\log(1/\varepsilon)\bigr)\,d\varepsilon, \tag{19}$$

where in the last inequality, we used the fact that $\eta_0 = q^{-L_0}$ so that $\log(\log(1/\varepsilon)) \geqslant \log(\log q)$. Putting (17), (19), and (18) together, we finally get that for any $\eta \leqslant \eta_0$, for any pair (u, v) such that $\delta(u, v) \leqslant \eta$,

$$\begin{aligned}|X_v - X_u| &\leqslant k_n\int_0^\eta \log^{n/2}\bigl(N_\delta(\varepsilon)\bigr)\,d\varepsilon + 2k_n\int_0^\eta \log^{n/2}\bigl(\log(1/\varepsilon)\bigr)\,d\varepsilon\\ &\leqslant k_n\int_0^\eta \log^{n/2}\bigl(N_\delta(\varepsilon)\bigr)\,d\varepsilon + 4k_n\eta\log^{n/2}\bigl(\log(1/\eta)\bigr),\end{aligned}$$

which finishes the proof of Theorem 5.3. □

Proof of Corollary 5.5. Since we only need to evaluate a modulus of continuity on compact subsets of E up to undetermined non-random constants, it is sufficient to assume that X is defined on a subset of the box $x_0 + [-M, M]^d$ where M can be as small as needed and the center x_0 of this box is arbitrary; one can then obviously tile any compact subset of E with such boxes defined by M fixed and a collection of centers x_0. Thus, without loss of generality, we assume $E = [-M, M]^d$ where we may choose $M > 0$ as small as desired. In view of the homogeneous upper bound $\delta(s,t) \leqslant \delta(|s-t|)$ in the corollary's assumptions, we may indeed use $x_0 = 0$ without loss of generality.

From Theorem 5.3, there exists an almost-surely positive random number η_0 and a universal constant k_n depending only on n such that

$$\sup_{\delta(u,v)\leqslant\eta} |X_v - X_u| \leqslant k_n \int_0^\eta \bigl(\log N_\delta(\varepsilon)\bigr)^{n/2}\, d\varepsilon + 4k_n g\bigl(\delta(h)\bigr),$$

for all $\eta < \eta_0$, where $N_\delta(\varepsilon)$ is the smallest number of δ-balls of radius ε needed to cover E. Denote by δ^{-1} the inverse function of the univariate δ. Since the set $\{|u-v| \leqslant \delta^{-1}(\eta)\}$ is smaller than the set $\{\delta(u,v) \leqslant \eta\}$, denoting $h = \delta^{-1}(\eta)$, the inequality above implies

$$\sup_{|u-v|\leqslant h} |X_v - X_u| \leqslant k_n \int_0^{\delta(h)} \bigl(\log N_\delta(\varepsilon)\bigr)^{n/2}\, d\varepsilon + 4k_n g\bigl(\delta(h)\bigr), \tag{20}$$

for all $h < \delta^{-1}(\eta_0)$.

The set $E = [-M, M]^d$ may be covered with balls of δ-radius ε by using instead Euclidean balls of radius $r := \delta^{-1}(\varepsilon)$ centered at the points of $(r\mathbf{Z})^d \cap E$. This is done with $(2M/r)^d$ Euclidean balls if $2M/r$ is an integer, which proves

$$N_\delta(\varepsilon) \leqslant \left(\frac{2M}{\delta^{-1}(\varepsilon)} + 1\right)^d.$$

Now choose $M = 1/4$. By choosing η small enough, we can make $h = \delta^{-1}(\eta)$ as small as we wish, since $\lim_0 \delta^{-1} = 0$; thus, also using the fact that δ^{-1} is increasing, choose η so that $\delta^{-1}(\varepsilon) \leqslant \delta^{-1}(\eta) \leqslant 1/2$. This guarantees that

$$N_\delta(\varepsilon) \leqslant \frac{1}{(\delta^{-1}(\varepsilon))^d}.$$

From (20), this proves that for some almost surely positive h_0, for $h < h_0$,

$$\sup_{|u-v|\leqslant h} |X_v - X_u| \leqslant k_n d^{n/2} \int_0^{\delta(h)} \left(\log \frac{1}{\delta^{-1}(\varepsilon)}\right)^{n/2} d\varepsilon + 4k_n g\bigl(\delta(h)\bigr). \tag{21}$$

We may now use a change of measure in the last integral above in the Riemann–Stieltjes sense, to transform the almost-sure uniform modulus of continuity for X we have just obtained in (21):

$$I_\delta(h) := \int_0^{\delta(h)} \left(\log \frac{1}{\delta^{-1}(\varepsilon)}\right)^{n/2} d\varepsilon = \int_0^h \left(\log \frac{1}{r}\right)^{n/2} d\delta(r).$$

By an integration by parts, we may now write

$$I_\delta(h) = \delta(h)\left(\log \frac{1}{h}\right)^{n/2} + n\int_0^h \frac{1}{2r}\left(\log \frac{1}{r}\right)^{n/2-1} \delta(r)\, dr.$$

The last term above is positive; moreover, by using the estimate $\delta(r) \leqslant \delta(h)$ therein, we get that this term is bounded above by $\delta(h)(\log h^{-1})^{n/2}$. Hence

$$I_\delta(h) \leqslant 2\delta(h)\left(\log \frac{1}{h}\right)^{n/2} = 2 f_\delta(h). \tag{22}$$

Combining (21) and (22), with $c' = 4k_n d^{n/2}$, we obtain

$$\sup_{|u-v|\leqslant h} |X_v - X_u| \leqslant c'\big(f_\delta(h) + g\big(\delta(h)\big)\big). \tag{23}$$

To complete the proof of the corollary, it is sufficient to show that $g(\delta(h)) = o(f_\delta(h))$. If we assume that $\delta(h)/h$ is bounded below by a positive constant, this follows from a straightforward calculation. This assumption is satisfied for all usual examples of regularity scales because there δ is concave, and therefore $\delta(h)/h$ increases to $\delta'(0)$ as $h \to 0$. Moreover it is not a restriction in any case. One can show that any random field for which $\delta(t, t+h)/h$ tends to 0 as $h \to 0$ uniformly for all t in an open set of the Euclidean space is actually constant on that set. If a random field satisfying the hypotheses of the corollary also had $\lim_{h\to 0} \delta(h)/h = \delta'(0) = 0$, then we would have, uniformly in t, $\lim_{h\to 0} \delta(t, t+h)/h = 0$, and the conclusion of the corollary would hold trivially for this constant random field [even though it would not be sharp]. Let now $c > c'$. By letting h be so small that $g(\delta(h)) < ((c-c')/c') f_\delta(h)$, (23) finishes the proof of the corollary. $\square$

The end of the above proof points to an inefficiency in the result, which also appears in all works in the literature which deal with almost-sure uniform moduli of continuity. Consider the example $X(t) = tW(t)$ where W is a standard Brownian motion. In this case, for all $s, t \in [0, 1]$, we may write that this Gaussian process has a canonical metric bounded as $\delta(s, t) \leqslant K|t - s|^{1/2}$ where K is a universal constant. Thus $f_\delta(h) = h^{1/2} \log^{1/2}(1/h)$ is an almost sure uniform modulus of continuity for X up to a universal constant, which is of course a sharp result, since lower bounds of the same order can also be obtained in this very simple case when looking at the entire interval [0, 1]. However, this hides the fact that at $t = 0$, X is almost-surely differentiable (indeed, its derivative is 0 almost surely!). Because of the extreme inhomogeneity of X near the origin, no statement on uniform moduli of continuity can accurately describe X's behavior there.

The result of Corollary 5.5 is nevertheless quite sharp, in the following sense. First note that the integral $I_\delta(h)$ in (21) is also bounded below by $f_\delta(h)$. Thus there is no hope to better the result of the corollary analytically. More specifically, one can show that if $\delta(u, v) = \delta(|u - v|)$, a *homogeneous* case, then $f_\delta(h)$ is in fact, an upper and lower bound for the entropy integral in (20), with constants 2 and 1, respectively; therefore in this case, $f_\delta(h)$ appears indeed as a sharp estimate of the modulus of continuity of X. Referring to the homogeneous Gaussian case for comparison, it was proved in [15] that f_δ is a uniform modulus of continuity if and only if the canonical metric of X is commensurate with δ. Our corollary above is a first step in proving the same result for sub-nth processes. To prove such a strong result for these processes, however, one would need a strong lower bound assumption on the tails of the processes' increments. Even in the sub-Gaussian case, such a problem is open.

We close this article with a note on the scope of condition (15). In the Gaussian case, it is well understood that this condition is sufficient for almost-sure boundedness and continuity, while in the corresponding homogeneous case with scalar parameter, when the condition is not satisfied, one can prove that the process is almost surely unbounded on any open interval; this is a consequence of the so-called Fernique lower bound (e.g. opposite inequality from Corollary 3.3 with $n = 1$). When $n > 1$, it is simple enough to conjecture that such sharpness should also hold; this would be easier to prove than the program outlined in the previous paragraph, but it is still not a straightforward endeavor, since even in the sub-Gaussian case, formulating a lower bound hypothesis sufficient to obtain an analogue Fernique's lower bound appears to be non-trivial. We may simply state at this stage that the following regularity scales all yield continuous sub-nth chaos processes; condition (15) is satisfied by the following scales (c is a generic constant which depends on n and the other regularity parameters H, β, γ given below):

1. $\delta(r) = r^H$, $H \in (0,1)$ (fractional Brownian scale): $f_\delta(r) \leqslant cr^H \log^{n/2}(1/r)$;
2. $\delta(r) = (\log \frac{1}{r})^{-\beta}$, $\beta > n/2$ (logarithmic Brownian scale): $f_\delta(r) \leqslant c \log^{-\beta+n/2}(1/r)$;
3. $\delta(r) = r \log^\gamma \frac{1}{r}$, $\gamma \geqslant 0$ (nearly Lipschitz case): $f_\delta(r) \leqslant cr \log^{\gamma+n/2}(1/r)$.

For more information on the stochastic analysis of Gaussian random fields with regularity scales 1 and 2, see [3,10], respectively.

Appendix A. The fractional exponential Poincaré inequality; proof of Proposition 4.2

Here we prove Proposition 4.2. In the proof below, we denote by $\|\cdot\|$ the norm of $L^2((\mathbf{R}_+)^k, d\bar{s})$ for any integer k. We will also make use of the notation $\|\|\cdot\|\|$ for iterated L^2 norms.

Step 1 (*Setup*). We only need to show that

$$\mathbf{E}\big[\exp\big(|Y/M|^{2/n}\big)\big] \leqslant K(n),$$

where the constant $K(n)$ depends only on n. Indeed, if $K(n) \leqslant 2$, we may take $C_n = 1$, and otherwise we must take $C_n > 1$. We can thus use Jensen's inequality to write

$$\begin{aligned}\mathbf{E}\left[\exp\left(\left|\frac{Y}{C_n M}\right|^{2/n}\right)\right] &= \mathbf{E}\big[\big(\exp\big(|Y/M|^{2/n}\big)\big)^{(C_n)^{-2/n}}\big] \\ &\leqslant K(n)^{(C_n)^{-2/n}}\end{aligned}$$

so that it is sufficient to take $C_n = (\log K(n)/\log 2)^{n/2}$.

Step 2 (*Poincaré inequality for fractional exponential moments*). We will use a coupling inequality of Üstünel for convex functions, namely Theorem 9.2.2 in [16]. Since the function $\exp(|x|^{2/n})$ is convex only for $|x|^{2/n} \geqslant (n-2)/2$, we write brutally

$$\mathbf{E}\big[\exp\big(|Y/M|^{2/n}\big)\big] \leqslant \mathbf{E}\big[\exp\big(|Y/M|^{2/n} \vee a_n\big)\big],$$

where $a_n := (n/2-1)$. Now we claim that Theorem 9.2.2 in [16], allows us to prove the following.

Lemma A.1 *(Fractional exponential Poincaré inequality). For any centered random variable $X \in L^2(\Omega)$, and any value $\alpha \in [0,1]$,*

$$\begin{aligned}\mathbf{E}_X\big[\exp\big(|X|^\alpha \vee a_{2/\alpha}\big)\big] &\leqslant \mathbf{E}_X \otimes \mathbf{E}_Z\left[\exp\left(\left|Z\frac{\pi}{2}\|DX\|\right|^\alpha \vee a_{2/\alpha}\right)\right] \\ &= \mathbf{E}_Z\left[\mathbf{E}_X\left[\exp\left(\left|Z\frac{\pi}{2}\|DX\|\right|^\alpha \vee a_{2/\alpha}\right)\right]\right],\end{aligned} \tag{A.1}$$

where under the measure $P_X \otimes P_Z$, Z and X are independent, and Z is standard normal.

Similarly, for a centered random field X on an index set I with $X(t) \in L^2(\Omega,\mathcal{F},\mathbf{P})$ for all $t \in I$, and with $\|\cdot\|_{\mathcal{F}}$ a norm on a function space $\mathcal{F}$ such that $X(\cdot) \in \mathcal{F}$ almost-surely,

$$\begin{aligned}\mathbf{E}_X\big[\exp\big(\|X\|_{\mathcal{F}}^\alpha \vee a_{2/\alpha}\big)\big] &\leqslant \mathbf{E}_X \otimes \mathbf{E}_Z\left[\exp\left(\left\|Z\frac{\pi}{2}\|DX\|\right\|_{\mathcal{F}}^\alpha \vee a_{2/\alpha}\right)\right] \\ &= \mathbf{E}_Z\left[\mathbf{E}_X\left[\exp\left(\left\|Z\frac{\pi}{2}\|DX\|\right\|_{\mathcal{F}}^\alpha \vee a_{2/\alpha}\right)\right]\right].\end{aligned} \tag{A.2}$$

Proof. To prove (A.1), simply note that Üstünel's notation $I_1(\nabla\varphi(w))(z) = \int_0^1 \frac{d}{dt}\nabla\varphi(w,t)\,dz_t$ is none other than $\int_0^1 (D_t\varphi)(\omega)\,dz_t$ where $\varphi = \varphi(\omega)$ is a random variable w.r.t. ω, and z is a Brownian motion on a space not related to ω. Therefore, with respect to the randomness of the Brownian motion z (and hence, with ω fixed), the random variable $I_1(\nabla\varphi(w))(z)$ has the same distribution as a Gaussian r.v. with variance $\int_0^1 |D_t\varphi|^2(\omega)\,dt$. Now we repeat precisely the proof of Theorem 9.2.3 in [16] with $U(x) = \exp(|x|^\alpha \vee a_{2/\alpha})$. Let $\tilde{X} := X(\tilde{\omega})$ be an independent copy of $X = X(\omega)$ and we write X as $X = X(\omega) - \mathbf{E}_{\tilde{X}} X(\tilde{\omega})$. Using Jensen's inequality w.r.t. $\mathbf{E}_{\tilde{X}}$ for the convex function $x \mapsto \exp(|X(\omega)-x|^\alpha \vee a_{2/\alpha})$

$$\begin{aligned}\mathbf{E}_X\big[\exp\big(|X|^\alpha \vee a_{2/\alpha}\big)\big] &= \mathbf{E}_X\big[\exp\big(\big|X(\omega) - \mathbf{E}_{\tilde{X}} X(\tilde{\omega})\big|^\alpha \vee a_{2/\alpha}\big)\big] \\ &\leqslant \mathbf{E}_X\big[\mathbf{E}_{\tilde{X}}\big[\exp\big(\big|X(\omega) - X(\tilde{\omega})\big|^\alpha \vee a_{2/\alpha}\big)\big]\big].\end{aligned}$$

Next we invoke Theorem 9.2.2 in [16] with the $U(x)$ and the remark above on the law of $I_1(\nabla\varphi(w))(z)$ for w fixed, yielding that the last expression above is

$$\begin{aligned}\mathbf{E}_X\big[\mathbf{E}_{\tilde{X}}\big[\exp\big(\big|X(\omega) - X(\tilde{\omega})\big|^\alpha \vee a_{2/\alpha}\big)\big]\big] &= \mathbf{E}_X\big[\mathbf{E}_{\tilde{X}}\big[U\big(X(\omega) - X(\tilde{\omega})\big)\big]\big] \\ &\leqslant \mathbf{E}_X\left[\mathbf{E}_Z\left[U\left(\frac{\pi}{2}I_1\big(\nabla X(\omega)\big)(z)\right)\right]\right] \\ &= \mathbf{E}_X \otimes \mathbf{E}_Z\left[\exp\left(\left|Z\frac{\pi}{2}\|DX\|\right|^\alpha \vee a_{2/\alpha}\right)\right],\end{aligned}$$

where Z is standard normal and which proves (A.1). To prove (A.2), we claim that Lemma 9.2.1 in [16] also holds for processes and fields.

Lemma *(Intermediate lemma). Let $X = \{X_s\colon s \in I\}$ be a Gaussian field on an index set I with values in $\mathbf{R}^d$ and let $U(X) = \exp(\|X\|^\alpha \vee a_{2/\alpha})$, where $\|\cdot\|$ is a norm on a function space $\mathcal{F}$ such that $X \in \mathcal{F}$ a.s. For any C^1-function $V : \mathbf{R}^d \to \mathbf{R}$, with V' its gradient, we have the following inequality*:

$$\mathbf{E}\big[U\big(V(X_\cdot) - V(Y_\cdot)\big)\big] \leqslant \mathbf{E}\bigg[U\bigg(\frac{\pi}{2}\big(V'(X_\cdot), Y_\cdot\big)_{\mathbf{R}^d}\bigg)\bigg],$$

where Y is an independent copy of X and $\mathbf{E}$ is the expectation with respect to the product measure.

The proof of this intermediate lemma follows the same proof in [16]. Indeed let $X_s(\theta) = X_s \sin\theta + Y_s \cos\theta$. Then

$$\begin{aligned} V(X_s) - V(Y_s) &= \int_0^{\pi/2} \frac{d}{d\theta} V\big(X_s(\theta)\big)\, d\theta \\ &= \int_0^{\pi/2} \big(V'\big(X_s(\theta)\big), X_s'(\theta)\big)_{\mathbf{R}^d}\, d\theta \\ &= \frac{\pi}{2} \int_0^{\pi/2} \big(V'\big(X_s(\theta)\big), X_s'(\theta)\big)_{\mathbf{R}^d}\, d\tilde{\theta}, \end{aligned}$$

where $d\tilde{\theta} = \frac{d\theta}{\pi/2}$. We have

$$\begin{aligned} U\big(V(X_\cdot) - V(Y_\cdot)\big) &= \exp\left(\left\| \int_0^{\pi/2} \frac{\pi}{2}\big(V'\big(X_\cdot(\theta)\big), X_\cdot'(\theta)\big)_{\mathbf{R}^d}\, d\tilde{\theta}\right\|^\alpha \vee a_{2/\alpha}\right) \\ &\leqslant \exp\left(\left(\int_0^{\pi/2} \left\|\frac{\pi}{2}\big(V'\big(X_\cdot(\theta)\big), X_\cdot'(\theta)\big)_{\mathbf{R}^d}\right\| d\tilde{\theta}\right)^\alpha \vee a_{2/\alpha}\right) \\ &\leqslant \int_0^{\pi/2} \exp\left(\left\|\frac{\pi}{2}\big(V'\big(X_\cdot(\theta)\big), X_\cdot'(\theta)\big)_{\mathbf{R}^d}\right\|^\alpha \vee a_{2/\alpha}\right) d\tilde{\theta} \\ &= \int_0^{\pi/2} U\left(\frac{\pi}{2}\big(V'\big(X_\cdot(\theta)\big), X_\cdot'(\theta)\big)_{\mathbf{R}^d}\right) d\tilde{\theta}, \end{aligned} \tag{A.3}$$

where in line (A.3), we used that the function $\|X\| \mapsto \exp(\|X\|^\alpha \vee a_{2/\alpha})$ is convex. Moreover $X_s(\theta)$ and $X_s'(\theta)$ are two independent Gaussian processes with the same law as X_s. Hence

$$\mathbf{E}\big[U\big(V(X.)-V(Y.)\big)\big] \leqslant \int_0^{\pi/2} \mathbf{E}\left[U\left(\frac{\pi}{2}\big(V'(X.),Y.\big)_{\mathbf{R}^d}\right)\right] d\tilde{\theta}$$
$$= \mathbf{E}\left[U\left(\frac{\pi}{2}\big(V'(X.),Y.\big)_{\mathbf{R}^d}\right)\right].$$

This proves the Intermediate lemma.

It immediately implies that Theorem 9.2.2 in [16] also holds for processes and fields. A similar argument as in the proof for (A.1) now proves (A.2) from this extension of Theorem 9.2.2 in [16]. This finishes the proof of the fractional exponential Poincaré inequality, Lemma A.1. □

Step 3 (*Iteration*). Now we simply apply inequality (A.2) to inequality (A.1) with $\alpha = 2/n$ and iterate. Assuming $X = \sum_{k=n}^{\infty} I_k(f_k)$, the iterated Malliavin derivatives of X up to order $n-1$ will have mean 0, justifying the repeated use of (A.2) below. In the first pair of iterations we have

$$\begin{aligned}
&\mathbf{E}_X\big[\exp\big(|X|^{2/n} \vee a_n\big)\big] \\
&\quad\leqslant \mathbf{E}_{Z_1}\left[\mathbf{E}_X\left[\exp\left(\left|Z_1\frac{\pi}{2}\right|^{2/n}\|DX\|^{2/n} \vee a_n\right)\right]\right] \\
&\quad\leqslant \mathbf{E}_{Z_1}\left[\mathbf{E}_{Z_2}\left[\mathbf{E}_X\left[\exp\left(\left(\left|Z_2Z_1\left(\frac{\pi}{2}\right)^2\right|^{2/n}\big\|\|DDX\|\big\|_{\mathcal{F}}^{2/n}\right) \vee a_n\right)\right]\right]\right] \\
&\quad= \mathbf{E}_{Z_1}\left[\mathbf{E}_{Z_2}\left[\mathbf{E}_X\left[\exp\left(\left(\left|Z_2Z_1\left(\frac{\pi}{2}\right)^2\right|^{2/n}\big\|D^{(2)}X\big\|^{2/n}\right) \vee a_n\right)\right]\right]\right],
\end{aligned}$$

where Z_1 and Z_2 are independent standard normals, and we used the fact that, in the notation of Lemma A.1, since $\mathcal{F}$ is $L^2([0,1])$, we may write $L^2([0,1];\mathcal{F}) = L^2([0,1]^2)$. Hence, iterating this procedure we get, using a measure $\mathbf{P}_Z$ under which $Z_1, Z_2, \ldots, Z_n$ are IID standard normals,

$$\mathbf{E}_X\big[\exp\big(|X|^{2/n} \vee a_n\big)\big] \leqslant \mathbf{E}_Z\left[\mathbf{E}_X\left[\exp\left(\left(|Z_n\cdots Z_2Z_1|^{2/n}\frac{\pi^2}{4}\big\|D^{(n)}X\big\|^{2/n}\right) \vee a_n\right)\right]\right].$$

Now, replacing X by Y/M and noting that our hypothesis on Y says $\|D^{(n)}X\| \leqslant 1$,

$$\begin{aligned}
\mathbf{E}_X\big[\exp\big(|Y/M|^{2/n} \vee a_n\big)\big] &\leqslant \mathbf{E}_Z\left[\exp\left(\left(\frac{\pi^2}{4}|Z_n\cdots Z_2Z_1|^{2/n}\right) \vee (n/2-1)\right)\right] \\
&=: K(n).
\end{aligned} \tag{A.4}$$

The latter being a universal constant depending only on n, the theorem is proved. In order to make the definition of the constant $K(n)$ more explicit, we may use the relation between algebraic and geometric means to get

$$K(n) \leqslant \left(\mathbf{E}_{Z_1}\left[\exp\left(\frac{\pi^2}{4n}Z_1^2 \vee (n/2-1)\right)\right]\right)^n. \tag{A.5}$$

The proof of Proposition 4.2 is complete.

References

[1] R.J. Adler, J. Taylor, Random Fields and Geometry, Monogr. Math., Springer, 2007.
[2] C. Borell, Tail probabilities in Gauss space, in: Vector Space Measures and Applications, Dublin, 1977, in: Lecture Notes in Math., vol. 644, Springer, 1978, pp. 71–82.
[3] P. Cheridito, D. Nualart, Stochastic integral of divergence type with respect to fractional Brownian motion with Hurst parameter $H \in (0, 1/2)$, Ann. Inst. H. Poincaré Probab. Statist. 41 (6) (2005) 1049–1081.
[4] R.M. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes, J. Funct. Anal. 1 (1967) 290–330.
[5] X. Fernique, Régularité des processus gaussienes, Invent. Math. 12 (1971) 304–320.
[6] X. Fernique, Régularité des trajectoires des fonctions aléatoires gaussiennes, in: Ecole d'Eté de Probabilités de Saint-Flour IV—1974, in: Lecture Notes in Math., vol. 480, Springer, 1975, pp. 1–96.
[7] M. Ledoux, Isoperimetry and Gaussian analysis, in: Lectures on Probability Theory and Statistics, Saint-Flour, 1994, in: Lecture Notes in Math., vol. 1648, Springer, 1996, pp. 165–294.
[8] M. Ledoux, M. Talagrand, Probability in Banach Spaces, Springer, 1991.
[9] P. Malliavin, Stochastic Analysis, Springer, 2002.
[10] O. Mocioalca, F. Viens, Skorohod integration and stochastic calculus beyond the fractional Brownian scale, J. Funct. Anal. 222 (2) (2004) 385–434.
[11] D. Nualart, The Malliavin Calculus and Related Topics, 2nd ed., Springer, 2006.
[12] D. Revuz, M. Yor, Continuous Martingales and Brownian Motion, 3rd ed., Springer, 1999.
[13] M. Talagrand, Regularity of Gaussian processes, Acta Math. 159 (1–2) (1987) 99–149.
[14] M. Talagrand, The Generic Chaining, Springer, 2005.
[15] S. Tindel, C.A. Tudor, F. Viens, Sharp Gaussian regularity on the circle and application to the fractional stochastic heat equation, J. Funct. Anal. 217 (2) (2004) 280–313.
[16] A.-S. Üstünel, An Introduction to Analysis on Wiener Space, Lecture Notes in Math., vol. 1610, Springer, 1995.
[17] A. Vizcarra, F. Viens, Some applications of the Malliavin calculus to sub-Gaussian and non-sub-Gaussian random fields, in: Seminar on Stochastic Analysis, Random Fields and Applications V, Progr. Probab., Birkhäuser, 2007, in press.

Available online at www.sciencedirect.com
ScienceDirect

Journal of Functional Analysis 248 (2007) 27–78

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

Complex-valued Ray–Singer torsion ☆

Dan Burghelea [a], Stefan Haller [b,*]

[a] *Department of Mathematics, The Ohio State University, 231 West 18th Avenue, Columbus, OH 43210, USA*
[b] *Department of Mathematics, University of Vienna, Nordbergstraße 15, A-1090 Vienna, Austria*

Received 22 April 2006; accepted 18 March 2007

Available online 10 May 2007

Communicated by J. Coron

Abstract

In the spirit of Ray and Singer we define a complex-valued analytic torsion using non-selfadjoint Laplacians. We establish an anomaly formula which permits to turn this into a topological invariant. Conjecturally this analytically defined invariant computes the complex-valued Reidemeister torsion, including its phase. We establish this conjecture in some non-trivial situations.

Keywords: Ray–Singer torsion; Reidemeister torsion; Analytic torsion; Combinatorial torsion; Anomaly formula; Bismut–Zhang, Cheeger, Müller theorem; Euler structures; Co-Euler structures; Asymptotic expansion; Heat kernel; Dirac operator

Contents

☆ Part of this work was done while the second author enjoyed the hospitality of the Max Planck Institute for Mathematics in Bonn. During the final stage of preparation of this manuscript the second author was partially supported by the Austrian Science Fund (FWF), grant P19392-N13, and the IK I008-N funded by the University of Vienna.
* Corresponding author.
E-mail addresses: burghele@mps.ohio-state.edu (D. Burghelea), stefan.haller@univie.ac.at (S. Haller).

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2007.03.027

1. Introduction

Let M be a closed connected smooth manifold with Riemannian metric g. Suppose E is a flat complex vector bundle over M. Let h be a Hermitian metric on E. Recall the de Rham differential $d_E : \Omega^*(M; E) \to \Omega^{*+1}(M; E)$ on the space of E-valued differential forms. Let $d^*_{E,g,h} : \Omega^{*+1}(M; E) \to \Omega^*(M; E)$ denote its formal adjoint with respect to the Hermitian scalar product on $\Omega^*(M; E)$ induced by g and h. Consider the Laplacian $\Delta_{E,g,h} = d_E d^*_{E,g,h} + d^*_{E,g,h} d_E : \Omega^*(M; E) \to \Omega^*(M; E)$. Recall the (inverse square of the) Ray–Singer torsion [29]

$$\prod_q \big(\det{}'(\Delta_{E,g,h,q})\big)^{(-1)^q q} \in \mathbb{R}^+.$$

Here $\det'(\Delta_{E,g,h,q})$ denotes the zeta regularized product of all non-zero eigen-values of the Laplacian acting in degree q. This is a positive real number which coincides, up to a computable correction term, with the absolute value of the Reidemeister torsion, see [2].

The aim of this paper is to introduce a complex-valued Ray–Singer torsion which, conjecturally, computes the Reidemeister torsion, including its phase. This is accomplished by replacing the Hermitian fiber metric h with a fiber wise non-degenerate symmetric bilinear form b on E. The bilinear form b permits to define a formal transposed $d^\sharp_{E,g,b}$ of d_E, and an in general not self-adjoint Laplacian $\Delta_{E,g,b} := d_E d^\sharp_{E,g,b} + d^\sharp_{E,g,b} d_E$. The (inverse square of the) complex-valued Ray–Singer torsion is then defined by

$$\prod_q \big(\det{}'(\Delta_{E,g,b,q})\big)^{(-1)^q q} \in \mathbb{C}^\times := \mathbb{C} \setminus \{0\}. \tag{1}$$

The main result proved here, see Theorem 4.2, is an anomaly formula for the complex-valued Ray–Singer torsion, i.e. we compute the variation of the quantity (1) through a variation of g and b. This ultimately permits to define a smooth invariant, the analytic torsion.[1]

The paper is roughly organized as follows. In Section 2 we recall Euler and coEuler structures. These are used to turn the Reidemeister torsion and the complex-valued Ray–Singer torsion into topological invariants referred to as combinatorial and analytic torsion, respectively. In Section 3 we discuss some finite-dimensional linear algebra and recall the combinatorial torsion which was also called Milnor–Turaev torsion in [9]. Section 4 contains the definition of the proposed complex-valued analytic torsion. In Section 5 we formulate a conjecture, see Conjecture 5.1, relating the complex-valued analytic torsion with the combinatorial torsion. We establish this conjecture in some non-trivial cases via analytic continuation from a result of Cheeger [16,17], Müller [28] and Bismut–Zhang [2]. Section 6 contains the derivation of the anomaly formula.

[1] The use of a fiberwise non-degenerate bilinear form instead of Hermitian fiber metric was suggested by W. Müller.

This proof is based on the computation of leading and subleading terms in the asymptotic expansion of the heat kernel associated with a certain class of Dirac operators. This asymptotic expansion is formulated and proved in Section 7, see Theorem 7.1. In Section 8 we apply this result to the Laplacians $\Delta_{E,g,b}$ and therewith complete the proof of the anomaly formula.

We restrict the presentation to the case of vanishing Euler–Poincaré characteristics to avoid geometric regularization, see [9,10]. With minor modifications everything can easily be extended to the general situation. This is sketched in Section 9. The analytic core of the results, Theorem 7.1 and its corollaries Propositions 6.1 and 6.2, are formulated and proved without any restriction on the Euler–Poincaré characteristics.

Let us also mention the series of recent preprints [3–7]. In these papers Braverman and Kappeler construct a “refined analytic torsion” based on the odd signature operator on odd-dimensional manifolds. Their torsion is closely related to the analytic torsion proposed in this paper. For a comparison result see [7, Theorem 1.4]. Some of the results below which partially establish Conjecture 5.1, have first appeared in [7], and were not contained in the first version of this paper. The proofs we will provide have been inspired by [7] but do not rely on the results therein.

Recently, in October 2006, two preprints [13,33] have been posted on the internet providing the proof of Conjecture 5.1. In [13] Witten–Helffer–Sjöstrand theory has been extended to the non-selfadjoint Laplacians discussed here, and used along the lines of [15], to establish Conjecture 5.1 for odd-dimensional manifolds, up to sign. Comments were made how to derive the conjecture in full generality on these lines. A few days earlier, by adapting the methods in [2] to the non-selfadjoint situation, Su and Zhang in [33] provided a proof of the conjecture.

The definition of the complex-valued analytic torsion was sketched in [12], the particular case $\operatorname{rank} E = 1$ is implicit in [11].

2. Preliminaries

Throughout this section M denotes a closed connected smooth manifold of dimension n. For simplicity we will also assume vanishing Euler–Poincaré characteristics, $\chi(M) = 0$. At the expense of a base point everything can easily be extended to the general situation, see [8–10] and Section 9.

2.1. Euler structures

Let M be a closed connected smooth manifold of dimension n with $\chi(M) = 0$. The set of *Euler structures with integral coefficients* $\mathfrak{Eul}(M;\mathbb{Z})$ is an *affine version* of $H_1(M;\mathbb{Z})$. That is, the homology group $H_1(M;\mathbb{Z})$ acts free and transitively on $\mathfrak{Eul}(M;\mathbb{Z})$ but in general there is no distinguished origin. Euler structures have been introduced by Turaev [34] in order to remove the ambiguities in the definition of the Reidemeister torsion. Below we will briefly recall a possible definition. For more details we refer to [9,10].

Recall that a vector field X is called non-degenerate if $X : M \to TM$ is transverse to the zero section. Denote its set of zeros by $\mathcal{X}$. Recall that every $x \in \mathcal{X}$ has a *Hopf index* $\mathrm{IND}_X(x) \in \{\pm 1\}$. Consider pairs (X, c) where X is a non-degenerate vector field and $c \in C_1^{\mathrm{sing}}(M;\mathbb{Z})$ is a singular 1-chain satisfying

$$\partial c = \mathrm{e}(X) := \sum_{x\in\mathcal{X}} \mathrm{IND}_X(x)x.$$

Every non-degenerate vector field admits such c since we assumed $\chi(M) = 0$.

We call two such pairs (X_1, c_1) and (X_2, c_2) equivalent if

$$c_2 - c_1 = \mathrm{cs}(X_1, X_2) \in C_1^{\mathrm{sing}}(M; \mathbb{Z})/\partial C_2^{\mathrm{sing}}(M; \mathbb{Z}).$$

Here $\mathrm{cs}(X_1, X_2) \in C_1^{\mathrm{sing}}(M; \mathbb{Z})/\partial C_2^{\mathrm{sing}}(M; \mathbb{Z})$ denotes the *Chern–Simons class* which is represented by the zero set of a generic homotopy connecting X_1 with X_2. It follows from $\mathrm{cs}(X_1, X_2) + \mathrm{cs}(X_2, X_3) = \mathrm{cs}(X_1, X_3)$ that this indeed is an equivalence relation.

Define $\mathfrak{Eul}(M; \mathbb{Z})$ as the set of equivalence classes $[X, c]$ of pairs considered above. The action of $[\sigma] \in H_1(M; \mathbb{Z})$ on $[X, c] \in \mathfrak{Eul}(M; \mathbb{Z})$ is simply given by $[X, c] + [\sigma] := [X, c + \sigma]$. Since $\mathrm{cs}(X, X) = 0$ this action is well defined and free. Because of $\partial\mathrm{cs}(X_1, X_2) = \mathrm{e}(X_2) - \mathrm{e}(X_1)$ it is transitive.

Replacing singular chains with integral coefficients by singular chains with real or complex coefficients we obtain in exactly the same way *Euler structures with real coefficients* $\mathfrak{Eul}(M; \mathbb{R})$ and *Euler structures with complex coefficients* $\mathfrak{Eul}(M; \mathbb{C})$. These are affine version of $H_1(M; \mathbb{R})$ and $H_1(M; \mathbb{C})$, respectively. There are obvious maps $\mathfrak{Eul}(M; \mathbb{Z}) \to \mathfrak{Eul}(M; \mathbb{R}) \to \mathfrak{Eul}(M; \mathbb{C})$ which are affine over the homomorphisms $H_1(M; \mathbb{Z}) \to H_1(M; \mathbb{R}) \to H_1(M; \mathbb{C})$. We refer to the image of $\mathfrak{Eul}(M; \mathbb{Z})$ in $\mathfrak{Eul}(M; \mathbb{R})$ or $\mathfrak{Eul}(M; \mathbb{C})$ as the *lattice of integral Euler structures*.

Since we have $\mathrm{e}(-X) = (-1)^n \mathrm{e}(X)$ and $\mathrm{cs}(-X_1, -X_2) = (-1)^n \mathrm{cs}(X_1, X_2)$, the assignment $\nu([X, c]) := [-X, (-1)^n c]$ defines *affine involutions* on $\mathfrak{Eul}(M; \mathbb{Z})$, $\mathfrak{Eul}(M; \mathbb{R})$ and $\mathfrak{Eul}(M; \mathbb{C})$. If n is even, then the involutions on $\mathfrak{Eul}(M; \mathbb{R})$ and $\mathfrak{Eul}(M; \mathbb{C})$ are affine over the identity and so we must have $\nu = \mathrm{id}$. If n is odd the involutions on $\mathfrak{Eul}(M; \mathbb{R})$ and $\mathfrak{Eul}(M; \mathbb{C})$ are affine over $-\mathrm{id}$ and thus must have a unique fixed point $\mathfrak{e}_{\mathrm{can}} \in \mathfrak{Eul}(M; \mathbb{R}) \subseteq \mathfrak{Eul}(M; \mathbb{C})$. This *canonic Euler structure* permits to naturally identify $\mathfrak{Eul}(M; \mathbb{R})$ respectively $\mathfrak{Eul}(M; \mathbb{C})$ with $H_1(M; \mathbb{R})$ respectively $H_1(M; \mathbb{C})$, provided n is odd. Note that in general none of these statements is true for the involution on $\mathfrak{Eul}(M; \mathbb{Z})$. This is due to the fact that in general $H_1(M; \mathbb{Z})$ contains non-trivial elements of order 2, and elements which are not divisible by 2.

Finally, observe that the assignment $[X, c] \mapsto [X, \bar{c}]$ defines a *conjugation* $\mathfrak{e} \mapsto \bar{\mathfrak{e}}$ on $\mathfrak{Eul}(M; \mathbb{C})$ which is affine over the complex conjugation $H_1(M; \mathbb{C}) \to H_1(M; \mathbb{C})$, $[\sigma] \mapsto [\bar{\sigma}]$. Clearly, the set of fixed points of this conjugation coincides with $\mathfrak{Eul}(M; \mathbb{R}) \subseteq \mathfrak{Eul}(M; \mathbb{C})$.

Lemma 2.1. *Let M be a closed connected smooth manifold with $\chi(M) = 0$, let $\mathfrak{e} \in \mathfrak{Eul}(M; \mathbb{Z})$ be an Euler structure, and let $x_0 \in M$ be a base point. Suppose X is a non-degenerate vector field on M with zero set $\mathcal{X} \neq \emptyset$. Then there exists a collection of paths σ_x, $\sigma_x(0) = x_0$, $\sigma_x(1) = x$, $x \in \mathcal{X}$, so that $\mathfrak{e} = [X, \sum_{x \in \mathcal{X}} \mathrm{IND}_X(x)\sigma_x]$.*

Proof. For every zero $x \in \mathcal{X}$ choose a path $\tilde{\sigma}_x$ with $\tilde{\sigma}_x(0) = x_0$ and $\tilde{\sigma}_x(1) = x$. Set $\tilde{c} := \sum_{x \in \mathcal{X}} \mathrm{IND}_X(x)\tilde{\sigma}_x$. Since $\chi(M) = 0$ we clearly have $\partial\tilde{c} = \mathrm{e}(X)$. So the pair $(X, \tilde{c})$ represents an Euler structure $\tilde{\mathfrak{e}} := [X, \tilde{c}] \in \mathfrak{Eul}(M; \mathbb{Z})$. Because $H_1(M; \mathbb{Z})$ acts transitively on $\mathfrak{Eul}(M; \mathbb{Z})$ we find $a \in H_1(M; \mathbb{Z})$ with $\tilde{\mathfrak{e}} + a = \mathfrak{e}$. Since the Huréwicz homomorphism is onto we can represent a by a closed path π with $\pi(0) = \pi(1) = x_0$. Choose $y \in \mathcal{X}$. Define σ_y as the concatenation of $\tilde{\sigma}_y$ with $\pi^{\mathrm{IND}_X(y)}$, and set $\sigma_x := \tilde{\sigma}_x$ for $x \neq y$. Then the pair $(x, \sum_{x \in \mathcal{X}} \mathrm{IND}_X(x)\sigma_x)$ represents $\tilde{\mathfrak{e}} + a = \mathfrak{e}$. $\square$

2.2. CoEuler structures

Let M be a closed connected smooth manifold of dimension n with $\chi(M) = 0$. The set of *coEuler structures* $\mathfrak{Eul}^*(M; \mathbb{C})$ is an affine version of $H^{n-1}(M; \mathcal{O}_M^{\mathbb{C}})$. That is the cohomology group $H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$ with values in the complexified orientation bundle $\mathcal{O}_M^{\mathbb{C}}$ acts free and transitively on $\mathfrak{Eul}^*(M;\mathbb{C})$. CoEuler structures are well suited to remove the metric dependence from the Ray–Singer torsion. Below we will briefly recall their definition, and discuss an affine version of Poincaré duality relating Euler with coEuler structures. For more details and the general situation we refer to [9] or [10].

Consider pairs (g,α), g a Riemannian metric on M, $\alpha\in\Omega^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$, which satisfy

$$d\alpha=\mathrm{e}(g).$$

Here $\mathrm{e}(g)\in\Omega^n(M;\mathcal{O}_M^{\mathbb{C}})$ denotes the *Euler form* associated with g. In view of the Gauss–Bonnet theorem every g admits such α for we assumed $\chi(M)=0$.

Two pairs (g_1,α_1) and (g_2,α_2) as above are called equivalent if

$$\alpha_2-\alpha_1=\mathrm{cs}(g_1,g_2)\in\Omega^{n-1}\big(M;\mathcal{O}_M^{\mathbb{C}}\big)/d\Omega^{n-2}\big(M;\mathcal{O}_M^{\mathbb{C}}\big).$$

Here $\mathrm{cs}(g_1,g_2)\in\Omega^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})/d\Omega^{n-2}(M;\mathcal{O}_M^{\mathbb{C}})$ denotes the *Chern–Simons class* [18] associated with g_1 and g_2. Since $\mathrm{cs}(g_1,g_2)+\mathrm{cs}(g_2,g_3)=\mathrm{cs}(g_1,g_3)$ this is indeed an equivalence relation.

Define the set of *coEuler structures with complex coefficients* $\mathfrak{Eul}^*(M;\mathbb{C})$ as the set of equivalence classes $[g,\alpha]$ of pairs considered above. The action of $[\beta]\in H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$ on $[g,\alpha]\in\mathfrak{Eul}^*(M;\mathbb{C})$ is defined by $[g,\alpha]+[\beta]:=[g,\alpha-\beta]$. Since $\mathrm{cs}(g,g)=0$ this action is well defined and free. Because of $d\mathrm{cs}(g_1,g_2)=\mathrm{e}(g_2)-\mathrm{e}(g_1)$ it is transitive too.

Replacing forms with values in $\mathcal{O}_M^{\mathbb{C}}$ by forms with values in the real orientation bundle $\mathcal{O}_M^{\mathbb{R}}$ we obtain in exactly the same way *coEuler structures with real coefficients* $\mathfrak{Eul}^*(M;\mathbb{R})$, an affine version of $H^{n-1}(M;\mathcal{O}_M^{\mathbb{R}})$. There is an obvious map $\mathfrak{Eul}^*(M;\mathbb{R})\to\mathfrak{Eul}^*(M;\mathbb{C})$ which is affine over the homomorphism $H^{n-1}(M;\mathcal{O}_M^{\mathbb{R}})\to H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$.

In view of $(-1)^n\mathrm{e}(g)=\mathrm{e}(g)$ and $(-1)^n\mathrm{cs}(g_1,g_2)=\mathrm{cs}(g_1,g_2)$ the assignment $\nu([g,\alpha]):=[g,(-1)^n\alpha]$ defines *affine involutions* on $\mathfrak{Eul}^*(M;\mathbb{R})$ and $\mathfrak{Eul}^*(M;\mathbb{C})$. For even n these involutions are affine over the identity and so we must have $\nu=\mathrm{id}$. For odd n they are affine over $-\mathrm{id}$ and thus must have a unique fixed point $\mathfrak{e}^*_{\mathrm{can}}\in\mathfrak{Eul}^*(M;\mathbb{R})\subseteq\mathfrak{Eul}^*(M;\mathbb{C})$. Since $\mathrm{e}(g)=0$ in this case, we have $\mathfrak{e}^*_{\mathrm{can}}=[g,0]$ where g is any Riemannian metric. This *canonic coEuler structure* provides a natural identification of $\mathfrak{Eul}^*(M;\mathbb{R})$ respectively $\mathfrak{Eul}^*(M;\mathbb{C})$ with $H^{n-1}(M;\mathcal{O}_M^{\mathbb{R}})$ respectively $H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$, provided the dimension is odd.

Finally, observe that the assignment $[g,\alpha]\mapsto[g,\bar\alpha]$ defines a *complex conjugation* $\mathfrak{e}^*\mapsto\bar{\mathfrak{e}}^*$ on $\mathfrak{Eul}^*(M;\mathbb{C})$ which is affine over the complex conjugation $H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})\to H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$, $[\beta]\mapsto[\bar\beta]$. Clearly, the set of fixed points of this conjugation coincides with the image of $\mathfrak{Eul}^*(M;\mathbb{R})\subseteq\mathfrak{Eul}^*(M;\mathbb{C})$.

2.3. Poincaré duality for Euler structures

Let M be a closed connected smooth manifold of dimension n with $\chi(M)=0$. There is a canonic isomorphism

$$P:\mathfrak{Eul}(M;\mathbb{C})\to\mathfrak{Eul}^*(M;\mathbb{C})\tag{2}$$

which is affine over the Poincaré duality $H_1(M;\mathbb{C}) \to H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$. If $[X,c] \in \mathfrak{Eul}(M;\mathbb{C})$ and $[g,\alpha] \in \mathfrak{Eul}^*(M;\mathbb{C})$ then $P([X,c]) = [g,\alpha]$ iff we have

$$\int_{M\setminus\mathcal{X}} \omega \wedge \big(X^*\Psi(g) - \alpha\big) = \int_c \omega \tag{3}$$

for all closed one forms ω which vanish in a neighborhood of $\mathcal{X}$, the zero set of X. Here $\Psi(g) \in \Omega^{n-1}(TM \setminus M; \pi^*\mathcal{O}_M^{\mathbb{C}})$ denotes the *Mathai–Quillen form* [26] associated with g, and $\pi : TM \to M$ denotes the projection. With a little work one can show that (3) does indeed define an assignment as in (2). Once this is established (2) is obviously affine over the Poincaré duality and hence an isomorphism. It follows immediately from $(-X)^*\Psi(g) = (-1)^n X^*\Psi(g)$ that P intertwines the involution on $\mathfrak{Eul}(M;\mathbb{C})$ with the involution on $\mathfrak{Eul}^*(M;\mathbb{C})$. Moreover, P obviously intertwines the complex conjugations on $\mathfrak{Eul}(M;\mathbb{C})$ and $\mathfrak{Eul}^*(M;\mathbb{C})$. Particularly, (2) restricts to an isomorphism

$$P : \mathfrak{Eul}(M;\mathbb{R}) \to \mathfrak{Eul}^*(M;\mathbb{R})$$

affine over the Poincaré duality $H_1(M;\mathbb{R}) \to H^{n-1}(M;\mathcal{O}_M^{\mathbb{R}})$.

2.4. *Kamber–Tondeur form*

Suppose E is a flat complex vector bundle over a smooth manifold M. Let ∇^E denote the flat connection on E. Suppose b is a fiber wise non-degenerate symmetric bilinear form on E. The *Kamber–Tondeur form* is the one form

$$\omega_{E,b} := -\tfrac{1}{2}\operatorname{tr}\big(b^{-1}\nabla^E b\big) \in \Omega^1(M;\mathbb{C}). \tag{4}$$

More precisely, for a vector field Y on M we have $\omega_{E,b}(Y) := \operatorname{tr}(b^{-1}\nabla^E_Y b)$. Here the derivative of b with respect to the induced flat connection on $(E \otimes E)'$ is considered as $\nabla^E_Y b : E \to E'$. Then $b^{-1}\nabla^E_Y b : E \to E$ and $\omega_{E,b}(Y)$ is obtained by taking the fiber wise trace.

The bilinear form b induces a non-degenerate bilinear form $\det b$ on $\det E := \Lambda^{\operatorname{rk}(E)}E$. From $\det b^{-1}\nabla^{\det E}(\det b) = \operatorname{tr}(b^{-1}\nabla^E b)$ we obtain

$$\omega_{\det E,\det b} = \omega_{E,b}. \tag{5}$$

Particularly, $\omega_{E,b}$ depends on the flat line bundle $\det E$ and the induced bilinear form $\det b$ only. Since ∇^E is flat, $\omega_{E,b}$ is a closed 1-form, cf. (5).

Suppose b_1 and b_2 are two fiber wise non-degenerate symmetric bilinear forms on E. Set $A := b_1^{-1}b_2 \in \operatorname{Aut}(E)$, i.e. $b_2(v,w) = b_1(Av,w)$ for all v, w in the same fiber of E. Then $\det b_2 = \det b_1 \det A$, hence

$$\nabla^{\det E}(\det b_2) = \nabla^{\det E}(\det b_1)\det A + (\det b_1) d \det A$$

and therefore

$$\omega_{E,b_2} = \omega_{E,b_1} + \det A^{-1} d \det A. \tag{6}$$

If $\det b_1$ and $\det b_2$ are homotopic as fiber wise non-degenerate bilinear forms on $\det E$, then the function $\det A : M \to \mathbb{C}^\times := \mathbb{C} \setminus \{0\}$ is homotopic to the constant function 1. So we find a function $\log\det A : M \to \mathbb{C}$ with $d\log\det A = \det A^{-1} d\det A$, and in view of (6) the cohomology classes of ω_{E,b_1} and ω_{E,b_2} coincide. We conclude that the cohomology class $[\omega_{E,b}] \in H^1(M;\mathbb{C})$ depends on the flat line bundle $\det E$ and the homotopy class $[\det b]$ of the induced non-degenerate bilinear form $\det b$ on $\det E$ only.

If E_1 and E_2 are two flat vector bundles with fiber wise non-degenerate symmetric bilinear forms b_1 and b_2 then

$$\omega_{E_1\oplus E_2, b_1\oplus b_2} = \omega_{E_1,b_1} + \omega_{E_2,b_2}. \tag{7}$$

If E' denotes the dual of a flat vector bundle E, and if b' denotes the bilinear form on E' induced from a fiber wise non-degenerate symmetric bilinear form b on E then clearly

$$\omega_{E',b'} = -\omega_{E,b}. \tag{8}$$

If $\bar{E}$ denotes the complex conjugate of a flat complex vector bundle E, and if $\bar{b}$ denotes the complex conjugate bilinear form of a fiber wise non-degenerate symmetric bilinear form b on E, then obviously

$$\omega_{\bar{E},\bar{b}} = \overline{\omega_{E,b}}. \tag{9}$$

Finally, if F is a real flat vector bundle and h is a fiber wise non-degenerate symmetric bilinear form on F one defines in exactly the same way a real Kamber–Tondeur form $\omega_{F,h} := -\frac{1}{2}\operatorname{tr}(h^{-1}\nabla^F h)$ which is closed too. If $F^{\mathbb{C}} := F \otimes \mathbb{C}$ denotes the complexification of F and $h^{\mathbb{C}}$ denotes the complexification of h then clearly

$$\omega_{F^{\mathbb{C}},h^{\mathbb{C}}} = \omega_{F,h} \tag{10}$$

in $\Omega^1(M;\mathbb{R}) \subseteq \Omega^1(M;\mathbb{C})$. Note that all such h give rise to the same cohomology class $[\omega_{F,h}] \in H^1(M;\mathbb{R})$, see (5) and (6). To see this also note that the induced fiber wise non-degenerate bilinear form $\det h$ on $\det F$ has to be positive definite or negative definite, but $\omega_{\det F,-\det h} = \omega_{\det F,\det h}$.

2.5. Holonomy

Suppose E is a flat complex vector bundle over a connected smooth manifold M. Let $x_0 \in M$ be a base point. Parallel transport along closed loops provides an anti homomorphism $\pi_1(M,x_0) \to \mathrm{GL}(E_{x_0})$, where E_{x_0} denotes the fiber of E over x_0. Composing with the inversion in $\mathrm{GL}(E_{x_0})$ we obtain the *holonomy representation* of E at x_0

$$\mathrm{hol}^E_{x_0} : \pi_1(M,x_0) \to \mathrm{GL}(E_{x_0}).$$

Applying this to the flat line bundle $\det E := \Lambda^{\mathrm{rk}(E)} E$ we obtain a homomorphism $\mathrm{hol}^{\det E}_{x_0}$: $\pi_1(M,x_0) \to \mathrm{GL}(\det E_{x_0}) = \mathbb{C}^\times$ which factors to a homomorphism

$$\theta_E : H_1(M;\mathbb{Z}) \to \mathbb{C}^\times. \tag{11}$$

Lemma 2.2. *Suppose b is a non-degenerate symmetric bilinear form on E. Then*

$$\theta_E(\sigma) = \pm e^{\langle [\omega_{E,b}], \sigma \rangle}, \quad \sigma \in H_1(M;\mathbb{Z}).$$

Here $\langle [\omega_{E,b}], \sigma \rangle \in \mathbb{C}$ denotes the natural pairing of the cohomology class $[\omega_{E,b}] \in H^1(M;\mathbb{C})$ and $\sigma \in H_1(M;\mathbb{Z})$.

Proof. Let $\tau : [0,1] \to M$ be a smooth path with $\tau(0) = \tau(1) = x_0$. Consider the flat vector bundle $(\det E)^{-2} := (\det E \otimes \det E)'$. Let $\beta : [0,1] \to (\det E)^{-2}$ be a section over τ which is parallel. Since $\det b$ defines a global nowhere vanishing section of $(\det E)^{-2}$ we find $\lambda : [0,1] \to \mathbb{C}$ so that $\beta = \lambda \det b$. Clearly,

$$\lambda(1)\, \mathrm{hol}_{x_0}^{(\det E)^{-2}}\big([\tau]\big) = \lambda(0). \tag{12}$$

Differentiating $\beta = \lambda \det b$ we obtain $0 = \lambda' \det b + \lambda \nabla_{\tau'}^{(\det E)^{-2}}(\det b)$. Using (5) this yields $0 = \lambda' - 2\lambda\omega_{E,b}(\tau')$. Integrating we get

$$\lambda(1) = \lambda(0) \exp\left(\int_0^1 2\omega_{E,b}\big(\tau'(t)\big)\, dt \right) = \lambda(0) e^{2\langle [\omega_{E,b}], [\tau] \rangle}.$$

Taking (12) into account we obtain $\mathrm{hol}_{x_0}^{(\det E)^{-2}}([\tau]) = e^{-2\langle [\omega_{E,b}], [\tau] \rangle}$, and this gives $\mathrm{hol}_{x_0}^{\det E}([\tau]) = \pm e^{\langle [\omega_{E,b}], [\tau] \rangle}$. $\square$

3. Reidemeister torsion

The combinatorial torsion is an invariant associated to a closed connected smooth manifold M, an Euler structure with integral coefficients $\mathfrak{e}$, and a flat complex vector bundle E over M. In the way we consider it here this invariant is a non-degenerate bilinear form $\tau_{E,\mathfrak{e}}^{\mathrm{comb}}$ on the complex line $\det H^*(M;E)$—the graded determinant line of the cohomology with values in (the local system of coefficients provided by) E. If $H^*(M;E)$ vanishes, then $\tau_{E,\mathfrak{e}}^{\mathrm{comb}}$ becomes a non-vanishing complex number. The aim of this section is to recall these definitions, and to provide some linear algebra which will be used in the analytic approach to this invariant in Section 4.

Throughout this section M denotes a closed connected smooth manifold of dimension n. For simplicity we will also assume vanishing Euler–Poincaré characteristics, $\chi(M) = 0$. At the expense of a base point everything can easily be extended to the general situation, see [8–10] and Section 9.

3.1. Finite-dimensional Hodge theory

Suppose C^* is a finite-dimensional graded complex over $\mathbb{C}$ with differential $d : C^* \to C^{*+1}$. Its cohomology is a finite-dimensional graded vector space and will be denoted by $H(C^*)$. Recall that there is a canonic isomorphism of complex lines

$$\det C^* = \det H(C^*). \tag{13}$$

Let us explain the terms appearing in (13) in more details. If V is a finite-dimensional vector space its *determinant line* is defined to be the top exterior product $\det V := \Lambda^{\dim(V)} V$. If V^* is a finite-dimensional graded vector space its *graded determinant line* is defined by $\det V^* := \det V^{\mathrm{even}} \otimes (\det V^{\mathrm{odd}})'$. Here $V^{\mathrm{even}} := \bigoplus_q V^{2q}$ and $V^{\mathrm{odd}} := \bigoplus_q V^{2q+1}$ are considered as ungraded vector spaces and $V' := L(V; \mathbb{C})$ denotes the dual space. For more details on determinant lines consult for instance [24]. Let us only mention that every short exact sequence of graded vector spaces $0 \to U^* \to V^* \to W^* \to 0$ provides a canonic isomorphism of determinant lines $\det U^* \otimes \det W^* = \det V^*$. The complex C^* gives rise to two short exact sequences

$$0 \to B^* \to Z^* \to H(C^*) \to 0 \quad \text{and} \quad 0 \to Z^* \to C^* \xrightarrow{d} B^{*+1} \to 0 \tag{14}$$

where B^* and Z^* denote the boundaries and cycles in C^*, respectively. The isomorphism (13) is then obtained from the isomorphisms of determinant lines induced by (14) together with the canonic isomorphism $\det B^* \otimes \det B^{*+1} = \det B^* \otimes (\det B^*)' = \mathbb{C}$.

Suppose our complex C^* is equipped with a *graded non-degenerate symmetric bilinear form* b. That is, we have a non-degenerate symmetric bilinear form on every homogeneous component C^q, and different homogeneous components are b-orthogonal. The bilinear form b will induce a non-degenerate bilinear form on $\det C^*$. Using (13) we obtain a non-degenerate bilinear form on $\det H(C^*)$ which is called the *torsion associated with* C^* *and* b. It will be denoted by $\tau_{C^*,b}$.

Remark 3.1. Note that a non-degenerate bilinear form on a complex line essentially is a non-vanishing complex number. If C^* happens to be acyclic, i.e. $H(C^*) = 0$, then canonically $\det H(C^*) = \mathbb{C}$ and $\tau_{C^*,b} \in \mathbb{C}^\times$ is a genuine non-vanishing complex number—the entry in the 1×1-matrix representing this bilinear form.

Example 3.2. Suppose $q \in \mathbb{Z}$, $n \in \mathbb{N}$ and $A \in \mathrm{GL}_n(\mathbb{C})$. Let C^* denote the acyclic complex $\mathbb{C}^n \xrightarrow{d=A} \mathbb{C}^n$ concentrated in degrees q and $q+1$. Let b denote the standard non-degenerate symmetric bilinear form on C^*. In this situation we have $\tau_{C^*,b} = (\det A)^{(-1)^{q+1}2} = (\det AA^t)^{(-1)^{q+1}}$.

The bilinear form b permits to define the *transposed* $d_b^\sharp$ of d

$$d_b^\sharp : C^{*+1} \to C^*, \quad b(dv, w) = b\big(v, d_b^\sharp w\big), \ v, w \in C^*.$$

Define the *Laplacian* $\Delta_b := dd_b^\sharp + d_b^\sharp d : C^* \to C^*$. Let us write $C_b^*(\lambda)$ for the generalized λ-eigen space of Δ_b. Clearly,

$$C^* = \bigoplus_\lambda C_b^*(\lambda). \tag{15}$$

Since Δ_b is symmetric with respect to b, different generalized eigen spaces of Δ are b-orthogonal. It follows that the restriction of b to $C_b^*(\lambda)$ is non-degenerate.

Since Δ_b commutes with d and $d_b^\sharp$ the latter two will preserve the decomposition (15). Hence every eigen space $C_b^*(\lambda)$ is a subcomplex of C^*. The inclusion $C_b^*(0) \to C^*$ induces an isomorphism in cohomology. Indeed, the Laplacian factors to an invertible map on $C^*/C_b^*(0)$ and thus induces an isomorphism on $H(C^*/C_b^*(0))$. On the other hand, the equation $\Delta_b = dd_b^\sharp + d_b^\sharp d$

tells that the Laplacian will induce the zero map on cohomology. Therefore $H(C^*/C_b^*(0))$ must vanish and $C_b^*(0) \to C^*$ is indeed a quasi isomorphism. Particularly, we obtain a canonic isomorphism of complex lines

$$\det H\big(C_b^*(0)\big) = \det H\big(C^*\big). \tag{16}$$

Lemma 3.3. *Suppose C^* is a finite-dimensional graded complex over $\mathbb{C}$ which is equipped with a graded non-degenerate symmetric bilinear form b. Then via* (16) *we have*

$$\tau_{C^*,b} = \tau_{C_b^*(0),b|_{C_b^*(0)}} \cdot \prod_q \big(\det{}'(\Delta_{b,q})\big)^{(-1)^q q}$$

where $\det'(\Delta_{b,q})$ *denotes the product over all non-vanishing eigen-values of the Laplacian acting in degree q, $\Delta_{b,q} := \Delta_b|_{C^q} : C^q \to C^q$.*

Proof. Suppose (C_1^*, b_1) and (C_2^*, b_2) are finite-dimensional complexes equipped with graded non-degenerate symmetric bilinear forms. Clearly, $H(C_1^* \oplus C_2^*) = H(C_1^*) \oplus H(C_2^*)$ and we obtain a canonic isomorphism of determinant lines

$$\det H\big(C_1^* \oplus C_2^*\big) = \det H\big(C_1^*\big) \otimes \det H\big(C_2^*\big).$$

It is not hard to see that via this identification we have

$$\tau_{C_1^* \oplus C_2^*, b_1 \oplus b_2} = \tau_{C_1^*, b_1} \otimes \tau_{C_2, b_2}. \tag{17}$$

In view of the b-orthogonal decomposition (15) we may therefore without loss of generality assume $\ker \Delta_b = 0$. Particularly, C^* is acyclic.

Then $\operatorname{img} d \cap \ker d_b^\sharp \subseteq \ker d \cap \ker d_b^\sharp \subseteq \ker \Delta_b = 0$. Since $\operatorname{img} d$ and $\ker d_b^\sharp$ are of complementary dimension we conclude $\operatorname{img} d \oplus \ker d_b^\sharp = C^*$. The acyclicity of C^* implies $\ker d_b^\sharp = \operatorname{img} d_b^\sharp$ and hence $\operatorname{img} d \oplus \operatorname{img} d_b^\sharp = C^*$. This decomposition is b-orthogonal and invariant under Δ_b. We obtain

$$\det{}'(\Delta_{b,q}) = \det(\Delta_{b,q}) = \det(\Delta_b|_{C^q \cap \operatorname{img} d}) \cdot \det(\Delta_b|_{C^q \cap \operatorname{img} d_b^\sharp}).$$

Since $d : C^q \cap \operatorname{img} d_b^\sharp \to C^{q+1} \cap \operatorname{img} d$ is an isomorphism commuting with Δ

$$\det(\Delta|_{C^q \cap \operatorname{img} d_b^\sharp}) = \det(\Delta|_{C^{q+1} \cap \operatorname{img} d}).$$

A telescoping argument then shows

$$\prod_q \big(\det{}'(\Delta_{b,q})\big)^{(-1)^q q} = \prod_q \det(\Delta_b|_{C^q \cap \operatorname{img} d})^{(-1)^q}. \tag{18}$$

On the other hand, the b-orthogonal decomposition of complexes

$$C^* = \bigoplus_q \big(C^q \cap \operatorname{img} d_b^\sharp \xrightarrow{d} C^{q+1} \cap \operatorname{img} d\big)$$

together with (17) and the computation in Example 3.2 imply

$$\tau_{C^*,b} = \prod_q \det\bigl(dd_b^\sharp|_{C^{q+1}\cap \operatorname{img} d}\bigr)^{(-1)^{q+1}}$$

which clearly coincides with (18) since $\Delta_b|_{\operatorname{img} d} = dd_b^\sharp|_{\operatorname{img} d}$. $\square$

Example 3.4. Suppose $0 \neq v \in \mathbb{C}^2$ satisfies $v^t v = 0$. Moreover, suppose $0 \neq z \in \mathbb{C}$ and set $w := zv^t$. Let C^* denote the acyclic complex $\mathbb{C} \xrightarrow{v} \mathbb{C}^2 \xrightarrow{w} \mathbb{C}$ concentrated in degrees 0, 1 and 2. Equip this complex with the standard symmetric bilinear form b. Then $\Delta_{b,0} = v^t v = 0$, $\Delta_{b,2} = ww^t = 0$, $\Delta_{b,1} = (1 + z^2)vv^t$, $(\Delta_{b,1})^2 = 0$. Thus all of this complex is contained in the generalized 0-eigen space of Δ_b. The torsion of the complex computes to $\tau_{C^*,b} = -z^2$. Observe that the kernel of Δ_b does not compute the cohomology; that the bilinear form becomes degenerate when restricted to the kernel of Δ_b; and that the torsion cannot be computed from the spectrum of Δ_b.

3.2. Morse complex

Let E be a flat complex vector bundle over a closed connected smooth manifold M of dimension n. Suppose $X = -\operatorname{grad}_g(f)$ is a *Morse–Smale vector field* on M, see [30]. Let $\mathcal{X}$ denote the zero set of X. Elements in $\mathcal{X}$ are called *critical points* of f. Every $x \in \mathcal{X}$ has a *Morse index* $\operatorname{ind}(x) \in \mathbb{N}$ which coincides with the dimension of the unstable manifold of x with respect to X. We will write $\mathcal{X}_q := \{x \in \mathcal{X} \mid \operatorname{ind}(x) = q\}$ for the set of critical points of index q.

Recall that the Morse–Smale vector field provides a *Morse complex* $C^*(X; E)$ with underlying finite-dimensional graded vector space

$$C^q(X; E) = \bigoplus_{x \in \mathcal{X}_q} E_x \otimes_{\{\pm 1\}} \mathcal{O}_x.$$

Here E_x denotes the fiber of E over x, and $\mathcal{O}_x$ denotes the set of orientations of the unstable manifold of x. The Smale condition tells that stable and unstable manifolds intersect transversally. It follows that for two critical points of index difference one there is only a finite number of unparametrized trajectories connecting them. The differential in $C^*(X; E)$ is defined with the help of these isolated trajectories and parallel transport in E along them.

Integration over unstable manifolds provides a homomorphism of complexes

$$\operatorname{Int}: \Omega^*(M; E) \to C^*(X; E) \tag{19}$$

where $\Omega^*(M; E)$ denotes the de Rham complex with values in E. It is a folklore fact that (19) induces an isomorphism on cohomology, see [30]. Particularly, we obtain a canonic isomorphism of complex lines

$$\det H^*(M; E) = \det H\bigl(C^*(X; E)\bigr). \tag{20}$$

Suppose $\chi(M) = 0$ and let $\mathfrak{e} \in \mathfrak{Eul}(M; \mathbb{Z})$ be an Euler structure. Choose a base point $x_0 \in M$. For every critical point $x \in \mathcal{X}$ choose a path σ_x with $\sigma(0) = x_0$ and $\sigma_x(1) = x$ so that $\mathfrak{e} = \bigl[-X, \sum_{x\in\mathcal{X}} (-1)^{\operatorname{ind}(x)} \sigma_x\bigr]$. This is possible in view of Lemma 2.1. Also note that

$\mathrm{IND}_{-X}(x) = (-1)^{\mathrm{ind}(x)}$. Choose a non-degenerate symmetric bilinear form b_{x_0} on the fiber E_{x_0} over x_0. For $x \in \mathcal{X}$ define a bilinear form b_x on E_x by parallel transport of b_{x_0} along σ_x. The collection of bilinear forms $\{b_x\}_{x\in\mathcal{X}}$ defines a non-degenerate symmetric bilinear form on the Morse complex $C^*(X;E)$. It is elementary to check that the induced bilinear form on $\det C^*(X;E)$ does not depend on the choice of $\{\sigma_x\}_{x\in\mathcal{X}}$, and because $\chi(M)=0$ it does not depend on x_0 or b_{x_0} either. Hence the corresponding torsion is a non-degenerate bilinear form on $\det H(C^*(X;E))$ depending on E, $\mathfrak{e}$ and X only. Using (20) we obtain a non-degenerate bilinear form on $\det H^*(M;E)$ which we will denote by $\tau^{\mathrm{comb}}_{E,\mathfrak{e},X}$. For the following non-trivial statement we refer to [27,34] or [25].

Theorem 3.5 *(Milnor, Turaev). The bilinear form $\tau^{\mathrm{comb}}_{E,\mathfrak{e},X}$ does not depend on X.*

In view of Theorem 3.5 we will denote $\tau^{\mathrm{comb}}_{E,\mathfrak{e},X}$ by $\tau^{\mathrm{comb}}_{E,\mathfrak{e}}$ from now on.

Definition 3.6 *(Combinatorial torsion).* The non-degenerate bilinear form $\tau^{\mathrm{comb}}_{E,\mathfrak{e}}$ on $\det H^*(M;E)$ is called the *combinatorial torsion* associated with the flat complex vector bundle E and the Euler structure $\mathfrak{e}\in\mathfrak{Eul}(M;\mathbb{Z})$.

Remark 3.7. The combinatorial torsion's dependence on the Euler structure is very simple. For $\mathfrak{e}\in\mathfrak{Eul}(M;\mathbb{Z})$ and $\sigma\in H_1(M;\mathbb{Z})$ we obviously have, see (11)

$$\tau^{\mathrm{comb}}_{E,\mathfrak{e}+\sigma} = \tau^{\mathrm{comb}}_{E,\mathfrak{e}}\cdot\theta_E(\sigma)^2.$$

The dependence on E, i.e. the dependence on the flat connection, is subtle and interesting. Let us only mention the following example.

Example 3.8 *(Torsion of mapping tori).* Consider a mapping torus

$$M = N\times[0,1]/_{(x,1)\sim(\varphi(x),0)}$$

where $\varphi: N\to N$ is a diffeomorphism. Let $\pi: M\to S^1=[0,1]/_{0\sim 1}$ denote the canonic projection. The set of vector fields which project to the vector field $-\frac{\partial}{\partial\theta}$ on S^1 is contractible and thus defines an Euler structure $\mathfrak{e}\in\mathfrak{Eul}(M;\mathbb{Z})$ represented by $[X,0]$ where X is any of these vector fields. Let $\tilde{E}^z$ denote the flat line bundle over S^1 with holonomy $z\in\mathbb{C}^\times$, i.e. $\theta_{\tilde{E}^z}: H_1(S^1;\mathbb{Z})=\mathbb{Z}\to\mathbb{C}^\times$, $\theta_{\tilde{E}^z}(k)=z^k$. Consider the flat line bundle $E^z := \pi^*\tilde{E}^z$ over M. It follows from the Wang sequence of the fibration $\pi: M\to S^1$ that for generic z we will have $H^*(M;E^z)=0$. In this case

$$\tau^{\mathrm{comb}}_{E^z,\mathfrak{e}} = \big(\zeta_\varphi(z)\big)^2$$

where

$$\begin{aligned}\zeta_\varphi(z) &= \exp\biggl(\sum_{k\geqslant 1}\mathrm{str}\bigl(H^*(N;\mathbb{Q})\xrightarrow{(\varphi^k)^*}H^*(N;\mathbb{Q})\bigr)\frac{z^k}{k}\biggr)\\ &= \mathrm{sdet}\bigl(H^*(N;\mathbb{C})\xrightarrow{1-z\varphi^*}H^*(N;\mathbb{C})\bigr)^{-1}\end{aligned}$$

denotes the Lefschetz zeta function of φ. Here we wrote str and sdet for the super trace and the super determinant, respectively. For more details and proofs we refer to [9,20].

Remark 3.9. Often the combinatorial torsion is considered as an element in (rather than a bilinear form on) $\det H^*(M;E)$. This element is one of the two unit vectors of $\tau^{\mathrm{comb}}_{E,\mathfrak{e}}$. It is a non-trivial task (and requires the choice of a homology orientation) to fix the sign, i.e. to describe which of the two unite vectors actually is the torsion [19]. Considering bilinear forms this sign issue disappears.

3.3. Basic properties of the combinatorial torsion

If E_1 and E_2 are two flat vector bundles over M then we have a canonic isomorphism $H^*(M;E_1\oplus E_2)=H^*(M;E_1)\oplus H^*(M;E_2)$ which induces a canonic isomorphism of complex lines $\det H^*(M;E_1\oplus E_2)=\det H^*(M;E_1)\otimes\det H^*(M;E_2)$. Via this identification we have

$$\tau^{\mathrm{comb}}_{E_1\oplus E_2,\mathfrak{e}}=\tau^{\mathrm{comb}}_{E_1,\mathfrak{e}}\otimes\tau^{\mathrm{comb}}_{E_2,\mathfrak{e}}. \tag{21}$$

This follows from $C^*(X;E_1\oplus E_2)=C^*(X;E_1)\oplus C^*(X;E_2)$ and (17).

If E' denotes the dual of a flat vector bundle E then Poincaré duality induces an isomorphism $H^*(M;E'\otimes\mathcal{O}_M)=H^{n-*}(M;E)'$ which induces a canonic isomorphism $\det H^*(M;E'\otimes\mathcal{O}_M)=(\det H^*(M;E))^{(-1)^{n+1}}$. Via this identification we have

$$\tau^{\mathrm{comb}}_{E'\otimes\mathcal{O}_M,\nu(\mathfrak{e})}=\bigl(\tau^{\mathrm{comb}}_{E,\mathfrak{e}}\bigr)^{(-1)^{n+1}} \tag{22}$$

where ν denotes the involution on $\mathfrak{Eul}(M;\mathbb{Z})$ discussed in Section 2. To see that use a Morse–Smale vector field X to compute $\tau^{\mathrm{comb}}_{E,\mathfrak{e}}$ and use the Morse–Smale vector field $-X$ to compute $\tau^{\mathrm{comb}}_{E'\otimes\mathcal{O}_M,\nu(\mathfrak{e})}$. Then there is an obvious isomorphism of complexes $C^*(-X;E'\otimes\mathcal{O}_M)=C^{n-*}(X;E)'$ which induces Poincaré duality on cohomology.

If V is a complex vector space let $\bar V$ denote the complex conjugate vector space. If b is a bilinear form on V let $\bar b$ denote the complex conjugate bilinear form on $\bar V$, that is $\bar b(v,w)=\overline{b(v,w)}$. Let $\bar E$ denote the complex conjugate of a flat vector bundle E. Then we have a canonic isomorphism $H^*(M;\bar E)=\overline{H^*(M;E)}$ which induces a canonic isomorphism of complex lines $\det H^*(M;\bar E)=\overline{\det H^*(M;E)}$. Via this identification we have

$$\tau^{\mathrm{comb}}_{\bar E,\mathfrak{e}}=\overline{\tau^{\mathrm{comb}}_{E,\mathfrak{e}}}. \tag{23}$$

This follows from $C^*(X;\bar E)=\overline{C^*(X;E)}$.

If V is a real vector space we let $V^{\mathbb{C}}:=V\otimes\mathbb{C}$ denote its complexification. If h is a real bilinear form on V we let $h^{\mathbb{C}}$ denote its complexification, more explicitly $h^{\mathbb{C}}(v_1\otimes z_1,v_2\otimes z_2)=h(v_1,v_2)z_1z_2$. If F is real flat vector bundle its torsion, defined analogously to the complex case, is a real non-degenerate bilinear form on $\det H^*(M;F)$. Let $F^{\mathbb{C}}=F\otimes\mathbb{C}$ denote the complexification of the flat vector bundle F. We have a canonic isomorphism $H^*(M;F^{\mathbb{C}})=H^*(M;F)^{\mathbb{C}}$ which induces a canonic isomorphism of complex lines $\det H^*(M;F^{\mathbb{C}})=(\det H^*(M;F))^{\mathbb{C}}$. Via this identification we have

$$\tau^{\mathrm{comb}}_{F^{\mathbb{C}},\mathfrak{e}}=\bigl(\tau^{\mathrm{comb}}_{F,\mathfrak{e}}\bigr)^{\mathbb{C}}. \tag{24}$$

This follows from $C^*(X; F^{\mathbb{C}}) = C^*(X; F)^{\mathbb{C}}$. Note that $\tau^{\mathrm{comb}}_{F,\mathfrak{e}}$ is positive definite.

4. Ray–Singer torsion

The analytic torsion defined below is an invariant associated to a closed connected smooth manifold M, a complex flat vector bundle E over M, a coEuler structure $\mathfrak{e}^*$ and a homotopy class $[b]$ of fiber wise non-degenerate symmetric bilinear forms on E. In the way considered below, this invariant is a non-degenerate symmetric bilinear form $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ on the complex line $\det H^*(M; E)$. If $H^*(M; E)$ vanishes, then $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ becomes a non-vanishing complex number.

Throughout this section M denotes a closed connected smooth manifold of dimension n. For simplicity we will also assume vanishing Euler–Poincaré characteristics, $\chi(M) = 0$. At the expense of a base point everything can easily be extended to the general situation, see [8–10] and Section 9.

4.1. Laplacians and spectral theory

Suppose M is a closed connected smooth manifold of dimension n. Let E be a flat vector bundle over M. We will denote the flat connection of E by ∇^E. Suppose there exists a *fiber wise non-degenerate symmetric bilinear* form b on E. Moreover, let g be a Riemannian metric on M. This permits to define a symmetric bilinear form $\beta_{g,b}$ on the space of E-valued differential forms $\Omega^*(M; E)$,

$$\beta_{g,b}(v, w) := \int_M v \wedge (\star_g \otimes b) w, \qquad v, w \in \Omega^*(M; E).$$

Here $\star_g \otimes b : \Omega^*(M; E) \to \Omega^{n-*}(M; E' \otimes \mathcal{O}_M)$ denotes the isomorphism induced by the Hodge star operator[2] $\star_g : \Omega^*(M; \mathbb{R}) \to \Omega^{n-*}(M; \mathcal{O}_M)$ and the isomorphism of vector bundles $b : E \to E'$. The wedge product is computed with respect to the canonic pairing of $E \otimes E' \to \mathbb{C}$.

Let $d_E : \Omega^*(M; E) \to \Omega^{*+1}(M; E)$ denote the de Rham differential. Let

$$d^\sharp_{E,g,b} : \Omega^{*+1}(M; E) \to \Omega^*(M; E)$$

denote its formal transposed with respect to $\beta_{g,b}$. A straightforward computation shows that $d^\sharp_{E,g,b} : \Omega^q(M; E) \to \Omega^{q-1}(M; E)$ is given by

$$d^\sharp_{E,g,b} = (-1)^q (\star_g \otimes b)^{-1} \circ d_{E' \otimes \mathcal{O}_M} \circ (\star_g \otimes b). \tag{25}$$

Define the *Laplacian* by

$$\Delta_{E,g,b} := d_E \circ d^\sharp_{E,g,b} + d^\sharp_{E,g,b} \circ d_E. \tag{26}$$

[2] The normalization of the Hodge star operator we are using is $\alpha_1 \wedge \star_g \alpha_2 = \langle \alpha_1, \alpha_2 \rangle_g \Omega_g$, where $\alpha_1, \alpha_2 \in \Omega(M; \mathbb{R})$, $\Omega_g \in \Omega^n(M; \mathcal{O}_M)$ denotes the volume density associated with g, and $\langle \alpha_1, \alpha_2 \rangle_g$ denotes the inner product on $\Lambda^* T^* M$ induced by g, see [23, Section 2.1]. Although we will frequently refer to [1] in the subsequent sections, the convention for the Hodge star operator we are using differs from the one in [1].

These are generalized Laplacians in the sense that their principal symbol coincides with the symbol of the Laplace–Beltrami operator.

In the next proposition we collect some well-known facts concerning the spectral theory of $\Delta_{E,g,b}$. For details we refer to [32], particularly Theorems 8.4 and 9.3 therein.

Proposition 4.1. *For the Laplacian $\Delta_{E,g,b}$ constructed above the following hold*:

(i) *The spectrum of $\Delta_{E,g,b}$ is discrete. For every $\theta > 0$ all but finitely many points of the spectrum are contained in the angle $\{z \in \mathbb{C} \mid -\theta < \arg(z) < \theta\}$.*
(ii) *If λ is in the spectrum of $\Delta_{E,g,b}$ then the image of the associated spectral projection is finite-dimensional and contains smooth forms only. We will refer to this image as the (generalized) λ-eigen space of $\Delta_{E,g,b}$ and denote it by $\Omega^*_{g,b}(M;E)(\lambda)$. There exists $N_\lambda \in \mathbb{N}$ such that*

$$(\Delta_{E,g,b} - \lambda)^{N_\lambda}|_{\Omega^*_{g,b}(M;E)(\lambda)} = 0.$$

We have a $\Delta_{E,g,b}$-invariant $\beta_{g,b}$-orthogonal decomposition

$$\Omega^*_{g,b}(M;E) = \Omega^*_{g,b}(M;E)(\lambda) \oplus \Omega^*_{g,b}(M;E)(\lambda)^{\perp_{\beta_{g,b}}}. \tag{27}$$

The restriction of $\Delta_{E,g,b} - \lambda$ to $\Omega^*_{g,b}(M;E)(\lambda)^{\perp_{\beta_{g,b}}}$ is invertible.
(iii) *The decomposition* (27) *is invariant under d_E and $d^\sharp_{E,g,b}$.*
(iv) *For $\lambda \neq \mu$ the eigen spaces $\Omega^*_{g,b}(M;E)(\lambda)$ and $\Omega^*_{g,b}(M;E)(\mu)$ are orthogonal with respect to $\beta_{g,b}$.*

In view of Proposition 4.1 the generalized 0-eigen space $\Omega^*_{g,b}(M;E)(0)$ is a finite-dimensional subcomplex of $\Omega^*(M;E)$. The inclusion

$$\Omega^*_{g,b}(M;E)(0) \to \Omega^*(M;E) \tag{28}$$

induces an isomorphism in cohomology. Indeed, in view of Proposition 4.1(ii) the Laplacian $\Delta_{E,g,b}$ induces an isomorphism on $\Omega^*_{g,b}(M;E)/\Omega^*_{g,b}(M;E)(0)$ and thus an isomorphism on $H(\Omega^*_{g,b}(M;E)/\Omega^*_{g,b}(M;E)(0))$. On the other hand, (26) tells that $\Delta_{E,g,b}$ induces 0 on cohomology, hence $H(\Omega^*_{g,b}(M;E)/\Omega^*_{g,b}(M;E)(0))$ must vanish and (28) is indeed a quasi-isomorphism. We obtain a canonic isomorphism of complex lines

$$\det H\big(\Omega^*_{g,b}(M;E)(0)\big) = \det H^*(M;E). \tag{29}$$

In view of Proposition 4.1(ii) the bilinear form $\beta_{g,b}$ restricts to a non-degenerate bilinear form on $\Omega^*_{g,b}(M;E)(0)$. Using the linear algebra discussed in Section 3 we obtain a non-degenerate bilinear form on $\det H(\Omega^*_{g,b}(M;E)(0))$. Via (29) this gives rise to a non-degenerate bilinear form on $\det H^*(M;E)$ which will be denoted by $\tau^{\mathrm{an}}_{E,g,b}(0)$.

Let $\Delta_{E,g,b,q}$ denote the Laplacian acting in degree q. Define the zeta regularized product of its non-vanishing eigen-values, as

$$\det{}'(\Delta_{E,g,b,q}) := \exp\left(-\frac{\partial}{\partial s}\Big|_{s=0} \operatorname{tr}\big(\big(\Delta_{E,g,b,q}|\Omega^q_{g,b}(M;E)(0)^{\perp_{\beta_{g,b}}}\big)^{-s}\big)\right).$$

Here the complex powers are defined with respect to any non-zero Agmon angle which avoids the spectrum of $\Delta_{E,g,b,q}|\Omega^q_{g,b}(M;E)(0)^{\perp_{\beta_{g,b}}}$, see Proposition 4.1(i). Recall that for $\Re(s)>n/2$ the operator $(\Delta_{E,g,b,q}|\Omega^q_{g,b}(M;E)(0)^{\perp_{\beta_{g,b}}})^{-s}$ is trace class. As a function in s this trace extends to a meromorphic function on the complex plane which is holomorphic at 0, see [31] or [32, Theorem 13.1]. It is clear from Proposition 4.1(i) that $\det'(\Delta_{E,g,b,q})$ does not depend on the Agmon angle used to define the complex powers.

Assume $\chi(M)=0$ and suppose $\alpha\in\Omega^{n-1}(M;\mathcal{O}^{\mathbb{C}}_M)$ such that $d\alpha=\mathrm{e}(g)$. Consider the non-degenerate bilinear form on $\det H^*(M;E)$ defined by, cf. (4),

$$\tau^{\mathrm{an}}_{E,g,b,\alpha}:=\tau^{\mathrm{an}}_{E,g,b}(0)\cdot\prod_q\bigl(\det{}'(\Delta_{E,g,b,q})\bigr)^{(-1)^q q}\cdot\exp\biggl(-2\int_M\omega_{E,b}\wedge\alpha\biggr).$$

In Section 6 we will provide a proof of the following result which can be interpreted as an anomaly formula for the complex-valued Ray–Singer torsion (1).

Theorem 4.2 *(Anomaly formula). Let M be a closed connected smooth manifold with vanishing Euler–Poincaré characteristics. Let E be a flat complex vector bundle over M. Suppose g_u is a smooth one-parameter family of Riemannian metrics on M, and $\alpha_u\in\Omega^{n-1}(M;\mathcal{O}^{\mathbb{C}}_M)$ is a smooth one-parameter family so that $[g_u,\alpha_u]$ represent the same coEuler structure in $\mathfrak{Eul}^*(M;\mathbb{C})$. Moreover, suppose b_u is a smooth one-parameter family of fiber wise non-degenerate symmetric bilinear forms on E. Then, as bilinear forms on* $\det H^*(M;E)$*, we have $\frac{\partial}{\partial u}\tau^{\mathrm{an}}_{E,g_u,b_u,\alpha_u}=0$.*

In view of Theorem 4.2 the bilinear form $\tau^{\mathrm{an}}_{E,g,b,\alpha}$ does only depend on the flat vector bundle E, the coEuler structure $\mathfrak{e}^*\in\mathfrak{Eul}^*(M;\mathbb{C})$ represented by (g,α), and the homotopy class $[b]$ of b. We will denote it by $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ from now on.

Definition 4.3 *(Analytic torsion).* The non-degenerate bilinear form $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ on $\det H^*(M;E)$ is called the *analytic torsion* associated to the flat complex vector bundle E, the coEuler structure $\mathfrak{e}^*\in\mathfrak{Eul}^*(M;\mathbb{C})$ and the homotopy class $[b]$ of fiber wise non-degenerate symmetric bilinear forms on E.

Remark 4.4. The analytic torsion's dependence on the coEuler structure is very simple. For $\mathfrak{e}^*\in\mathfrak{Eul}^*(M;\mathbb{C})$ and $\beta\in H^{n-1}(M;\mathcal{O}^{\mathbb{C}}_M)$ we obviously have:

$$\tau^{\mathrm{an}}_{E,\mathfrak{e}^*+\beta,[b]}=\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}\cdot\bigl(e^{\langle[\omega_{E,b}]\cup\beta,[M]\rangle}\bigr)^2.$$

Here $\langle[\omega_{E,b}]\cup\beta,[M]\rangle\in\mathbb{C}$ denotes the evaluation of $[\omega_{E,b}]\cup\beta\in H^n(M;\mathcal{O}^{\mathbb{C}}_M)$ on the fundamental class $[M]\in H_n(M;\mathcal{O}_M)$.

Remark 4.5. Recall from Section 2 that for odd n there is a canonic coEuler structure $\mathfrak{e}^*_{\mathrm{can}}\in\mathfrak{Eul}^*(M;\mathbb{C})$ given by $\mathfrak{e}^*_{\mathrm{can}}=[g,0]$. The corresponding analytic torsion is:

$$\tau^{\mathrm{an}}_{E,\mathfrak{e}^*_{\mathrm{can}},[b]}=\tau^{\mathrm{an}}_{E,g,b}(0)\cdot\prod_q\bigl(\det{}'(\Delta_{E,g,b,q})\bigr)^{(-1)^q q}.$$

Note however that in general this does depend on the homotopy class $[b]$, see for instance the computation for the circle in Section 5 below. This is related to the fact that $\mathfrak{e}^*_{\mathrm{can}}$ in general is not integral, cf. Remark 5.3 below.

4.2. Basic properties of the analytic torsion

Suppose E_1 and E_2 are two flat vector bundles with fiber wise non-degenerate symmetric bilinear forms b_1 and b_2. Via the canonic isomorphism of complex lines $\det H^*(M; E_1 \oplus E_2) = \det H^*(M; E_1) \otimes \det H^*(M; E_2)$ we have:

$$\tau^{\mathrm{an}}_{E_1\oplus E_2,\mathfrak{e}^*,[b_1\oplus b_2]} = \tau^{\mathrm{an}}_{E_1,\mathfrak{e}^*,[b_1]} \otimes \tau^{\mathrm{an}}_{E_2,\mathfrak{e}^*,[b_2]}. \tag{30}$$

For this note that via the identification $\Omega^*(M; E_1 \oplus E_2) = \Omega^*(M; E_1) \oplus \Omega^*(M; E_2)$ we have $\Delta_{E_1\oplus E_2,g,b_1\oplus b_2} = \Delta_{E_1,g,b_1} \oplus \Delta_{E_2,g,b_2}$, hence $\det'(\Delta_{E_1\oplus E_2,g,b_1\oplus b_2,q}) = \det'(\Delta_{E_1,g,b_1,q}) \times \det'(\Delta_{E_2,g,b_2,q})$. Moreover, recall (7) for the correction terms.

Suppose E' is the dual of a flat vector bundle E. Let b' denote the bilinear form on E' dual to the non-degenerate symmetric bilinear form b on E. The bilinear form b' induces a fiber wise non-degenerate symmetric bilinear form on the flat vector bundle $E' \otimes \mathcal{O}_M$ which will be denoted by b' too. Via the canonic isomorphism of complex lines $\det H^*(M; E' \otimes \mathcal{O}_M) = (\det H^*(M; E))^{(-1)^{n+1}}$ induced by Poincaré duality we have

$$\tau^{\mathrm{an}}_{E'\otimes\mathcal{O}_M,\nu(\mathfrak{e}^*),[b']} = \left(\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}\right)^{(-1)^{n+1}} \tag{31}$$

where ν denotes the involution introduced in Section 2. This follows from the fact that $\star_g \otimes b : \Omega^q(M; E) \to \Omega^{n-q}(M; E' \otimes \mathcal{O}_M)$ intertwines the Laplacians $\Delta_{E,g,b,q}$ and $\Delta_{E'\otimes\mathcal{O}_M,g,b',n-q}$, see (25). Therefore $\Delta_{E,g,b,q}$ and $\Delta_{E'\otimes\mathcal{O}_M,g,b',n-q}$ are isospectral and thus $\det'(\Delta_{E,g,b,q}) = \det'(\Delta_{E'\otimes\mathcal{O}_M,g,b',n-q})$. Here one also has to use $\prod_q (\det'(\Delta_{E,g,b,q}))^{(-1)^q} = 1$, and (8).

Let $\bar{E}$ denote the complex conjugate of a flat vector bundle E. Let $\bar{b}$ denote the complex conjugate of a fiber wise non-degenerate symmetric bilinear form on E. Via the canonic isomorphism of complex lines $\det H^*(M; \bar{E}) = \overline{\det H^*(M; E)}$ we obviously have

$$\tau^{\mathrm{an}}_{\bar{E},\bar{\mathfrak{e}}^*,[\bar{b}]} = \overline{\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}} \tag{32}$$

where $\mathfrak{e}^* \mapsto \bar{\mathfrak{e}}^*$ denotes the complex conjugation of coEuler structures introduced in Section 2. For this note that $\Delta_{\bar{E},g,\bar{b}} = \Delta_{E,g,b}$ but the spectrum of $\Delta_{\bar{E},g,\bar{b}}$ is complex conjugate to the spectrum of $\Delta_{E,g,b}$ and thus $\det'(\Delta_{\bar{E},g,\bar{b},q}) = \overline{\det'(\Delta_{E,g,b,q})}$. Also recall (9).

Suppose F is a flat real vector bundle over M. Let $\mathfrak{e}^* \in \mathfrak{Eul}(M; \mathbb{R})$ be a coEuler structure with real coefficients. Let h be a fiber wise non-degenerate symmetric bilinear form on F. Proceeding exactly as in the complex case we obtain a non-degenerate bilinear form $\tau^{\mathrm{an}}_{F,\mathfrak{e},[h]}$ on the real line $\det H^*(M; F)$. Note that although the Laplacians $\Delta_{F,g,h}$ need not be selfadjoint their spectra are invariant under complex conjugation and hence $\det'(\Delta_{F,g,h,q})$ will be real. Let $F^{\mathbb{C}}$ denote the complexification of the flat bundle F, and let $h^{\mathbb{C}}$ denote the complexification of h, a non-degenerate symmetric bilinear form on $F^{\mathbb{C}}$. Via the canonic isomorphism of complex lines $\det H^*(M; F^{\mathbb{C}}) = (\det H^*(M; F))^{\mathbb{C}}$ we have:

$$\tau^{\mathrm{an}}_{F^{\mathbb{C}},\mathfrak{e}^*,[h^{\mathbb{C}}]} = \left(\tau^{\mathrm{an}}_{F,\mathfrak{e}^*,[h]}\right)^{\mathbb{C}}. \tag{33}$$

For this note that via $\Omega^*(M;F^{\mathbb{C}})=\Omega^*(M;F)^{\mathbb{C}}$ we have $\Delta_{F^{\mathbb{C}},g,h^{\mathbb{C}}}=(\Delta_{F,g,h})^{\mathbb{C}}$ and thus $\det'(\Delta_{F^{\mathbb{C}},g,h^{\mathbb{C}},q})=\det'(\Delta_{F,g,h,q})$, and also recall (10). If n is odd, $H^*(M;F)=0$, and if h is positive definite, then $\tau^{\mathrm{an}}_{F,\mathfrak{e}^*_{\mathrm{can}},[h]}$ is the square of the analytic torsion considered in [29], see Remark 4.5.

Remark 4.6. Not every flat complex vector bundle E admits a fiber wise non-degenerate symmetric bilinear form b. However, since E is flat all rational Chern classes of E must vanish. Since M is compact, the Chern character induces an isomorphism on rational K-theory, and hence E is trivial in rational K-theory. Thus there exists $N\in\mathbb{N}$ so that $E^N=E\oplus\cdots\oplus E$ is a trivial vector bundle. Particularly, there exists a fiber wise non-degenerate bilinear form b on E^N. In view of (30) the non-degenerate bilinear form $(\tau^{\mathrm{an}}_{E^N,\mathfrak{e}^*,[b]})^{1/N}$ on $\det H^*(M;E)$ is a reasonable candidate for the analytic torsion of E. Note however, that this is only defined up to a root of unity.

4.3. Rewriting the analytic torsion

Instead of just treating the 0-eigen space by means of finite-dimensional linear algebra one can equally well do this with finitely many eigen spaces of $\Delta_{E,g,b}$. Proposition 4.7 below makes this precise. We will make use of this formula when computing the variation of the analytic torsion through a variation of g and b. This is necessary since the dimension of the 0-eigen space need not be locally constant through such a variation. Note that this kind of problem does not occur in the selfadjoint situation, i.e. when instead of a non-degenerate symmetric bilinear form we have a Hermitian structure.

Suppose γ is a simple closed curve around 0, avoiding the spectrum of $\Delta_{E,g,b}$. Let $\Omega^*_{g,b}(M;E)(\gamma)$ denote the sum of eigen spaces corresponding to eigen-values in the interior of γ. Using Proposition 4.1 we see that the inclusion $\Omega^*_{g,b}(M;E)(\gamma)\to\Omega^*(M;E)$ is a quasi isomorphism. We obtain a canonic isomorphism of determinant lines

$$\det H\big(\Omega^*_{g,b}(M;E)(\gamma)\big)=\det H^*(M;E). \tag{34}$$

Moreover, the restriction of $\beta_{g,b}$ to $\Omega^*_{g,b}(M;E)(\gamma)$ is non-degenerate. Hence the torsion provides us with a non-degenerate bilinear form on $\det H(\Omega^*_{g,b}(M;E)(\gamma))$ and via (34) we get a non-degenerate bilinear form $\tau^{\mathrm{an}}_{E,g,b}(\gamma)$ on $\det H^*(M;E)$. Moreover, introduce

$$\det{}^{\gamma}(\Delta_{E,g,b,q}):=\exp\bigg(-\frac{\partial}{\partial s}\bigg|_{s=0}\operatorname{tr}\Big(\big(\Delta_{E,g,b,q}\big|\Omega^q_{g,b}(M;E)(\gamma)^{\perp_{\beta_{g,b}}}\big)^{-s}\Big)\bigg),$$

the zeta regularized product of eigen-values in the exterior of γ.

Proposition 4.7. *In this situation, as bilinear forms on* $\det H^*(M;E)$*, we have*:

$$\tau^{\mathrm{an}}_{E,g,b}(0)\cdot\prod_q\big(\det{}'(\Delta_{E,g,b,q})\big)^{(-1)^q q}=\tau^{\mathrm{an}}_{E,g,b}(\gamma)\cdot\prod_q\big(\det{}^{\gamma}(\Delta_{E,g,b,q})\big)^{(-1)^q q}.$$

Proof. Let $C^*\subseteq\Omega^*_{g,b}(M;E)(\gamma)$ denote the sum of the eigen spaces of $\Delta_{E,g,b}$ corresponding to non-zero eigen-values in the interior of γ. Clearly, for every q we have

$$\det{}'(\Delta_{E,g,b,q})=\det(\Delta_{E,g,b,q}|_{C^q})\cdot\det{}^{\gamma}(\Delta_{E,g,b,q}).$$

Particularly,

$$\prod_q \bigl(\det'(\Delta_{E,g,b,q})\bigr)^{(-1)^q q} = \prod_q \bigl(\det(\Delta_{E,g,b,q}|_{C^q})\bigr)^{(-1)^q q} \cdot \prod_q \bigl(\det^{\gamma}(\Delta_{E,g,b,q})\bigr)^{(-1)^q q}. \tag{35}$$

Applying Lemma 3.3 to the finite-dimensional complex $\Omega^*_{g,b}(M;E)(\gamma)$ we obtain

$$\tau^{\mathrm{an}}_{E,g,b}(\gamma) = \tau^{\mathrm{an}}_{E,g,b}(0) \cdot \prod_q \bigl(\det(\Delta_{E,g,b,q}|_{C^q})\bigr)^{(-1)^q q}. \tag{36}$$

Multiplying (35) with $\tau^{\mathrm{an}}_{E,g,b}(0)$ and using (36) we obtain the statement. □

5. A Bismut–Zhang, Cheeger, Müller type formula

The conjecture below asserts that the complex-valued analytical torsion defined in Section 4 coincides with the combinatorial torsion from Section 3. It should be considered as a complex-valued version of a theorem of Cheeger [16,17], Müller [28] and Bismut–Zhang [2].

Conjecture 5.1. *Let M be a closed connected smooth manifold with vanishing Euler–Poincaré characteristics. Let E be a flat complex vector bundle over M, and suppose b is a fiber wise non-degenerate symmetric bilinear form on E. Let $\mathfrak{e} \in \mathfrak{Eul}(M;\mathbb{Z})$ be an Euler structure. Then, as bilinear forms on the complex line* $\det H^*(M;E)$*, we have*:

$$\tau^{\mathrm{comb}}_{E,\mathfrak{e}} = \tau^{\mathrm{an}}_{E,P(\mathfrak{e}),[b]}.$$

Here we slightly abuse notation and let P also denote the composition $\mathfrak{Eul}(M;\mathbb{Z}) \to \mathfrak{Eul}(M;\mathbb{C}) \xrightarrow{P} \mathfrak{Eul}^*(M;\mathbb{C})$, see Section 2.

We will establish this conjecture in several special cases, see Remark 5.8, Theorem 5.10, Corollaries 5.13, 5.14 and the discussion for the circle below. Some of these results have been established by Braverman, Kappeler [7] and were not contained in the first version of this manuscript. The proofs we provide below have been inspired by a trick used in [7] but do not rely on the results therein.

Remark 5.2. If Conjecture 5.1 holds for one Euler structure $\mathfrak{e} \in \mathfrak{Eul}(M;\mathbb{Z})$ then it will hold for all Euler structures. This follows immediately from Remarks 4.4, 3.7 and Lemma 2.2.

Remark 5.3. If Conjecture 5.1 holds, and if $\mathfrak{e}^* \in \mathfrak{Eul}^*(M;\mathbb{C})$ is integral, then $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ is independent of $[b]$. This is not obvious from the definition of the analytic torsion.

Remark 5.4. If Conjecture 5.1 holds, $\mathfrak{e} \in \mathfrak{Eul}(M;\mathbb{Z})$ and $\mathfrak{e}^* \in \mathfrak{Eul}^*(M;\mathbb{C})$ then:

$$\tau^{\mathrm{comb}}_{E,\mathfrak{e}} = \tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]} \cdot \bigl(e^{\langle [\omega_{E,b}] \cup (P(\mathfrak{e}) - \mathfrak{e}^*), [M]\rangle}\bigr)^2.$$

This follows from Remark 4.4.

Remark 5.5. If Conjecture 5.1 holds, and $\mathfrak{e}^* \in \mathfrak{Eul}^*(M;\mathbb{C})$, then $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ does only depend on E, $\mathfrak{e}^*$ and the induced homotopy class $[\det b]$ of non-degenerate bilinear forms on $\det E$. This follows from Remark 5.4 and the fact that the cohomology class $[\omega_{E,b}]$ does depend on $\det E$ and the homotopy class $[\det b]$ on $\det E$ only, see Section 2.

5.1. Relative torsion

In the situation above, consider the non-vanishing complex number

$$\mathcal{S}_{E,\mathfrak{e},[b]} := \frac{\tau^{\mathrm{an}}_{E,P(\mathfrak{e}),[b]}}{\tau^{\mathrm{comb}}_{E,\mathfrak{e}}} \in \mathbb{C}^\times.$$

It follows from Remarks 4.4, 3.7 and Lemma 2.2 that this does not depend on $\mathfrak{e} \in \mathfrak{Eul}(M;\mathbb{Z})$. We will thus denote it by $\mathcal{S}_{E,[b]}$. The number $\mathcal{S}_{E,[b]}$ will be referred to as the *relative torsion* associated with the flat complex vector bundle E and the homotopy class $[b]$. Conjecture 5.1 asserts that $\mathcal{S}_{E,[b]} = 1$.

Similarly, if F is a real flat vector bundle over M equipped with a fiber wise non-degenerate symmetric bilinear form h, we set

$$\mathcal{S}_{F,[h]} := \frac{\tau^{\mathrm{an}}_{F,P(\mathfrak{e}),[h]}}{\tau^{\mathrm{comb}}_{F,\mathfrak{e}}} \in \mathbb{R}^\times := \mathbb{R}\setminus\{0\}$$

where $\mathfrak{e} \in \mathfrak{Eul}(M;\mathbb{Z})$ is any Euler structure. The combinatorial torsion $\tau^{\mathrm{comb}}_{F,\mathfrak{e}}$ and the analytic torsion $\tau^{\mathrm{an}}_{F,P(\mathfrak{e}),[h]}$ on $\det H^*(M;F)$ have been introduced in Sections 3 and 4, respectively. It follows via complexification from the corresponding statements for complex vector bundles that this does indeed only depend on F and the homotopy class of h, see (24) and (33).

Remark 5.6. If F is a flat real vector bundle equipped with a positive definite symmetric bilinear form h, then the Bismut–Zhang theorem [2, Theorem 0.2] asserts that $\mathcal{S}_{F,[h]} = 1$. This follows from the formula in Proposition 5.11 below (applied to a simple closed curve whose interior contains the eigen-value 0 only) which, via complexification, provides an analogous formula for flat real vector bundles. For the relation of the first factor in this formula with the statement in [2, Theorem 0.2] see (45).

Proposition 5.7. *The following properties hold*:

(i) $\mathcal{S}_{E_1\oplus E_2,[b_1\oplus b_2]} = \mathcal{S}_{E_1,[b_1]}\cdot\mathcal{S}_{E_2,[b_2]}$;
(ii) $\mathcal{S}_{E'\otimes\mathcal{O}_M,[b']} = (\mathcal{S}_{E,[b]})^{(-1)^{n+1}}$;
(iii) $\mathcal{S}_{\bar E,[\bar b]} = \overline{\mathcal{S}_{E,[b]}}$;
(iv) $\mathcal{S}_{F^{\mathbb{C}},[h^{\mathbb{C}}]} = \mathcal{S}_{F,[h]}$.

Proof. This follows immediately from the basic properties of analytic and combinatorial torsion discussed in Sections 3 and 4. For (ii) and (iii) one also has to use $P(\nu(\mathfrak{e})) = \nu(P(\mathfrak{e}))$ and $\overline{P(\mathfrak{e})} = P(\bar{\mathfrak{e}})$, see Section 2. $\square$

Remark 5.8. Proposition 5.7(ii) permits to verify Conjecture 5.1, up to sign, for even-dimensional orientable manifolds and parallel bilinear forms. More precisely, let M be an even-dimensional closed connected orientable smooth manifold with vanishing Euler–Poincaré characteristics. Let E be a flat complex vector bundle over M and suppose b is a parallel fiber wise non-degenerate symmetric bilinear form on E. Let $\mathfrak{e} \in \mathfrak{Eul}(M;\mathbb{Z})$ be an Euler structure. Then

$$\tau^{\mathrm{comb}}_{E,\mathfrak{e}} = \pm\tau^{\mathrm{an}}_{E,P(\mathfrak{e}),[b]} \tag{37}$$

i.e. in this situation Conjecture 5.1 holds up to sign. To see this, note that the parallel bilinear form b and the choice of an orientation provides an isomorphism of flat vector bundles $b : E \to E' \otimes \mathcal{O}_M$ which maps b to b'. Thus $\mathcal{S}_{E'\otimes\mathcal{O}_M,[b']} = \mathcal{S}_{E,[b]}$. Combining this with Proposition 5.7(ii) we obtain $(\mathcal{S}_{E,[b]})^2 = 1$, and hence (37). Note, however, that in this situation the arguments used to establish (31) immediately yield

$$\prod_q \bigl(\det{}'(\Delta_{E,g,b,q})\bigr)^{(-1)^q q} = 1.$$

Corollary 5.9 below has been established by Braverman and Kappeler see [7, Theorem 5.3] by comparing $\tau^{\mathrm{an}}_{E,P(\mathfrak{e}),[b]}$ with their refined analytic torsion, see [7, Theorem 1.4]. We will give an elementary proof relying on Proposition 5.7 and a trick similar to the one used in the proof of Theorem 1.4 in [7].

Corollary 5.9. *Let M be a closed connected smooth orientable manifold of odd dimension. Suppose E is a flat complex vector bundle over M equipped with a non-degenerate symmetric bilinear form b. Let $\mathfrak{e}^* \in \mathfrak{Eul}^*(M;\mathbb{C})$ be an integral coEuler structure. Then, up to sign, $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ is independent of* $[b]$*, cf. Remark* 5.3.

Proof. It suffices to show $(\mathcal{S}_{E,[b]})^2$ is independent of $[b]$. The choice of an orientation provides an isomorphism of flat vector bundles $E' \cong E' \otimes \mathcal{O}_M$ from which we obtain

$$\mathcal{S}_{E',[b']} = \mathcal{S}_{E'\otimes\mathcal{O}_M,[b']} = \mathcal{S}_{E,[b]}$$

where the latter equality follows from Proposition 5.7(ii). Together with Proposition 5.7(i) we thus obtain

$$(\mathcal{S}_{E,[b]})^2 = \mathcal{S}_{E,[b]} \cdot \mathcal{S}_{E',[b']} = \mathcal{S}_{E\oplus E',[b\oplus b']}. \tag{38}$$

Observe that on $E \oplus E'$ there exists a canonic (independent of b) symmetric non-degenerate bilinear form b_{can} defined by

$$b_{\mathrm{can}}\bigl((x_1,\alpha_1),(x_2,\alpha_2)\bigr) := \alpha_1(x_2) + \alpha_2(x_1), \quad x_1, x_2 \in E,\ \alpha_1, \alpha_2 \in E'.$$

This bilinear form b_{can} is homotopic to $b \oplus b'$, and thus

$$\mathcal{S}_{E\oplus E',[b\oplus b']} = \mathcal{S}_{E\oplus E',[b_{\mathrm{can}}]}.$$

Hence $\mathcal{S}_{E\oplus E',[b\oplus b']}$ does not depend on $[b]$. In view of (38) the same holds for $(\mathcal{S}_{E,[b]})^2$, and the proof is complete.

To see that $b\oplus b'$ is indeed homotopic to b_{can} let us consider b as an isomorphism $b : E \to E'$. For $t \in \mathbb{R}$ consider the endomorphisms

$$\Phi_t \in \operatorname{end}(E \oplus E'), \quad \Phi_t := \begin{pmatrix} \mathrm{id}_E \cos t & -b^{-1}\sin t \\ b \sin t & \mathrm{id}_{E'} \cos t \end{pmatrix}.$$

From $\Phi_{t+s} = \Phi_t \Phi_s$ we conclude that every Φ_t is invertible. Consider the curve of non-degenerate symmetric bilinear forms

$$b_t := \Phi_t^* b_{\mathrm{can}}, \quad b_t(X_1, X_2) = b_{\mathrm{can}}(\Phi_t X_1, \Phi_t X_2), \quad X_1, X_2 \in E \oplus E'.$$

Then clearly $b_0 = b_{\mathrm{can}}$. An easy calculation shows $b_{\pi/4} = b \oplus (-b')$. Clearly, $b \oplus (-b')$ is homotopic to $b \oplus b'$. So we see that b_{can} is homotopic to $b \oplus b'$. □

Using a result of Cheeger [16,17], Müller [28] and Bismut–Zhang [2] we will next show that the absolute value of the relative torsion is always one. In odd dimensions this has been established by Braverman and Kappeler, see [7, Theorem 1.10]. We will again use a trick similar to the one in [7].

Theorem 5.10. *Suppose M is a closed connected smooth manifold with vanishing Euler–Poincaré characteristics. Let E be a flat complex vector bundle over M equipped with a non-degenerate symmetric bilinear form b. Then $|\mathcal{S}_{E,[b]}| = 1$.*

Proof. Note first that in view of Proposition 5.7(iii) and (i) we have

$$|\mathcal{S}_{E,[b]}|^2 = \mathcal{S}_{E,[b]} \cdot \overline{\mathcal{S}_{E,[b]}} = \mathcal{S}_{E\oplus\bar{E},[b\oplus\bar{b}]}. \tag{39}$$

Set $k := \operatorname{rank} E$, and observe that b provides a reduction of the structure group of E to $O_k(\mathbb{C})$. Since the inclusion $O_k(\mathbb{R}) \subseteq O_k(\mathbb{C})$ is a homotopy equivalence, the structure group can thus be further reduced to $O_k(\mathbb{R})$. In other words, there exists a complex anti-linear involution $\nu : E \to E$ such that

$$\nu^2 = \mathrm{id}_E, \qquad b(\nu x, y) = \overline{b(x, \nu y)}, \quad b(x, \nu x) \geqslant 0,\ x, y \in E.$$

Then

$$\mu : E \otimes E \to \mathbb{C}, \quad \mu(x, y) := b(x, \nu y)$$

is a fiber wise positive definite Hermitian structure on E, anti-linear in the second variable. Define a non-degenerate symmetric bilinear form b^μ on $E \oplus \bar{E}$ by

$$b^\mu\big((x_1, y_1), (x_2, y_2)\big) := \mu(x_1, y_2) + \mu(x_2, y_1).$$

We claim that the symmetric bilinear form b^μ is homotopic to $b \oplus \bar{b}$. To see this, consider $\nu : E \to \bar{E}$ as a complex linear isomorphism. For $t \in \mathbb{R}$, define

$$\Phi_t \in \operatorname{end}(E \oplus \bar{E}), \quad \Phi_t := \begin{pmatrix} \mathrm{id}_E \cos t & -\nu^{-1} \sin t \\ \nu \sin t & \mathrm{id}_{\bar{E}} \cos t \end{pmatrix}.$$

From $\Phi_{t+s} = \Phi_t \Phi_s$ we conclude that every Φ_t is invertible. Consider the curve of non-degenerate symmetric bilinear forms

$$b_t := \Phi_t^* b^\mu, \qquad b_t(X_1, X_2) = b^\mu(\Phi_t X_1, \Phi_t X_2), \quad X_1, X_2 \in E \oplus \bar{E}.$$

Clearly, $b_0 = b^\mu$. An easy computation shows $b_{\pi/4} = b \oplus (-\bar{b})$. Since $b \oplus (-\bar{b})$ is homotopic to $b \oplus \bar{b}$ we see that b^μ is indeed homotopic to $b \oplus \bar{b}$. Together with (39) we conclude

$$|\mathcal{S}_{E,[b]}|^2 = \mathcal{S}_{E \oplus \bar{E}, b^\mu}. \tag{40}$$

Next, recall that there is a canonic isomorphism of flat vector bundles

$$\psi\colon\ E^{\mathbb{C}} \cong E \oplus \bar{E}, \quad \psi(x + \mathbf{i}y) := (x + \mathbf{i}y, x - \mathbf{i}y), \quad x, y \in E.$$

Consider the fiber wise positive definite symmetric real bilinear form $h := \Re\mu$ on $E^{\mathbb{R}}$, the underlying real vector bundle. Its complexification $h^{\mathbb{C}}$ is a non-degenerate symmetric bilinear form on $E^{\mathbb{C}}$. A simple computations shows $\psi^* b^\mu = 2h^{\mathbb{C}}$. Together with (40) we obtain

$$|\mathcal{S}_{E,[b]}|^2 = \mathcal{S}_{E^{\mathbb{C}}, 2h^{\mathbb{C}}} = \mathcal{S}_{E^{\mathbb{R}}, 2h}$$

where the last equation follows from Proposition 5.7(iv). The Bismut–Zhang theorem [2, Theorem 0.2] asserts that $\mathcal{S}_{E^{\mathbb{R}}, 2h} = 1$, see Remark 5.6, and the proof is complete. $\square$

5.2. Analyticity of the relative torsion

In this section we will show that the relative torsion $\mathcal{S}_{E,[b]}$ depends holomorphically on the flat connection, see Proposition 5.12 below. Combined with Theorem 5.10 this implies that $\mathcal{S}_{E,[b]}$ is locally constant on the space of flat connections on a fixed vector bundle, see Corollary 5.13 below. We start by establishing an explicit formula for the relative torsion, see Proposition 5.11.

Suppose $f : C_1 \to C_2$ is a homomorphism of finite-dimensional complexes. Consider the mapping cone $C_2^{*-1} \oplus C_1^*$ with differential $\left(\begin{smallmatrix} d & f \\ 0 & -d \end{smallmatrix}\right)$. If C_1^* and C_2^* are equipped with graded non-degenerate symmetric bilinear forms b_1 and b_2 we equip the mapping cone with the bilinear form $b_2 \oplus b_1$. The resulting torsion $\tau(f, b_1, b_2) := \tau_{C_2^{*-1} \oplus C_1^*, b_2 \oplus b_1}$ is called the relative torsion of f. It is a non-degenerate bilinear form on the determinant line $\det H(C_2^{*-1} \oplus C_1^*)$. Recall that if f is a quasi isomorphism then $C_2^{*-1} \oplus C_1^*$ is acyclic and

$$\tau(f, b_1, b_2) = \frac{(\det H(f))(\tau_{C_1^*, b_1})}{\tau_{C_2^*, b_2}} \tag{41}$$

where $\det H(f) : \det H(C_1^*) \to \det H(C_2^*)$ denotes the isomorphism of complex lines induces from the isomorphism in cohomology $H(f) : H(C_1^*) \to H(C_2^*)$.

Let us apply this to the integration homomorphism

$$\mathrm{Int}\colon \Omega^*_{g,b}(M;E)(\gamma)\to C^*(X;E) \tag{42}$$

where the notation is as in Proposition 4.7. Equip $\Omega^*_{g,b}(M;E)(\gamma)$ with the restriction of $\beta_{g,b}$, and equip $C^*(X;E)$ with the bilinear form $b|_{\mathcal{X}}$ obtained by restricting b to the fibers over $\mathcal{X}$. Since (42) is a quasi isomorphism the mapping cone is acyclic and the corresponding relative torsion is a non-vanishing complex number we will denote by

$$\tau\big(\Omega^*_{g,b}(M;E)(\gamma)\xrightarrow{\mathrm{Int}} C^*_b(X;E)\big)\in\mathbb{C}^\times.$$

Proposition 5.11. *Let M be a closed connected smooth manifold with vanishing Euler–Poincaré characteristics. Let E be a flat complex vector bundle over M. Let g be a Riemannian metric, and let X be a Morse–Smale vector field on M. Suppose b is a fiber wise non-degenerate symmetric bilinear form on E which is parallel in a neighborhood of the critical points $\mathcal{X}$. Moreover, let γ be a simple closed curve around 0 which avoids the spectrum of $\Delta_{E,g,b}$. Then*:

$$\begin{aligned}\mathcal{S}_{E,[b]}&=\tau\big(\Omega^*_{g,b}(M;E)(\gamma)\xrightarrow{\mathrm{Int}} C^*_b(X;E)\big)\\ &\quad\times\prod_q\big(\det{}^\gamma(\Delta_{E,g,b,q})\big)^{(-1)^q q}\cdot\exp\bigg(-2\int_{M\setminus\mathcal{X}}\omega_{E,b}\wedge(-X)^*\Psi(g)\bigg).\end{aligned}$$

The integral is absolutely convergent since $\omega_{E,b}$ vanishes in a neighborhood of $\mathcal{X}$.

Proof. Let $x_0\in M$ be a base point. For every critical point $x\in\mathcal{X}$ choose a path σ_x with $\sigma_x(0)=x_0$ and $\sigma_x(1)=x$. Set $c:=\sum_{x\in\mathcal{X}}(-1)^{\mathrm{ind}(x)}\sigma_x$ and consider the Euler structure $\mathfrak{e}:=[-X,c]\in\mathfrak{Eul}(M;\mathbb{Z})$. For the dual coEuler structure $P(\mathfrak{e})=[g,\alpha]$ we have, see (3),

$$\int_{M\setminus\mathcal{X}}\omega_{E,b}\wedge\big((-X)^*\Psi(g)-\alpha\big)=\int_c\omega_{E,b}. \tag{43}$$

Let b_{x_0} denote the bilinear form on the fiber E_{x_0} obtained by restricting b. For $x\in\mathcal{X}$ let $\tilde b_x$ denote the bilinear form obtained from b_{x_0} by parallel transport along σ_x. Let $\tilde b_{\det C^*(X;E)}$ denote the induced bilinear form on $\det C^*(X;E)$. This is the bilinear form used in the definition of the combinatorial torsion. We want to compare it with the bilinear form $b_{\det C^*(X;E)}$ on $\det C^*(X;E)$ induced by the restriction $b|_{\mathcal{X}}$ of b to the fibers over $\mathcal{X}$. A simple computation similar to the proof of Lemma 2.2 yields

$$\tilde b_{\det C^*(X;E)}=\exp\bigg(2\int_c\omega_{E,b}\bigg)\cdot b_{\det C^*(X;E)}. \tag{44}$$

Let $\tau_{C^*(X;E),b|_{\mathcal{X}}}$ denote the non-degenerate bilinear form on $\det H^*(M;E)$ obtained from the torsion of the complex $C^*(X;E)$ equipped with the bilinear form $b|_{\mathcal{X}}$ via the isomorphism $\det H^*(M;E)=\det H(C^*(X;E))$, see (19) and (20). Then, using (41),

$$\frac{\tau^{\mathrm{an}}_{E,g,b}(\gamma)}{\tau_{C^*(X;E),b|_{\mathcal{X}}}}=\tau\big(\Omega^*_{g,b}(M;E)(\gamma)\xrightarrow{\mathrm{Int}} C^*_b(X;E)\big). \tag{45}$$

Moreover, (44) implies

$$\tau^{\mathrm{comb}}_{E,\mathfrak{e}} = \tau_{C^*(X;E),b|_{\mathcal{X}}} \cdot \exp\biggl(2\int_c \omega_{E,b}\biggr). \tag{46}$$

From Proposition 4.7 we obtain

$$\tau^{\mathrm{an}}_{E,P(\mathfrak{e}),[b]} = \tau^{\mathrm{an}}_{E,g,b}(\gamma)\cdot\prod_q\bigl(\det{}^{\gamma}(\Delta_{E,g,b,q})\bigr)^{(-1)^q q}\cdot\exp\biggl(-2\int_M \omega_{E,b}\wedge\alpha\biggr). \tag{47}$$

Combining (43), (45)–(47) we obtain the statement of the proposition. □

Consider an open subset $U \subseteq \mathbb{C}$ and a family of flat complex vector bundles $\{E^z\}_{z\in U}$. Such a family is called *holomorphic* if the underlying vector bundles are the same for all $z \in U$ and the mapping $z \mapsto \nabla^{E^z}$ is holomorphic into the affine Fréchet space of linear connections equipped with the C^∞-topology.

Proposition 5.12. *Let M be a closed connected smooth manifold with vanishing Euler–Poincaré characteristics. Let $\{E^z\}_{z\in U}$ be a holomorphic family of flat complex vector bundles over M, and let b^z be a holomorphic family of fiber wise non-degenerate symmetric bilinear forms on E^z. Then $\mathcal{S}_{E^z,[b^z]}$ depends holomorphically on z.*

Proof. Let X be a Morse–Smale vector field on M. Let g be a Riemannian metric on M. In view of Theorem 4.2 we may without loss of generality assume $\nabla^{E^z} b^z = 0$ in a neighborhood of $\mathcal{X}$. Without loss of generality we may assume that there exists a simple closed curve γ around 0 so that the spectrum of Δ_{E^z,g,b^z} avoids γ for all $z \in U$. From Proposition 5.11 we know:

$$\begin{aligned}\mathcal{S}_{E^z,[b^z]} = {} & \tau\bigl(\Omega^*_{g,b^z}(M;E^z)(\gamma) \xrightarrow{\mathrm{Int}} C^*_{b^z}(X;E^z)\bigr)\\ & \times\prod_q\bigl(\det{}^{\gamma}(\Delta_{E^z,g,b^z,q})\bigr)^{(-1)^q q}\cdot\exp\biggl(-2\int_{M\setminus\mathcal{X}}\omega_{E^z,b^z}\wedge(-X)^*\Psi(g)\biggr).\end{aligned}$$

Since Δ_{E^z,g,b^z} depends holomorphically on z, each of the three factors in this expression for $S_{E^z,[b^z]}$ will depend holomorphically on z too. □

In odd dimensions the following result has been established by Braverman and Kappeler, see [7, Theorem 1.10].

Corollary 5.13. *Let M be a closed connected smooth manifold with vanishing Euler–Poincaré characteristics. Let E be a complex vector bundle over M, and let b be a fiber wise non-degenerate symmetric bilinear form on E. Then the assignment $\nabla \mapsto \mathcal{S}_{(E,\nabla),[b]}$ is locally constant, and of absolute value one, on the space of flat connections on E.*

Proof. Note that in view of Theorem 5.10 and Proposition 5.12 the relative torsion $\mathcal{S}_{(E,\nabla^z),[b]}$ is constant along every holomorphic path of flat connections $z \mapsto \nabla^z$ on E. Moreover, note that two flat connections, contained in the same connected component, can always be joined by a piecewise holomorphic path of flat connections. □

Using the Bismut–Zhang, Cheeger, Müller theorem again, we are able to verify Conjecture 5.1 for flat connections contained in particular connected components of the space of flat connections on a fixed complex vector bundle. More precisely, we have[3]

Corollary 5.14. *Let M be a closed connected smooth manifold with vanishing Euler–Poincaré characteristics. Let (F, ∇^F) be a flat real vector bundle over M equipped with a fiber wise Hermitian structure h. Let (E, ∇^E) denote the flat complex vector bundle obtained by complexifying (F, ∇^F), and let b denote the fiber wise non-degenerate symmetric bilinear form on E obtained by complexifying h. Then, for every flat connection ∇ on E which is contained in the connected component of ∇^E, we have $\mathcal{S}_{(E,\nabla),[b]} = 1$.*

Proof. In view of Corollary 5.13 it suffices to show $\mathcal{S}_{(E,\nabla^E),[b]} = 1$. From Proposition 5.7(iv) we have $\mathcal{S}_{(E,\nabla^E),[b]} = \mathcal{S}_{(F,\nabla^F),[h]}$. In view of [2, Theorem 0.2], see Remark 5.6, we indeed have $\mathcal{S}_{(F,\nabla^F),[h]} = 1$, and the statement follows. □

5.3. The circle, a simple explicit example

Consider $M := S^1$. In this case it is possible to explicitly compute the combinatorial and analytic torsion, see below. It turns out that Conjecture 5.1 holds true for every flat vector bundle over the circle.

We think of S^1 as $\{z \in \mathbb{C} \mid |z| = 1\}$. Equip S^1 with the standard Riemannian metric g of circumference 2π. Orient S^1 in the standard way. Let θ denote the angular 'coordinate.' Let $\frac{\partial}{\partial\theta}$ denote the corresponding vector field which is of length 1 and induces the orientation. For the dual 1-form we write $d\theta$.

Let $k \in \mathbb{N}$ and suppose $a \in C^\infty(S^1, \mathfrak{gl}_k(\mathbb{C}))$. Let E^a denote the trivial vector bundle $S^1 \times \mathbb{C}^k$ equipped with the flat connection $\nabla = \frac{\partial}{\partial\theta} + a$. Here and in what follows we use the identifications $\Omega^0(M; E^a) = C^\infty(S^1; \mathbb{C}^k) = \Omega^1(M; E^a)$ where the latter stems from the global coframe $d\theta$.

Let $b \in C^\infty(S^1, \mathrm{Sym}_k^\times(\mathbb{C}))$ where $\mathrm{Sym}_k^\times(\mathbb{C})$ denotes the space of complex non-degenerate symmetric $k \times k$-matrices. We consider b as a fiber wise non-degenerate symmetric bilinear form on E^a. For the induced bilinear form on $\Omega^*(S^1; E^a)$ we have:

$$\beta_{g,b}(v, w) = \int_{S^1} v^t b w \, d\theta, \quad v, w \in \Omega^0\big(S^1; E^a\big) = C^\infty\big(S^1, \mathbb{C}^k\big),$$

$$\beta_{g,b}(v, w) = \int_{S^1} v^t b w \, d\theta, \quad v, w \in \Omega^1\big(S^1; E^a\big) = C^\infty\big(S^1, \mathbb{C}^k\big).$$

A straightforward computations yields:

---

[3] In a recent preprint [22] R.-T. Huang verified a similar statement for flat connections whose connected component contains a flat connection which admits a parallel Hermitian structure.

$$
\begin{aligned}
d_{E^a} &= \frac{\partial}{\partial\theta} + a, \\
d^\sharp_{E^a,g,b} &= -\frac{\partial}{\partial\theta} - b^{-1}b' + b^{-1}a^t b, \\
\Delta_{E^a,g,b,0} &= -\left(\frac{\partial}{\partial\theta}\right)^2 + \left(b^{-1}a^t b - b^{-1}b' - a\right)\frac{\partial}{\partial\theta} + \left(b^{-1}a^t ba - b^{-1}b'a - a'\right), \\
\Delta_{E^a,g,b,1} &= -\left(\frac{\partial}{\partial\theta}\right)^2 + \left(b^{-1}a^t b - b^{-1}b' - a\right)\frac{\partial}{\partial\theta} \\
&\quad + \left(\left(b^{-1}a^t b\right)' - \left(b^{-1}b'\right)' - ab^{-1}b' + ab^{-1}a^t b\right), \\
b^{-1}\nabla_{\frac{\partial}{\partial\theta}} b &= b^{-1}b' - b^{-1}a^t b - a, \\
\omega_{E^a,b} &= -\frac{1}{2}\operatorname{tr}\left(b^{-1}b' - b^{-1}a^t b - a\right)d\theta = -\frac{1}{2}\left(\operatorname{tr}(b^{-1}b') - 2\operatorname{tr}(a)\right)d\theta.
\end{aligned}
$$

Here $b' := \frac{\partial}{\partial\theta} b$ and $a' := \frac{\partial}{\partial\theta} a$.

Let us write $A \in \mathrm{GL}_k(\mathbb{C})$ for the holonomy in E^a along the standard generator of $\pi_1(S^1)$. Recall that $\det A = \exp(\int_{S^1} \operatorname{tr}(a)\, d\theta)$. Using the explicit formula in [14, Theorem 1] we get:

$$
\begin{aligned}
\det(\Delta_{E^a,g,b,1}) &= \mathbf{i}^{2k} \exp\left(\frac{\mathbf{i}}{2}\int_{S^1} \operatorname{tr}\left(\mathbf{i}\left(b^{-1}a^t b - b^{-1}b' - a\right)\right) d\theta\right) \det\left(1 - \begin{pmatrix} A^{-1} & * \\ 0 & A^t \end{pmatrix}\right) \\
&= \exp\left(\frac{1}{2}\int_{S^1} \operatorname{tr}\left(b^{-1}b'\right) d\theta\right) \det(A-1)^2 \det A^{-1} \\
&= \exp\left(\frac{1}{2}\int_{S^1} \left(\operatorname{tr}\left(b^{-1}b'\right) - 2\operatorname{tr}(a)\right) d\theta\right) \det(A-1)^2.
\end{aligned}
$$

Consider the Euler structure $\mathfrak{e} := [-\frac{\partial}{\partial\theta}, 0] \in \mathfrak{Eul}(S^1;\mathbb{Z})$, and the coEuler structure $\mathfrak{e}^* := [g, \frac{1}{2}] \in \mathfrak{Eul}^*(S^1;\mathbb{C})$. Then $P(\mathfrak{e}) = \mathfrak{e}^*$, see (3). Assuming acyclicity, i.e. 1 is not an eigen-value of A, we conclude:

$$
\tau^{\mathrm{an}}_{E^a,\mathfrak{e}^*,[b]} = \det(A-1)^{-2}. \tag{48}
$$

Observe that this is independent of $[b]$, cf. Remark 5.3.

Considering a Morse–Smale vector field X with two critical points and the Euler structure $\mathfrak{e}$ we obtain a Morse complex $C^*(X;E^a)$ isomorphic to

$$
\mathbb{C}^k \xrightarrow{A-1} \mathbb{C}^k
$$

equipped with the standard bilinear form. From Example 3.2 we obtain

$$
\tau^{\mathrm{comb}}_{E^a,\mathfrak{e}} = \det(A-1)^{-2}
$$

which coincides with (48). So we see that $\tau^{\mathrm{comb}}_{E^a,\mathfrak{e}} = \tau^{\mathrm{an}}_{E^a,\mathfrak{e}^*,[b]}$, i.e. $\mathcal{S}_{E^a,[b]} = 1$, whenever E^a is acyclic. From Proposition 5.12 we conclude $\mathcal{S}_{E^a,[b]} = 1$ for all, not necessarily acyclic, E^a. Thus Conjecture 5.1 holds for $M = S^1$.

Remark 5.15. Recall the canonic coEuler structure $\mathfrak{e}^*_{\mathrm{can}} = [g, 0]$ defined as the unique fixed point of the involution on $\mathfrak{Eul}^*(S^1; \mathbb{C})$, see Section 2. Note that $\mathfrak{e}^*_{\mathrm{can}}$ is not integral. The computations above show that for the analytic torsion we have

$$\tau^{\mathrm{an}}_{E^a,\mathfrak{e}^*_{\mathrm{can}},[b]} = s_{[b]} \det A \det(A-1)^{-2}$$

where

$$s_{[b]} = \exp\left(-\frac{1}{2}\int_{S^1} \mathrm{tr}\left(b^{-1}b'\right)\right) \in \{\pm 1\}$$

does depend on b. Note that this sign $s_{[b]}$ appears, although we consider the torsion as a bilinear form, i.e. we essentially consider the square of what is traditionally called the torsion.

On odd-dimensional manifolds one often considers the analytic torsion without a correction term, i.e. one considers $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*_{\mathrm{can}},[b]}$. Let us give two reasons why this is not such a natural choice as it might seem. First, the celebrated fact that the Ray–Singer torsion on odd-dimensional manifolds does only depend on the flat connection, is no longer true in the complex setting as the appearance of the sign $s_{[b]}$ shows. Of course a different definition of complex-valued analytic torsion might circumvent this problem. More serious is the second point. One would expect that the analytic torsion as considered above is the square of a rational function on the space of acyclic representations of the fundamental group. As the computation for the circle shows, this cannot be true for $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*_{\mathrm{can}},[b]}$, simply because $\sqrt{\det A}$ cannot be rational in $A \in \mathrm{GL}_k(\mathbb{C})$. Any reasonable definition of complex-valued analytic torsion will have to face this problem.

If one is willing to consider $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ where $\mathfrak{e}^*$ is an integral coEuler structure both problems disappear, assuming E admits a non-degenerate symmetric bilinear form and Conjecture 5.1 is true. Then $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ is indeed independent of $[b]$, see Remark 5.3, and the dependence on $\mathfrak{e}^*$ is very simple, see Remark 4.4. More importantly, $\tau^{\mathrm{an}}_{E,\mathfrak{e}^*,[b]}$ is the square of a rational function on the space of acyclic representations of the fundamental group. This follows from the fact that $\tau^{\mathrm{comb}}_{E,\mathfrak{e}}$ with $P(\mathfrak{e}) = \mathfrak{e}^*$ is the square of such a rational function, see [9].

6. Proof of the anomaly formula

We continue to use the notation of Section 4. The proof of Theorem 4.2 is based on the following two results whose proof we postpone till Section 8.

Proposition 6.1. *Suppose $\phi \in \Gamma(\mathrm{end}(E))$. Then*

$$\operatorname*{LIM}_{t\to 0} \mathrm{str}\left(\phi e^{-t\Delta_{E,g,b}}\right) = \int_M \mathrm{tr}(\phi)\mathrm{e}(g).$$

Here LIM *denotes the renormalized limit, see* [1, Section 9.6]*, which in this case is actually an ordinary limit.*

Proposition 6.2. *Suppose $\xi \in \Gamma(\mathrm{end}(TM))$ is symmetric with respect to g, and let $\Lambda^*\xi \in \mathrm{end}(\Lambda^*T^*M)$ denote its extension to a derivation on Λ^*T^*M. Then*

$$\operatorname*{LIM}_{t\to 0}\operatorname{str}\left(\left(\Lambda^*\xi - \frac{1}{2}\operatorname{tr}(\xi)\right)e^{-t\Delta_{E,g,b}}\right) = \int_M \operatorname{tr}\bigl(b^{-1}\nabla^E b\bigr)\wedge(\partial_2\mathrm{cs})(g, g\xi).$$

Again LIM *denotes the renormalized limit, which in this case is just the constant term of the asymptotic expansion for $t \to 0$. Moreover, we use the notation $(\partial_2\mathrm{cs})(g, g\xi) := \frac{\partial}{\partial t}|_0\mathrm{cs}(g, g+tg\xi)$.*

Let us now give a proof of Theorem 4.2. Suppose g_u and b_u depend smoothly on a real parameter u. Let γ be a simple closed curve around 0 and assume without loss of generality that u varies in an open set U so that the spectrum of $\Delta_u := \Delta_{E,g_u,b_u}$ avoids the curve γ for all $u \in U$. Let Q_u denote the spectral projection onto the eigen spaces corresponding to eigen-values in the exterior of γ, and $Q_{u,q}$ the part acting in degree q. Let us write $\dot{\Delta}_u := \frac{\partial}{\partial u}\Delta_u$, and $\dot{\Delta}_{u,q}$ for the part acting in degree q. From the variation formula for the determinant of generalized Laplacians, see for instance [1, Proposition 9.38], we obtain

$$\begin{aligned}
\frac{\partial}{\partial u}\log\prod_q\bigl(\det{}^\gamma(\Delta_{u,q})\bigr)^{(-1)^q q} &= \sum_q(-1)^q q\left(\frac{\partial}{\partial u}\log\det{}^\gamma(\Delta_{u,q})\right)\\
&= \sum_q(-1)^q q\Bigl(\operatorname*{LIM}_{t\to 0}\operatorname{tr}\bigl(\dot{\Delta}_{u,q}(\Delta_{u,q})^{-1}Q_{u,q}e^{-t\Delta_{u,q}}\bigr)\Bigr)\\
&= \operatorname*{LIM}_{t\to 0}\operatorname{str}\bigl(N\dot{\Delta}_u\Delta_u^{-1}Q_u e^{-t\Delta_u}\bigr)
\end{aligned}\tag{49}$$

where N denotes the grading operator which acts by multiplication with q on $\Omega^q(M;E)$.

Choose $u_0 \in U$ and define $G_u \in \Gamma(\mathrm{Aut}(TM))$ by

$$g_u(a,b) = g_{u_0}(G_u a, b) = g_{u_0}(a, G_u b)$$

and similarly $B_u \in \Gamma(\mathrm{Aut}(E))$ by

$$b_u(e,f) = b_{u_0}(B_u e, f) = b_{u_0}(e, B_u f).$$

Let $\Lambda^*G_u^{-1}$ denote the natural extension of G_u^{-1} to $\Gamma(\mathrm{Aut}(\Lambda^*T^*M))$ and define

$$A_u = \det(G_u)^{1/2}\cdot\Lambda^*G_u^{-1}\otimes B_u \in \Gamma\bigl(\mathrm{Aut}\bigl(\Lambda^*T^*M\otimes E\bigr)\bigr).$$

Then

$$\beta_{g_u,b_u}(v,w) = \beta_{g_{u_0},b_{u_0}}(A_u v, w) = \beta_{g_{u_0},b_{u_0}}(v, A_u w), \quad v,w \in \Omega(M;E).\tag{50}$$

Abbreviating $d_u^\sharp := d_{E,g_u,b_u}^\sharp$ we immediately get

$$d_u^\sharp := A_u^{-1}d_{u_0}^\sharp A_u.$$

Writing $\dot{A}_u := \frac{\partial}{\partial u} A_u$, $\dot{g}_u := \frac{\partial}{\partial u} g_u$ and $\dot{b}_u := \frac{\partial}{\partial u} b_u$ we have

$$A_u^{-1}\dot{A}_u = \left(-\Lambda^*\big(g_u^{-1}\dot{g}_u\big) + \frac{1}{2}\operatorname{tr}\big(g_u^{-1}\dot{g}_u\big)\right) \otimes 1 + 1 \otimes \big(b_u^{-1}\dot{b}_u\big) \in \Gamma\big(\operatorname{end}\big(\Lambda^* T^*M \otimes E\big)\big) \tag{51}$$

where $\Lambda^*(g_u^{-1}\dot{g}_u)$ denotes the extension of $g_u^{-1}\dot{g}_u \in \Gamma(\operatorname{end}(TM))$ to a derivation on Λ^*T^*M.

Let us write $d := d_E$ and $\dot{d}_u^\sharp := \frac{\partial}{\partial u} d_u^\sharp$. Using the obvious relations $\dot{\Delta}_u = [d, \dot{d}_u^\sharp]$, $[N, d] = d$, $[d, \Delta_u] = 0$, $[d, Q_u] = 0$, $\dot{d}_u^\sharp = [d_u^\sharp, A_u^{-1}\dot{A}_u]$ and the fact that the super trace vanishes on super commutators we get:

$$\begin{aligned}
\operatorname{str}\big(N\dot{\Delta}_u\Delta_u^{-1}Q_u e^{-t\Delta_u}\big) &= \operatorname{str}\big(Nd\dot{d}_u^\sharp\Delta_u^{-1}Q_u e^{-t\Delta_u}\big) + \operatorname{str}\big(N\dot{d}_u^\sharp d\Delta_u^{-1}Q_u e^{-t\Delta_u}\big) \\
&= \operatorname{str}\big(d\dot{d}_u^\sharp\Delta_u^{-1}Q_u e^{-t\Delta_u}\big) \\
&= \operatorname{str}\big(dd_u^\sharp A_u^{-1}\dot{A}_u\Delta_u^{-1}Q_u e^{-t\Delta_u}\big) \\
&\quad - \operatorname{str}\big(dA_u^{-1}\dot{A}_u d_u^\sharp\Delta_u^{-1}Q_u e^{-t\Delta_u}\big) \\
&= \operatorname{str}\big(A_u^{-1}\dot{A}_u\big(dd_u^\sharp + d_u^\sharp d\big)\Delta_u^{-1}Q_u e^{-t\Delta_u}\big) \\
&= \operatorname{str}\big(A_u^{-1}\dot{A}_u Q_u e^{-t\Delta_u}\big).
\end{aligned}$$

Together with (49) this gives

$$\frac{\partial}{\partial u}\log\prod_q\big(\det{}^\gamma(\Delta_{u,q})\big)^{(-1)^q q} = \operatorname*{LIM}_{t\to 0}\operatorname{str}\big(A_u^{-1}\dot{A}_u Q_u e^{-t\Delta_u}\big). \tag{52}$$

Let us write $\Omega_u^* := \Omega^*_{E,g_u,b_u}(M;E)(\gamma)$. Note that this is a family of finite-dimensional complexes smoothly parametrized by $u \in U$. Let $P_u = 1 - Q_u$ denote the spectral projection of Δ_u onto Ω_u^*. Note that since $\operatorname{str} P_u P_u = \text{const}$ we have $\operatorname{str} P_u \dot{P}_u = 0$. For sufficiently small $w - u$ the restriction of the spectral projection $P_w|_{\Omega_u^*} : \Omega_u^* \to \Omega_w^*$ is an isomorphism of complexes. We get a commutative diagram of determinant lines:

$$\begin{tikzcd}
\det\Omega_u^* \arrow[r] \arrow[d, "\det(P_w|_{\Omega_u^*})"'] & \det H(\Omega_u^*) \arrow[r] \arrow[d, "\det H(P_w|_{\Omega_u^*})"] & \det H^*(M;E) \arrow[d, "\det H(P_w)=1"] \\
\det\Omega_w^* \arrow[r] & \det H(\Omega_w^*) \arrow[r] & \det H^*(M;E).
\end{tikzcd}$$

Writing $\beta_u := \beta_{E,g_u,b_u}$ and $\tau_u^{\mathrm{an}}(\gamma) := \tau^{\mathrm{an}}_{E,g_u,b_u}(\gamma)$, we obtain, for sufficiently small $w - u$,

$$\frac{\tau_w^{\mathrm{an}}(\gamma)}{\tau_u^{\mathrm{an}}(\gamma)} = \operatorname{sdet}\big((\beta_u|_{\Omega_u^*})^{-1}(P_w|_{\Omega_u^*})^*\beta_w\big). \tag{53}$$

Here the two non-degenerate bilinear forms $\beta_u|_{\Omega_u^*}$ and $(P_w|_{\Omega_u^*})^*\beta_w$ on Ω_u^* are considered as isomorphisms from Ω_u^* to its dual, hence $(\beta_u|_{\Omega_u^*})^{-1}(P_w|_{\Omega_u^*})^*\beta_w$ is an automorphism of Ω_u^*. Using (50) we find

$$(\beta_u|_{\Omega_u^*})^{-1}(P_w|_{\Omega_u^*})^*\beta_w = P_u A_u^{-1} A_w P_w|_{\Omega_u^*}.$$

Using (53) we thus obtain

$$\frac{\tau_w^{\mathrm{an}}(\gamma)}{\tau_u^{\mathrm{an}}(\gamma)} = \operatorname{sdet}\bigl(P_u A_u^{-1} A_w P_w|_{\Omega_u^*}\bigr).$$

In view of $\operatorname{str}(P_u \dot{P}_u) = 0$ we get

$$\frac{\partial}{\partial w}\bigg|_u \left(\frac{\tau_w^{\mathrm{an}}(\gamma)}{\tau_u^{\mathrm{an}}(\gamma)}\right) = \operatorname{str}\bigl(P_u A_u^{-1} \dot{A}_u P_u + P_u A_u^{-1} A_u \dot{P}_u\bigr) = \operatorname{str}\bigl(A_u^{-1} \dot{A}_u P_u\bigr).$$

Combining this with (52) and Proposition 4.7 we obtain

$$\frac{\partial}{\partial w}\bigg|_u \left(\frac{\tau_w^{\mathrm{an}}(0) \cdot \prod_q (\det' \Delta_{w,q})^{(-1)^q q}}{\tau_u^{\mathrm{an}}(0) \cdot \prod_q (\det' \Delta_{u,q})^{(-1)^q q}}\right) = \operatorname*{LIM}_{t\to 0} \operatorname{str}\bigl(A_u^{-1} \dot{A}_u e^{-t\Delta_u}\bigr). \tag{54}$$

Applying Proposition 6.1 to $\phi = b_u^{-1}\dot{b}_u$ we obtain

$$\operatorname*{LIM}_{t\to 0} \operatorname{str}\bigl(b_u^{-1}\dot{b}_u e^{-t\Delta_u}\bigr) = \int_M \operatorname{tr}\bigl(b_u^{-1}\dot{b}_u\bigr)\mathrm{e}(g_u).$$

Using Proposition 6.2 with $\xi = g_u^{-1}\dot{g}_u$ we get

$$\operatorname*{LIM}_{t\to 0} \operatorname{str}\left(\left(\Lambda^*\bigl(g_u^{-1}\dot{g}_u\bigr) - \frac{1}{2}\operatorname{tr}\bigl(g_u^{-1}\dot{g}_u\bigr)\right)e^{-t\Delta_u}\right) = \int_M \operatorname{tr}\bigl(b_u^{-1}\nabla^E b_u\bigr) \wedge (\partial_2 \mathrm{cs})(g_u, \dot{g}_u).$$

Using (51) we conclude

$$\operatorname*{LIM}_{t\to 0} \operatorname{str}\bigl(A_u^{-1}\dot{A}_u e^{-t\Delta_u}\bigr) = \int_M \operatorname{tr}\bigl(b_u^{-1}\dot{b}_u\bigr)\mathrm{e}(g_u) - \int_M \operatorname{tr}\bigl(b_u^{-1}\nabla^E b_u\bigr) \wedge (\partial_2 \mathrm{cs})(g_u, \dot{g}_u). \tag{55}$$

Let us finally turn to the correction term. If $[g_u, \alpha_u] \in \mathfrak{Eul}^*(M;\mathbb{C})$ represent the same coEuler structure then $\alpha_w - \alpha_u = \mathrm{cs}(g_u, g_w)$ and thus

$$\frac{\partial}{\partial u}\alpha_u = \frac{\partial}{\partial w}\bigg|_u \mathrm{cs}(g_u, g_w) = (\partial_2 \mathrm{cs})(g_u, \dot{g}_u).$$

Moreover, we have

$$\begin{aligned}
\frac{\partial}{\partial u}\operatorname{tr}\bigl(b_u^{-1}\nabla^E b_u\bigr) &= \operatorname{tr}\bigl(-b_u^{-1}\dot{b}_u b_u^{-1}\nabla^E b_u\bigr) + \operatorname{tr}\bigl(b_u^{-1}\nabla^E \dot{b}_u\bigr) \\
&= \operatorname{tr}\bigl(-b_u^{-1}\bigl(\nabla^E b_u\bigr)b_u^{-1}\dot{b}_u\bigr) + \operatorname{tr}\bigl(b_u^{-1}\nabla^E \dot{b}_u\bigr) \\
&= \operatorname{tr}\bigl(\bigl(\nabla^E b_u^{-1}\bigr)\dot{b}_u\bigr) + \operatorname{tr}\bigl(b_u^{-1}\nabla^E \dot{b}_u\bigr) \\
&= \operatorname{tr}\bigl(\nabla^E\bigl(b_u^{-1}\dot{b}_u\bigr)\bigr) \\
&= d\operatorname{tr}\bigl(b_u^{-1}\dot{b}_u\bigr).
\end{aligned}$$

Using $-2\omega_{E,b_u} = \mathrm{tr}(b_u^{-1}\nabla^E b_u)$, $d\alpha_u = \mathrm{e}(g_u)$ and Stokes' theorem we get

$$\frac{\partial}{\partial u}\int_M -2\omega_{E,b_u}\wedge\alpha_u = \int_M d\,\mathrm{tr}\big(b_u^{-1}\dot{b}_u\big)\wedge\alpha_u + \int_M \mathrm{tr}\big(b_u^{-1}\nabla^E b_u\big)\wedge(\partial_2\mathrm{cs})(g_u,\dot{g}_u)$$
$$= -\int_M \mathrm{tr}\big(b_u^{-1}\dot{b}_u\big)\mathrm{e}(g_u) + \int_M \mathrm{tr}\big(b_u^{-1}\nabla^E b_u\big)\wedge(\partial_2\mathrm{cs})(g_u,\dot{g}_u). \tag{56}$$

Combining (54)–(56) we obtain

$$\frac{\partial}{\partial w}\bigg|_u \frac{\tau^{\mathrm{an}}_{E,g_w,b_w,\alpha_w}}{\tau^{\mathrm{an}}_{E,g_u,b_u,\alpha_u}} = 0.$$

This completes the proof of Theorem 4.2.

7. Asymptotic expansion of the heat kernel

In this section we will consider Dirac operators associated to a class of Clifford super connections. The main result Theorem 7.1 below computes the leading and subleading terms of the asymptotic expansion of the corresponding heat kernels. In Section 8 we will apply these results to the Laplacians introduced in Section 4 which are squares of such Dirac operators. We refer to [1] for background on the Clifford super connection formalism.

Let (M,g) be a closed Riemannian manifold of dimension n. Let $\mathrm{Cl} = \mathrm{Cl}(T^*M,g)$ denote the corresponding Clifford bundle. Recall that $\mathrm{Cl} = \mathrm{Cl}^+ \oplus \mathrm{Cl}^-$ is a bundle of $\mathbb{Z}_2$-graded filtered algebras, and let us write Cl_k for the subbundle of filtration degree k. Recall that we have the symbol map

$$\sigma : \mathrm{Cl} \to \Lambda^* T^*M, \quad \sigma(a) := c(a)\cdot 1$$

where c denotes the usual Clifford action on Λ^*T^*M. Explicitly, for $a \in T_x^*M \subseteq \mathrm{Cl}_x$ and $\alpha \in \Lambda^*T_x^*M$ we have $c(a)\cdot\alpha = a\wedge\alpha - i_{\sharp a}\alpha$, where $\sharp a = g^{-1}a \in T_xM$ and $i_{\sharp a}$ denotes contraction with $\sharp a$. Here the metric is considered as an isomorphism $g : TM \to T^*M$ and g^{-1} denotes its inverse. Recall that σ is an isomorphism of filtered $\mathbb{Z}_2$-graded vector bundles inducing an isomorphism on the associated graded bundles of algebras.

Let $\mathcal{E} = \mathcal{E}^+ \oplus \mathcal{E}^-$ be a $\mathbb{Z}_2$-graded complex Clifford module over M. The forms with values in $\mathcal{E}$ inherit a $\mathbb{Z}_2$-grading which will be denoted by

$$\Omega(M;\mathcal{E}) = \Omega(M;\mathcal{E})^+ \oplus \Omega(M;\mathcal{E})^-.$$

We have $\Omega(M;\mathcal{E})^+ = \Omega^{\mathrm{even}}(M;\mathcal{E}^+)\oplus\Omega^{\mathrm{odd}}(M;\mathcal{E}^-)$ and similarly for $\Omega(M;\mathcal{E})^-$. Let us write $\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})$ for the bundle of algebras of endomorphisms of $\mathcal{E}$ which (super) commute with the Clifford action, and let us indicate its $\mathbb{Z}_2$-grading by

$$\mathrm{end}_{\mathrm{Cl}}(\mathcal{E}) = \mathrm{end}^+_{\mathrm{Cl}}(\mathcal{E}) \oplus \mathrm{end}^-_{\mathrm{Cl}}(\mathcal{E}).$$

Recall that we have a canonic isomorphism of bundles of $\mathbb{Z}_2$-graded algebras

$$\mathrm{end}(\mathcal{E}) = \mathrm{Cl}\otimes\mathrm{end}_{\mathrm{Cl}}(\mathcal{E}). \tag{57}$$

Suppose $\mathbb{A}: \Omega(M;\mathcal{E})^{\pm} \to \Omega(M;\mathcal{E})^{\mp}$ is a Clifford super connection, see [1, Definition 3.39]. Recall that with respect to (57) its curvature $\mathbb{A}^2 \in \Omega(M;\mathrm{end}(\mathcal{E}))^+$ decomposes as

$$\mathbb{A}^2 = R^{\mathrm{Cl}} \otimes 1 + 1 \otimes F_{\mathbb{A}}^{\mathcal{E}/S} \tag{58}$$

where $R^{\mathrm{Cl}} \in \Omega^2(M;\mathrm{Cl}^2)$ with $\mathrm{Cl}^2 := \sigma^{-1}(\Lambda^2 T^*M) \subseteq \mathrm{Cl}^+$ is a variant of the Riemannian curvature

$$R^{\mathrm{Cl}}(X,Y) = \frac{1}{4}\sum_{i,j} g(R_{X,Y}e_i, e_j)c^i c^j \tag{59}$$

and $F_{\mathbb{A}}^{\mathcal{E}/S} \in \Omega(M;\mathrm{end}_{\mathrm{Cl}}(\mathcal{E}))^+$ is called the twisting curvature, see [1, Proposition 3.43]. Here e_i is a local orthonormal frame of TM, $e^i := ge_i$ denotes its dual local coframe and $c^i = c(e^i)$ denotes Clifford multiplication with e^i.

Recall that the Dirac operator $D_{\mathbb{A}}$ associated to the Clifford super connection $\mathbb{A}$ is given by the composition

$$\Gamma(\mathcal{E}) \xrightarrow{\mathbb{A}} \Omega(M;\mathcal{E}) = \Gamma\big(\Lambda^* T^*M \otimes \mathcal{E}\big) \xrightarrow{\sigma^{-1}\otimes 1_{\mathcal{E}}} \Gamma(\mathrm{Cl} \otimes \mathcal{E}) \xrightarrow{c} \Gamma(\mathcal{E})$$

where $c: \mathrm{Cl} \otimes \mathcal{E} \to \mathcal{E}$ denotes Clifford multiplication.

We will from now on restrict to very special Clifford super connections on $\mathcal{E}$ which are of the form

$$\mathbb{A} = \nabla + A$$

where $\nabla: \Omega^*(M;\mathcal{E}^{\pm}) \to \Omega^{*+1}(M;\mathcal{E}^{\pm})$ is a Clifford connection on $\mathcal{E}$, and

$$A \in \Omega^0\big(M;\mathrm{end}^-_{\mathrm{Cl}}(\mathcal{E})\big).$$

For the associated Dirac operator acting on $\Gamma(\mathcal{E})$ we have

$$D_{\mathbb{A}} = D_{\nabla} + A.$$

Consider the induced connection $\nabla: \Omega^*(M;\mathrm{end}^{\pm}(\mathcal{E})) \to \Omega^{*+1}(M;\mathrm{end}^{\pm}(\mathcal{E}))$. Since ∇ is a Clifford connection this induced connection preserves the subbundle $\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})$. Moreover, we have $[D_{\nabla}, A] = c(\nabla A)$ and thus

$$D_{\mathbb{A}}^2 = D_{\nabla}^2 + c(\nabla A) + A^2. \tag{60}$$

Here $\nabla A \in \Omega^1(M;\mathrm{end}^-_{\mathrm{Cl}}(\mathcal{E}))$, $A^2 \in \Omega^0(M;\mathrm{end}^+_{\mathrm{Cl}}(\mathcal{E}))$, and the Clifford action $c(B)$ of $B \in \Omega(M;\mathrm{end}(\mathcal{E}))$ on $\Gamma(\mathcal{E})$ is given by the composition:

$$\Gamma(\mathcal{E}) \xrightarrow{B} \Omega(M;\mathcal{E}) = \Gamma\big(\Lambda^* T^*M \otimes \mathcal{E}\big) \xrightarrow{\sigma^{-1}\otimes 1_{\mathcal{E}}} \Gamma(\mathrm{Cl} \otimes \mathcal{E}) \xrightarrow{c} \Gamma(\mathcal{E}).$$

Note that for $B \in \Omega^0(M;\mathrm{end}(\mathcal{E}))$ the Clifford action coincides with the usual action $c(B) = B$.

Theorem 7.1. *Let $\mathcal{E}$ be a $\mathbb{Z}_2$-graded complex Clifford bundle over a closed Riemannian manifold (M,g) of dimension n. Suppose ∇ is a Clifford connection on $\mathcal{E}$ and $A \in \Omega^0(M;\mathrm{end}^-_{\mathrm{Cl}}(\mathcal{E}))$. Consider the Clifford super connection $\mathbb{A} = \nabla + A$ and the associated Dirac operator $D_{\mathbb{A}}$ acting on $\Gamma(\mathcal{E})$. Let $\Omega_g \in \Omega^n(M;\mathcal{O}^{\mathbb{C}}_M)$ denote the volume density associated with the Riemannian metric g. Let $k_t \in \Gamma(\mathrm{end}(\mathcal{E}))$ so that $k_t\Omega_g$ is the restriction of the kernel of $e^{-tD^2_{\mathbb{A}}}$ to the diagonal in $M \times M$. Consider its asymptotic expansion*

$$k_t \sim (4\pi t)^{-n/2}\sum_{i\geqslant 0} t^i \tilde{k}_i \quad \textit{as } t \to 0 \tag{61}$$

with $\tilde{k}_i \in \Gamma(\mathrm{end}(\mathcal{E}))$, see [1, Theorem 2.30]. *Then*

$$\tilde{k}_i \in \Gamma\big(\mathrm{Cl}_{2i} \otimes \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})\big) \subseteq \Gamma\big(\mathrm{Cl} \otimes \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})\big) = \Gamma\big(\mathrm{end}(\mathcal{E})\big). \tag{62}$$

Moreover, with the help of the symbol map

$$\sigma : \Gamma\big(\mathrm{end}(\mathcal{E})\big) = \Gamma\big(\mathrm{Cl} \otimes \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})\big) \xrightarrow{\sigma\otimes 1} \Omega^*\big(M;\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})\big)$$

and writing $\alpha_{[j]}$ for the j-form piece of α we have

$$\sum_{i\geqslant 0} \sigma(\tilde{k}_i)_{[2i]} = \hat{A}_g \wedge \exp\big(-F^{\mathcal{E}/S}_\nabla\big). \tag{63}$$

Here $\hat{A}_g \in \Omega^{4*}(M;\mathbb{R})$ denotes the $\hat{A}$-genus

$$\hat{A}_g = \det{}^{1/2}\left(\frac{R/2}{\sinh(R/2)}\right)$$

and $R \in \Omega^2(M;\mathrm{end}(TM))$ the Riemannian curvature. Moreover, we have

$$\sum_{i\geqslant 0} \sigma(\tilde{k}_i)_{[2i-1]} = -\nabla\left(\hat{A}_g \wedge \left(\frac{e^{\mathrm{ad}(-F^{\mathcal{E}/S}_\nabla)} - 1}{\mathrm{ad}(-F^{\mathcal{E}/S}_\nabla)} A\right) \wedge \exp\big(-F^{\mathcal{E}/S}_\nabla\big)\right) \tag{64}$$

where $\mathrm{ad}(F^{\mathcal{E}/S}_\nabla) : \Omega^*(M;\mathrm{end}^\pm_{\mathrm{Cl}}(\mathcal{E})) \to \Omega^{*+2}(M;\mathrm{end}^\pm_{\mathrm{Cl}}(\mathcal{E}))$*, is given by*

$$\mathrm{ad}\big(F^{\mathcal{E}/S}_\nabla\big)\phi := F^{\mathcal{E}/S}_\nabla \wedge \phi - \phi \wedge F^{\mathcal{E}/S}_\nabla.$$

Remark 7.2. Note that (62) and (63) tell that on this level the asymptotic expansions for $e^{-tD^2_{\mathbb{A}}}$ and $e^{-tD^2_\nabla}$ are the same.

Proof. The proof below parallels the one of Theorem 4.1 in [1] where the case $A = 0$ is treated. It too is based on Getzler's scaling techniques, see [21]. In order to prove Theorem 7.1 we need to compute one more term in the asymptotic expansion of the rescaled operator.

The calculation is local. Let $x_0 \in M$. Use normal coordinates, i.e. the exponential mapping of g, to identify a convex neighborhood U of $0 \in T_{x_0}M$ with a neighborhood of x_0. Choose an orthonormal basis $\{\partial_i\}$ of $T_{x_0}M$ and linear coordinates $\mathbf{x} = (x^1,\dots,x^n)$ on $T_{x_0}M$ such that $\{dx^i\}$ is dual to $\{\partial_i\}$. Let $\mathcal{R} := \sum_i x^i\partial_i$ denote the radial vector field. Note that every affinely parametrized line through the origin in $T_{x_0}M$ is a geodesic. Let $\{e_i\}$ denote the local orthonormal frame of TM obtained from $\{\partial_i\}$ by parallel transport along $\mathcal{R}$, i.e. $\nabla^g_{\mathcal{R}} e_i = 0$ and $e_i(x_0) = \partial_i$. Let $\{e^i\}$ denote the dual local coframe.

Trivialize $\mathcal{E}$ with the help of radial parallel transport by ∇. Use this trivialization to identify $\Gamma(\mathcal{E}|_U)$ with $C^\infty(U, \mathcal{E}_0)$, where $\mathcal{E}_0 := \mathcal{E}_{x_0}$. Let $\omega \in \Omega^1(U; \mathrm{end}(\mathcal{E}_0))$ denote the connection one form of this trivialization, i.e. $\nabla_{\partial_i} = \partial_i + \omega(\partial_i)$. For the curvature F of ∇ we then have $F = d\omega + \omega \wedge \omega \in \Omega^2(M; \mathrm{end}^+(\mathcal{E}_0))$. By the choice of trivialization of $\mathcal{E}|_U$ we have $i_{\mathcal{R}}\omega = 0$ and thus $i_{\mathcal{R}}F = i_{\mathcal{R}}(d\omega + \omega \wedge \omega) = i_{\mathcal{R}}\, d\omega$. Contracting this with ∂_i and using $[\partial_i, \mathcal{R}] = \partial_i$ we obtain

$$-F(\partial_i, \mathcal{R}) = F(\mathcal{R}, \partial_i) = (d\omega)(\mathcal{R}, \partial_i) = (L_{\mathcal{R}} + 1)\big(\omega(\partial_i)\big) \tag{65}$$

where $L_{\mathcal{R}}$ denotes Lie derivative with respect to the vector field $\mathcal{R}$. Let $\omega(\partial_i) \sim \sum_\alpha \frac{\partial_\alpha \omega(\partial_i)_{x_0}}{\alpha!} x^\alpha$ denote the Taylor expansion of $\omega(\partial_i)$ at x_0, written with the help of multi index notation. Using $L_{\mathcal{R}} x^\alpha = |\alpha| x^\alpha$ we obtain the following Taylor expansion $(L_{\mathcal{R}} + 1)(\omega(\partial_i)) \sim \sum_\alpha \big(|\alpha| + 1\big) \frac{\partial_\alpha \omega(\partial_i)_{x_0}}{\alpha!} x^\alpha$. If $F(\partial_i, \partial_j) \sim \sum_\alpha \frac{\partial_\alpha F(\partial_i, \partial_j)_{x_0}}{\alpha!} x^\alpha$ denotes the Taylor expansion of $F(\partial_i, \partial_j)$ at x_0 then we obtain the Taylor expansion $F(\partial_i, \mathcal{R}) \sim \sum_{j,\alpha} \frac{\partial_\alpha F(\partial_i, \partial_j)_{x_0}}{\alpha!} x^j x^\alpha$. Comparing the Taylor expansions of both sides of (65) we obtain the Taylor expansion, cf. [1, Proposition 1.18],

$$\nabla_{\partial_i} - \partial_i = \omega(\partial_i) \sim -\sum_{j,\alpha} \frac{\partial_\alpha F(\partial_i, \partial_j)_{x_0}}{(|\alpha| + 2)\alpha!} x^j x^\alpha.$$

For the first few terms this gives:

$$\nabla_{\partial_i} = \partial_i - \frac{1}{2} \sum_j F(\partial_i, \partial_j)_{x_0} x^j - \frac{1}{3} \sum_{j,k} \partial_k F(\partial_i, \partial_j)_{x_0} x^j x^k + O\big(|\mathbf{x}|^3\big). \tag{66}$$

Let $c^i := c(e^i) \in \Gamma(\mathcal{E}|_U) = C^\infty(U, \mathrm{end}(\mathcal{E}_0))$ denote Clifford multiplication with e^i. Since $\nabla^g_{\mathcal{R}} e^i = 0$ and since ∇ is a Clifford connection we have $\nabla_{\mathcal{R}} c^i = c(\nabla^g_{\mathcal{R}} e^i) = 0$. So we see that c^i is actually a constant in $\mathrm{end}(\mathcal{E}_0)$, cf. [1, Lemma 4.14]. Particularly, our trivialization of $\mathcal{E}|_U$ identifies $\Gamma(\mathrm{end}_{\mathrm{Cl}}(\mathcal{E}|_U))$ with $C^\infty(U, \mathrm{end}_{\mathrm{Cl}}(\mathcal{E}_0))$. Recall that

$$F(\partial_i, \partial_j) = \frac{1}{4} \sum_{l,m} g(R_{\partial_i, \partial_j} e_l, e_m) c^l c^m + F^{\mathcal{E}/S}_\nabla(\partial_i, \partial_j)$$

with $F^{\mathcal{E}/S}_\nabla \in \Omega^2(U; \mathrm{end}^+_{\mathrm{Cl}}(\mathcal{E}_0))$. From (66) we thus obtain, cf. [1, Lemma 4.15],

$$\begin{aligned}
\nabla_{\partial_i} = \partial_i &- \frac{1}{8} \sum_{j,l,m} g(R_{\partial_i, \partial_j} e_l, e_m)_{x_0} x^j c^l c^m \\
&- \frac{1}{12} \sum_{j,k,l,m} \partial_k g(R_{\partial_i, \partial_j} e_l, e_m)_{x_0} x^j x^k c^l c^m \\
&+ \sum_{l,m} u_{ilm}(\mathbf{x}) c^l c^m + v_i(\mathbf{x})
\end{aligned} \tag{67}$$

with $u_{ilm}(\mathbf{x}) = O(|\mathbf{x}|^3) \in C^\infty(U)$ and $v_i(\mathbf{x}) = O(|\mathbf{x}|) \in C^\infty(U, \mathrm{end}_{\mathrm{Cl}}(\mathcal{E}_0))$.

Let Δ denote the connection Laplacian given by the composition

$$\Gamma(\mathcal{E}) \xrightarrow{\nabla} \Gamma\big(T^*M \otimes \mathcal{E}\big) \xrightarrow{\nabla^g \otimes 1 + 1 \otimes \nabla} \Gamma\big(T^*M \otimes T^*M \otimes \mathcal{E}\big) \xrightarrow{-\operatorname{tr}_g} \Gamma(\mathcal{E}).$$

Let r denote the scalar curvature of g and recall Lichnerowicz' formula [1, Theorem 3.52]

$$D_\nabla^2 = \Delta + c\big(F_\nabla^{\mathcal{E}/S}\big) + \frac{r}{4}.$$

Recall our Clifford super connection $\mathbb{A} = \nabla + A$ with $A \in \Omega^0(M; \operatorname{end}^-_{\mathrm{Cl}}(\mathcal{E}))$. Since $D^2_{\mathbb{A}} = D^2_\nabla + c(\nabla A) + A^2$ we obtain

$$D^2_{\mathbb{A}} = \Delta + c\big(F_\nabla^{\mathcal{E}/S} + \nabla A\big) + A^2 + \frac{r}{4}. \tag{68}$$

Use the symbol map to identify $\operatorname{end}(\mathcal{E}_0) = \Lambda^* T^*_{x_0} M \otimes \operatorname{end}_{\mathrm{Cl}}(\mathcal{E}_0)$. For $0 < u \leqslant 1$ and $\alpha \in C^\infty(\mathbb{R}^+ \times U, \Lambda^* T^*_{x_0} M \otimes \operatorname{end}_{\mathrm{Cl}}(\mathcal{E}_0))$ define Getzler's rescaling

$$(\delta_u \alpha)(t, \mathbf{x}) := \sum_i u^{-i/2} \alpha\big(ut, u^{1/2}\mathbf{x}\big)_{[i]}.$$

Consider the kernel $p \in C^\infty(\mathbb{R}^+ \times U, \Lambda^* T^*_{x_0} M \otimes \operatorname{end}_{\mathrm{Cl}}(\mathcal{E}_0))$ of $e^{-tD^2_{\mathbb{A}}}$, $p(t, \mathbf{x}) = p_t(\mathbf{x}, x_0)$. Note that $p(t, 0) = k_t(x_0)$. Define the rescaled kernel $r_u := u^{n/2} \delta_u p$ and the rescaled operator $L_u := u \delta_u D^2_{\mathbb{A}} \delta_u^{-1}$. Since $(\partial_t + D^2_{\mathbb{A}})p = 0$ and $\delta_u \partial_t \delta_u^{-1} = u^{-1}\partial_t$ we have

$$(\partial_t + L_u) r_u = 0. \tag{69}$$

Note that setting $t = 1$ and $x = 0$ and using (61) we get an asymptotic expansion

$$r_u(1, 0) \sim (4\pi)^{-n/2} \sum_{j \geqslant -n} u^{j/2} \sum_{i \geqslant 0} \sigma\big(\tilde{k}_i(x_0)\big)_{[2i-j]} \quad \text{as } u \to 0. \tag{70}$$

The claim (62) just states that the terms for $-n \leqslant j < 0$ vanish, i.e. there are no Laurent terms in (70). Statements (63) and (64) are explicit expressions for the term $j = 0$ and $j = 1$ in (70).

Let us compute the first terms in the asymptotic expansion of L_u in powers of $u^{1/2}$. Let us write ε^j for the exterior multiplication with e^j, and ι^j for the contraction with e_j. Note that

$$\delta_u \varepsilon^j \delta_u^{-1} = u^{-1/2} \varepsilon^j, \qquad \delta_u \iota^j \delta_u^{-1} = u^{1/2} \iota^j, \qquad \delta_u \partial_i \delta_u^{-1} = u^{-1/2} \partial_i$$

and recall that $c^j = \varepsilon^j - \iota^j$. Let us look at the simplest part first. Clearly,

$$u \delta_u \left(A^2 + \frac{r}{4}\right) \delta_u^{-1} = O(u) \quad \text{as } u \to 0. \tag{71}$$

Next we have $\nabla A = \sum_i (\nabla_{e_i} A) e^i$, hence $c(\nabla A) = \sum_i (\nabla_{e_i} A) c^i$ and therefore

$$u \delta_u c(\nabla A) \delta_u^{-1} = u^{1/2} \mathsf{A}' + O\big(u^{3/2}\big) \quad \text{as } u \to 0 \tag{72}$$

where $\mathsf{A}' := \sum_i (\nabla_{\partial_i} A)_{x_0} \varepsilon^i$. Moreover, $F_\nabla^{\mathcal{E}/S} = \frac{1}{2}\sum_{i,j} F_\nabla^{\mathcal{E}/S}(e_i, e_j) e^i \wedge e^j$, hence $c(F_\nabla^{\mathcal{E}/S}) = \frac{1}{2}\sum_{i,j} F_\nabla^{\mathcal{E}/S}(e_i, e_j) c^i c^j$, and thus

$$u\delta_u c\big(F_\nabla^{\mathcal{E}/S}\big)\delta_u^{-1} = \mathsf{F} + O(u) \quad \text{as } u \to 0 \tag{73}$$

where $\mathsf{F} := \frac{1}{2}\sum_{i,j} F_\nabla^{\mathcal{E}/S}(\partial_i, \partial_j)_{x_0} \varepsilon^i \varepsilon^j$. From (67) we easily get

$$u^{1/2}\delta_u \nabla_{\partial_i} \delta_u^{-1} = \partial_i - \frac{1}{4}\sum_j \mathsf{R}_{ij} x^j + u^{1/2} R_i' + O(u) \quad \text{as } u \to 0$$

where $\mathsf{R}_{ij} := \frac{1}{2}\sum_{l,m} g(R_{\partial_i,\partial_j} e_l, e_m)_{x_0} \varepsilon^l \varepsilon^m$ and R_i' is an operator which acts on $C^\infty(U, \Lambda^{\text{even}/\text{odd}} T_{x_0}^* M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0))$ in a way which preserves the parity of the form degree. Using the formula $\Delta = -\sum_i ((\nabla_{e_i})^2 - \nabla_{\nabla^g_{e_i} e_i})$ and the fact that $\nabla^g_{e_i} e_i$ vanishes at x_0 we obtain

$$u\delta_u \Delta \delta_u^{-1} = -\sum_i \left(\partial_i - \frac{1}{4}\sum_j \mathsf{R}_{ij} x^j\right)^2 + u^{1/2} K^{\text{even}} + O(u) \quad \text{as } u \to 0 \tag{74}$$

where K^{even} acts on $C^\infty(U, \Lambda^{\text{even}/\text{odd}} T_{x_0}^* M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0))$ in a parity preserving way. Let us write

$$\mathsf{K} := -\sum_i \left(\partial_i - \frac{1}{4}\sum_j \mathsf{R}_{ij} x^j\right)^2 + \mathsf{F}.$$

Then (71)–(74) together with (68) finally give

$$L_u = \mathsf{K} + u^{1/2}\big(\mathsf{A}' + K^{\text{even}}\big) + O(u) \quad \text{as } u \to 0. \tag{75}$$

Recall, see [1, Lemma 4.19], that there exist $\Lambda^* T^* M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0)$-valued polynomials $\tilde{r}_i$ on $\mathbb{R} \times U$ so that we have an asymptotic expansion

$$r_u(t, \mathbf{x}) \sim q_t(\mathbf{x}) \sum_{j \geqslant -n} u^{j/2} \tilde{r}_j(t, \mathbf{x}) \quad \text{as } u \to 0 \tag{76}$$

where $q_t(\mathbf{x}) = (4\pi t)^{-n/2} e^{-|\mathbf{x}|^2/4t}$. Moreover, the initial condition for the heat kernel translates to

$$\tilde{r}_j(0, 0) = \delta_{j,0}. \tag{77}$$

Setting $t = 1$, $\mathbf{x} = 0$ in (76) we get

$$r_u(1, 0) \sim (4\pi)^{-n/2} \sum_{j \geqslant -n} u^{j/2} \tilde{r}_j(1, 0) \quad \text{as } u \to 0. \tag{78}$$

Comparing this with (70) we obtain

$$\tilde{r}_j(1, 0) = \sum_{i \geqslant 0} \sigma(\tilde{k}_i)_{[2i-j]}(x_0). \tag{79}$$

Expanding the equation $(\partial_t + L_u)r_u = 0$ in a power series in $u^{1/2}$ with the help of (76) and (75) the leading term $q\tilde{r}_l$ satisfies $(\partial_t + \mathsf{K})(q\tilde{r}_l) = 0$. Because of the initial condition (77) and the uniqueness of formal solutions [1, Theorem 4.13] we must have $l \geqslant 0$ and thus $\tilde{r}_j = 0$ for $j < 0$. In view of (79) this proves (62).

So $q\tilde{r}_0$ satisfies $(\partial_t + \mathsf{K})(q\tilde{r}_0) = 0$ with initial condition $\tilde{r}_0(0,0) = 1$, see (77). Mehler's formula [1, Theorem 4.13] provides an explicit solution:

$$\begin{aligned} &q_t(\mathbf{x})\tilde{r}_0(t,\mathbf{x}) \\ &\quad = (4\pi t)^{-n/2}\det{}^{1/2}\left(\frac{t\mathsf{R}/2}{\sinh(t\mathsf{R}/2)}\right) \wedge \exp\left(-\frac{1}{4t}\left\langle \mathbf{x} \,\middle|\, \frac{t\mathsf{R}}{2}\coth\left(\frac{t\mathsf{R}}{2}\right) \,\middle|\, \mathbf{x}\right\rangle\right) \wedge \exp(-t\mathsf{F}). \end{aligned}$$

Setting $t = 1$, $\mathbf{x} = 0$ we obtain

$$\tilde{r}_0(1,0) = \det{}^{1/2}\left(\frac{\mathsf{R}/2}{\sinh(\mathsf{R}/2)}\right) \wedge \exp(-\mathsf{F}). \tag{80}$$

In view of (79) we thus have established (63).

The term $q\tilde{r}_1$ satisfies $(\partial_t + \mathsf{K})(q\tilde{r}_1) = -(\mathsf{A}' + K^{\text{even}})(\tilde{q}r_0)$. Let us write

$$\tilde{r}_1(t,\mathbf{x}) = \tilde{r}_1^{\text{even}}(t,\mathbf{x}) + \tilde{r}_1^{\text{odd}}(t,\mathbf{x})$$

with $\tilde{r}_1^{\text{even}}(t,\mathbf{x}) \in \Lambda^{\text{even}} T^*_{x_0}M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0)$ and $\tilde{r}_1^{\text{odd}}(t,\mathbf{x}) \in \Lambda^{\text{odd}} T^*_{x_0}M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0)$. Note that in view of (79) we have

$$\tilde{r}_1(1,0) = \sum_{i \geqslant 0} \sigma(\tilde{k}_i)_{[2i-1]}(x_0) = \tilde{r}_1^{\text{odd}}(1,0). \tag{81}$$

It thus suffices to determine $\tilde{r}_1^{\text{odd}}$. Since

$$\begin{aligned} \big(K^{\text{even}}(q\tilde{r}_0)\big)(t,\mathbf{x}) &\in \Lambda^{\text{even}} T^*_{x_0}M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0), \\ \big(\mathsf{A}'(q\tilde{r}_0)\big)(t,\mathbf{x}) &\in \Lambda^{\text{odd}} T^*_{x_0}M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0), \end{aligned}$$

we must have $(\partial_t + \mathsf{K})(q\tilde{r}_1^{\text{odd}}) = -\mathsf{A}'(q\tilde{r}_0)$. We make the following ansatz, and suppose that $\tilde{r}_1^{\text{odd}} = B\tilde{r}_0$ with $B \in C^\infty(\mathbb{R}, \Lambda^{\text{odd}} T^*_{x_0}M \otimes \text{end}_{\text{Cl}}(\mathcal{E}_0))$. Then

$$\begin{aligned} (\partial_t + \mathsf{K})\big(q\tilde{r}_1^{\text{odd}}\big) &= (\partial_t B)q\tilde{r}_0 + B\partial_t(q\tilde{r}_0) + \mathsf{K}(Bq\tilde{r}_0) \\ &= (\partial_t B)q\tilde{r}_0 - B\mathsf{K}(q\tilde{r}_0) + \mathsf{K}(Bq\tilde{r}_0) \\ &= (\partial_t B)q\tilde{r}_0 - B\mathsf{F}q\tilde{r}_0 + \mathsf{F}Bq\tilde{r}_0 \\ &= \big(\partial_t B + \text{ad}(\mathsf{F})B\big)q\tilde{r}_0. \end{aligned}$$

Hence we have to solve $\partial_t B = \text{ad}(-\mathsf{F})B - \mathsf{A}'$ with initial condition $B(0) = 0$. This is easily carried out and we find the solution:

$$B(t) = -\frac{e^{\text{ad}(-t\mathsf{F})} - 1}{\text{ad}(-\mathsf{F})}\mathsf{A}'.$$

Thus $qB\tilde r_0$ satisfies $(\partial_t+\mathsf{K})(qB\tilde r_0)=-\mathsf{A}'(q\tilde r_0)$ with initial condition $(B\tilde r_0)(0,0)=0$. Again, the uniqueness of formal solutions of the heat equation guarantees that we actually have $\tilde r_1^{\mathrm{odd}}=B\tilde r_0$. Setting $t=1$, $\mathbf{x}=0$ and using (80) we get

$$\tilde r_1^{\mathrm{odd}}(1,0)=B(1)\tilde r_0(1,0)=-\det{}^{1/2}\biggl(\frac{\mathsf{R}/2}{\sinh(\mathsf{R}/2)}\biggr)\wedge\biggl(\frac{e^{\mathrm{ad}(-\mathsf{F})}-1}{\mathrm{ad}(-\mathsf{F})}\mathsf{A}'\biggr)\wedge\exp(-\mathsf{F}).$$

Using (81) we conclude

$$\sum_{i\geqslant 0}\sigma(\tilde k_i)_{[2i-1]}=-\hat A_g\wedge\biggl(\frac{e^{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}-1}{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}\nabla A\biggr)\wedge\exp\bigl(-F_\nabla^{\mathcal{E}/S}\bigr).$$

The Bianchi identity $\nabla F_\nabla^{\mathcal{E}/S}=0$ implies $\nabla\exp(-F_\nabla^{\mathcal{E}/S})=0$, $\nabla\,\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})=0$, and similarly $d\hat A_g=0$, from which we finally obtain (64). $\square$

7.1. Certain heat traces

Since $\mathcal{E}$ is $\mathbb{Z}_2$-graded we have a *super trace* $\mathrm{str}_{\mathcal{E}}:\Gamma(\mathrm{end}(\mathcal{E}))\to\Omega^0(M;\mathbb{C})$. If n is even we will also make use of the so called *relative super trace*, see [1, Definition 3.28], $\mathrm{str}_{\mathcal{E}/S}:\Gamma(\mathrm{end}_{\mathrm{Cl}}(\mathcal{E}))\to\Omega^0(M;\mathcal{O}_M^{\mathbb{C}})$

$$\mathrm{str}_{\mathcal{E}/S}(b):=2^{-n/2}\,\mathrm{str}_{\mathcal{E}}\bigl(c(\Gamma)b\bigr).$$

Here $\Gamma\in\Gamma(\mathrm{Cl}\otimes\mathcal{O}_M^{\mathbb{C}})$ denotes the chirality element, see [1, Lemma 3.17]. With respect to a local orthonormal frame $\{e_i\}$ of TM and its dual local coframe $\{e^i\}$ the chirality element Γ is given as $\mathbf{i}^{n/2}e^1\cdots e^n$ times the orientation of $(e_1,\dots,e_n)$. This relative super trace gives rise to

$$\mathrm{str}_{\mathcal{E}/S}:\Omega^*\bigl(M;\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})\bigr)\to\Omega^*\bigl(M;\mathcal{O}_M^{\mathbb{C}}\bigr)$$

which will be denoted by the same symbol. For every $\phi\in\Gamma(\mathrm{end}(\mathcal{E}))$ we have

$$\bigl(\mathrm{str}_{\mathcal{E}}(\phi)\bigr)\cdot\Omega_g=(\mathbf{i}/2)^{-n/2}\,\mathrm{str}_{\mathcal{E}/S}\bigl(\sigma(\phi)_{[n]}\bigr)\tag{82}$$

where $\Omega_g\in\Omega^n(M;\mathcal{O}_M^{\mathbb{C}})$ denotes the volume density associated with g. To see (82) note first that

$$\mathrm{Cl}_{n-1}=[\mathrm{Cl},\mathrm{Cl}]\tag{83}$$

where Cl_k denotes the filtration on Cl, see [1, Proof of Proposition 3.21]. Hence both sides of (82) vanish on $\Gamma(\mathrm{Cl}_{n-1}\otimes\mathrm{end}_{\mathrm{Cl}}(\mathcal{E}))$. It remains to check (82) on sections of $\mathrm{Cl}/\mathrm{Cl}_{n-1}\otimes\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})$, but for these the desired equality follows immediately from the definition of the relative super trace.

Lemma 7.3. *Let $D_{\mathbb{A}}$ be a Dirac operator and $\tilde{k}_i \in \Gamma(\mathrm{end}(\mathcal{E}))$ as in Theorem* 7.1. *Moreover, let $\Phi \in \Gamma(\mathrm{end}(\mathcal{E}))$. Then, for even n, we have*

$$\operatorname*{LIM}_{t\to 0} \mathrm{str}\big(\Phi e^{-tD^2_{\mathbb{A}}}\big) = (2\pi\mathbf{i})^{-n/2} \int_M \mathrm{str}_{\mathcal{E}/S}\big(\sigma(\Phi\tilde{k}_{n/2})_{[n]}\big)$$

whereas $\mathrm{LIM}_{t\to 0}\,\mathrm{str}(\Phi e^{-tD^2_{\mathbb{A}}}) = 0$ *if n is odd. Here* LIM *denotes the renormalized limit* [1, Section 9.6] *which in this case is just the constant term in the asymptotic expansion for* $t \to 0$.

Proof. For odd n this follows immediately from (61). So assume n is even. Recall from [1, Proposition 2.32] that

$$\mathrm{str}\big(\Phi e^{-tD^2_{\mathbb{A}}}\big) = \int_M \mathrm{str}_{\mathcal{E}}(\Phi k_t)\cdot \Omega_g. \tag{84}$$

Combining this with (82) we obtain

$$\mathrm{str}\big(\Phi e^{-tD^2_{\mathbb{A}}}\big) = (\mathbf{i}/2)^{-n/2} \int_M \mathrm{str}_{\mathcal{E}/S}\big(\sigma(\Phi k_t)_{[n]}\big).$$

We thus get an asymptotic expansion, see (61),

$$\mathrm{str}\big(\Phi e^{-tD^2_{\mathbb{A}}}\big) \sim (2\pi\mathbf{i}t)^{-n/2} \sum_{i\geqslant 0} t^i \int_M \mathrm{str}_{\mathcal{E}/S}\big(\sigma(\Phi\tilde{k}_i)_{[n]}\big) \quad \text{as } t\to 0,$$

from which the desired formula follows at once. □

Corollary 7.4. *Let $D_{\mathbb{A}}$ be a Dirac operator as in Theorem* 7.1. *Moreover, let $U \in \Gamma(\mathrm{end}_{\mathrm{Cl}}(\mathcal{E}))$. Then, for even n, we have*

$$\operatorname*{LIM}_{t\to 0} \mathrm{str}\big(Ue^{-tD^2_{\mathbb{A}}}\big) = (2\pi\mathbf{i})^{-n/2} \int_M \hat{A}_g \wedge \mathrm{str}_{\mathcal{E}/S}\big(U\exp\big(-F^{\mathcal{E}/S}_{\nabla}\big)\big), \tag{85}$$

whereas $\mathrm{LIM}_{t\to 0}\,\mathrm{str}(Ue^{-tD^2_{\mathbb{A}}}) = 0$ *if n is odd.*

Proof. For odd n this follows immediately from Lemma 7.3. So assume n is even. Since $\sigma(U\tilde{k}_i)_{[n]} = U\sigma(\tilde{k}_i)_{[n]}$ Theorem 7.1 yields

$$\mathrm{str}_{\mathcal{E}/S}\big(\sigma(U\tilde{k}_{n/2})_{[n]}\big) = \big(\hat{A}_g \wedge \mathrm{str}_{\mathcal{E}/S}\big(U\exp\big(-F^{\mathcal{E}/S}_{\nabla}\big)\big)\big)_{[n]}.$$

Equation (85) then follows from Lemma 7.3. □

Corollary 7.5. *Let $D_{\mathbb{A}}$ be a Dirac operator as in Theorem* 7.1. *Moreover, suppose $V \in \Omega^1(M; \mathrm{end}_{\mathrm{Cl}}(\mathcal{E}))$, let $c(V) \in \Gamma(\mathrm{end}(\mathcal{E}))$ denote Clifford multiplication with V, and consider $\nabla V \in \Omega^2(M; \mathrm{end}_{\mathrm{Cl}}(\mathcal{E}))$. Then, for even n, we have*

$$\begin{aligned} &\operatorname*{LIM}_{t\to 0} \mathrm{str}\big(c(V)e^{-tD_{\mathbb{A}}^2}\big) \\ &\quad = -(2\pi\mathbf{i})^{-n/2} \int_M \hat{A}_g \wedge \mathrm{str}_{\mathcal{E}/S}\left(\left(\frac{e^{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}-1}{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}A\right) \wedge \exp\big(-F_\nabla^{\mathcal{E}/S}\big) \wedge \nabla V\right), \end{aligned} \tag{86}$$

whereas $\mathrm{LIM}_{t\to 0}\,\mathrm{str}(c(V)e^{-tD_{\mathbb{A}}^2}) = 0$ if n is odd.

Proof. If n is odd the statement follows immediately from Lemma 7.3. So assume n is even. Since $\sigma(c(V)\tilde{k}_i)_{[n]} = V \wedge \sigma(\tilde{k}_i)_{[n-1]}$ Theorem 7.1 yields

$$\begin{aligned} &\mathrm{str}_{\mathcal{E}/S}\big(\sigma\big(c(V)\tilde{k}_{n/2}\big)_{[n]}\big) \\ &\quad = -\,\mathrm{str}_{\mathcal{E}/S}\left(V \wedge \nabla\left(\hat{A}_g \wedge \left(\frac{e^{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}-1}{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}A\right) \wedge \exp\big(-F_\nabla^{\mathcal{E}/S}\big)\right)\right)_{[n]} \\ &\quad = -\,\mathrm{str}_{\mathcal{E}/S}\left(\nabla\left(\hat{A}_g \wedge \left(\frac{e^{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}-1}{\mathrm{ad}(-F_\nabla^{\mathcal{E}/S})}A\right) \wedge \exp\big(-F_\nabla^{\mathcal{E}/S}\big)\right) \wedge V\right)_{[n]}. \end{aligned}$$

Applying Lemma 7.3 and using Stokes' theorem we obtain (86). □

8. Application to Laplacians

Below we will see that the Laplacians $\Delta_{E,g,b}$ introduced in Section 4 are the squares of Dirac operators of the kind considered in Section 7. Applying Corollaries 7.4 and 7.5 will lead to a proofs of Propositions 6.1 and 6.2, respectively.

8.1. The exterior algebra as Clifford module

Let (M, g) be a Riemannian manifold of dimension n. In order to understand the Clifford module structure of $\Lambda := \Lambda^* T^* M$ we first note that Λ is a Clifford module for $\hat{\mathrm{Cl}} := \mathrm{Cl}(T^*M, -g)$ too. Let us write $\hat{c}$ for the Clifford multiplication of $\hat{\mathrm{Cl}}$ on Λ. Explicitly, for $a \in T_x^* M \subseteq \hat{\mathrm{Cl}}$ and $\alpha \in \Lambda^* T_x^* M$ we have $\hat{c}(a)\alpha = a \wedge \alpha + i_{\sharp a}\alpha$, where $\sharp a := g^{-1}a \in T_x M$ and $i_{\sharp a}$ denotes contraction with $\sharp a$. It follows from this formula that every $\hat{c}(a)$ commutes with the Clifford action of Cl. We thus obtain an isomorphism of $\mathbb{Z}_2$-graded filtered algebras

$$\hat{c} : \hat{\mathrm{Cl}} \to \mathrm{end}_{\mathrm{Cl}}(\Lambda).$$

Let us write

$$\hat{\sigma} : \hat{\mathrm{Cl}} \to \Lambda, \quad \hat{\sigma}(a) := \hat{c}(a) \cdot 1$$

for the symbol map of $\hat{\mathrm{Cl}}$.

As in (59) define $R^{\hat{\mathrm{Cl}}} \in \Omega^2(M;\hat{\mathrm{Cl}})$ by

$$R^{\hat{\mathrm{Cl}}}(X,Y) := -\frac{1}{4}\sum_{i,j} g\big(R(X,Y)e_i, e_j\big)\hat{c}^i\hat{c}^j$$

where X and Y are two vector fields, $\{e_i\}$ is a local orthonormal frame, $\{e^i\}$ denotes its dual local coframe, and $\hat{c}^i := \hat{c}(e^i)$. For the twisting curvature $F^{\Lambda/S}_{\nabla^g} \in \Omega^2(M;\mathrm{end}_{\mathrm{Cl}}(\Lambda))$ we then have, see [1, p. 145],

$$F^{\Lambda/S}_{\nabla^g} = (1\otimes\hat{c})\big(R^{\hat{\mathrm{Cl}}}\big) \in \Omega^2\big(M;\mathrm{end}_{\mathrm{Cl}}(\Lambda)\big) \tag{87}$$

where $(1\otimes\hat{c}) : \Omega(M;\hat{\mathrm{Cl}}) \to \Omega(M;\mathrm{end}_{\mathrm{Cl}}(\Lambda))$. Indeed, the curvature of Λ,

$$R^{\Lambda} \in \Omega^2\big(M;\mathrm{end}(\Lambda)\big),$$

can be written as

$$R^{\Lambda}(X,Y) = \sum_{i,j} g\big(R(X,Y)e_i, e_j\big)\frac{1}{2}\big(\varepsilon^j\iota^i - \varepsilon^i\iota^j\big) \in \Gamma\big(\mathrm{end}(\Lambda)\big)$$

where $\varepsilon^j \in \Gamma(\mathrm{end}(\Lambda))$ denotes exterior multiplication with e^j, and $\iota^i \in \Gamma(\mathrm{end}(\Lambda))$ denotes contraction with e_i. Using $\varepsilon^i = \frac{1}{2}(c^i + \hat{c}^i)$ and $\iota^i = -\frac{1}{2}(c^i - \hat{c}^i)$ one easily deduces

$$\frac{1}{2}\big(\varepsilon^j\iota^i - \varepsilon^i\iota^j\big) = \frac{1}{4}\left(\frac{1}{2}\big(c^ic^j - c^jc^i\big)\right) - \frac{1}{4}\left(\frac{1}{2}\big(\hat{c}^i\hat{c}^j - \hat{c}^j\hat{c}^i\big)\right)$$

from which we read off (87), see (58). Also note that we have

$$(1\otimes\hat{\sigma})\big(R^{\hat{\mathrm{Cl}}}\big) = -\frac{1}{2}R \in \Omega^2\big(M;\Lambda^2T^*M\big) \tag{88}$$

where $(1\otimes\hat{\sigma}) : \Omega(M;\hat{\mathrm{Cl}}) \to \Omega(M;\Lambda)$.

If n is even then the relative super trace

$$\mathrm{str}_{\Lambda/S} : \mathrm{end}_{\mathrm{Cl}}(\Lambda) \to \mathcal{O}^{\mathbb{C}}_M$$

is given by

$$\mathrm{str}_{\Lambda/S}\big(\hat{c}(a)\big) = (\mathbf{i}/2)^{-n/2}T\big(\hat{\sigma}(a)\big), \quad a \in \hat{\mathrm{Cl}} \tag{89}$$

where $T : \Lambda \to \mathcal{O}^{\mathbb{C}}_M$ denotes the Berezin integration associated with g. Indeed, since $[\hat{\mathrm{Cl}},\hat{\mathrm{Cl}}] = \hat{\mathrm{Cl}}_{n-1}$, see (83), both sides of (89) vanish for $a \in \hat{\mathrm{Cl}}_{n-1}$. Checking (89) on $\hat{\mathrm{Cl}}/\hat{\mathrm{Cl}}_{n-1}$ is straightforward. We will also make use of the formula

$$\mathrm{str}_{\Lambda/S}\big(\exp\big((1\otimes\hat{c})a\big)\big) = (\mathbf{i}/2)^{-n/2}T\big(\exp_{\Lambda}\big((1\otimes\hat{\sigma})a\big)\big), \quad a \in \Omega^2(M;\hat{\mathrm{Cl}}_2) \tag{90}$$

where $1\otimes\hat c : \Omega(M;\hat{\mathrm{Cl}}) \to \Omega(M;\mathrm{end}_{\mathrm{Cl}}(\Lambda))$, $1\otimes\hat\sigma : \Omega(M;\hat{\mathrm{Cl}}) \to \Omega(M;\Lambda)$ and $T : \Omega(M;\Lambda) \to \Omega(M;\mathcal{O}^{\mathbb{C}}_M)$ denotes Berezin integration. To check this equation note that the assumption on the form degree and the filtration degree of a implies:

$$\mathrm{str}_{\Lambda/S}\big(\exp\big((1\otimes\hat c)a\big)\big) = \mathrm{str}_{\Lambda/S}\bigg(\frac{1}{n!}\big((1\otimes\hat c)a\big)^{n/2}\bigg),$$
$$T\big(\exp_\Lambda\big((1\otimes\hat\sigma)a\big)\big) = T\bigg(\frac{1}{n!}\big((1\otimes\hat\sigma)a\big)^{n/2}\bigg).$$

Using the fact that $1\otimes\hat c$ is an algebra isomorphism and (89) we obtain

$$\begin{aligned}\mathrm{str}_{\Lambda/S}\big(\big((1\otimes\hat c)a\big)^{n/2}\big) &= \mathrm{str}_{\Lambda/S}\big((1\otimes\hat c)\big(a^{n/2}\big)\big)\\ &= (\mathbf{i}/2)^{-n/2}T\big((1\otimes\hat\sigma)\big(a^{n/2}\big)\big) = (\mathbf{i}/2)^{-n/2}T\big(\big((1\otimes\hat\sigma)a\big)^{n/2}\big)\end{aligned}$$

where we made use of the fact that $1\otimes\hat\sigma$ induces an isomorphism on the level of associated graded algebras, for the last equality. Combined with the previous two equations this proves Eq. (90).

Lemma 8.1. *Let (M,g) be a Riemannian manifold of even dimension n. Then*[4]

$$\mathrm{e}(g) = (2\pi\mathbf{i})^{-n/2}\,\mathrm{str}_{\Lambda/S}\big(\exp\big(-F^{\Lambda/S}_{\nabla^g}\big)\big).$$

Proof. Consider the negative of the Riemannian curvature $-R\in\Omega^2(M;\Lambda^2T^*M)$ and its exponential $\exp_\Lambda(-R)\in\Omega(M;\Lambda)$. Recall that

$$\mathrm{e}(g) := (2\pi)^{-n/2}T\big(\exp_\Lambda(-R)\big)\in\Omega^n\big(M;\mathcal{O}^{\mathbb{C}}_M\big).$$

Using (88), (90) and (87) we conclude:

$$\begin{aligned}\mathrm{e}(g) &= (2\pi)^{-n/2}T\big(\exp_\Lambda\big((1\otimes\hat\sigma)\big(2R^{\hat{\mathrm{Cl}}}\big)\big)\big)\\ &= (-\pi)^{-n/2}T\big(\exp_\Lambda\big((1\otimes\hat\sigma)\big(-R^{\hat{\mathrm{Cl}}}\big)\big)\big)\\ &= (2\pi\mathbf{i})^{-n/2}\,\mathrm{str}_{\Lambda/S}\big(\exp\big(-(1\otimes\hat c)\big(R^{\hat{\mathrm{Cl}}}\big)\big)\big)\\ &= (2\pi\mathbf{i})^{-n/2}\,\mathrm{str}_{\Lambda/S}\big(\exp\big(-F^{\Lambda/S}_{\nabla^g}\big)\big). \qquad \square\end{aligned}$$

Lemma 8.2. *Let (M,g) be a Riemannian manifold of even dimension n. Suppose $\tilde\xi\in\Gamma(T^*M\otimes T^*M)$ is symmetric, use $1\otimes\hat c : T^*M\otimes T^*M\to T^*M\otimes\mathrm{end}_{\mathrm{Cl}}(\Lambda)$ to define $V := \frac{1}{2}(1\otimes\hat c)(\tilde\xi)\in\Omega^1(M;\mathrm{end}_{\mathrm{Cl}}(\Lambda))$, and consider $\nabla^gV\in\Omega^2(M;\mathrm{end}_{\mathrm{Cl}}(\Lambda))$. Then, for every closed one form $\omega\in\Omega^1(M;\mathbb{C})$, we have*

$$\omega\wedge(\partial_2\mathrm{cs})(g,\tilde\xi) = \frac{1}{2}(2\pi\mathbf{i})^{-n/2}\,\mathrm{str}_{\Lambda/S}\big(\hat c(\omega)\wedge\exp\big(-F^{\Lambda/S}_{\nabla^g}\big)\wedge\nabla^gV\big)$$

in $\Omega^n(M;\mathcal{O}^{\mathbb{C}}_M)/d\Omega^{n-1}(M;\mathcal{O}^{\mathbb{C}}_M)$.

[4] Since the degree 0 part of $\hat A_g$ is 1, this formula is easily seen to be equivalent to $\mathrm{e}(g) = (2\pi\mathbf{i})^{-n/2}\hat A_g\wedge\mathrm{str}_{\Lambda/S}(\exp(-F^{\Lambda/S}_{\nabla^g}))$ which can be found in [1, Proposition 4.6].

Proof. Set $\tilde{M} := M \times \mathbb{R}$ and consider the two natural projections $p : \tilde{M} \to M$ and $t : \tilde{M} \to \mathbb{R}$. Consider the bundle $\tilde{T}M := p^*TM$ over $\tilde{M}$, and equip it with the fiber metric $\tilde{g} := p^*(g + t\tilde{\xi})$. For sufficiently small t, this will indeed be non-degenerate. For $t \in \mathbb{R}$ let $\mathrm{inc}_t : M \to \tilde{M}$ denote the inclusion $x \mapsto (x, t)$. Define a connection $\tilde{\nabla}$ on $\tilde{T}M$ so that $\mathrm{inc}_t^* \tilde{\nabla} = \nabla^{g+t\tilde{\xi}}$ for sufficiently small t, where $\nabla^{g+t\tilde{\xi}}$ denotes the Levi-Civita connection of $g + t\tilde{\xi}$, and so that $\tilde{\nabla}_{\partial t} = \partial_t + \frac{1}{2}\tilde{g}^{-1}(p^*\tilde{\xi})$. It is not hard to check that $\tilde{g}$ is parallel with respect to $\tilde{\nabla}$, i.e. $\tilde{\nabla}\tilde{g} = 0$. Let $\mathrm{e}(\tilde{T}M, \tilde{g}, \tilde{\nabla}) \in \Omega^n(\tilde{M}; \mathcal{O}_{\tilde{T}M})$ denote the Euler form of this Euclidean bundle. Recall that

$$\mathrm{cs}(g, g + \tau\tilde{\xi}) = \int_0^\tau \mathrm{inc}_t^* i_{\partial_t} \mathrm{e}(\tilde{T}M, \tilde{g}, \tilde{\nabla})\, dt$$

and thus

$$\begin{aligned}
(\partial_2 \mathrm{cs})(g, \tilde{\xi}) &= \mathrm{inc}_0^* i_{\partial_t} \mathrm{e}(\tilde{T}M, \tilde{g}, \tilde{\nabla}) \\
&= (2\pi)^{-n/2} \cdot \mathrm{inc}_0^* i_{\partial_t} T\bigl(\exp_\Lambda\bigl(-R^{\tilde{\nabla}}\bigr)\bigr) \\
&= (-2\pi)^{-n/2} \cdot T\bigl(\mathrm{inc}_0^* i_{\partial_t} \exp_\Lambda\bigl(R^{\tilde{\nabla}}\bigr)\bigr) \\
&= (-2\pi)^{-n/2} \cdot T\bigl(\mathrm{inc}_0^*\bigl(\exp_\Lambda\bigl(R^{\tilde{\nabla}}\bigr) \wedge i_{\partial_t} R^{\tilde{\nabla}}\bigr)\bigr) \\
&= (-2\pi)^{-n/2} \cdot T\bigl(\exp_\Lambda\bigl(\mathrm{inc}_0^* R^{\tilde{\nabla}}\bigr) \wedge \mathrm{inc}_0^* i_{\partial_t} R^{\tilde{\nabla}}\bigr)
\end{aligned}$$

where $R^{\tilde{\nabla}} \in \Omega^2(\tilde{M}; \Lambda^2\tilde{T}M)$ denotes the curvature of $\tilde{\nabla}$. Let

$$S : \Lambda \otimes \Lambda \to \Lambda \otimes \Lambda, \quad S(\alpha \otimes \beta) := (-1)^{|\alpha||\beta|}\beta \otimes \alpha$$

denote the isomorphism of graded algebras obtained by interchanging variables. Consider $\tilde{\xi} \in \Omega^1(M; T^*M)$, $\nabla^g\tilde{\xi} \in \Omega^2(M; T^*M)$ and $S(\nabla^g\tilde{\xi}) \in \Omega^1(M; \Lambda^2T^*M)$. With this notation we have

$$\begin{aligned}
\mathrm{inc}_0^* R^{\tilde{\nabla}} &= R \in \Omega^2\bigl(M; \Lambda^2T^*M\bigr), \\
\mathrm{inc}_0^* i_{\partial_t} R^{\tilde{\nabla}} &= S\biggl(\frac{1}{2}\nabla^g\tilde{\xi}\biggr) \in \Omega^1\bigl(M; \Lambda^2T^*M\bigr)
\end{aligned}$$

where R denotes the Riemannian curvature of g. We obtain

$$(\partial_2 \mathrm{cs})(g, \tilde{\xi}) = (-2\pi)^{-n/2} T\biggl(\exp_\Lambda(R) \wedge S\biggl(\frac{1}{2}\nabla^g\tilde{\xi}\biggr)\biggr)$$

and wedging with ω we get

$$\omega \wedge (\partial_2 \mathrm{cs})(g, \tilde{\xi}) = (-2\pi)^{-n/2} T\biggl(\exp_\Lambda(R) \wedge \omega \wedge S\biggl(\frac{1}{2}\nabla^g\tilde{\xi}\biggr)\biggr).$$

Next, note that for $a \in \Lambda^nT^*M \otimes \Lambda^nT^*M$ we have $T(S(a)) = T(a)$, for n is supposed to be even. Together with the symmetries of the Riemann curvature, $S(R) = R$, we obtain

$$\begin{aligned}
\omega \wedge (\partial_2 \mathrm{cs})(g, \tilde{\xi}) &= (-2\pi)^{-n/2} T\left(S\left(\exp_\Lambda(R) \wedge \omega \wedge S\left(\frac{1}{2}\nabla^g \tilde{\xi}\right)\right)\right) \\
&= (-2\pi)^{-n/2} T\left(S\big(\exp_\Lambda(R)\big) \wedge S(\omega) \wedge \frac{1}{2}\nabla^g \tilde{\xi}\right) \\
&= (-2\pi)^{-n/2} T\left(\exp_\Lambda(R) \wedge (1 \otimes \omega) \wedge \frac{1}{2}\nabla^g \tilde{\xi}\right) \\
&= (-2\pi)^{-n/2} \frac{\partial}{\partial s}\bigg|_{s=0} T\left(\exp_\Lambda\left(R + s(1\otimes\omega) \wedge \frac{1}{2}\nabla^g \tilde{\xi}\right)\right).
\end{aligned}$$

In view of (88) we have:

$$R + s(1\otimes\omega) \wedge \frac{1}{2}\big(\nabla^g \tilde{\xi}\big) = (1 \otimes \hat{\sigma})\left(-2R^{\hat{\mathrm{Cl}}} + s(1\otimes\omega)\wedge \frac{1}{2}\nabla^g\tilde{\xi}\right).$$

Moreover, using (87) and $(1\otimes\hat{c})(\frac{1}{2}\nabla^g\tilde{\xi}) = \nabla^g V$ we also have:

$$(1\otimes\hat{c})\left(-2R^{\hat{\mathrm{Cl}}} + s(1\otimes\omega)\wedge\frac{1}{2}\nabla^g\tilde{\xi}\right) = -2F^{\Lambda/S}_{\nabla^g} + s\hat{c}(\omega)\wedge\nabla^g V.$$

Using these two equations and applying (90) we obtain

$$\begin{aligned}
\omega \wedge (\partial_2 \mathrm{cs})(g, \tilde{\xi}) &= (4\pi\mathbf{i})^{-n/2} \frac{\partial}{\partial s}\bigg|_{s=0} \mathrm{str}_{\Lambda/S}\big(\exp\big(-2F^{\Lambda/S}_{\nabla^g} + s\hat{c}(\omega)\wedge\nabla^g V\big)\big) \\
&= (4\pi\mathbf{i})^{-n/2} \mathrm{str}_{\Lambda/S}\big(\exp\big(-2F^{\Lambda/S}_{\nabla^g}\big)\wedge\hat{c}(\omega)\wedge\nabla^g V\big) \\
&= \frac{1}{2}(2\pi\mathbf{i})^{-n/2} \mathrm{str}_{\Lambda/S}\big(\exp\big(-F^{\Lambda/S}_{\nabla^g}\big)\wedge\hat{c}(\omega)\wedge\nabla^g V\big) \\
&= \frac{1}{2}(2\pi\mathbf{i})^{-n/2} \mathrm{str}_{\Lambda/S}\big(\hat{c}(\omega)\wedge\exp\big(-F^{\Lambda/S}_{\nabla^g}\big)\wedge\nabla^g V\big). \qquad \square
\end{aligned}$$

8.2. The Laplacians as squares of Dirac operators

Let E be a flat complex vector bundle equipped with a fiber wise non-degenerate symmetric bilinear form b. Let ∇^E denote the flat connection on E. Consider $b^{-1}\nabla^E b \in \Omega^1(M; \mathrm{end}(E))$ and introduce the connection, cf. [2, Section 4],

$$\nabla^{E,b} := \nabla^E + \frac{1}{2}b^{-1}\nabla^E b$$

on E. Consider the Clifford bundle $\mathcal{E} := \Lambda \otimes E$ with Clifford connection

$$\nabla^{E,g,b} := \nabla^g \otimes 1_E + 1_\Lambda \otimes \nabla^{E,b}.$$

Since $(\nabla^{E,g,b})^2 = (\nabla^g)^2 + (\nabla^{E,b})^2$ the twisting curvature is

$$F^{\mathcal{E}/S}_{\nabla^{E,g,b}} = F^{\Lambda/S}_{\nabla^g} + \big(\nabla^{E,b}\big)^2. \tag{91}$$

Since the two summands commute we obtain

$$\exp\bigl(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\bigr)=\exp\bigl(-F^{\Lambda/S}_{\nabla^g}\bigr)\wedge\exp\bigl(-\bigl(\nabla^{E,b}\bigr)^2\bigr). \tag{92}$$

An easy computation shows that the Dirac operator associated to the Clifford connection $\nabla^{E,g,b}$ is

$$D_{\nabla^{E,g,b}}=d_E+d^{\sharp}_{E,g,b}+\hat{c}\left(\frac{1}{2}b^{-1}\nabla^E b\right). \tag{93}$$

Setting

$$A_{E,g,b}:=-\hat{c}\left(\frac{1}{2}b^{-1}\nabla^E b\right)\in\Omega^0\bigl(M;\mathrm{end}^-_{\mathrm{Cl}}(\mathcal{E})\bigr) \tag{94}$$

we obtain a Clifford super connection

$$\mathbb{A}_{E,g,b}:=\nabla^{E,g,b}+A_{E,g,b}. \tag{95}$$

For the associated Dirac operator $D_{\mathbb{A}_{E,g,b}}=d_E+d^{\sharp}_{E,g,b}$ we find

$$(D_{\mathbb{A}_{E,g,b}})^2=\bigl(d_E+d^{\sharp}_{E,g,b}\bigr)^2=\Delta_{E,g,b}. \tag{96}$$

So we see that the Laplacians introduced in Section 4 are indeed squares of Dirac operators of the type considered in Theorem 7.1.

8.3. Proof of Proposition 6.1

For odd n the statement follows immediately from Lemma 7.3. So let us assume that n is even. We will apply Corollary 7.4 to the Clifford super connection (95) and $U:=\phi$. From (92) and Lemma 8.1 we get:

$$\begin{aligned}\mathrm{str}_{\mathcal{E}/S}\bigl(\phi\exp\bigl(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\bigr)\bigr)&=\mathrm{str}_{\mathcal{E}/S}\bigl(\phi\exp\bigl(-F^{\Lambda/S}_{\nabla^g}\bigr)\wedge\exp\bigl(-\bigl(\nabla^{E,b}\bigr)^2\bigr)\bigr)\\&=\mathrm{str}_{\Lambda/S}\bigl(\exp\bigl(-F^{\Lambda/S}_{\nabla^g}\bigr)\bigr)\wedge\mathrm{tr}_E\bigl(\phi\exp\bigl(-\bigl(\nabla^{E,b}\bigr)^2\bigr)\bigr)=(2\pi\mathbf{i})^{n/2}\mathrm{e}(g)\,\mathrm{tr}(\phi).\end{aligned}$$

Here we also used the fact that the form $\mathrm{str}_{\Lambda/S}(\exp(-F^{\Lambda/S}_{\nabla^g}))=(2\pi\mathbf{i})^{n/2}\mathrm{e}(g)$ has degree n, and thus the only contributing part of $\mathrm{tr}_E(\phi\exp(-(\nabla^{E,b})^2))$ is the one of form degree 0, which is just $\mathrm{tr}(\phi)$. Using again the fact that $\mathrm{e}(g)$ has maximal form degree, we conclude

$$(2\pi\mathbf{i})^{-n/2}\hat{A}_g\wedge\mathrm{str}_{\mathcal{E}/S}\bigl(\phi\exp\bigl(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\bigr)\bigr)=\mathrm{tr}(\phi)\mathrm{e}(g),$$

since the degree 0 part of $\hat{A}_g$ is just 1. Proposition 6.1 now follows from Corollary 7.4 and (96).

8.4. Proof of Proposition 6.2

For odd n the statement follows immediately from Lemma 7.3. So let us assume n is even. Consider $\tilde{\xi} := g\xi \in \Gamma(T^*M \otimes T^*M)$, and use the bundle map $1 \otimes \hat{c} : T^*M \otimes T^*M \to T^*M \otimes \mathrm{end}^-_{\mathrm{Cl}}(\Lambda)$ to define

$$V := \frac{1}{2}(1 \otimes \hat{c})(\tilde{\xi}) \in \Omega^1\big(M; \mathrm{end}^-_{\mathrm{Cl}}(\Lambda)\big).$$

We claim

$$c(V) = \Lambda^*\xi - \frac{1}{2}\operatorname{tr}(\xi). \tag{97}$$

To check this let $\{e_i\}$ be a local orthonormal frame and let $\{e^i\}$ be its dual local coframe. Then

$$\Lambda^*\xi = \sum_{i,j} g(\xi e_i, e_j)\frac{1}{2}\big(\varepsilon^i \iota^j + \varepsilon^j \iota^i\big)$$

where $\varepsilon^i \in \Gamma(\mathrm{end}(\Lambda))$ denotes exterior multiplication with e^i, and $\iota^i \in \Gamma(\mathrm{end}(\Lambda))$ denotes contraction with e_i. Writing $c^i := c(e^i)$, $\hat{c}^i := \hat{c}(e^i)$ and using $\varepsilon^i = \frac{1}{2}(c^i + \hat{c}^i)$ as well as $\iota^i = -\frac{1}{2}(c^i - \hat{c}^i)$ one easily checks

$$\frac{1}{2}\big(\varepsilon^i \iota^j + \varepsilon^j \iota^i\big) = \frac{1}{4}\big(c^i \hat{c}^j + c^j \hat{c}^i\big) + \frac{1}{2}\delta^{ij}.$$

We conclude

$$\Lambda^*\xi = \sum_{i,j} g(\xi e_i, e_j)\left(\frac{1}{4}\big(c^i \hat{c}^j + c^j \hat{c}^i\big) + \frac{1}{2}\delta^{ij}\right) = \sum_{i,j} g(\xi e_i, e_j)\frac{1}{2}c^i \hat{c}^j + \frac{1}{2}\operatorname{tr}(\xi).$$

On the other hand, we clearly have $c(V) = \sum_{i,j} g(\xi e_i, e_j)\frac{1}{2}c^i\hat{c}^j$ and thus (97) is established. We will apply Corollary 7.5 to the Clifford super connection (95) and this V.

Next we claim that for all integers $k \geqslant 1$ and $l \geqslant 0$ we have

$$\mathrm{str}_{\mathcal{E}/S}\Big(\Big(\big(\mathrm{ad}\big(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\big)\big)^k A_{E,g,b}\Big) \wedge \big(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\big)^l \wedge \nabla V\Big) = 0. \tag{98}$$

To see this let us write $\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})_i$ for the subspace of $\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})$ which via the isomorphism $\hat{c} \otimes 1 : \hat{\mathrm{Cl}} \otimes \mathrm{end}(E) \to \mathrm{end}_{\mathrm{Cl}}(\Lambda) \otimes \mathrm{end}(E) = \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})$ corresponds to the filtration subspace $\hat{\mathrm{Cl}}_i \otimes \mathrm{end}(E)$. Then $-F^{\mathcal{E}/S}_{\nabla^{E,g,b}} \in \Omega^2(M; \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})_2)$, $\nabla V \in \Omega^2(M; \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})_1)$ and $A_{E,g,b} \in \Omega^0(M; \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})_1)$. Looking at the form degree, we see that (98) holds whenever $2k + 2l + 2 > n$. Moreover, since $k \geqslant 1$ we have $(\mathrm{ad}(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}))^k A_{E,g,b} \in \Omega^{2k}(M; \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})_{2k})$, for $[\hat{\mathrm{Cl}}_2, \hat{\mathrm{Cl}}_1] \subseteq \hat{\mathrm{Cl}}_2$. Thus, considering the filtration degree, we see that (98) holds whenever $2k + 2l + 1 < n$, for $\mathrm{str}_{\mathcal{E}/S}$ vanishes on $\Omega(M; \mathrm{end}_{\mathrm{Cl}}(\mathcal{E})_{n-1})$. This establishes (98). We conclude

$$\begin{aligned}&\mathrm{str}_{\mathcal{E}/S}\biggl(\biggl(\frac{e^{\mathrm{ad}(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}})}-1}{\mathrm{ad}(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}})}A_{E,g,b}\biggr)\wedge\exp\bigl(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\bigr)\wedge\nabla V\biggr)\\&\quad=\mathrm{str}_{\mathcal{E}/S}\bigl(A_{E,g,b}\wedge\exp\bigl(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\bigr)\wedge\nabla^g V\bigr).\end{aligned}\tag{99}$$

Here we wrote $\nabla V=\nabla^g V$ to emphasize that this form does not depend on the flat connection on E, but only on the Levi-Civita connection. Using (92) and $(\nabla^{E,b})^2\in\Omega^2(M;\mathrm{end}_{\mathrm{Cl}}(\mathcal{E})_0)$ and considering form and filtration degree we easily obtain:

$$\begin{aligned}&\mathrm{str}_{\mathcal{E}/S}\bigl(A_{E,g,b}\wedge\exp\bigl(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\bigr)\wedge\nabla^g V\bigr)\\&\quad=\mathrm{str}_{\mathcal{E}/S}\bigl(A_{E,g,b}\wedge\exp\bigl(-F^{\Lambda/S}_{\nabla^g}\bigr)\wedge\nabla^g V\bigr)\\&\quad=\mathrm{str}_{\Lambda/S}\bigl(\mathrm{tr}_E(A_{E,g,b})\wedge\exp\bigl(-F^{\Lambda/S}_{\nabla^g}\bigr)\wedge\nabla^g V\bigr).\end{aligned}\tag{100}$$

Using (94) and applying Lemma 8.2 to the closed one-form $\mathrm{tr}(b^{-1}\nabla^E b)$ we find

$$\begin{aligned}&\mathrm{str}_{\Lambda/S}\bigl(\mathrm{tr}_E(A_{E,g,b})\wedge\exp\bigl(-F^{\Lambda/S}_{\nabla^g}\bigr)\wedge\nabla^g V\bigr)\\&\quad=-\frac{1}{2}\,\mathrm{str}_{\Lambda/S}\bigl(\hat{c}\bigl(\mathrm{tr}\bigl(b^{-1}\nabla^E b\bigr)\bigr)\wedge\exp\bigl(-F^{\Lambda/S}_{\nabla^g}\bigr)\wedge\nabla^g V\bigr)\\&\quad=-(2\pi\mathbf{i})^{n/2}\,\mathrm{tr}\bigl(b^{-1}\nabla^E b\bigr)\wedge(\partial_2\mathrm{cs})(g,\tilde{\xi}).\end{aligned}\tag{101}$$

Combining (99)–(101) we conclude:

$$\begin{aligned}&-(2\pi\mathbf{i})^{-n/2}\hat{A}_g\wedge\mathrm{str}_{\mathcal{E}/S}\biggl(\biggl(\frac{e^{\mathrm{ad}(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}})}-1}{\mathrm{ad}(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}})}A_{E,g,b}\biggr)\wedge\exp\bigl(-F^{\mathcal{E}/S}_{\nabla^{E,g,b}}\bigr)\wedge\nabla V\biggr)\\&\quad=\mathrm{tr}\bigl(b^{-1}\nabla^E b\bigr)\wedge(\partial_2\mathrm{cs})(g,g\xi).\end{aligned}$$

Now apply Corollary 7.5 and use (97) as well as (96) to complete the proof of Proposition 6.2.

9. The case of non-vanishing Euler–Poincaré characteristics

It is not necessary to restrict to manifolds with vanishing Euler characteristics. In the general situation [9,10] Euler structures, coEuler structures, the combinatorial torsion and the analytic torsion depend on the choice of a base point. Given a path connecting two such base points everything associated with the first base point identifies in an equivariant way with the everything associated to the other base point. However, these identifications do depend on the homotopy class of such a path. Below we sketch a natural way to conveniently deal with this situation.

In general the set of Euler structures $\mathfrak{Eul}_{x_0}(M;\mathbb{Z})$ depends on a base point $x_0\in M$. One defines the set of Euler structures based at x_0 as equivalence classes $[X,c]$ where X is a vector field with non-degenerate zeros and $c\in C_1^{\mathrm{sing}}(M;\mathbb{Z})$ is such that $\partial c=\mathrm{e}(X)-\chi(M)x_0$. Two such pairs (X_1,c_1) and (X_2,c_2) are equivalent iff $c_2-c_1=\mathrm{cs}(X_1,X_2)$ mod boundaries. Again this is an affine version of $H_1(M;\mathbb{Z})$, the action is defined as in Section 2. Given a path σ from x_0 to x_1, the assignment $[X,c]\mapsto[X,c-\chi(M)\sigma]$ defines an $H_1(M;\mathbb{Z})$-equivariant isomorphism from $\mathfrak{Eul}_{x_0}(M;\mathbb{Z})$ to $\mathfrak{Eul}_{x_1}(M;\mathbb{Z})$. Since this isomorphism depends on the homotopy class of

σ only, we can consider the set of Euler structures as a flat principal bundle $\mathfrak{Eul}(M;\mathbb{Z})$ over M with structure group $H_1(M;\mathbb{Z})$. Its fiber over x_0 is just $\mathfrak{Eul}_{x_0}(M;\mathbb{Z})$, and its holonomy is given by the composition

$$\pi_1(M)\to H_1(M;\mathbb{Z})\xrightarrow{-\chi(M)} H_1(M;\mathbb{Z}).$$

Similarly, the set of Euler structures with complex coefficients can be considered as a flat principal bundle $\mathfrak{Eul}(M;\mathbb{C})$ over M with structure group $H_1(M;\mathbb{C})$ and holonomy given by the composition

$$\pi_1(M)\to H_1(M;\mathbb{Z})\xrightarrow{-\chi(M)} H_1(M;\mathbb{Z})\to H_1(M;\mathbb{C}).$$

There is an obvious parallel homomorphism of flat principal bundles over M

$$\iota:\mathfrak{Eul}(M;\mathbb{Z})\to\mathfrak{Eul}(M;\mathbb{C}) \tag{102}$$

which is equivariant over the homomorphism of structure groups $H_1(M;\mathbb{Z})\to H_1(M;\mathbb{C})$.

The set of coEuler structures $\mathfrak{Eul}^*_{x_0}(M;\mathbb{C})$ depends on the choice of a base point $x_0\in M$. It can be defined as the set of equivalence classes $[g,\alpha]$, where g is a Riemannian metric and $\alpha\in\Omega^{n-1}(M\setminus\{x_0\};\mathcal{O}_M^{\mathbb{C}})$ is such that $\mathrm{e}(g)=d\alpha$ on $M\setminus\{x_0\}$. Two such pairs $[g_1,\alpha_1]$ and $[g_2,\alpha_2]$ are equivalent iff $\alpha_2-\alpha_1=\mathrm{cs}(g_1,g_2)$ mod coboundaries, see [9, Section 3.2]. Every homotopy class of paths connecting x_0 and x_1 provides an identification between $\mathfrak{Eul}^*_{x_0}(M;\mathbb{C})$ and $\mathfrak{Eul}^*_{x_1}(M;\mathbb{C})$. Again, one can consider the set of coEuler structures as a flat principal bundle $\mathfrak{Eul}^*(M;\mathbb{C})$ over M with structure group $H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$. Its fiber over x_0 is $\mathfrak{Eul}^*_{x_0}(M;\mathbb{C})$, and its holonomy is given by the composition

$$\pi_1(M)\to H_1(M;\mathbb{Z})\xrightarrow{-\chi(M)} H_1(M;\mathbb{Z})\to H_1(M;\mathbb{C})\to H^{n-1}\big(M;\mathcal{O}_M^{\mathbb{C}}\big)$$

where the last arrow indicates Poincaré duality.

The affine version of Poincaré duality introduced in Section 2 can be consider as a parallel isomorphism of flat principal bundles over M

$$P:\mathfrak{Eul}(M;\mathbb{C})\to\mathfrak{Eul}^*(M;\mathbb{C}) \tag{103}$$

which is equivariant over the homomorphism of structure groups $H_1(M;\mathbb{C})\to H^{n-1}(M;\mathcal{O}_M^{\mathbb{C}})$ provided by Poincaré duality. We have $P([X,c])=[g,\alpha]$ iff

$$\int\limits_{M\setminus(\mathcal{X}\cup\{x_0\})}\omega\wedge\big(X^*\Psi(g)-\alpha\big)=\int\limits_c\omega$$

for all closed one forms ω which vanish in a neighborhood of $\mathcal{X}\cup\{x_0\}$.

If E is a flat complex vector bundle over M we consider the flat line bundle

$$\mathrm{Det}(M;E):=\det H^*(M;E)\otimes(\det E)^{-\chi(M)}.$$

Let $\mathrm{Det}^\times(M;E)$ denote its frame bundle, a flat principal bundle over M with structure group $\mathbb{C}^\times$ and holonomy given by

$$\pi_1(M) \to H_1(M;\mathbb{Z}) \xrightarrow{(\theta_E)^{\chi(M)}} \mathbb{C}^\times.$$

We will also consider the flat principal bundle $\mathrm{Det}^\times(M;E)^{-2}$ over M with structure group $\mathbb{C}^\times$ and holonomy given by the composition

$$\pi_1(M) \to H_1(M;\mathbb{Z}) \xrightarrow{(\theta_E)^{-2\chi(M)}} \mathbb{C}^\times.$$

Note that elements in $\mathrm{Det}^\times(M;E)^{-2}$ can be considered as non-degenerate bilinear forms on the corresponding fiber of $\mathrm{Det}(M;E)$.

The combinatorial torsion defines a parallel homomorphism of flat principal bundles

$$\tau_E^{\mathrm{comb}} : \mathfrak{Eul}(M;\mathbb{Z}) \to \mathrm{Det}^\times(M;E)^{-2} \tag{104}$$

which is equivariant over the homomorphism of structure groups

$$(\theta_E)^2 : H_1(M;\mathbb{Z}) \to \mathbb{C}^\times.$$

This formulation encodes in a rather natural way the combinatorial torsion's dependence on the Euler structure and its base point. Concerning the definition of (104), recall that the corresponding construction in Section 3 assigns to an Euler structure $\mathfrak{e}_{x_0} \in \mathfrak{Eul}_{x_0}(M;\mathbb{Z})$ and a bilinear form b_{x_0} on E_{x_0} a bilinear form on $\det H^*(M;E)$. Tensorizing this with the bilinear form on $(\det E_{x_0})^{-\chi(M)}$ induced by b_{x_0}, we obtain an element of $\mathrm{Det}^\times_{x_0}(M;E)^{-2}$ which does not depend on the choice of b_{x_0}. By definition this is the combinatorial torsion $\tau_E^{\mathrm{comb}}(\mathfrak{e}_{x_0})$ in (104).

If b is a fiber wise non-degenerate symmetric bilinear form on E, its analytic torsion provides a parallel homomorphism of flat principal bundles

$$\tau^{\mathrm{an}}_{E,[b]} : \mathfrak{Eul}^*(M;\mathbb{C}) \to \mathrm{Det}^\times(M;E)^{-2} \tag{105}$$

which is equivariant over the homomorphism of structure groups

$$H^{n-1}\big(M;\mathcal{O}_M^{\mathbb{C}}\big) \to \mathbb{C}^\times, \quad \beta \mapsto \big(e^{\langle[\omega_{E,b}]\cup\beta,[M]\rangle}\big)^2.$$

The definition of (105) is essentially the same as in Section 4. To be more precise, we represent the coEuler structure $\mathfrak{e}^*_{x_0} \in \mathfrak{Eul}^*_{x_0}(M;\mathbb{C})$ as $\mathfrak{e}^*_{x_0} = [g,\alpha]$, where $\alpha \in \Omega^{n-1}(M \setminus \{x_0\};\mathcal{O}_M^{\mathbb{C}})$ is such that $\mathrm{e}(g) = d\alpha$. We write $b_{(\det E_{x_0})^{-\chi(M)}}$ for the induced bilinear form on $(\det E_{x_0})^{-\chi(M)}$, and set

$$\tau^{\mathrm{an}}_{E,g,b,\alpha} := \tau^{\mathrm{an}}_{E,g,b}(0) \cdot \prod_q \big(\det{}'(\Delta_{E,g,b,q})\big)^{(-1)^q q} \cdot \exp\bigg(-2\int_M \omega_{E,b} \wedge \alpha\bigg) \otimes b_{(\det E_{x_0})^{-\chi(M)}}.$$

If $\chi(M) \neq 0$, then α will be singular at x_0 and the integral $\int_M \omega_{E,b} \wedge \alpha$ has to be regularized, see [9,10]. Due to this regularization the additional term $\chi(M)\,\mathrm{tr}(b_u^{-1}\dot{b}_u)(x_0)$ will appear on the right-hand side of (56) and cancel the variation of $b_{(\det E_{x_0})^{-\chi(M)}}$. Other than that the proof of

Theorem 4.2 remains the same. Thus $\tau_{E,g,b,\alpha}$ depends on E, $\mathfrak{e}^*_{x_0}$ and $[b]$ only. By definition this is the analytic torsion $\tau^{\mathrm{an}}_{E,[b]}(\mathfrak{e}^*_{x_0})$ in (105).

In this language the extension of Conjecture 5.1 to non-vanishing Euler–Poincaré characteristics asserts that for all b we have

$$\tau^{\mathrm{an}}_{E,[b]} \circ P \circ \iota = \tau^{\mathrm{comb}}_E$$

as an equality of homomorphism of principal bundles over M, see (102)–(105).

As in Section 5 one defines the relative torsion as the quotient of analytic and combinatorial torsion. This is a non-vanishing complex number independent of the Euler structure and its base point. Its properties in Proposition 5.7 remain true as stated. With little more effort one shows that the relative torsion in general is given by the formula in Proposition 5.11. Proving the generalization of Conjecture 5.1 thus amounts to show that the right-hand side of the equation in Proposition 5.11 equals 1, even if $\chi(M) \neq 0$. In view of the anomaly formula it suffices to check this for a single Riemannian metric and any representative of the homotopy class $[b]$.

Acknowledgments

We thank the referees for useful remarks and for pointing out several sign mistakes.

References

[1] N. Berline, E. Getzler, M. Vergne, Heat Kernels and Dirac Operators, Grundlehren Math. Wiss., vol. 298, Springer, Berlin, 1992.
[2] J.M. Bismut, W. Zhang, An extension of a theorem by Cheeger and Müller, Astérisque 205 (1992).
[3] M. Braverman, T. Kappeler, A Refinement of the Ray–Singer torsion, C. R. Math. Acad. Sci. Paris 341 (2005) 497–502.
[4] M. Braverman, T. Kappeler, Refined analytic torsion, preprint, math.DG/0505537.
[5] M. Braverman, T. Kappeler, Refined analytic torsion as an element of the determinant line, preprint, math.GT/0510532.
[6] M. Braverman, T. Kappeler, Ray–Singer type theorem for the refined analytic torsion, preprint, math.GT/0603638.
[7] M. Braverman, T. Kappeler, Comparison of the refined analytic and the Burghelea–Haller torsions, preprint, math.DG/0606398.
[8] D. Burghelea, Removing metric anomalies from Ray–Singer torsion, Lett. Math. Phys. 47 (1999) 149–158.
[9] D. Burghelea, S. Haller, Euler structures, variety of representations and the Milnor–Turaev torsion, Geom. Topol. 10 (2006) 1185–1238.
[10] D. Burghelea, S. Haller, A Riemannian invariant, Euler structures and some topological applications, in: Proceedings of the Conference C^*-Algebras and Elliptic Theory, Będlewo, Poland, 2004, in: Trends Math., Birkhäuser, Basel, 2006, pp. 37–60.
[11] D. Burghelea, S. Haller, Dynamics, Laplace transform and spectral geometry, preprint, math.DG/0508216.
[12] D. Burghelea, S. Haller, Torsion, as function on the space of representations, preprint, math.DG/0507587.
[13] D. Burghelea, S. Haller, Complex valued Ray–Singer torsion II, preprint, math.DG/0610875.
[14] D. Burghelea, L. Friedlander, T. Kappeler, On the determinant of elliptic differential and finite difference operators in vector bundles over S^1, Comm. Math. Phys. 138 (1991) 1–18.
[15] D. Burghelea, L. Friedlander, T. Kappeler, Asymptotic expansion of the Witten deformation of the analytic torsion, J. Funct. Anal. 137 (1996) 320–363.
[16] J. Cheeger, Analytic torsion and Reidemeister torsion, Proc. Natl. Acad. Sci. USA 74 (1977) 2651–2654.
[17] J. Cheeger, Analytic torsion and the heat equation, Ann. of Math. (2) 109 (1979) 259–322.
[18] S.S. Chern, J. Simons, Characteristic forms and geometric invariants, Ann. of Math. (2) 99 (1974) 48–69.
[19] M. Farber, V. Turaev, Poincaré–Reidemeister metric, Euler structures and torsion, J. Reine Angew. Math. 520 (2000) 195–225.

[20] D. Fried, Lefschetz formulas for flows, in: The Lefschetz Centennial Conference, Part III, Mexico City, 1984, in: Contemp. Math., vol. 58, Amer. Math. Soc., Providence, RI, 1987, pp. 19–69.
[21] E. Getzler, A short proof of the local Atiyah–Singer index theorem, Topology 25 (1986) 111–117.
[22] R.-T. Huang, A note on Burghelea–Haller conjecture, preprint, 2006.
[23] J. Jost, Riemannian Geometry and Geometric Analysis, Universitext, Springer, Berlin, 2005.
[24] F.F. Knudson, D. Mumford, The projectivity of the moduli space of stable curves, I: Preliminaries on "det" and "div", Math. Scand. 39 (1976) 19–55.
[25] F. Laudenbach, On the Thom–Smale complex, Astérisque 205 (1992); an appendix to [2].
[26] V. Mathai, D. Quillen, Superconnections, Thom classes, and equivariant differential forms, Topology 25 (1986) 85–110.
[27] J. Milnor, Whitehead torsion, Bull. Amer. Math. Soc. 72 (1966) 358–426.
[28] W. Müller, Analytic torsion and R-torsion of Riemannian manifolds, Adv. Math. 28 (1978) 233–305.
[29] D.B. Ray, I.M. Singer, R-torsion and the Laplacian on Riemannian manifolds, Adv. Math. 7 (1971) 145–210.
[30] M. Schwarz, Morse Homology, Progr. Math., vol. 111, Birkhäuser, Basel, 1993.
[31] R.T. Seeley, Complex powers of an elliptic operator, in: Proc. Sympos. Pure Math., vol. 10, Amer. Math. Soc., Providence, RI, 1967, pp. 288–307.
[32] M.A. Shubin, Pseudodifferential Operators and Spectral Theory, second ed., Springer, Berlin, 2001; translated from the 1978 Russian original by S.I. Andersson.
[33] G. Su, W. Zhang, A Cheeger–Müller theorem for symmetric bilinear torsions, preprint, math.DG/0610577.
[34] V. Turaev, Euler structures, nonsingular vector fields, and Reidemeister-type torsions, Math. USSR–Izv. 34 (1990) 627–662.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 248 (2007) 79–106
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

The dilation property of modulation spaces and their inclusion relation with Besov spaces

Mitsuru Sugimoto, Naohito Tomita *

Department of Mathematics, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

Received 5 June 2006; accepted 17 March 2007

Available online 7 May 2007

Communicated by C. Kenig

Abstract

We consider the dilation property of the modulation spaces $M^{p,q}$. Let $D_\lambda : f(t) \mapsto f(\lambda t)$ be the dilation operator, and we consider the behavior of the operator norm $\|D_\lambda\|_{M^{p,q} \to M^{p,q}}$ with respect to λ. Our result determines the best order for it, and as an application, we establish the optimality of the inclusion relation between the modulation spaces and Besov spaces, which was proved by Toft [J. Toft, Continuity properties for modulation spaces, with applications to pseudo-differential calculus, I, J. Funct. Anal. 207 (2004) 399–429].

Keywords: Modulation spaces; Besov spaces; Dilation; Inclusion

1. Introduction

The modulation spaces $M^{p,q}$ were first introduced by Feichtinger [3,4] and generalized by Feichtinger and Gröchenig [6]. The exact definition will be given in the next section, but the main idea is to consider the decaying property of a function with respect to the space variable and the variable of its Fourier transform simultaneously. That is exactly the heart of the matter of the time–frequency analysis which is originated in signal analysis or quantum mechanics.

* Corresponding author.
E-mail addresses: sugimoto@math.wani.osaka-u.ac.jp (M. Sugimoto), tomita@gaia.math.wani.osaka-u.ac.jp (N. Tomita).

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2007.03.015

Based on a similar idea, Sjöstrand [15] independently introduced a symbol class which assures the L^2-boundedness of corresponding pseudo-differential operators. In the last decade, the theory of the modulation spaces has been developed, and its usefulness for the theory of pseudo-differential operators is getting realized gradually. Nowadays Sjöstrand's symbol class is recognized as a special case of the modulation spaces, and many authors used these spaces, as a powerful tool, to discuss the boundedness or compactness properties of pseudo-differential operators. See, for example, Boulkhemair [1], Gröchenig [11], Gröchenig and Heil [12,13], and Toft [18,19]. Consult Feichtinger [5], Gröchenig [10], and Teofanov [17] for further and detailed history of this research fields. Some arguments in these works have their origin in the field of phase space analysis. See also Dimassi and Sjöstrand [2] and Folland [7] for this direction.

Now we are in a situation to start showing fundamental properties of the modulation spaces, in order to apply them to many other problems. Actually in Toft's recent work [18], he investigated the mapping property of convolutions, and showed Young-type results for the modulation spaces. As an application, he showed an inclusion relation between the modulation spaces and Besov spaces. We also mention that some extensions to weighted modulation spaces of the inclusion can be found in Toft [19,20]. We remark that Besov spaces are used in various problems of partial differential equations, and his result will help us to understand how they are translated into the terminology of modulation spaces.

Among many other important properties to be shown, we focus on the dilation property of the modulation spaces in this article. Since $M^{2,2}(\mathbb{R}^n) = L^2(\mathbb{R}^n)$, we have easily $\|f_\lambda\|_{M^{2,2}} = \lambda^{-n/2}\|f\|_{M^{2,2}}$ by the change of variables $t \mapsto \lambda^{-1}t$, where $f_\lambda(t) = f(\lambda t)$ and $t \in \mathbb{R}^n$. But it is not clear how $\|f_\lambda\|_{M^{p,q}}$ behaves like with respect to λ except for the case $(p,q) = (2,2)$. Our objective is to draw the complete picture of the best order of λ for every pair of (p,q) (Theorem 1.1).

We can expect various kinds of applications of this consideration. In fact, this kind of dilation property is frequently used in the "scaling argument," which is a popular tool to know the best possible order of the conditions in problems of partial differential equations. Actually, in this article, we also show the best possibility of Toft's inclusion relation mentioned above, as a side product of the main argument (Theorem 1.2).

In order to state our main results, we introduce several indices. For $1 \leqslant p \leqslant \infty$, we denote by p' the conjugate exponent of p (that is, $1/p + 1/p' = 1$). We define subsets of $(1/p, 1/q) \in [0,1] \times [0,1]$ in the following way:

$$\begin{aligned} &I_1\colon \max(1/p, 1/p') \leqslant 1/q, && I_1^*\colon \min(1/p, 1/p') \geqslant 1/q, \\ &I_2\colon \max(1/q, 1/2) \leqslant 1/p', && I_2^*\colon \min(1/q, 1/2) \geqslant 1/p', \\ &I_3\colon \max(1/q, 1/2) \leqslant 1/p, && I_3^*\colon \min(1/q, 1/2) \geqslant 1/p. \end{aligned}$$

Let us consider Fig. 1. In [18], Toft introduced the indices

$$\begin{aligned} \nu_1(p,q) &= \max\{0, 1/q - \min(1/p, 1/p')\}, \\ \nu_2(p,q) &= \min\{0, 1/q - \max(1/p, 1/p')\}. \end{aligned}$$

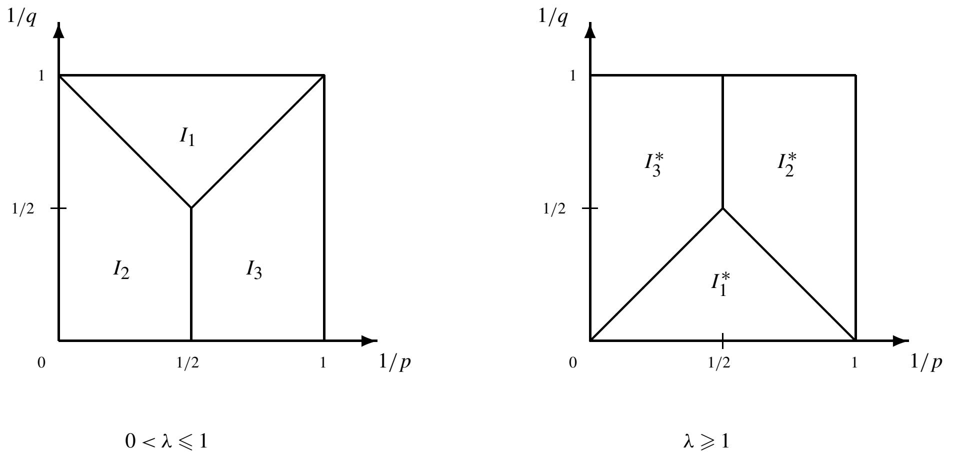

Fig. 1.

Note that

$$\nu_1(p,q)=\begin{cases}0 & \text{if } (1/p,1/q)\in I_1^*,\\ 1/p+1/q-1 & \text{if } (1/p,1/q)\in I_2^*,\\ -1/p+1/q & \text{if } (1/p,1/q)\in I_3^*,\end{cases}$$

and

$$\nu_2(p,q)=\begin{cases}0 & \text{if } (1/p,1/q)\in I_1,\\ 1/p+1/q-1 & \text{if } (1/p,1/q)\in I_2,\\ -1/p+1/q & \text{if } (1/p,1/q)\in I_3.\end{cases}$$

We also introduce the indices

$$\mu_1(p,q)=\nu_1(p,q)-1/p, \qquad \mu_2(p,q)=\nu_2(p,q)-1/p.$$

Then we have

$$\mu_1(p,q)=\begin{cases}-1/p & \text{if } (1/p,1/q)\in I_1^*,\\ 1/q-1 & \text{if } (1/p,1/q)\in I_2^*,\\ -2/p+1/q & \text{if } (1/p,1/q)\in I_3^*,\end{cases}$$

and

$$\mu_2(p,q)=\begin{cases}-1/p & \text{if } (1/p,1/q)\in I_1,\\ 1/q-1 & \text{if } (1/p,1/q)\in I_2,\\ -2/p+1/q & \text{if } (1/p,1/q)\in I_3.\end{cases}$$

Our first main result is on the dilation property of the modulation spaces. For a function (or tempered distribution) f on $\mathbb{R}^n$ and $\lambda>0$, we use the notation f_λ which is defined by $f_\lambda(t)=f(\lambda t)$, $t\in\mathbb{R}^n$.

Theorem 1.1. *Let $1 \leqslant p, q \leqslant \infty$. Then the following are true*:

(1) *There exists a constant $C > 0$ such that*

$$C^{-1}\lambda^{n\mu_2(p,q)}\|f\|_{M^{p,q}} \leqslant \|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{n\mu_1(p,q)}\|f\|_{M^{p,q}} \tag{1.1}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $\lambda \geqslant 1$. Conversely, if there exist constants $C > 0$ and $\alpha, \beta \in \mathbb{R}$ such that

$$C^{-1}\lambda^{\beta}\|f\|_{M^{p,q}} \leqslant \|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{\alpha}\|f\|_{M^{p,q}}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $\lambda \geqslant 1$, then $\alpha \geqslant n\mu_1(p,q)$ and $\beta \leqslant n\mu_2(p,q)$.

(2) *There exists a constant $C > 0$ such that*

$$C^{-1}\lambda^{n\mu_1(p,q)}\|f\|_{M^{p,q}} \leqslant \|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{n\mu_2(p,q)}\|f\|_{M^{p,q}} \tag{1.2}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $0 < \lambda \leqslant 1$. Conversely, if there exist constants $C > 0$ and $\alpha, \beta \in \mathbb{R}$ such that

$$C^{-1}\lambda^{\alpha}\|f\|_{M^{p,q}} \leqslant \|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{\beta}\|f\|_{M^{p,q}}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $0 < \lambda \leqslant 1$, then $\alpha \geqslant n\mu_1(p,q)$ and $\beta \leqslant n\mu_2(p,q)$.

Since the Gauss function $\varphi(t) = e^{-|t|^2}$ does not change its form under the Fourier transformation, the modulation norm of it can have a "good" property. In this sense, it is reasonable to believe that the Gauss function $f = \varphi$ attains the critical order of $\|f_\lambda\|_{M^{p,q}}$ with respect to λ. But it is not true because $\|\varphi_\lambda\|_{M^{p,q}} \sim \lambda^{n(1/q-1)}$ in the case $\lambda \geqslant 1$ and $\|\varphi_\lambda\|_{M^{p,q}} \sim \lambda^{-n/p}$ in the case $0 < \lambda \leqslant 1$ (see Lemma 2.1). Theorem 1.1 says that they are not critical orders for every pair of (p,q).

It should be pointed out here that the behavior of $\|f_\lambda\|_{M^{p,q}}$ with respect to λ might depend on the choice of $f \in M^{p,q}(\mathbb{R}^n)$. In fact, $f(t) = \sum_{k\in\mathbb{Z}^n} e^{ik\cdot t}\psi(t-k)$, where ψ is an appropriate Schwartz function, has the property $\|f_\lambda\|_{M^{p,\infty}} \sim \lambda^{-2n/p}$ $(0 < \lambda \leqslant 1)$ in the case $1 \leqslant p \leqslant 2$ (Lemma 3.10), while the Gauss function has the different behavior $\|\varphi_\lambda\|_{M^{p,\infty}} \sim \lambda^{-n/p}$ $(0 < \lambda \leqslant 1)$ as mentioned above. On the other hand, the L^p-norm never has such a property since $\|f_\lambda\|_{L^p} = \lambda^{-n/p}\|f\|_{L^p}$ for all $f \in L^p(\mathbb{R}^n)$. That is one of great differences between the modulation spaces and L^p-spaces.

Our second main result is on the optimality of the inclusion relation between the modulation spaces and Besov spaces. In [18, Theorem 3.1], Toft proved the inclusions

$$B^{p,q}_{n\nu_1(p,q)}(\mathbb{R}^n) \hookrightarrow M^{p,q}(\mathbb{R}^n) \hookrightarrow B^{p,q}_{n\nu_2(p,q)}(\mathbb{R}^n)$$

for $1 \leqslant p, q \leqslant \infty$. See also [19, Theorem 2.10] for the case of weighted modulation spaces, and some related results can be seen in Gröbner [9] and Okoudjou [14]. Toft also remarked that the left inclusion is optimal in the case $1 \leqslant p = q \leqslant 2$, that is, if $B^{p,p}_{s_1}(\mathbb{R}^n) \hookrightarrow M^{p,p}(\mathbb{R}^n)$ then $s_1 \geqslant n\nu_1(p,p)$. The same is true for the right inclusion in the case $2 \leqslant p = q \leqslant \infty$, that is, if $M^{p,p}(\mathbb{R}^n) \hookrightarrow B^{p,p}_{s_2}(\mathbb{R}^n)$ then $s_2 \leqslant n\nu_2(p,p)$ [18, Remark 3.11]. The next theorem says that Toft's inclusion result is optimal in the above meaning for every pair of (p,q).

Theorem 1.2. *Let* $1 \leqslant p, q \leqslant \infty$ *and* $s \in \mathbb{R}$. *Then the following are true*:

(1) *If* $B_s^{p,q}(\mathbb{R}^n) \hookrightarrow M^{p,q}(\mathbb{R}^n)$, *then* $s \geqslant n\nu_1(p,q)$.
(2) *If* $M^{p,q}(\mathbb{R}^n) \hookrightarrow B_s^{p,q}(\mathbb{R}^n)$ *and* $1 \leqslant p, q < \infty$, *then* $s \leqslant n\nu_2(p,q)$.

During the evaluation process of this paper, a preprint of the independent work by Wang and Huang [22] was sent to the authors, where we can find a related result of Theorem 1.2.

We end this introduction by explaining the plan of this article. In Section 2, we give the precise definition and basic properties of the modulation spaces and Besov spaces. In Sections 3 and 4, we prove Theorems 1.1 and 1.2, respectively.

2. Preliminaries

We introduce the modulation spaces based on Gröchenig [10]. Let $\mathcal{S}(\mathbb{R}^n)$ and $\mathcal{S}'(\mathbb{R}^n)$ be the Schwartz spaces of rapidly decreasing smooth functions and tempered distributions, respectively. We define the Fourier transform $\widehat{f}$ and the inverse Fourier transform $\mathcal{F}^{-1}f$ of $f \in \mathcal{S}(\mathbb{R}^n)$ by

$$\widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{-i\xi\cdot x} f(x)\,dx \quad \text{and} \quad \mathcal{F}^{-1}f(x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{ix\cdot\xi} f(\xi)\,d\xi.$$

Fix a function $\varphi \in \mathcal{S}(\mathbb{R}^n) \setminus \{0\}$ (called the *window function*). Then the short-time Fourier transform $V_\varphi f$ of $f \in \mathcal{S}'(\mathbb{R}^n)$ with respect to φ is defined by

$$V_\varphi f(x,\xi) = \langle f, M_\xi T_x \varphi\rangle \quad \text{for } x, \xi \in \mathbb{R}^n,$$

where $M_\xi\varphi(t) = e^{i\xi\cdot t}\varphi(t)$, $T_x\varphi(t) = \varphi(t-x)$, and $\langle\cdot,\cdot\rangle$ is the inner product on $L^2(\mathbb{R}^n)$. We can express it in a form of the integral

$$V_\varphi f(x,\xi) = \int_{\mathbb{R}^n} f(t)\,\overline{\varphi(t-x)}e^{-i\xi\cdot t}\,dt,$$

which has actually the meaning for an appropriate function f on $\mathbb{R}^n$. We note that, for $f \in \mathcal{S}'(\mathbb{R}^n)$, $V_\varphi f$ is continuous on $\mathbb{R}^{2n}$ and $|V_\varphi f(x,\xi)| \leqslant C(1+|x|+|\xi|)^N$ for some constants $C, N \geqslant 0$ [10, Theorem 11.2.3]. Let $1 \leqslant p, q \leqslant \infty$. Then the modulation space $M^{p,q}(\mathbb{R}^n)$ consists of all $f \in \mathcal{S}'(\mathbb{R}^n)$ such that

$$\|f\|_{M^{p,q}} = \|V_\varphi f\|_{L^{p,q}} = \left\{\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\left|V_\varphi f(x,\xi)\right|^p dx\right)^{q/p} d\xi\right\}^{1/q} < \infty,$$

(with usual modification when $p = \infty$ or $q = \infty$). We note that $M^{2,2}(\mathbb{R}^n) = L^2(\mathbb{R}^n)$ [10, Proposition 11.3.1] and $M^{p,q}(\mathbb{R}^n)$ is a Banach space [10, Proposition 11.3.5]. The definition of $M^{p,q}(\mathbb{R}^n)$ is independent of the choice of the window function $\varphi \in \mathcal{S}(\mathbb{R}^n) \setminus \{0\}$, that is, different window functions yield equivalent norms [10, Proposition 11.3.2].

We also introduce Besov spaces. Let $1 \leqslant p, q \leqslant \infty$ and $s \in \mathbb{R}$. Suppose that $\varphi_0, \varphi \in \mathcal{S}(\mathbb{R}^n)$ satisfy $\operatorname{supp}\varphi_0 \subset \{\xi\colon |\xi| \leqslant 2\}$, $\operatorname{supp}\varphi \subset \{\xi\colon 1/2 \leqslant |\xi| \leqslant 2\}$ and $\varphi_0(\xi) + \sum_{j=1}^\infty \varphi(\xi/2^j) = 1$ for all $\xi \in \mathbb{R}^n$. Set $\varphi_j = \varphi(\cdot/2^j)$ if $j \geqslant 1$. Then the Besov space $B_s^{p,q}(\mathbb{R}^n)$ consists of all $f \in \mathcal{S}'(\mathbb{R}^n)$ such that

$$\|f\|_{B_s^{p,q}} = \left(\sum_{j=0}^{\infty} 2^{jsq} \|\mathcal{F}^{-1}[\varphi_j \widehat{f}]\|_{L^p}^q\right)^{1/q} = \left(\sum_{j=0}^{\infty} 2^{jsq} \|\Phi_j * f\|_{L^p}^q\right)^{1/q} < \infty,$$

where $\Phi_j = \mathcal{F}^{-1}\varphi_j$ (with usual modification again when $q = \infty$). We remark that $B_s^{p,q}(\mathbb{R}^n)^* = B_{-s}^{p',q'}(\mathbb{R}^n)$ for $1 \leqslant p, q < \infty$.

Finally, we list below the lemmas which will be used in the subsequent section. In this article, we frequently use the Gauss function $\varphi(t) = e^{-|t|^2}$.

Lemma 2.1. *(See [18, Lemma 1.8].) Let φ be the Gauss function. Then*

$$\big\| V_\varphi(\varphi_\lambda)\big\|_{L^{p,q}} = \pi^{n(1/p+1/q+1)/2} p^{-n/2p} q^{-n/2q} 2^{n/q} \lambda^{-n/p} \big(1+\lambda^2\big)^{n(1/p+1/q-1)/2}.$$

Lemma 2.1 says that $\|\varphi_\lambda\|_{M^{p,q}} \sim \lambda^{n(1/q-1)}$ in the case $\lambda \geqslant 1$ and $\|\varphi_\lambda\|_{M^{p,q}} \sim \lambda^{-n/p}$ in the case $0 < \lambda \leqslant 1$.

Lemma 2.2. *(See [10, Corollary 11.2.7].) Let $f \in \mathcal{S}'(\mathbb{R}^n)$ and $\varphi, \psi, \gamma \in \mathcal{S}(\mathbb{R}^n)$. Then*

$$\langle f, \varphi\rangle = \frac{1}{\langle \gamma, \psi\rangle} \int_{\mathbb{R}^{2n}} V_\psi f(x,\xi) \overline{V_\gamma \varphi(x,\xi)}\, dx\, d\xi \quad \textit{for all } \varphi \in \mathcal{S}\big(\mathbb{R}^n\big).$$

We remark that Lemma 2.2 is also found in Folland [7, Proposition 1.92].

Lemma 2.3. *(See [10, Lemma 11.3.3].) Let $f \in \mathcal{S}'(\mathbb{R}^n)$ and $\varphi, \psi, \gamma \in \mathcal{S}(\mathbb{R}^n)$. Then*

$$\big|V_\varphi f(x,\xi)\big| \leqslant \frac{1}{|\langle \gamma, \psi\rangle|} \big(|V_\psi f| * |V_\varphi \gamma|\big)(x,\xi) \quad \textit{for all } x, \xi \in \mathbb{R}^n.$$

Lemma 2.4. *(See [10, Proposition 11.3.4, Theorem 11.3.6].) Let $1 \leqslant p, q < \infty$. Then $\mathcal{S}(\mathbb{R}^n)$ is dense in $M^{p,q}(\mathbb{R}^n)$ and $M^{p,q}(\mathbb{R}^n)^* = M^{p',q'}(\mathbb{R}^n)$ under the duality*

$$\langle f, g\rangle_M = \frac{1}{\|\varphi\|_{L^2}^2} \int_{\mathbb{R}^{2n}} V_\varphi f(x,\xi) \overline{V_\varphi g(x,\xi)}\, dx\, d\xi$$

for $f \in M^{p,q}(\mathbb{R}^n)$ and $g \in M^{p',q'}(\mathbb{R}^n)$.

By Lemmas 2.2 and 2.4, if $1 < p, q \leqslant \infty$ and $f \in M^{p,q}(\mathbb{R}^n)$ then

$$\|f\|_{M^{p,q}} = \sup\big|\langle f, g\rangle_M\big| = \sup\big|\langle f, g\rangle\big|, \tag{2.1}$$

where the supremum is taken over all $g \in \mathcal{S}(\mathbb{R}^n)$ such that $\|g\|_{M^{p',q'}} = 1$.

Lemma 2.5. *(See [3, Corollary 2.3].) Let* $1 \leqslant p_1, p_2, q_1, q_2 \leqslant \infty$ *and* $p_2, q_2 < \infty$. *If* T *is a linear operator such that*

$$\|Tf\|_{M^{p_1,q_1}} \leqslant A_1 \|f\|_{M^{p_1,q_1}} \quad \text{for all } f \in M^{p_1,q_1}(\mathbb{R}^n)$$

and

$$\|Tf\|_{M^{p_2,q_2}} \leqslant A_2 \|f\|_{M^{p_2,q_2}} \quad \text{for all } f \in M^{p_2,q_2}(\mathbb{R}^n),$$

then

$$\|Tf\|_{M^{p,q}} \leqslant C A_1^{1-\theta} A_2^{\theta} \|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n),$$

where $1/p = (1-\theta)/p_1 + \theta/p_2$, $1/q = (1-\theta)/q_1 + \theta/q_2$, $0 \leqslant \theta \leqslant 1$ *and* C *is independent of* T.

Remark 2.6. Lemma 2.5 with the cases $p_2 = \infty$ or $q_2 = \infty$ is treated in [18, Remark 3.2], which says that it is true under a modification.

3. The dilation property of modulation spaces

In this section, we prove Theorem 1.1 which appeared in Section 1. We remark that the left-hand sides of inequalities in Theorem 1.1 are obtained from the right-hand sides of them.

Theorem 3.1. *Let* $1 \leqslant p, q \leqslant \infty$. *Then the following are true*:

(1) *There exists a constant* $C > 0$ *such that*

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{n\mu_1(p,q)} \|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } \lambda \geqslant 1.$$

Conversely, if there exist constants $C > 0$ *and* $\alpha \in \mathbb{R}$ *such that*

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{\alpha} \|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } \lambda \geqslant 1,$$

then $\alpha \geqslant n\mu_1(p,q)$.

(2) *There exists a constant* $C > 0$ *such that*

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{n\mu_2(p,q)} \|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1.$$

Conversely, if there exist constants $C > 0$ *and* $\beta \in \mathbb{R}$ *such that*

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{\beta} \|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1,$$

then $\beta \leqslant n\mu_2(p,q)$.

Indeed, if $0 < \lambda \leqslant 1$, then the first part of Theorem 3.1(1) gives

$$\|f\|_{M^{p,q}} = \left\|(f_\lambda)_{1/\lambda}\right\|_{M^{p,q}} \leqslant C\lambda^{-n\mu_1(p,q)}\|f_\lambda\|_{M^{p,q}}$$

which proves the left-hand side of (1.2) in Theorem 1.1. The others in Theorem 1.1 are given by Theorem 3.1 in a similar way. We also remark that Boulkhemair [1, Proposition 3.2] proved the first part of Theorem 3.1(2) with $(p,q) = (\infty, 1)$.

Now we prove Theorem 3.1. We begin with the following preparing lemma which might be well known.

Lemma 3.2. *Let $1 \leqslant p, q \leqslant \infty$. Then there exists a constant $C > 0$ which only depends on the window functions in the modulation space norms such that*

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{-n(1/p-1/q+1)}\left(1+\lambda^2\right)^{n/2}\|f\|_{M^{p,q}}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $\lambda > 0$.

Although the proof of Lemma 3.2 might be found in some literature, we provide it for reader's convenience. Here (and also in other situations) we may assume that the window function is given by the Gauss function $\varphi(t) = e^{-|t|^2}$.

Proof. Let φ be the Gauss function, that is, $\varphi(t) = e^{-|t|^2}$. By a change of variable, we have

$$\|f_\lambda\|_{M^{p,q}} = \left\|V_\varphi(f_\lambda)\right\|_{L^{p,q}} = \lambda^{-n(1/p-1/q+1)}\|V_{\varphi_{1/\lambda}} f\|_{L^{p,q}}.$$

From Lemma 2.3 it follows that

$$\left|V_{\varphi_{1/\lambda}} f(x,\xi)\right| \leqslant \|\varphi\|_{L^2}^{-2}\left(|V_\varphi f| * |V_{\varphi_{1/\lambda}}\varphi|\right)(x,\xi).$$

Hence, by Young's inequality and Lemma 2.1, we get

$$\begin{aligned}
\|f_\lambda\|_{M^{p,q}} &\leqslant \lambda^{-n(1/p-1/q+1)}\|\varphi\|_{L^2}^{-2}\|V_{\varphi_{1/\lambda}}\varphi\|_{L^{1,1}}\|V_\varphi f\|_{L^{p,q}} \\
&= \lambda^{-n(1/p-1/q+1)}\|\varphi\|_{L^2}^{-2}\left\|V_\varphi(\varphi_{1/\lambda})\right\|_{L^{1,1}}\|V_\varphi f\|_{L^{p,q}} \\
&= \lambda^{-n(1/p-1/q+1)}\|\varphi\|_{L^2}^{-2}\left(\pi^{3n/2}2^n\left(\lambda^{-1}\right)^{-n}\left(1+\lambda^{-2}\right)^{n/2}\right)\|f\|_{M^{p,q}} \\
&= C_{n,\varphi}\lambda^{-n(1/p-1/q+1)}\left(1+\lambda^2\right)^{n/2}\|f\|_{M^{p,q}}.
\end{aligned}$$

The proof is complete. □

We are now ready to prove Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$ and Theorem 3.1(2) with $(1/p, 1/q) \in I_1$.

Proof of Theorem 3.1(2) with $(1/p, 1/q) \in I_1$. Let $(1/p, 1/q) \in I_1$. Then $\mu_2(p,q) = -1/p$. By Lemma 3.2, we have

$$\|f_\lambda\|_{M^{r,1}} \leqslant C\lambda^{-n/r}\|f\|_{M^{r,1}} \quad \text{for all } f \in M^{r,1}\left(\mathbb{R}^n\right) \text{ and } 0 < \lambda \leqslant 1, \tag{3.1}$$

where $1 \leqslant r \leqslant \infty$. On the other hand, since $M^{2,2}(\mathbb{R}^n) = L^2(\mathbb{R}^n)$, we have

$$\|f_\lambda\|_{M^{2,2}} \leqslant C\lambda^{-n/2}\|f\|_{M^{2,2}} \quad \text{for all } f \in M^{2,2}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1. \tag{3.2}$$

Take $1 \leqslant r \leqslant \infty$ and $0 \leqslant \theta \leqslant 1$ such that $1/p = (1-\theta)/r + \theta/2$ and $1/q = (1-\theta)/1 + \theta/2$. Then, by interpolation (Lemma 2.5), (3.1) and (3.2) give

$$\|f_\lambda\|_{M^{p,q}} \leqslant C(\lambda^{-n/r})^{1-\theta}(\lambda^{-n/2})^\theta\|f\|_{M^{p,q}}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $0 < \lambda \leqslant 1$. Since $(1-\theta)/r = 1/p + 1/q - 1$ and $\theta/2 = -1/q + 1$, we get

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{-n/p}\|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1. \tag{3.3}$$

This is the first part of Theorem 3.1(2) with $(1/p, 1/q) \in I_1$.

We next prove the second part of Theorem 3.1(2) with $(1/p, 1/q) \in I_1$. Let $(1/p, 1/q) \in I_1$. Assume that there exist constants $C > 0$ and $\beta \in \mathbb{R}$ such that

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^\beta\|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1.$$

Let φ be the Gauss function. We note that the Gauss function belongs to $M^{p,q}(\mathbb{R}^n)$. Then, by Lemma 2.1 and our assumption, we have

$$\begin{aligned} C_{p,q}\lambda^{-n/p} &\leqslant C_{p,q}\lambda^{-n/p}(1+\lambda^2)^{n(1/p+1/q-1)/2} \\ &= \|V_\varphi(\varphi_\lambda)\|_{L^{p,q}} = \|\varphi_\lambda\|_{M^{p,q}} \leqslant C\lambda^\beta\|\varphi\|_{M^{p,q}} \end{aligned}$$

for all $0 < \lambda \leqslant 1$. This is possible only if $\beta \leqslant -n/p$. The proof is complete. □

Proof of Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$. We recall that $\mu_1(p,q) = -1/p$ if $(1/p, 1/q) \in I_1^*$. Let $(1/p, 1/q) \in I_1^*$. Then $(1/p', 1/q') \in I_1$. We first consider the case $p \neq 1$. Since $1 < p, q \leqslant \infty$, by duality (2.1) and Theorem 3.1(2) with $(1/p', 1/q') \in I_1$, we have

$$\begin{aligned} \|f_\lambda\|_{M^{p,q}} &= \sup|\langle f_\lambda, g\rangle| = \lambda^{-n}\sup|\langle f, g_{1/\lambda}\rangle| \\ &\leqslant \lambda^{-n}\sup\|f\|_{M^{p,q}}\|g_{1/\lambda}\|_{M^{p',q'}} \\ &\leqslant \lambda^{-n}\sup\|f\|_{M^{p,q}}\left(C(\lambda^{-1})^{-n/p'}\|g\|_{M^{p',q'}}\right) = C\lambda^{-n/p}\|f\|_{M^{p,q}} \end{aligned}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $\lambda \geqslant 1$, where the supremum is taken over all $g \in \mathcal{S}(\mathbb{R}^n)$ such that $\|g\|_{M^{p',q'}} = 1$. In the case $p = 1$, by Lemma 3.2, we see that

$$\|f_\lambda\|_{M^{1,\infty}} \leqslant C\lambda^{-n}\|f\|_{M^{1,\infty}} \quad \text{for all } f \in M^{1,\infty}(\mathbb{R}^n) \text{ and } \lambda \geqslant 1.$$

Hence, we obtain the first part of Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$.

We consider the second part of Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$. Let $(1/p, 1/q) \in I_1^*$ and $q < \infty$. Note that $1 < p, q < \infty$. Assume that there exist constants $C > 0$ and $\alpha \in \mathbb{R}$ such that

$$\|g_\lambda\|_{M^{p,q}} \leqslant C\lambda^\alpha \|g\|_{M^{p,q}} \quad \text{for all } g \in M^{p,q}(\mathbb{R}^n) \text{ and } \lambda \geqslant 1.$$

Then, by duality and our assumption, we have

$$\begin{aligned}
\|f_\lambda\|_{M^{p',q'}} &= \sup\bigl|\langle f_\lambda, g\rangle\bigr| = \lambda^{-n} \sup\bigl|\langle f, g_{1/\lambda}\rangle\bigr| \\
&\leqslant \lambda^{-n} \sup \|f\|_{M^{p',q'}} \|g_{1/\lambda}\|_{M^{p,q}} \\
&\leqslant \lambda^{-n} \sup \|f\|_{M^{p',q'}} \bigl(C\bigl(\lambda^{-1}\bigr)^\alpha \|g\|_{M^{p,q}}\bigr) = C\lambda^{-n-\alpha} \|f\|_{M^{p',q'}}
\end{aligned}$$

for all $f \in M^{p',q'}(\mathbb{R}^n)$ and $0 < \lambda \leqslant 1$, where the supremum is taken over all $g \in \mathcal{S}(\mathbb{R}^n)$ such that $\|g\|_{M^{p,q}} = 1$. Since $(1/p', 1/q') \in I_1$, by Theorem 3.1(2) with $(1/p', 1/q') \in I_1$, we get $-n - \alpha \leqslant -n/p'$. This implies $\alpha \geqslant -n/p$.

We next consider the case $q = \infty$. Let $1 \leqslant r \leqslant \infty$. Assume that there exist constants $C > 0$ and $\alpha \in \mathbb{R}$ such that

$$\|f_\lambda\|_{M^{r,\infty}} \leqslant C\lambda^\alpha \|f\|_{M^{r,\infty}} \quad \text{for all } f \in M^{r,\infty}(\mathbb{R}^n) \text{ and } \lambda \geqslant 1, \tag{3.4}$$

where $\alpha < -n/r$. Since $M^{2,2}(\mathbb{R}^n) = L^2(\mathbb{R}^n)$, we have

$$\|f_\lambda\|_{M^{2,2}} \leqslant C\lambda^{-n/2} \|f\|_{M^{2,2}} \quad \text{for all } f \in M^{2,2}(\mathbb{R}^n) \text{ and } \lambda \geqslant 1. \tag{3.5}$$

Then, by interpolation, (3.4) and (3.5) give

$$\|f_\lambda\|_{M^{p,q}} \leqslant \begin{cases} C\lambda^{(\alpha r+n)(1/p-1/q)-n/p} \|f\|_{M^{p,q}}, & \text{if } 1 \leqslant r < \infty, \\ C\lambda^{\alpha(1-2/q)-n/p} \|f\|_{M^{p,q}}, & \text{if } r = \infty \end{cases}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $\lambda \geqslant 1$, where $1/p = (1-\theta)/r + \theta/2$, $1/q = (1-\theta)/\infty + \theta/2$ and $0 < \theta < 1$. Note that $(1/p, 1/q) \in I_1^*$ and $2 < q < \infty$. However, since $(\alpha r + n)(1/p - 1/q) < 0$ if $1 \leqslant r < \infty$ and $\alpha(1-2/q) < 0$ if $r = \infty$, this contradicts Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$ and $2 < q < \infty$. Therefore, α must satisfy $\alpha \geqslant -n/r$. The proof is complete. □

Our next goal is to prove Theorem 3.1(1) with $(1/p, 1/q) \in I_2^*$ and Theorem 3.1(2) with $(1/p, 1/q) \in I_2$.

Lemma 3.3. *Let $1 \leqslant p, q \leqslant \infty$ be such that $(1/p, 1/q) \in I_2^*$ and $1/p \geqslant 1/q$. Then there exists a constant $C > 0$ such that*

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{-n(2/p-1/q)}\bigl(1+\lambda^2\bigr)^{n(1/p-1/2)} \|f\|_{M^{p,q}}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $\lambda > 0$.

Proof. Let $1 \leqslant r \leqslant \infty$. By Lemma 3.2, we have

$$\|f_\lambda\|_{M^{1,r}} \leqslant C\lambda^{n(1/r-2)}\bigl(1+\lambda^2\bigr)^{n/2}\|f\|_{M^{1,r}} \tag{3.6}$$

for all $f \in M^{1,r}(\mathbb{R}^n)$ and $\lambda > 0$. Then, by interpolation, (3.2), (3.5) and (3.6) give Lemma 3.3. □

The proof of the following lemma is based on that of [21, Theorem 3].

Lemma 3.4. *Suppose that $\varphi \in \mathcal{S}(\mathbb{R}^n)$ is a real-valued function satisfying $\varphi \geqslant C$ on $[-1/2, 1/2]^n$ for some constant $C > 0$, $\operatorname{supp}\varphi \subset [-1,1]^n$, $\varphi(t) = \varphi(-t)$ and $\sum_{k\in\mathbb{Z}^n}\varphi(t-k) = 1$ for all $t \in \mathbb{R}^n$. Then*

$$\sup_{k\in\mathbb{Z}^n}\bigl\|(M_k\Phi) * f\bigr\|_{L^2} \leqslant \|V_\Phi f\|_{L^{2,\infty}} \leqslant 5^n\|\Phi\|_{L^1}\sup_{k\in\mathbb{Z}^n}\bigl\|(M_k\Phi) * f\bigr\|_{L^2}$$

for all $f \in M^{2,\infty}(\mathbb{R}^n)$, where $\Phi = \mathcal{F}^{-1}\varphi$ and $M_k\Phi(t) = e^{ik\cdot t}\Phi(t)$.

Proof. Let $f \in M^{2,\infty}(\mathbb{R}^n)$. Since Φ is a real-valued function and $\Phi(t) = \Phi(-t)$ for all t, we have

$$\begin{aligned}\bigl|V_\Phi f(x,\xi)\bigr| &= \left|\int_{\mathbb{R}^n} f(t)\overline{\Phi(t-x)}e^{-i\xi\cdot t}\,dt\right| \\ &= \left|\int_{\mathbb{R}^n} f(t)\Phi(x-t)e^{i\xi\cdot(x-t)}\,dt\right| = \bigl|(M_\xi\Phi) * f(x)\bigr|. \end{aligned}\tag{3.7}$$

We first prove

$$\operatorname*{ess\,sup}_{\xi\in\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\bigl|V_\Phi f(x,\xi)\bigr|^2\,dx\right)^{1/2} = \sup_{\xi\in\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\bigl|V_\Phi f(x,\xi)\bigr|^2\,dx\right)^{1/2}. \tag{3.8}$$

To prove (3.8), it is enough to show that $(\int_{\mathbb{R}^n}|V_\Phi f(x,\xi)|^2\,dx)^{1/2}$ is continuous with respect to ξ. Since $\operatorname{ess\,sup}_{\xi\in\mathbb{R}^n}(\int_{\mathbb{R}^n}|V_\Phi f(x,\xi)|^2\,dx)^{1/2} < \infty$, for each $k \in \mathbb{Z}^n$ there exists $\xi_k \in k/2 + [-1/4, 1/4]^n$ such that $(\int_{\mathbb{R}^n}|V_\Phi f(x,\xi_k)|^2\,dx)^{1/2} < \infty$. Then, by (3.7), we have

$$\frac{1}{(2\pi)^{n/2}}\bigl\|\varphi(\cdot-\xi_k)\widehat{f}\bigr\|_{L^2} = \bigl\|(M_{\xi_k}\Phi) * f\bigr\|_{L^2} = \left(\int_{\mathbb{R}^n}\bigl|V_\Phi f(x,\xi_k)\bigr|^2\,dx\right)^{1/2} < \infty.$$

Since $k/2 + [-1/4, 1/4]^n \subset \xi_k + [-1/2, 1/2]^n$ and $\varphi(\cdot-\xi_k) \geqslant C > 0$ on $\xi_k + [-1/2, 1/2]^n$, we see that $|\widehat{f}|^2$ is integrable on $k/2+[-1/4, 1/4]^n$. The arbitrariness of $k \in \mathbb{Z}^n$ gives $\widehat{f} \in L^2_{\mathrm{loc}}(\mathbb{R}^n)$. By the Lebesgue dominated convergence theorem, we see that $\|\varphi(\cdot-\xi)\widehat{f}\|_{L^2}$ is continuous with respect to ξ. Hence, $(\int_{\mathbb{R}^n}|V_\Phi f(x,\xi)|^2\,dx)^{1/2}$ is continuous with respect to ξ. We obtain (3.8). Then, from (3.7) and (3.8) it follows that

$$\sup_{k\in\mathbb{Z}^n}\left\|(M_k\Phi)*f\right\|_{L^2}\leqslant\sup_{\xi\in\mathbb{R}^n}\left\|(M_\xi\Phi)*f\right\|_{L^2}=\sup_{\xi\in\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\left|V_\Phi f(x,\xi)\right|^2dx\right)^{1/2}$$
$$=\operatorname*{ess\,sup}_{\xi\in\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\left|V_\Phi f(x,\xi)\right|^2dx\right)^{1/2}=\|V_\Phi f\|_{L^{2,\infty}}.$$

We next prove $\|V_\Phi f\|_{L^{2,\infty}}\leqslant(5^n\|\Phi\|_{L^1})\sup_{k\in\mathbb{Z}^n}\|(M_k\Phi)*f\|_{L^2}$. Let $\xi=(\xi_1,\dots,\xi_n)\in\mathbb{R}^n$. Since

$$M_\xi\Phi=\mathcal{F}^{-1}\left[\varphi(\cdot-\xi)\right]=\mathcal{F}^{-1}\left[\varphi(\cdot-\xi)\left(\sum_{k\in\mathbb{Z}^n}\varphi(\cdot-k)\right)\right]$$
$$=\sum_{\substack{|k_i-\xi_i|\leqslant 2,\\ i=1,\dots,n}}\mathcal{F}^{-1}\left[\varphi(\cdot-\xi)\varphi(\cdot-k)\right]=\sum_{\substack{|k_i-\xi_i|\leqslant 2,\\ i=1,\dots,n}}(M_\xi\Phi)*(M_k\Phi),$$

by (3.7), we have

$$\left|V_\Phi f(x,\xi)\right|=\left|(M_\xi\Phi)*f(x)\right|\leqslant\sum_{\substack{|k_i-\xi_i|\leqslant 2,\\ i=1,\dots,n}}\left|(M_\xi\Phi)*(M_k\Phi)*f(x)\right|.$$

Hence, by (3.8), we get

$$\|V_\Phi f\|_{L^{2,\infty}}=\sup_{\xi\in\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\left|V_\Phi f(x,\xi)\right|^2dx\right)^{1/2}$$
$$\leqslant\sup_{\xi\in\mathbb{R}^n}\sum_{\substack{|k_i-\xi_i|\leqslant 2,\\ i=1,\dots,n}}\left\|(M_\xi\Phi)*(M_k\Phi)*f\right\|_{L^2}$$
$$\leqslant\sup_{\xi\in\mathbb{R}^n}\sum_{\substack{|k_i-\xi_i|\leqslant 2,\\ i=1,\dots,n}}\|M_\xi\Phi\|_{L^1}\left\|(M_k\Phi)*f\right\|_{L^2}$$
$$\leqslant\|\Phi\|_{L^1}\left(\sup_{\ell\in\mathbb{Z}^n}\left\|(M_\ell\Phi)*f\right\|_{L^2}\right)\left(\sup_{\xi\in\mathbb{R}^n}\sum_{\substack{|k_i-\xi_i|\leqslant 2,\\ i=1,\dots,n}}1\right)$$
$$\leqslant 5^n\|\Phi\|_{L^1}\sup_{\ell\in\mathbb{Z}^n}\left\|(M_\ell\Phi)*f\right\|_{L^2}.$$

The proof is complete. □

We remark that Lemma 3.2 implies

$$\|f_\lambda\|_{M^{2,\infty}}\leqslant C\lambda^{-3n/2}\|f\|_{M^{2,\infty}}\quad\text{for all } f\in M^{2,\infty}\left(\mathbb{R}^n\right)\text{ and } 0<\lambda\leqslant 1.$$

This is not our desired order of λ in the case $(p,q)=(2,\infty)$. But we have

Lemma 3.5. *There exists a constant $C > 0$ such that*

$$\|f_\lambda\|_{M^{2,\infty}} \leqslant C\lambda^{-n}\|f\|_{M^{2,\infty}} \quad \text{for all } f \in M^{2,\infty}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1.$$

Proof. Let $\Phi = \mathcal{F}^{-1}\varphi$, where φ is as in Lemma 3.4. Suppose that $f \in M^{2,\infty}(\mathbb{R}^n)$. We note that $\widehat{f} \in L^2_{\mathrm{loc}}(\mathbb{R}^n)$ (see the proof of Lemma 3.4). Then, by Lemma 3.4, we see that

$$\begin{aligned}
\big\|V_\Phi(f_\lambda)\big\|_{M^{2,\infty}} &\leqslant 5^n\|\Phi\|_{L^1} \sup_{k\in\mathbb{Z}^n}\big\|(M_k\Phi) * f_\lambda\big\|_{L^2} \\
&= 5^n\|\Phi\|_{L^1} \sup_{k\in\mathbb{Z}^n}(2\pi)^{-n/2}\big\|\varphi(\cdot - k)\widehat{f_\lambda}\big\|_{L^2} \\
&= C_n\lambda^{-n/2}\sup_{k\in\mathbb{Z}^n}\big\|\varphi(\lambda\cdot -k)\widehat{f}\big\|_{L^2} \\
&= C_n\lambda^{-n/2}\sup_{k\in\mathbb{Z}^n}\bigg\|\varphi(\lambda\cdot -k)\bigg(\sum_{\ell\in\mathbb{Z}^n}\varphi(\cdot - \ell)\bigg)\widehat{f}\bigg\|_{L^2}.
\end{aligned}$$

Since

$$\begin{aligned}
\bigg|\varphi(\lambda t - k)\bigg(\sum_{\ell\in\mathbb{Z}^n}\varphi(t-\ell)\bigg)\widehat{f}(t)\bigg|^2 &\leqslant 4^n\sum_{\ell\in\mathbb{Z}^n}\big|\varphi(\lambda t - k)\varphi(t-\ell)\widehat{f}(t)\big|^2 \\
&= 4^n\sum_{\substack{|\ell_i - k_i/\lambda|\leqslant 2/\lambda,\\ i=1,\dots,n}}\big|\varphi(\lambda t - k)\varphi(t-\ell)\widehat{f}(t)\big|^2,
\end{aligned}$$

we have

$$\begin{aligned}
&\bigg\|\varphi(\lambda\cdot -k)\bigg(\sum_{\ell\in\mathbb{Z}^n}\varphi(\cdot-\ell)\bigg)\widehat{f}\bigg\|_{L^2} \\
&\quad\leqslant\bigg(4^n\sum_{\substack{|\ell_i - k_i/\lambda|\leqslant 2/\lambda,\\ i=1,\dots,n}}\int_{\mathbb{R}^n}\big|\varphi(\lambda t - k)\varphi(t-\ell)\widehat{f}(t)\big|^2\,dt\bigg)^{1/2} \\
&\quad\leqslant\bigg(4^n(2\pi)^n\|\varphi\|_{L^\infty}^2\sum_{\substack{|\ell_i - k_i/\lambda|\leqslant 2/\lambda,\\ i=1,\dots,n}}\big\|(M_\ell\Phi) * f\big\|_{L^2}^2\bigg)^{1/2} \\
&\quad\leqslant\bigg(4^n(2\pi)^n\|\varphi\|_{L^\infty}^2\Big(\sup_{m\in\mathbb{Z}^n}\big\|(M_m\Phi) * f\big\|_{L^2}\Big)^2\sum_{\substack{|\ell_i - k_i/\lambda|\leqslant 2/\lambda,\\ i=1,\dots,n}}1\bigg)^{1/2} \\
&\quad\leqslant\Big(C_n\|\varphi\|_{L^\infty}^2\lambda^{-n}\Big(\sup_{m\in\mathbb{Z}^n}\big\|(M_m\Phi) * f\big\|_{L^2}\Big)^2\Big)^{1/2} \\
&\quad= C_n\|\varphi\|_{L^\infty}\lambda^{-n/2}\sup_{m\in\mathbb{Z}^n}\big\|(M_m\Phi) * f\big\|_{L^2}.
\end{aligned}$$

Hence, by Lemma 3.4, we get

$$\|f_\lambda\|_{M^{2,\infty}} \leqslant C_n\lambda^{-n} \sup_{m\in\mathbb{Z}^n} \big\|(M_m\Phi) * f\big\|_{L^2} \leqslant C_n\lambda^{-n}\|f\|_{M^{2,\infty}}.$$

The proof is complete. □

Lemma 3.6. *Let* $1 \leqslant p \leqslant \infty$. *Then the following are true*:

(1) *If* $p \leqslant 2$, *then there exists a constant* $C > 0$ *such that*

$$\|f_\lambda\|_{M^{p,1}} \leqslant C\|f\|_{M^{p,1}} \quad \textit{for all } f \in M^{p,1}\big(\mathbb{R}^n\big) \textit{ and } \lambda \geqslant 1.$$

(2) *If* $p \geqslant 2$, *then there exists a constant* $C > 0$ *such that*

$$\|f_\lambda\|_{M^{p,1}} \leqslant C\lambda^{-n(2/p-1)}\|f\|_{M^{p,1}} \quad \textit{for all } f \in M^{p,1}\big(\mathbb{R}^n\big) \textit{ and } \lambda \geqslant 1.$$

Proof. We first consider the case $p \leqslant 2$. By Lemmas 2.2, 2.4 and 3.5, we have

$$\begin{aligned}\|f_\lambda\|_{M^{2,1}} &= \sup\big|\langle f_\lambda, g\rangle_M\big| = \sup\big|\langle f_\lambda, g\rangle\big| \\ &= \lambda^{-n}\sup\big|\langle f, g_{1/\lambda}\rangle\big| \leqslant \lambda^{-n}\sup\|f\|_{M^{2,1}}\|g_{1/\lambda}\|_{M^{2,\infty}} \\ &\leqslant \lambda^{-n}\sup\|f\|_{M^{2,1}}\big(C(1/\lambda)^{-n}\|g\|_{M^{2,\infty}}\big) = C\|f\|_{M^{2,1}}\end{aligned}$$

for all $f \in \mathcal{S}(\mathbb{R}^n)$ and $\lambda \geqslant 1$, where the supremum is taken over all $g \in M^{2,\infty}(\mathbb{R}^n)$ such that $\|g\|_{M^{2,\infty}} = 1$. Since $\mathcal{S}(\mathbb{R}^n)$ is dense in $M^{2,1}(\mathbb{R}^n)$, this gives

$$\|f_\lambda\|_{M^{2,1}} \leqslant C\|f\|_{M^{2,1}} \quad \text{for all } f \in M^{2,1}\big(\mathbb{R}^n\big) \text{ and } \lambda \geqslant 1. \tag{3.9}$$

On the other hand, by Lemma 3.2, we see that

$$\|f_\lambda\|_{M^{1,1}} \leqslant C\|f\|_{M^{1,1}} \quad \text{for all } f \in M^{1,1}\big(\mathbb{R}^n\big) \text{ and } \lambda \geqslant 1. \tag{3.10}$$

Hence, by interpolation, (3.9) and (3.10) give Lemma 3.6(1).

We next consider the case $p \geqslant 2$. By Lemma 3.2, we have

$$\|f_\lambda\|_{M^{\infty,1}} \leqslant C\lambda^n\|f\|_{M^{\infty,1}} \quad \text{for all } f \in M^{\infty,1}\big(\mathbb{R}^n\big) \text{ and } \lambda \geqslant 1. \tag{3.11}$$

Therefore, by interpolation, (3.9) and (3.11) give Lemma 3.6(2). □

Lemma 3.7. *Let* $1 \leqslant p \leqslant \infty$. *Then the following are true*:

(1) *If* $p \leqslant 2$, *then there exists a constant* $C > 0$ *such that*

$$\|f_\lambda\|_{M^{p,\infty}} \leqslant C\lambda^{-2n/p}\|f\|_{M^{p,\infty}} \quad \textit{for all } f \in M^{p,\infty}\big(\mathbb{R}^n\big) \textit{ and } 0 < \lambda \leqslant 1.$$

(2) *If* $p \geqslant 2$, *then there exists a constant* $C > 0$ *such that*

$$\|f_\lambda\|_{M^{p,\infty}} \leqslant C\lambda^{-n}\|f\|_{M^{p,\infty}} \quad \textit{for all } f \in M^{p,\infty}\big(\mathbb{R}^n\big) \textit{ and } 0 < \lambda \leqslant 1.$$

Proof. Let $1 < p \leqslant 2$. By duality and Lemma 3.6(2), we have

$$\begin{aligned}\|f_\lambda\|_{M^{p,\infty}} &= \sup\bigl|\langle f_\lambda, g\rangle\bigr| = \lambda^{-n}\sup\bigl|\langle f, g_{1/\lambda}\rangle\bigr| \\ &\leqslant \lambda^{-n}\sup\|f\|_{M^{p,\infty}}\bigl(C(1/\lambda)^{-n(2/p'-1)}\|g\|_{M^{p',1}}\bigr) = C\lambda^{-2n/p}\|f\|_{M^{p,\infty}}\end{aligned}$$

for all $f \in M^{p,\infty}(\mathbb{R}^n)$ and $0 < \lambda \leqslant 1$, where the supremum is taken over all $g \in \mathcal{S}(\mathbb{R}^n)$ such that $\|g\|_{M^{p',1}} = 1$. In the case $p = 1$, by Lemma 3.2, we have

$$\|f_\lambda\|_{M^{1,\infty}} \leqslant C\lambda^{-2n}\|f\|_{M^{1,\infty}} \quad \text{for all } f \in M^{1,\infty}\bigl(\mathbb{R}^n\bigr) \text{ and } 0 < \lambda \leqslant 1.$$

Hence, we obtain Lemma 3.7(1). In the same way, we can prove Lemma 3.7(2). □

Lemma 3.8. *Let $1 \leqslant p, q \leqslant \infty$, $(p,q) \neq (1,\infty), (\infty,1)$ and $\epsilon > 0$. Set*

$$f(t) = \sum_{k\neq 0} |k|^{-n/q-\epsilon} e^{ik\cdot t}\varphi(t) \quad \text{in } \mathcal{S}'\bigl(\mathbb{R}^n\bigr),$$

where φ is the Gauss function. Then $f \in M^{p,q}(\mathbb{R}^n)$ and there exists a constant $C > 0$ such that $\|f_\lambda\|_{M^{p,q}} \geqslant C\lambda^{n(1/q-1)+\epsilon}$ for all $0 < \lambda \leqslant 1$.

Proof. We first prove $f \in M^{p,q}(\mathbb{R}^n)$. Although this fact is an immediate consequence of the discretization properties of the modulation spaces (see Feichtinger and Gröchenig [6], or Gröchenig [10, Theorem 12.2.4]), here we give the proof for reader's convenience. Since

$$\begin{aligned}&\left|\int_{\mathbb{R}^n} e^{ik\cdot t}\varphi(t)\varphi(t-x)e^{-i\xi\cdot t}\,dt\right| \\ &\quad= \left|\int_{\mathbb{R}^n}\varphi(t)\varphi(x-t)\bigl\{\bigl(1+|\xi-k|^2\bigr)^{-n}(I-\Delta_t)^n e^{-i(\xi-k)\cdot t}\bigr\}\,dt\right| \\ &\quad= \bigl(1+|\xi-k|^2\bigr)^{-n}\left|\sum_{|\beta_1+\beta_2|\leqslant 2n} C_{\beta_1,\beta_2}\int_{\mathbb{R}^n}\bigl(\partial^{\beta_1}\varphi\bigr)(t)\bigl(\partial^{\beta_2}\varphi\bigr)(x-t)e^{-i(\xi-k)\cdot t}\,dt\right| \\ &\quad\leqslant C\bigl(1+|\xi-k|^2\bigr)^{-n}\sum_{|\beta_1+\beta_2|\leqslant 2n}\bigl|\partial^{\beta_1}\varphi\bigr| * \bigl|\partial^{\beta_2}\varphi\bigr|(x),\end{aligned}$$

we see that

$$\begin{aligned}\|f\|_{M^{p,q}} &= \|V_\varphi f\|_{L^{p,q}} \\ &= \left\{\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\left|\sum_{k\neq 0}|k|^{-n/q-\epsilon}\int_{\mathbb{R}^n} e^{ik\cdot t}\varphi(t)\varphi(t-x)e^{-i\xi\cdot t}\,dt\right|dx\right)^{q/p}d\xi\right\}^{1/q} \\ &\leqslant C\left\{\int_{\mathbb{R}^n}\left(\sum_{k\neq 0}|k|^{-n/q-\epsilon}\bigl(1+|\xi-k|^2\bigr)^{-n}\right)^q d\xi\right\}^{1/q}\end{aligned}$$

$$= C\left\{\sum_{\ell\in\mathbb{Z}^n}\int\limits_{\ell+[-1/2,1/2]^n}\left(\sum_{k\neq 0}|k|^{-n/q-\epsilon}\left(1+|\xi-k|^2\right)^{-n}\right)^q d\xi\right\}^{1/q}$$

$$\leqslant C\left\{\sum_{\ell\in\mathbb{Z}^n}\left(\sum_{k\neq 0}|k|^{-n/q-\epsilon}\left(1+|\ell-k|^2\right)^{-n}\right)^q\right\}^{1/q}.$$

Since $\{|k|^{-n/q-\epsilon}\}_{k\neq 0}\in \ell^q(\mathbb{Z}^n)$, by Young's inequality, we have $f\in M^{p,q}(\mathbb{R}^n)$.

We next consider the second part. Since $\varphi\in M^{p',q'}(\mathbb{R}^n)$, by duality, we see that

$$\begin{aligned}
\|f_\lambda\|_{M^{p,q}} &= \sup_{\|g\|_{M^{p',q'}}=1}\left|\langle f_\lambda, g\rangle_M\right| \geqslant \left|\langle f_\lambda,\varphi\rangle\right| \\
&= \left|\sum_{k\neq 0}|k|^{-n/q-\epsilon}\int\limits_{\mathbb{R}^n} e^{i(\lambda k)\cdot t}\varphi(\lambda t)\varphi(t)\,dt\right| \\
&= C\left(1+\lambda^2\right)^{-n/2}\sum_{k\neq 0}|k|^{-n/q-\epsilon}e^{-\frac{\lambda^2|k|^2}{4(1+\lambda^2)}} \\
&\geqslant C\sum_{\substack{0<|k_i|\leqslant 1/\lambda,\\ i=1,\dots,n}}|k|^{-n/q-\epsilon}e^{-\frac{\lambda^2|k|^2}{4(1+\lambda^2)}} \\
&\geqslant C\lambda^{n/q+\epsilon}\sum_{\substack{0<|k_i|\leqslant 1/\lambda,\\ i=1,\dots,n}}1 \geqslant C\lambda^{n(1/q-1)+\epsilon}
\end{aligned}$$

for all $0<\lambda\leqslant 1$. The proof is complete. □

We are now ready to prove Theorem 3.1(1) with $(1/p,1/q)\in I_2^*$ and Theorem 3.1(2) with $(1/p,1/q)\in I_2$.

Proof of Theorem 3.1(2) with $(1/p,1/q)\in I_2$. We recall that $\mu_2(p,q)=1/q-1$ if $(1/p,1/q)\in I_2$. Let $(1/p,1/q)\in I_2$ and $1/p\leqslant 1/q$. If $q=1$ then $p=\infty$, and we have already proved this case in Theorem 3.1(2) with $(1/p,1/q)\in I_1$. Hence, we may assume $1<q\leqslant\infty$. Note that $1\leqslant p',q'<\infty$. Since $(1/p',1/q')\in I_2^*$ and $1/p'\geqslant 1/q'$, by duality and Lemma 3.3, we have

$$\begin{aligned}
\|f_\lambda\|_{M^{p,q}} &= \sup\left|\langle f_\lambda, g\rangle\right| = \lambda^{-n}\sup\left|\langle f, g_{1/\lambda}\rangle\right| \\
&\leqslant \lambda^{-n}\sup\|f\|_{M^{p,q}}\|g_{1/\lambda}\|_{M^{p',q'}} \\
&\leqslant \lambda^{-n}\sup\|f\|_{M^{p,q}}\left(C\left(\lambda^{-1}\right)^{-n(2/p'-1/q')}\left(1+\lambda^{-2}\right)^{n(1/p'-1/2)}\|g\|_{M^{p',q'}}\right) \\
&\leqslant C\lambda^{n(1/q-1)}\|f\|_{M^{p,q}} \qquad (3.12)
\end{aligned}$$

for all $f \in M^{p,q}(\mathbb{R}^n)$ and $0 < \lambda \leqslant 1$, where the supremum is taken over all $g \in \mathcal{S}(\mathbb{R}^n)$ such that $\|g\|_{M^{p',q'}} = 1$. This is the first part of Theorem 3.1(2) with $(1/p, 1/q) \in I_2$ and $1/p \leqslant 1/q$. We next consider the case $(1/p, 1/q) \in I_2$, $1/p \geqslant 1/q$ and $q < \infty$. From (3.12) it follows that

$$\|f_\lambda\|_{M^{r,r}} \leqslant C\lambda^{n(1/r-1)}\|f\|_{M^{r,r}} \quad \text{for all } f \in M^{r,r}\big(\mathbb{R}^n\big) \text{ and } 0 < \lambda \leqslant 1, \tag{3.13}$$

where $2 \leqslant r \leqslant \infty$. Then, by interpolation, Lemma 3.5 and (3.13) give

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{n(1/q-1)}\|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}\big(\mathbb{R}^n\big) \text{ and } 0 < \lambda \leqslant 1,$$

where $(1/p, 1/q) \in I_2$, $1/p \geqslant 1/q$ and $q < \infty$. In the case $q = \infty$, by Lemma 3.7(2), we have nothing to prove. Hence, we get the first part of Theorem 3.1(2) with $(1/p, 1/q) \in I_2$ and $1/p \geqslant 1/q$.

We next consider the second part of Theorem 3.1(2) with $(1/p, 1/q) \in I_2$. Let $(1/p, 1/q) \in I_2$. Since $(1/\infty, 1/1) \in I_1$, we may assume $(p, q) \neq (\infty, 1)$. Assume that there exist constants $C > 0$ and $\beta \in \mathbb{R}$ such that

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{\beta}\|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}\big(\mathbb{R}^n\big) \text{ and } 0 < \lambda \leqslant 1,$$

where $\beta > n(1/q - 1)$. Then we can take $\epsilon > 0$ such that $n(1/q - 1) + \epsilon < \beta$. For this ϵ, we set

$$f(t) = \sum_{k \neq 0} |k|^{-n/q-\epsilon} e^{ik\cdot t}\varphi(t),$$

where φ is the Gauss function. Then, by Lemma 3.8, we see that $f \in M^{p,q}(\mathbb{R}^n)$ and there exists a constant $C' > 0$ such that $\|f_\lambda\|_{M^{p,q}} \geqslant C'\lambda^{n(1/q-1)+\epsilon}$ for all $0 < \lambda \leqslant 1$. Hence,

$$C'\lambda^{n(1/q-1)+\epsilon} \leqslant \|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{\beta}\|f\|_{M^{p,q}}$$

for all $0 < \lambda \leqslant 1$. However, since $n(1/q - 1) + \epsilon < \beta$, this is a contradiction. Therefore, β must satisfy $\beta \leqslant n(1/q - 1)$. The proof is complete. □

Proof of Theorem 3.1(1) with $(1/p, 1/q) \in I_2^*$. We recall that $\mu_1(p, q) = 1/q - 1$ if $(1/p, 1/q) \in I_2^*$. In every case except for $(p, q) \neq (1, \infty)$, by duality, Theorem 3.1(2) with $(1/p', 1/q') \in I_2$ and the same argument as in the proof of Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$, we can prove Theorem 3.1(1) with $(1/p, 1/q) \in I_2^*$. For the case $(p, q) = (1, \infty)$, we have already proved in Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$. □

Our last goal of this section is to prove Theorem 3.1(1) with $(1/p, 1/q) \in I_3^*$ and Theorem 3.1(2) with $(1/p, 1/q) \in I_3$. In the following lemma, we use the fact that there exists $\varphi \in \mathcal{S}(\mathbb{R}^n)$ such that $\operatorname{supp}\varphi \subset [-1/8, 1/8]^n$ and $|\hat{\varphi}| \geqslant 1$ on $[-2, 2]^n$ (see, for example, the proof of [8, Theorem 2.6]).

Lemma 3.9. *Let* $1 \leqslant p \leqslant \infty$, $1 \leqslant q < \infty$ *and* $\epsilon > 0$. *Suppose that* $\varphi, \psi \in \mathcal{S}(\mathbb{R}^n)$ *satisfy* $\operatorname{supp}\varphi \subset [-1/8, 1/8]^n$, $\operatorname{supp}\psi \subset [-1/2, 1/2]^n$, $|\hat{\varphi}| \geqslant 1$ *on* $[-2, 2]^n$ *and* $\psi = 1$ *on* $[-1/4, 1/4]^n$. *Set*

$$f(t) = \sum_{k \neq 0} |k|^{-n/q-\epsilon} e^{ik\cdot t}\psi(t-k) \quad \textit{in } \mathcal{S}'\big(\mathbb{R}^n\big).$$

Then $f \in M^{p,q}(\mathbb{R}^n)$ *and there exists a constant* $C > 0$ *such that*

$$\left\| V_\varphi(f_\lambda) \right\|_{L^{p,q}} \geqslant C\lambda^{-n(2/p-1/q)+\epsilon} \quad \textit{for all } 0 < \lambda \leqslant 1.$$

Proof. In the same way as in the proof of Lemma 3.8, we can prove $f \in M^{p,q}(\mathbb{R}^n)$ (see also [6] or [10]). We consider the second part. It is enough to show that $\|V_{\varphi_{1/\lambda}} f\|_{L^{p,q}} \geqslant C\lambda^{-n/p+n+\epsilon}$ for all $0 < \lambda \leqslant 1$, since $\|V_\varphi(f_\lambda)\|_{L^{p,q}} = \lambda^{-n(1/p-1/q+1)}\|V_{\varphi_{1/\lambda}} f\|_{L^{p,q}}$. We note that $\operatorname{supp}\varphi((\cdot - x)/\lambda) \subset \ell + [-1/4, 1/4]^n$ for all $0 < \lambda \leqslant 1$, $\ell \in \mathbb{Z}^n$ and $x \in \ell + [-1/8, 1, 8]^n$. Since $\operatorname{supp}\psi(\cdot - k) \subset k + [-1/2, 1/2]^n$ and $\psi(t - k) = 1$ if $t \in k + [-1/4, 1/4]^n$, it follows that

$$\begin{aligned}
&\left(\int_{\mathbb{R}^n} \left| V_{\varphi_{1/\lambda}} f(x,\xi) \right|^p dx \right)^{1/p} \\
&= \left(\int_{\mathbb{R}^n} \left| \sum_{k \neq 0} |k|^{-n/q-\epsilon} \int_{\mathbb{R}^n} e^{ik\cdot t} \psi(t-k) \overline{\varphi\left(\frac{t-x}{\lambda}\right)} e^{-i\xi\cdot t}\, dt \right|^p dx \right)^{1/p} \\
&\geqslant \left(\sum_{\ell \neq 0} \int_{\ell+[-1/8,1/8]^n} \left| \sum_{k \neq 0} |k|^{-n/q-\epsilon} \int_{\mathbb{R}^n} e^{-i(\xi-k)\cdot t} \psi(t-k) \overline{\varphi\left(\frac{t-x}{\lambda}\right)}\, dt \right|^p dx \right)^{1/p} \\
&= \left(\sum_{\ell \neq 0} \int_{\ell+[-1/8,1/8]^n} \left| |\ell|^{-n/q-\epsilon} \int_{\mathbb{R}^n} e^{-i(\xi-\ell)\cdot t} \overline{\varphi\left(\frac{t-x}{\lambda}\right)}\, dt \right|^p dx \right)^{1/p} \\
&= 4^{-n/p} \left(\sum_{\ell \neq 0} \left| |\ell|^{-n/q-\epsilon} \lambda^n \hat{\varphi}\big(-\lambda(\xi-\ell)\big) \right|^p \right)^{1/p}.
\end{aligned}$$

Hence, using $|\hat{\varphi}| \geqslant 1$ on $[-2, 2]^n$, we get

$$\begin{aligned}
\|V_{\varphi_{1/\lambda}} f\|_{L^{p,q}} &\geqslant 4^{-n/p}\lambda^n \left\{ \int_{\mathbb{R}^n} \left(\sum_{\ell \neq 0} \left| |\ell|^{-n/q-\epsilon} \hat{\varphi}\big(-\lambda(\xi-\ell)\big) \right|^p \right)^{q/p} d\xi \right\}^{1/q} \\
&= 4^{-n/p}\lambda^{n-n/q} \left\{ \int_{\mathbb{R}^n} \left(\sum_{\ell \neq 0} \left| |\ell|^{-n/q-\epsilon} \hat{\varphi}(\xi + \lambda\ell) \right|^p \right)^{q/p} d\xi \right\}^{1/q} \\
&\geqslant 4^{-n/p}\lambda^{n-n/q} \left\{ \int_{[-1,1]^n} \left(\sum_{\substack{0<|\ell_i|\leqslant 1/\lambda, \\ i=1,\ldots,n}} \left| |\ell|^{-n/q-\epsilon} \hat{\varphi}(\xi + \lambda\ell) \right|^p \right)^{q/p} d\xi \right\}^{1/q} \\
&\geqslant 4^{-n/p} 2^{n/q} \lambda^{n-n/q} \left(\sum_{\substack{0<|\ell_i|\leqslant 1/\lambda, \\ i=1,\ldots,n}} |\ell|^{-(n/q+\epsilon)p} \right)^{1/p} \\
&\geqslant C_n \lambda^{n-n/q} \lambda^{n/q+\epsilon} \left(\sum_{\substack{0<|\ell_i|\leqslant 1/\lambda, \\ i=1,\ldots,n}} 1 \right)^{1/p} \geqslant C_n \lambda^{-n/p+n+\epsilon}
\end{aligned}$$

for all $0 < \lambda \leqslant 1$. The proof is complete. $\square$

For Lemma 3.9, we do not need $\epsilon > 0$ in the case $q = \infty$.

Lemma 3.10. *Let $1 \leqslant p \leqslant \infty$. Suppose that $\varphi, \psi \in \mathcal{S}(\mathbb{R}^n)$ are as in Lemma 3.9. Set*

$$f(t) = \sum_{k \in \mathbb{Z}^n} e^{ik\cdot t}\psi(t-k) \quad \text{in } \mathcal{S}'\big(\mathbb{R}^n\big).$$

Then $f \in M^{p,\infty}(\mathbb{R}^n)$ and there exists a constant $C > 0$ such that

$$\big\| V_\varphi(f_\lambda)\big\|_{L^{p,\infty}} \geqslant C\lambda^{-2n/p} \quad \text{for all } 0 < \lambda \leqslant 1.$$

In particular, if $1 \leqslant p \leqslant 2$ then there exist constants $C, C' > 0$ such that

$$C\lambda^{-2n/p} \leqslant \|f_\lambda\|_{M^{p,\infty}} \leqslant C'\lambda^{-2n/p} \quad \text{for all } 0 < \lambda \leqslant 1.$$

Proof. In the same way as in the proof of Lemma 3.8, we can prove

$$\left| \int_{\mathbb{R}^n} e^{ik\cdot t}\psi(t-k)\overline{\varphi(t-x)}e^{-i\xi\cdot t}\,dt \right| \leqslant C\big(1+|x-k|^2\big)^{-n}\big(1+|\xi-k|^2\big)^{-n}.$$

Hence,

$$\begin{aligned}
\big|V_\varphi f(x,\xi)\big| &= \left| \sum_{k\in\mathbb{Z}^n} \int_{\mathbb{R}^n} e^{ik\cdot t}\psi(t-k)\overline{\varphi(t-x)}e^{-i\xi\cdot t}\,dt \right| \\
&\leqslant C\sum_{k\in\mathbb{Z}^n}\big(1+|x-k|^2\big)^{-n}\big(1+|\xi-k|^2\big)^{-n} \leqslant C\big(1+|x-\xi|^2\big)^{-n}
\end{aligned}$$

for all $x, \xi \in \mathbb{R}^n$. This implies $f \in M^{p,\infty}(\mathbb{R}^n)$ (which is also a consequence of [6] or [10]).

We next consider the second part. Since $\|V_{\varphi_{1/\lambda}} f(\cdot,\xi)\|_{L^p}$ is continuous with respect to $\xi \in \mathbb{R}^n$, we see that $\|V_{\varphi_{1/\lambda}} f\|_{L^{p,\infty}} = \sup_{\xi\in\mathbb{R}^n}\|V_{\varphi_{1/\lambda}} f(\cdot,\xi)\|_{L^p}$ for each $0 < \lambda \leqslant 1$. Hence, by the same argument as in the proof of Lemma 3.9, we have

$$\begin{aligned}
\big\|V_\varphi(f_\lambda)\big\|_{L^{p,\infty}} &= \lambda^{-n(1/p+1)}\|V_{\varphi_{1/\lambda}} f\|_{L^{p,\infty}} \geqslant \lambda^{-n(1/p+1)}\big\|V_{\varphi_{1/\lambda}} f(\cdot,0)\big\|_{L^p} \\
&\geqslant C\lambda^{-n(1/p+1)}\left(\sum_{\ell\in\mathbb{Z}^n}\big|\lambda^n\hat{\varphi}(\lambda\ell)\big|^p\right)^{1/p} \\
&\geqslant C\lambda^{-n/p}\left(\sum_{\substack{|\ell_i|\leqslant 1/\lambda,\\ i=1,\dots,n}}\big|\hat{\varphi}(\lambda\ell)\big|^p\right)^{1/p} \geqslant C\lambda^{-2n/p}
\end{aligned}$$

for all $0 < \lambda \leqslant 1$.

By Lemma 3.7(1), if $1 \leqslant p \leqslant 2$, then $\|f\|_{M^{p,\infty}} \sim \lambda^{-2n/p}$ in the case $0 < \lambda \leqslant 1$. The proof is complete. □

We are now ready to prove Theorem 3.1(1) with $(1/p, 1/q) \in I_3^*$ and (2) with $(1/p, 1/q) \in I_3$.

Proof of Theorem 3.1(2) with $(1/p, 1/q) \in I_3$. We recall that $\mu_2(p,q) = -2/p + 1/q$ if $(1/p, 1/q) \in I_3$. Let $(1/p, 1/q) \in I_3$ and $1/p + 1/q \geqslant 1$. We note that, if $(1/p, 1/q) \in I_3$ and $1/p + 1/q \geqslant 1$, then $(1/p, 1/q) \in I_2^*$ and $1/p \geqslant 1/q$. Then, by Lemma 3.3, there exists a constant $C > 0$ such that

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{-n(2/p-1/q)} \|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1. \tag{3.14}$$

This is the first part of Theorem 3.1(2) with $(1/p, 1/q) \in I_3$ and $1/p+1/q \geqslant 1$. We next consider the case $(1/p, 1/q) \in I_3$, $1/p + 1/q \leqslant 1$ and $q < \infty$. (3.14) implies

$$\|f_\lambda\|_{M^{r,r'}} \leqslant C\lambda^{-n(2/r-1/r')} \|f\|_{M^{r,r'}} = C\lambda^{-n(3/r-1)} \|f\|_{M^{r,r'}} \tag{3.15}$$

for all $f \in M^{r,r'}(\mathbb{R}^n)$ and $0 < \lambda \leqslant 1$, where $1 \leqslant r \leqslant 2$. Then, by interpolation, Lemma 3.5 and (3.15) give

$$\|f_\lambda\|_{M^{p,q}} \leqslant C\lambda^{-n(2/p-1/q)} \|f\|_{M^{p,q}} \quad \text{for all } f \in M^{p,q}(\mathbb{R}^n) \text{ and } 0 < \lambda \leqslant 1,$$

where $(1/p, 1/q) \in I_3$, $1/p + 1/q \leqslant 1$ and $q < \infty$. In the case $q = \infty$, by Lemma 3.7(1), we have nothing to prove. Hence, we obtain the first part of Theorem 3.1(2) with $(1/p, 1/q) \in I_3$ and $1/p + 1/q \leqslant 1$.

By using Lemma 3.9 (or 3.10), we can prove the second part of Theorem 3.1(2) with $(1/p, 1/q) \in I_3$ in the same way as in the proof of the second part of Theorem 3.1(2) with $(1/p, 1/q) \in I_2$. We omit the proof. □

Proof of Theorem 3.1(1) with $(1/p, 1/q) \in I_3^*$. We recall that $\mu_1(p,q) = -2/p + 1/q$ if $(1/p, 1/q) \in I_3^*$. In every case except for $(p,q) \neq (\infty, 1)$, by duality, Theorem 3.1(2) with $(1/p', 1/q') \in I_3$ and the same argument as in the proof of Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$, we can prove Theorem 3.1(1) with $(1/p, 1/q) \in I_3^*$.

For the first part of Theorem 3.1(1) with $(p,q) = (\infty, 1)$, by (3.11), we have nothing to prove. By using interpolation, we can prove the second part in the same way as in the proof of Theorem 3.1(1) with $(1/p, 1/q) \in I_1^*$ and $q = \infty$. □

4. The inclusion between Besov spaces and modulation spaces

In this section, we prove Theorem 1.2 which appeared in Section 1. It is sufficient to prove the first statement only because the first one implies the second one by the duality argument and the elementary relation

$$\nu_2(p,q) = -\nu_1(p', q').$$

See also Section 2 for the dual spaces of the modulation spaces (Lemma 2.4) and Besov spaces.

For the preparation to prove Theorem 1.2(1) with $(1/p, 1/q) \in I_1^*$, we show three lemmas in the below. We denote by B the tensor product of B-spline of degree 2, that is

$$B(t) = \prod_{i=1}^{n} \chi_{[-1/2,1/2]} * \chi_{[-1/2,1/2]}(t_i),$$

where $t = (t_1, \dots, t_n) \in \mathbb{R}^n$. We note that $\operatorname{supp} B \subset [-1, 1]^n$ and $\mathcal{F}^{-1}B \in M^{p,q}(\mathbb{R}^n)$ for all $1 \leqslant p, q \leqslant \infty$.

Lemma 4.1. *Let $1 \leqslant p, q \leqslant \infty$, $(p, q) \neq (1, \infty), (\infty, 1)$ and $\epsilon > 0$. Suppose that $\psi \in \mathcal{S}(\mathbb{R}^n)$ satisfies $\psi = 1$ on $\{\xi\colon |\xi| \leqslant 1/2\}$ and $\operatorname{supp}\psi \subset \{\xi\colon |\xi| \leqslant 1\}$. Set*

$$f(t) = \sum_{\ell \neq 0} |\ell|^{-n/p-\epsilon} \Psi(t - \ell) \quad \text{in } \mathcal{S}'(\mathbb{R}^n),$$

where $\Psi = \mathcal{F}^{-1}\psi$. Then $f \in M^{p,q}(\mathbb{R}^n)$ and there exists a constant $C > 0$ such that $\|f_\lambda\|_{M^{p,q}} \geqslant C\lambda^{-n/p-\epsilon}$ for all $\lambda \geqslant 2\sqrt{n}$.

Proof. In the same way as in the proof of Lemma 3.8, we can prove $f \in M^{p,q}(\mathbb{R}^n)$ (see also [6] or [10]). We consider the second part. Let $\lambda \geqslant 2\sqrt{n}$. Since $\psi(\cdot/\lambda) = 1$ on $[-1, 1]^n$, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} \Psi(\lambda t - \ell)\big(\mathcal{F}^{-1}B\big)(t)\,dt &= (2\pi)^{-n}\lambda^{-n} \int_{\mathbb{R}^n} e^{-i(\ell/\lambda)\cdot t} \psi(t/\lambda)\, B(t)\,dt \\
&= (2\pi)^{-n}\lambda^{-n} \int_{\mathbb{R}^n} e^{-i(\ell/\lambda)\cdot t}\, B(t)\,dt \\
&= (2\pi)^{-n}\lambda^{-n} \prod_{i=1}^{n} \left(\frac{\sin \ell_i/2\lambda}{\ell_i/2\lambda}\right)^2.
\end{aligned}$$

We note that $\prod_{i=1}^n \{(\sin\xi_i)/\xi_i\}^2 \geqslant C$ on $[-\pi/2, \pi/2]^n$ for some constant $C > 0$. Since $\mathcal{F}^{-1}B \in M^{p',q'}(\mathbb{R}^n)$, by Lemmas 2.2 and 2.4, we get

$$\begin{aligned}
\|f_\lambda\|_{M^{p,q}} &= \sup_{\|g\|_{M^{p',q'}}=1} \big|\langle f_\lambda, g\rangle_M\big| \geqslant \big\|\mathcal{F}^{-1}B\big\|_{M^{p',q'}}^{-1} \big|\big\langle f_\lambda, \mathcal{F}^{-1}B\big\rangle_M\big| \\
&= C\left|\sum_{\ell\neq 0} |\ell|^{-n/p-\epsilon} \frac{1}{\|\Phi\|_{L^2}^2} \int_{\mathbb{R}^{2n}} V_\Phi\big[\Psi(\lambda\cdot-\ell)\big](x,\xi)\,\overline{V_\Phi\big[\mathcal{F}^{-1}B\big](x,\xi)}\,dx\,d\xi\right| \\
&= C\left|\sum_{\ell\neq 0} |\ell|^{-n/p-\epsilon} \int_{\mathbb{R}^n} \Psi(\lambda t - \ell)\big(\mathcal{F}^{-1}B\big)(t)\,dt\right| \\
&= C\lambda^{-n}\left|\sum_{\ell\neq 0} |\ell|^{-n/p-\epsilon} \prod_{i=1}^n \left(\frac{\sin\ell_i/2\lambda}{\ell_i/2\lambda}\right)^2\right| \\
&\geqslant C\lambda^{-n} \sum_{\substack{0<|\ell_i|\leqslant\lambda\pi,\\ i=1,\dots,n}} |\ell|^{-n/p-\epsilon} \prod_{i=1}^n \left(\frac{\sin\ell_i/2\lambda}{\ell_i/2\lambda}\right)^2 \\
&\geqslant C\lambda^{-n}\lambda^{-n/p-\epsilon} \sum_{\substack{0<|\ell_i|\leqslant\lambda\pi,\\ i=1,\dots,n}} 1 \geqslant C\lambda^{-n/p-\epsilon}
\end{aligned}$$

for all $\lambda \geqslant 2\sqrt{n}$. The proof is complete. □

Lemma 4.2. *Suppose that* $1 \leqslant p, q \leqslant \infty$, $(p,q) \neq (1,\infty), (\infty,1)$ *and* $\epsilon > 0$. *Let* $\psi \in \mathcal{S}(\mathbb{R}^n)$ *be as in Lemma* 4.1. *Set*

$$f(t) = e^{8it_1} \sum_{\ell \neq 0} |\ell|^{-n/p-\epsilon} \Psi(t-\ell) \quad \textit{in } \mathcal{S}'(\mathbb{R}^n), \tag{4.1}$$

where $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$ *and* $\Psi = \mathcal{F}^{-1}\psi$. *Then* $f \in M^{p,q}(\mathbb{R}^n)$ *and there exists a constant* $C > 0$ *such that* $\|f_\lambda\|_{M^{p,q}} \geqslant C\lambda^{-n/p-\epsilon}$ *for all* $\lambda \geqslant 2\sqrt{n}$.

Proof. Let $g(t) = \sum_{\ell \neq 0} |\ell|^{-n/p-\epsilon} \Psi(t-\ell)$. Since $f = M_{8e_1} g$ and $f_\lambda = M_{8\lambda e_1} g_\lambda$, we have $V_\Phi(f_\lambda)(x,\xi) = V_\Phi(g_\lambda)(x, \xi - 8\lambda e_1)$, where $e_1 = (1, 0, \ldots, 0)$. This gives $\|f_\lambda\|_{M^{p,q}} = \|g_\lambda\|_{M^{p,q}}$. Hence, by Lemma 4.1, we obtain Lemma 4.2. □

Lemma 4.3. *Suppose that* $1 \leqslant p, q \leqslant \infty$, $s \in \mathbb{R}$ *and* $\epsilon > 0$. *Let* f *be defined by* (4.1). *Then there exists a constant* $C > 0$ *such that* $\|f_{2^k}\|_{B_s^{p,q}} \leqslant C 2^{k(s-n/p)}$ *for all* $k \in \mathbb{Z}_+$.

Proof. Let $k \in \mathbb{Z}_+$. Since $\operatorname{supp}\varphi_0 \subset \{\xi\colon |\xi| \leqslant 2\}$, $\operatorname{supp}\varphi_j \subset \{\xi\colon 2^{j-1} \leqslant |\xi| \leqslant 2^{j+1}\}$ if $j \geqslant 1$ (see Section 2), and $\operatorname{supp}\psi(\cdot/2^k - 8e_1) \subset \{\xi\colon |\xi - 2^{k+3}e_1| \leqslant 2^k\}$, we see that

$$\begin{aligned}
&\int_{\mathbb{R}^n} \Phi_j(x-t)\bigl(e^{8i(2^k t_1)} \Psi\bigl(2^k t - \ell\bigr)\bigr)\, dt \\
&\quad = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot t} \varphi_j(t) \bigl(2^{-kn} e^{-i\ell\cdot(t/2^k - 8e_1)} \psi\bigl(t/2^k - 8e_1\bigr)\bigr)\, dt \\
&\quad = \begin{cases} (2\pi)^{-n} e^{8i\ell_1} \int_{\mathbb{R}^n} e^{i(2^k x - \ell)\cdot t} \varphi_j(2^k t) \psi(t - 8e_1)\, dt, & \text{if } k+2 \leqslant j \leqslant k+4, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

Hence,

$$\begin{aligned}
&\bigl|\Phi_j * (f_{2^k})(x)\bigr| \\
&\quad = \left| \sum_{\ell \neq 0} |\ell|^{-n/p-\epsilon} \int_{\mathbb{R}^n} \Phi_j(x-t)\bigl(e^{8i(2^k t_1)} \Psi\bigl(2^k t - \ell\bigr)\bigr)\, dt \right| \\
&\quad \leqslant C \sum_{\ell \neq 0} |\ell|^{-n/p-\epsilon} \left| \int_{\mathbb{R}^n} \bigl\{\bigl(1 + \bigl|2^k x - \ell\bigr|^2\bigr)^{-n} (I - \Delta_t)^n e^{i(2^k x - \ell)\cdot t}\bigr\} \varphi_j\bigl(2^k t\bigr) \psi(t - 8e_1)\, dt \right| \\
&\quad \leqslant C \sum_{\ell \neq 0} |\ell|^{-n/p-\epsilon} \bigl(1 + \bigl|2^k x - \ell\bigr|^2\bigr)^{-n},
\end{aligned}$$

where $k+2 \leqslant j \leqslant k+4$. On the other hand, $\Phi_j * (f_{2^k}) = 0$ if $j < k+2$ or $j > k+4$. Thus, $\|\Phi_j * (f_{2^k})\|_{L^p} \leqslant C 2^{-kn/p}$ if $k+2 \leqslant j \leqslant k+4$, and $\|\Phi_j * (f_{2^k})\|_{L^p} = 0$ if $j < k+2$ or $j > k+4$. Therefore,

$$\|f_{2^k}\|_{B_s^{p,q}} = \left(\sum_{j=k+2}^{k+4} 2^{jsq} \left\| \Phi_j * (f_{2^k}) \right\|_{L^p}^q \right)^{1/q}$$

$$\leqslant C 2^{-kn/p} \left(\sum_{j=k+2}^{k+4} 2^{jsq} \right)^{1/q} \leqslant C 2^{k(s-n/p)}.$$

The proof is complete. □

We are now ready to prove Theorem 1.2(1) with $(1/p, 1/q) \in I_1^*$.

Proof of Theorem 1.2(1) with $(1/p, 1/q) \in I_1^*$. Let $(1/p, 1/q) \in I_1^*$ and $(p, q) \neq (1, \infty)$. Then $\nu_1(p, q) = 0$. We assume that $B_s^{p,q}(\mathbb{R}^n) \hookrightarrow M^{p,q}(\mathbb{R}^n)$, where $s < 0$. Set $s = -\epsilon$, where $\epsilon > 0$. For this ϵ, we define f by

$$f(t) = e^{8it_1} \sum_{\ell \neq 0} |\ell|^{-n/p-\epsilon/2} \Psi(t - \ell),$$

where $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$, $\Psi = \mathcal{F}^{-1}\psi$ and ψ is as in Lemma 4.1. Then, by Lemmas 4.2 and 4.3, we have

$$C_1 2^{-k(n/p+\epsilon/2)} \leqslant \|f_{2^k}\|_{M^{p,q}} \leqslant C_2 \|f_{2^k}\|_{B_s^{p,q}} \leqslant C_3 2^{k(s-n/p)} = C_3 2^{-k(n/p+\epsilon)}$$

for any large integer k. However, this is a contradiction. Hence, s must satisfy $s \geqslant 0$.

We next consider the case $(p, q) = (1, \infty)$. Assume $B_s^{1,\infty}(\mathbb{R}^n) \hookrightarrow M^{1,\infty}(\mathbb{R}^n)$. Let $\psi \in \mathcal{S}(\mathbb{R}^n) \setminus \{0\}$ be such that $\operatorname{supp} \psi \subset \{\xi\colon 1/2 \leqslant |\xi| \leqslant 2\}$. Since $M^{1,\infty}(\mathbb{R}^n) \hookrightarrow \mathcal{F}L^\infty(\mathbb{R}^n)$ [18, Proposition 1.7], we see that

$$2^{-kn} \|\psi\|_{L^\infty} = \left\| \mathcal{F}[\Psi_{2^k}] \right\|_{L^\infty} \leqslant C \|\Psi_{2^k}\|_{M^{1,\infty}} \quad \text{for all } k \in \mathbb{Z}_+,$$

where $\Psi = \mathcal{F}^{-1}\psi$. On the other hand, it is easy to show that

$$\|\Psi_{2^k}\|_{B_s^{1,\infty}} \leqslant C 2^{k(s-n)} \quad \text{for all } k \in \mathbb{Z}_+.$$

Hence, by our assumption, we get

$$2^{-kn} \|\psi\|_{L^\infty} \leqslant C_1 \|\Psi_{2^k}\|_{M^{1,\infty}} \leqslant C_2 \|\Psi_{2^k}\|_{B_s^{1,\infty}} \leqslant C_3 2^{k(s-n)}$$

for all $k \in \mathbb{Z}_+$. This implies $s \geqslant 0$. The proof is complete. □

Our next goal is to prove Theorem 1.2(1) with $(1/p, 1/q) \in I_2^*$. We remark the following fact, and give the proof for reader's convenience.

Lemma 4.4. *(See [16, Proposition 1.1].) Let $1 \leqslant p, q \leqslant \infty$ and $s > 0$. Then there exists a constant $C > 0$ such that*

$$\|f_\lambda\|_{B_s^{p,q}} \leqslant C \lambda^{s-n/p} \|f\|_{B_s^{p,q}} \quad \textit{for all } f \in B_s^{p,q}\big(\mathbb{R}^n\big) \textit{ and } \lambda \geqslant 1.$$

Proof. Let $j_0 \in \mathbb{Z}_+$ be such that $2^{j_0} \leqslant \lambda < 2^{j_0+1}$. Since $\sum_{j=0}^{\infty} \varphi_j(\xi) = 1$ for all $\xi \in \mathbb{R}^n$, we see that

$$\varphi_j(\lambda\xi) = \sum_{\ell=-2}^{1} \varphi_j(\lambda\xi)\,\varphi_{j+\ell}\big(2^{j_0}\xi\big) \quad \text{for all } \xi \in \mathbb{R}^n \text{ and } j \in \mathbb{Z}_+,$$

where $\varphi_{j+\ell} = 0$ if $j+\ell < 0$. Hence, by Young's inequality, we have

$$\begin{aligned}
\|f_\lambda\|_{B_s^{p,q}} &= \left(\sum_{j=0}^{\infty} 2^{jsq} \big\|\mathcal{F}^{-1}[\varphi_j \widehat{f_\lambda}]\big\|_{L^p}^q\right)^{1/q} \\
&= \lambda^{-n/p}\left(\sum_{j=0}^{\infty} 2^{jsq} \big\|\mathcal{F}^{-1}\big[\varphi_j(\lambda\cdot)\widehat{f}\big]\big\|_{L^p}^q\right)^{1/q} \\
&\leqslant \lambda^{-n/p} \sum_{\ell=-2}^{1}\left(\sum_{j=0}^{\infty} 2^{jsq} \big\|\mathcal{F}^{-1}\big[\varphi_j(\lambda\cdot)\varphi_{j+\ell}\big(2^{j_0}\cdot\big)\widehat{f}\big]\big\|_{L^p}^q\right)^{1/q} \\
&\leqslant \lambda^{-n/p} \sum_{\ell=-2}^{1}\left\{\sum_{j=0}^{\infty} 2^{jsq} \big(\big\|\mathcal{F}^{-1}\big[\varphi_j(\lambda\cdot)\big]\big\|_{L^1} \big\|\mathcal{F}^{-1}\big[\varphi_{j+\ell}\big(2^{j_0}\cdot\big)\hat{f}\big]\big\|_{L^p}\big)^q\right\}^{1/q} \\
&\leqslant C\lambda^{-n/p}\left(\sum_{j=0}^{\infty} 2^{jsq} \big\|\mathcal{F}^{-1}\big[\varphi_j\big(2^{j_0}\cdot\big)\widehat{f}\big]\big\|_{L^p}^q\right)^{1/q} \\
&= C\lambda^{-n/p}\left\{\left(\sum_{j=0}^{j_0} + \sum_{j=j_0+1}^{\infty}\right) 2^{jsq} \big\|\mathcal{F}^{-1}\big[\varphi_j\big(2^{j_0}\cdot\big)\widehat{f}\big]\big\|_{L^p}^q\right\}^{1/q}.
\end{aligned}$$

For the first term, we see that

$$\begin{aligned}
\sum_{j=0}^{j_0} 2^{jsq} \big\|\mathcal{F}^{-1}\big[\varphi_j\big(2^{j_0}\cdot\big)\widehat{f}\big]\big\|_{L^p}^q &= \sum_{j=0}^{j_0} 2^{jsq} \big\|\mathcal{F}^{-1}\big[\varphi_j\big(2^{j_0}\cdot\big)(\varphi_0+\varphi_1)\widehat{f}\big]\big\|_{L^p}^q \\
&\leqslant C\sum_{j=0}^{j_0} 2^{jsq} \big\|\mathcal{F}^{-1}\big[(\varphi_0+\varphi_1)\widehat{f}\big]\big\|_{L^p}^q \\
&\leqslant C\big(2^{j_0 s}\|f\|_{B_s^{p,q}}\big)^q \leqslant C\big(\lambda^s\|f\|_{B_s^{p,q}}\big)^q.
\end{aligned}$$

For the second term, we have

$$\sum_{j=j_0+1}^{\infty} 2^{jsq} \big\|\mathcal{F}^{-1}\big[\varphi_j\big(2^{j_0}\cdot\big)\widehat{f}\big]\big\|_{L^p}^q = \sum_{j=j_0+1}^{\infty} 2^{jsq} \big\|\mathcal{F}^{-1}[\varphi_{j-j_0}\widehat{f}]\big\|_{L^p}^q \leqslant \big(\lambda^s\|f\|_{B_s^{p,q}}\big)^q.$$

Combining these estimates, we obtain the desired result. $\square$

We are now ready to prove Theorem 1.2(1) with $(1/p, 1/q) \in I_2^*$.

Proof of Theorem 1.2(1) with $(1/p, 1/q) \in I_2^*$. Let $(1/p, 1/q) \in I_2^*$. Then $\nu_1(p,q) = 1/p + 1/q - 1$. If $(1/p, 1/q) \in I_2^*$ and $1/p + 1/q = 1$ then $(1/p, 1/q) \in I_1^*$, and we have already proved this case in Theorem 1.2(1) with $(1/p, 1/q) \in I_1^*$. Hence, we may assume $1/p + 1/q > 1$. Suppose that $B_s^{p,q}(\mathbb{R}^n) \hookrightarrow M^{p,q}(\mathbb{R}^n)$, where $s < n(1/p + 1/q - 1)$. Then, since $n(1/p + 1/q - 1) > 0$, we can take $s_0 > 0$ such that $s \leqslant s_0 < n(1/p + 1/q - 1)$. Let φ be the Gauss function. By Lemma 2.1, we see that $\|\varphi_\lambda\|_{M^{p,q}} \geqslant C\lambda^{n(1/q-1)}$ for all $\lambda \geqslant 1$. On the other hand, by Lemma 4.4, we have

$$\|\varphi_\lambda\|_{B_{s_0}^{p,q}} \leqslant C\lambda^{s_0 - n/p}\|\varphi\|_{B_{s_0}^{p,q}} \quad \text{for all } \lambda \geqslant 1.$$

Hence, using $B_{s_0}^{p,q}(\mathbb{R}^n) \hookrightarrow B_s^{p,q}(\mathbb{R}^n) \hookrightarrow M^{p,q}(\mathbb{R}^n)$, we get

$$C_1\lambda^{n(1/q-1)} \leqslant \|\varphi_\lambda\|_{M^{p,q}} \leqslant C_2\|\varphi_\lambda\|_{B_{s_0}^{p,q}} \leqslant C_3\lambda^{s_0-n/p}\|\varphi\|_{B_{s_0}^{p,q}}$$

for all $\lambda \geqslant 1$. However, since $s_0 - n/p < n(1/q - 1)$, this is a contradiction. Therefore, s must satisfy $s \geqslant n(1/p + 1/q - 1)$. The proof is complete. □

Our last goal is to prove Theorem 1.2(1) with $(1/p, 1/q) \in I_3^*$.

Lemma 4.5. *Let $1 \leqslant p \leqslant \infty$, $1 \leqslant q < \infty$ and $\epsilon > 0$. Suppose that $\varphi, \psi \in \mathcal{S}(\mathbb{R}^n) \setminus \{0\}$ satisfy $\operatorname{supp}\varphi \subset [-1/8, 1/8]^n$, $\operatorname{supp}\psi \subset [-1/2, 1/2]^n$ and $\psi = 1$ on $[-1/4, 1/4]^n$. For $j \in \mathbb{Z}_+$, set*

$$f^j(t) = 2^{-jn/p} \sum_{\substack{0<|k_i|\leqslant 2^j,\\ i=1,\ldots,n}} |k|^{-n/p-\epsilon} e^{ik\cdot t/2^j}\Psi\bigl(t/2^j - k\bigr), \tag{4.2}$$

where $\Psi = \mathcal{F}^{-1}\psi$. Then $f^j \in M^{p,q}(\mathbb{R}^n)$ and there exists a constant $C > 0$ such that

$$\bigl\|V_\Phi\bigl[\bigl(f^j\bigr)_{2^j}\bigr]\bigr\|_{L^{p,q}} \geqslant C2^{-jn(2/p-1/q)-j\epsilon} \quad \textit{for all } j \in \mathbb{Z}_+,$$

where $\Phi = \mathcal{F}^{-1}\varphi$.

Proof. Since $f^j \in \mathcal{S}(\mathbb{R}^n)$, we have $f^j \in M^{p,q}(\mathbb{R}^n)$. We consider the second part. Note that $\operatorname{supp}\varphi(\cdot - \xi) \subset \ell + [-1/4, 1/4]^n$ for all $\ell \in \mathbb{Z}^n$ and $\xi \in \ell + [-1/8, 1, 8]^n$. Since $\operatorname{supp}\psi(\cdot - k) \subset k + [-1/2, 1/2]^n$ and $\psi(t-k) = 1$ if $t \in k + [-1/4, 1/4]^n$, it follows that

$$\begin{aligned}
&\bigl\|V_\Phi\bigl[\bigl(f^j\bigr)_{2^j}\bigr]\bigr\|_{L^{p,q}} \\
&\quad\geqslant \Biggl\{\sum_{\ell\in\mathbb{Z}^n}\int_{\ell+[-1/8,1/8]^n}\Biggl(\int_{\mathbb{R}^n}\Biggl|2^{-jn/p}\sum_{\substack{0<|k_i|\leqslant 2^j,\\ i=1,\ldots,n}}|k|^{-n/p-\epsilon} \\
&\qquad\times\int_{\mathbb{R}^n} e^{ik\cdot t}\Psi(t-k)\overline{\Phi(t-x)}e^{-i\xi\cdot t}\,dt\Biggr|^p dx\Biggr)^{q/p} d\xi\Biggr\}^{1/q}
\end{aligned}$$

$$\geqslant (2\pi)^{-n} 2^{-jn/p} \Biggl\{ \sum_{\substack{0<|\ell_i|\leqslant 2^j,\\ i=1,\dots,n}} \int_{\ell+[-1/8,1/8]^n} \Biggl(\int_{\mathbb{R}^n} \Biggl| \sum_{\substack{0<|k_i|\leqslant 2^j,\\ i=1,\dots,n}} |k|^{-n/p-\epsilon} e^{i|k|^2}$$

$$\times \int_{\mathbb{R}^n} e^{-ik\cdot t} \psi(t-k) \overline{\varphi(t-\xi)} e^{ix\cdot t}\, dt \Biggr|^p dx \Biggr)^{q/p} d\xi \Biggr\}^{1/q}$$

$$= (2\pi)^{-n} 2^{-jn/p}$$

$$\times \Biggl\{ \sum_{\substack{0<|\ell_i|\leqslant 2^j,\\ i=1,\dots,n}} \int_{\ell+[-1/8,1/8]^n} \Biggl(\int_{\mathbb{R}^n} \Biggl| |\ell|^{-n/p-\epsilon} \int_{\mathbb{R}^n} e^{i(x-\ell)\cdot t} \overline{\varphi(t-\xi)}\, dt \Biggr|^p dx \Biggr)^{q/p} d\xi \Biggr\}^{1/q}$$

$$= 2^{-jn/p} \Biggl\{ \sum_{\substack{0<|\ell_i|\leqslant 2^j,\\ i=1,\dots,n}} |\ell|^{-(n/p+\epsilon)q} \int_{\ell+[-1/8,1/8]^n} \bigl\| \Phi(-\cdot+\ell) \bigr\|_{L^p}^q \, d\xi \Biggr\}^{1/q}$$

$$= 4^{-n/q} \|\Phi\|_{L^p} 2^{-jn/p} \Biggl\{ \sum_{\substack{0<|\ell_i|\leqslant 2^j,\\ i=1,\dots,n}} |\ell|^{-(n/p+\epsilon)q} \Biggr\}^{1/q}$$

$$\geqslant C_n 2^{-jn/p} 2^{-j(n/p+\epsilon)} \Biggl\{ \sum_{\substack{0<|\ell_i|\leqslant 2^j,\\ i=1,\dots,n}} 1 \Biggr\}^{1/q} \geqslant C_n 2^{-jn(2/p-1/q)-j\epsilon}$$

for all $j \in \mathbb{Z}_+$. The proof is complete. □

Lemma 4.6. *Suppose that $1 \leqslant p, q \leqslant \infty$ and $s > 0$. Let f^j be defined by* (4.2). *Then there exists a constant $C > 0$ such that $\|(f^j)_{2^j}\|_{B_s^{p,q}} \leqslant C 2^{j(s-n/p)}$ for all $j \in \mathbb{Z}_+$.*

Proof. By Lemma 4.4, we have $\|(f^j)_{2^j}\|_{B_s^{p,q}} \leqslant C 2^{j(s-n/p)} \|f^j\|_{B_s^{p,q}}$ for all $j \in \mathbb{Z}_+$. Hence, it is enough to prove that $\sup_{j\in\mathbb{Z}_+} \|f^j\|_{B_s^{p,q}} < \infty$. Since

$$\widehat{f^j}(\xi) = 2^{jn(1-1/p)} \sum_{\substack{0<|k_i|\leqslant 2^j,\\ i=1,\dots,n}} |k|^{-n/p-\epsilon}\, e^{-ik\cdot(2^j\xi-k)} \psi\bigl(2^j\xi-k\bigr)$$

and $\operatorname{supp} \psi(2^j\cdot -k) \subset k/2^j + [-2^{-(j+1)}, 2^{-(j+1)}]^n$, we see that $\operatorname{supp} \widehat{f^j} \subset \{\xi\colon |\xi| \leqslant 2\sqrt{n}\}$. Let ℓ_0 be such that $2^{\ell_0-1} \geqslant 2\sqrt{n}$. Then,

$$\|f^j\|_{B_s^{p,q}} = \Biggl(\sum_{\ell=0}^{\ell_0-1} 2^{\ell sq} \bigl\| \Phi_\ell * f^j \bigr\|_{L^p}^q \Biggr)^{1/q}$$

$$\leqslant \Biggl(\sum_{\ell=0}^{\ell_0-1} 2^{\ell sq} \bigl(\|\Phi_\ell\|_{L^1} \|f^j\|_{L^p} \bigr)^q \Biggr)^{1/q} = C_n \|f^j\|_{L^p}.$$

Therefore, it is enough to show that $\sup_{j\in\mathbb{Z}_+}\|f^j\|_{L^p}<\infty$. By a change of variable, we have

$$\begin{aligned}\|f^j\|_{L^p} &= \biggl(\int_{\mathbb{R}^n}\biggl|\sum_{\substack{0<|k_i|\leqslant 2^j,\\ i=1,\dots,n}}|k|^{-n/p-\epsilon}e^{ik\cdot t}\Psi(t-k)\biggr|^p\,dt\biggr)^{1/p}\\ &\leqslant \biggl\{\sum_{m\in\mathbb{Z}^n}\int_{m+[-1/2,1/2]^n}\biggl(\sum_{k\neq 0}|k|^{-n/p-\epsilon}\bigl|\Psi(t-k)\bigr|\biggr)^p\,dt\biggr\}^{1/p}\\ &\leqslant C\biggl\{\sum_{m\in\mathbb{Z}^n}\biggl(\sum_{k\neq 0}|k|^{-n/p-\epsilon}\bigl(1+|m-k|\bigr)^{-n-1}\biggr)^p\biggr\}^{1/p}<\infty\end{aligned}$$

for all $j\in\mathbb{Z}_+$. The proof is complete. □

We are now ready to prove Theorem 1.2(1) with $(1/p,1/q)\in I_3^*$.

Proof of Theorem 1.2(1) with $(1/p,1/q)\in I_3^*$. Let $(1/p,1/q)\in I_3^*$. Then $\nu_1(p,q)=-1/p+1/q$. If $(1/p,1/q)\in I_3^*$ and $p=q$ then $(1/p,1/q)\in I_1^*$, and we have already proved this case in Theorem 1.2(1) with $(1/p,1/q)\in I_1^*$. Hence, we may assume $1/q>1/p$. Note that $q\neq\infty$. Suppose that $B_s^{p,q}(\mathbb{R}^n)\hookrightarrow M^{p,q}(\mathbb{R}^n)$, where $s<-n(1/p-1/q)$. Then, since $-n(1/p-1/q)>0$, we can take $s_0>0$ such that $s\leqslant s_0<-n(1/p-1/q)$. Set $s_0=-n(1/p-1/q)-\epsilon$, where $\epsilon>0$. For this ϵ, we define f^j by

$$f^j(t)=2^{-jn/p}\sum_{\substack{0<|k_i|\leqslant 2^j,\\ i=1,\dots,n}}|k|^{-n/p-\epsilon/2}e^{ik\cdot t/2^j}\Psi\bigl(t/2^j-k\bigr),$$

where $j\in\mathbb{Z}_+$, $\Psi=\mathcal{F}^{-1}\psi$ and ψ is as in Lemma 4.5. Then, since $B_{s_0}^{p,q}(\mathbb{R}^n)\hookrightarrow B_s^{p,q}(\mathbb{R}^n)\hookrightarrow M^{p,q}(\mathbb{R}^n)$, by Lemmas 4.5 and 4.6, we get

$$\begin{aligned}C_1 2^{-jn(2/p-1/q)-j\epsilon/2} &\leqslant \bigl\|V_\Phi\bigl[\bigl(f^j\bigr)_{2^j}\bigr]\bigr\|_{L^{p,q}}\leqslant C_2\bigl\|\bigl(f^j\bigr)_{2^j}\bigr\|_{M^{p,q}}\\ &\leqslant C_3\bigl\|\bigl(f^j\bigr)_{2^j}\bigr\|_{B_{s_0}^{p,q}}\leqslant C_4 2^{j(s_0-n/p)}=C_4 2^{-jn(2/p-1/q)-j\epsilon}\end{aligned}$$

for all $j\in\mathbb{Z}_+$, where $\Phi=\mathcal{F}^{-1}\varphi$ and φ is as in Lemma 4.5. However, this is a contradiction. Therefore, s must satisfy $s\geqslant -n(1/p-1/q)$. The proof is complete. □

Acknowledgments

The authors would like to thank the referee for valuable comments and remarks. Especially, the lower estimates in Theorem 1.1, which are equivalent to the upper estimates, are added by his/her suggestion. They also thank Professor Joachim Toft for his kind remarks on the inclusions between Besov and modulation spaces.

References

[1] A. Boulkhemair, Remarks on a Wiener type pseudodifferential algebra and Fourier integral operators, Math. Res. Lett. 4 (1997) 53–67.

[2] M. Dimassi, J. Sjöstrand, Spectral Asymptotics in the Semi-Classical Limit, London Math. Soc. Lecture Note Ser., vol. 268, Cambridge Univ. Press, Cambridge, 1999.

[3] H.G. Feichtinger, Banach spaces of distributions of Wiener's type and interpolation, in: P. Butzer, B.Sz. Nagy, E. Görlich (Eds.), Proc. Conf. Oberwolfach, Functional Analysis and Approximation, August 1980, in: Int. Ser. Num. Math., vol. 60, Birkhäuser, Basel, 1981, pp. 153–165.

[4] H.G. Feichtinger, Modulation spaces on locally compact abelian groups, in: M. Krishna, R. Radha, S. Thangavelu (Eds.), Wavelets and Their Applications, Chennai, India, Allied Publ., New Delhi, 2003, pp. 99–140, updated version of a technical report, University of Vienna, 1983.

[5] H.G. Feichtinger, Modulation spaces: Looking back and ahead, Sampl. Theory Signal Image Process. 5 (2006) 109–140.

[6] H.G. Feichtinger, K. Gröchenig, Banach spaces related to integrable group representations and their atomic decompositions, I, J. Funct. Anal. 86 (1989) 307–340.

[7] G.B. Folland, Harmonic Analysis in Phase Space, Princeton Univ. Press, Princeton, NJ, 1989.

[8] M. Frazier, B. Jawerth, Decomposition of Besov spaces, Indiana Univ. Math. J. 34 (1985) 777–799.

[9] P. Gröbner, Banachräume Glatter Funktionen und Zerlegungsmethoden, Thesis, University of Vienna, Vienna, 1983.

[10] K. Gröchenig, Foundations of Time–Frequency Analysis, Birkhäuser Boston, Boston, MA, 2001.

[11] K. Gröchenig, Time–frequency analysis of Sjöstrand's class, Rev. Mat. Iberoamericana 22 (2006) 703–724.

[12] K. Gröchenig, C. Heil, Modulation spaces and pseudodifferential operators, Integral Equations Operator Theory 34 (1999) 439–457.

[13] K. Gröchenig, C. Heil, Modulation spaces as symbol classes for pseudodifferential operators, in: M. Krishna, R. Radha, S. Thangavelu (Eds.), Wavelets and Their Applications, Chennai, India, Allied Publ., New Delhi, 2003, pp. 151–170.

[14] K. Okoudjou, Embedding of some classical Banach spaces into modulation spaces, Proc. Amer. Math. Soc. 132 (2004) 1639–1647.

[15] J. Sjöstrand, An algebra of pseudodifferential operators, Math. Res. Lett. 1 (1994) 185–192.

[16] M. Sugimoto, Pseudo-differential operators on Besov spaces, Tsukuba J. Math. 12 (1988) 43–63.

[17] N. Teofanov, Ultramodulation Spaces and Pseudodifferential Operators, Endowment Andrejević, Beograd, 2003.

[18] J. Toft, Continuity properties for modulation spaces, with applications to pseudo-differential calculus, I, J. Funct. Anal. 207 (2004) 399–429.

[19] J. Toft, Continuity properties for modulation spaces, with applications to pseudo-differential calculus, II, Ann. Global Anal. Geom. 26 (2004) 73–106.

[20] J. Toft, Convolution and embeddings for weighted modulation spaces, in: R. Ashino, P. Boggiatto, M.W. Wong (Eds.), Advances in Pseudo-Differential Operators, in: Oper. Theory Adv. Appl., vol. 155, Birkhäuser, Basel, 2004, pp. 165–186.

[21] H. Triebel, Modulation spaces on the Euclidean n-spaces, Z. Anal. Anwendungen 2 (1983) 443–457.

[22] B. Wang, C. Huang, Frequency-uniform decomposition method for the generalized BO, KdV and NLS equations, preprint.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 248 (2007) 107–127
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

C^*- and JB^*-algebras generated by a nonself-adjoint idempotent ☆

Julio Becerra Guerrero [a], Ángel Rodríguez Palacios [b,*]

[a] *Universidad de Granada, Facultad de Ciencias, Departamento de Matemática Aplicada, 18071 Granada, Spain*
[b] *Universidad de Granada, Facultad de Ciencias, Departamento de Análisis Matemático, 18071 Granada, Spain*

Received 13 June 2006; accepted 15 March 2007

Available online 11 May 2007

Communicated by Paul Malliavin

Abstract

Let A be a C^*-algebra generated by a nonself-adjoint idempotent e, and put $K := \mathrm{sp}(\sqrt{e^*e}) \setminus \{0\}$. It is known that K is a compact subset of $[1, \infty[$ whose maximum element is greater than 1, and that, in general, no more can be said about K. We prove that, if 1 does not belong to K, then A is $*$-isomorphic to the C^*-algebra $C(K, M_2(\mathbb{C}))$ of all continuous functions from K to the C^*-algebra $M_2(\mathbb{C})$ (of all 2×2 complex matrices), and that, if 1 belongs to K, then A is $*$-isomorphic to a distinguished proper C^*-subalgebra of $C(K, M_2(\mathbb{C}))$. By replacing C^*-algebra with JB^*-algebra, $\mathrm{sp}(\sqrt{e^*e}) \setminus \{0\}$ with the triple spectrum $\sigma(e)$ of e, and $M_2(\mathbb{C})$ with the three-dimensional spin factor C_3, similar results are obtained.

Keywords: C^*-algebra; JB^*-algebra; Idempotent

1. Introduction

Let A be a C^*-algebra generated by a nonself-adjoint idempotent e, and put

$$K := \mathrm{sp}\big(\sqrt{e^*e}\big) \setminus \{0\},$$

☆ Partially supported by Junta de Andalucía grants FQM 0199 and FQM 1215, and Project I+D MCYT MTM-2004-03882.

* Corresponding author.
E-mail addresses: juliobg@ugr.es (J.B. Guerrero), apalacio@ugr.es (Á.R. Palacios).

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2007.03.026

where $\mathrm{sp}(\cdot)$ means spectrum. We proved in [1] that K is a compact subset of $[1, \infty[$ whose maximum element is greater than 1, and that, in general, no more can be said about K. Moreover, we got an "almost description" of A (collected in Proposition 2.4 of the present paper) in terms of a Banach $*$-algebra $\mathcal{A}(K)$, which consists of all 2×2 matrices over $C(K)$ with an unusual but natural multiplication. In the present paper we obtain a complete description of A. We prove that, if 1 does not belong to K, then A is $*$-isomorphic to the C^*-algebra $C(K, M_2(\mathbb{C}))$ of all continuous functions from K to the C^*-algebra $M_2(\mathbb{C})$ of all 2×2 complex matrices (Theorem 2.8). To study the case that 1 belongs to K, we need to introduce a distinguished proper C^*-subalgebra of $C(K, M_2(\mathbb{C}))$, namely the one (denoted by $C_p(K, M_2(\mathbb{C}))$) consisting of all elements $\alpha \in C(K, M_2(\mathbb{C}))$ such that $\alpha(1)$ belongs to $\mathbb{C}p$, for a given self-adjoint idempotent $p \in M_2(\mathbb{C})$ different from 0 and 1. It is easy to see that $C_p(K, M_2(\mathbb{C}))$ does not depend structurally on p. We prove that, if 1 belongs to K, then A is $*$-isomorphic to $C_p(K, M_2(\mathbb{C}))$, for p as above (Theorem 3.3).

Among the consequences of the results reviewed in the preceding paragraph, we emphasize the one asserting that a C^*-algebra contains a non-central self-adjoint idempotent if and only if it contains a copy of either $M_2(\mathbb{C})$ or $C_p([1, 2], M_2(\mathbb{C}))$ for any self-adjoint idempotent $p \in M_2(\mathbb{C})$ different from 0 and 1 (Corollary 4.3 and Remark 4.4). It is also worth mentioning the fact that, if a C^*-algebra A contains a non-central idempotent e, then there exists a continuous mapping $r \to e_r$ from $[1, \infty[$ to the set of idempotents of A satisfying $e_{\|e\|} = e$ and $\|e_r\| = r$ for every $r \in [1, \infty[$ (Proposition 4.5).

The concluding sections of the paper (Sections 5 and 6) are devoted to prove the appropriate variants, for JB^*-algebras, of the results previously obtained for C^*-algebras. We show that, by replacing C^*-algebra with JB^*-algebra, $\mathrm{sp}(\sqrt{e^*e}) \setminus \{0\}$ with the triple spectrum $\sigma(e)$ of e (for a given idempotent e), and $M_2(\mathbb{C})$ with the three-dimensional (complex) spin factor C_3, all results reviewed above remain true. As a consequence, a JB-algebra contains a non-central idempotent if and only if it contains a copy of either $\mathcal{S}_3$ or $C_p([1, 2], \mathcal{S}_3)$ for any idempotent $p \in \mathcal{S}_3$ different from 0 and 1, where $\mathcal{S}_3$ stands for the three-dimensional real spin factor (Corollary 6.8).

Turning out to the world of C^*-algebras, let us review the fact, proved in Corollary 4.7, that a C^*-algebra contains a non-central self-adjoint idempotent if and only if it contains a non-normal partial isometry. In the case of JB^*-algebras, we have been able to prove the "only if" part of the appropriate variant of the fact just reviewed (see Corollary 6.6), but have been unable to prove or disprove the "if" part. We note that, if such an "if" part were proved, then, in particular, we would be provided with an affirmative answer to the unsolved question whether every JB^*-algebra containing a non-zero tripotent must contain a non-zero self-adjoint idempotent.

2. The case of C^*-algebras: the first theorem

Let A be an associative complex algebra. In the case that A has not a unit, we denote by A_1 the algebra obtained by adjoining a unit to A. Otherwise, we put $A_1 := A$. As usual, for $a \in A$, we define the spectrum of a (relative to A) as the subset $\mathrm{sp}(A, a)$ of $\mathbb{C}$ given by

$$\mathrm{sp}(A, a) := \{\lambda \in \mathbb{C}\colon\ a - \lambda \text{ is not invertible in } A_1\},$$

and we recall that, if A is in fact a Banach algebra, then $\mathrm{sp}(A, a)$ is a non-empty compact subset of $\mathbb{C}$.

From now on, $M_2(\mathbb{C})$ will denote the C^*-algebra of all 2×2 matrices with entries in $\mathbb{C}$.

Lemma 2.1. *Let e be an idempotent in $M_2(\mathbb{C})$ different from 0 and 1, and put $e_{11} := \|e\|^{-2}e^*e$, $e_{12} := \|e\|^{-1}e^*$, $e_{21} := \|e\|^{-1}e$, and $e_{22} := \|e\|^{-2}ee^*$. Then, for $i, j, k, l \in \{1, 2\}$, we have $e_{ij}^* = e_{ji}$, $e_{ij}e_{kl} = e_{il}$ if $j = k$, and $e_{ij}e_{kl} = \|e\|^{-1}e_{il}$ if $j \neq k$.*

Proof. The equality $e_{ij}^* = e_{ji}$ is clear. On the other hand, we have $\mathrm{sp}(M_2(\mathbb{C}), e^*e) = \{0, \|e\|^2\}$, and hence $(e^*e - \|e\|^2)e^*e = 0$, which reads as $e_{11}^2 = e_{11}$. Analogously, $e_{22}^2 = e_{22}$. Now we have

$$\begin{aligned}\big(ee^*e - \|e\|^2e\big)^*\big(ee^*e - \|e\|^2e\big) &= \big(e^*ee^* - \|e\|^2e^*\big)\big(ee^*e - \|e\|^2e\big)\\ &= \big(e^*e\big)^3 - 2\|e\|^2\big(e^*e\big)^2 + \|e\|^4e^*e = 0,\end{aligned}$$

and hence $ee^*e - \|e\|^2e = 0$, which reads as both $e_{21}e_{11} = e_{21}$ and $e_{22}e_{21} = e_{21}$. By taking adjoints, we deduce $e_{11}e_{12} = e_{12}$ and $e_{12}e_{22} = e_{12}$. The remaining assertions in the lemma are either obvious or easily deducible from the above computations. $\square$

The mapping $\eta : [1, \infty[\to M_2(\mathbb{C})$, which is introduced in Lemma 2.2 immediately below, will play a crucial role through the paper.

Lemma 2.2. *Let t be in $[1, \infty[$, and let $\eta(t)$ denote the element of $M_2(\mathbb{C})$ defined by*

$$\eta(t) := \frac{1}{2}\begin{pmatrix} 1 & t + \sqrt{t^2 - 1} \\ t - \sqrt{t^2 - 1} & 1 \end{pmatrix}.$$

Then $\eta(t)$ is an idempotent satisfying $\|\eta(t)\| = t$. As a consequence, putting $\eta_{11}(t) := t^{-2}\eta(t)^*\eta(t)$, $\eta_{12}(t) := t^{-1}\eta(t)^*$, $\eta_{21}(t) := t^{-1}\eta(t)$, and $\eta_{22}(t) := t^{-2}\eta(t)\eta(t)^*$, we have $\eta_{ij}(t)^* = \eta_{ji}(t)$, $\eta_{ij}(t)\eta_{kl}(t) = \eta_{il}(t)$ if $j = k$, and $\eta_{ij}(t)\eta_{kl}(t) = t^{-1}\eta_{il}(t)$ if $j \neq k$.

Proof. That $\eta(t)$ is an idempotent in $M_2(\mathbb{C})$ is straightforward. Moreover, computing its norm accordingly to the formula in the introduction of [4], we have $\|\eta(t)\| = t$. The consequence follows from Lemma 2.1. $\square$

Let K be a subset of $[1, \infty[$. We denote by η_K the restriction to K of the continuous mapping $t \to \eta(t)$ from $[1, \infty[$ to $M_2(\mathbb{C})$, given by Lemma 2.2. Moreover, for $i, j \in \{1, 2\}$, we denote by η_{ij}^K the restriction to K of the continuous mapping $t \to \eta_{ij}(t)$ from $[1, \infty[$ to $M_2(\mathbb{C})$, given by that lemma.

Now, let K be a compact subset of $[1, \infty[$. Let u stand for the element of $C(K)$ defined by $u(t) := t$ for every $t \in K$. We denote by $\mathcal{A}(K)$ the complex Banach $*$-algebra whose vector space is that of all 2×2 matrices with entries in $C(K)$, whose (bilinear) product is determined by the equalities $(f[ij])(g[kl]) := (fg)[il]$ if $j = k$ and $(f[ij])(g[kl]) := (u^{-1}fg)[il]$ if $j \neq k$, whose norm is given by $\|(f_{ij})\| := \|f_{11}\| + \|f_{12}\| + \|f_{21}\| + \|f_{22}\|$, and whose (conjugate-linear) involution $*$ is determined by $(f[ij])^* := \overline{f}[ji]$. Here, as usual, for $f \in C(K)$ and $i, j \in \{1, 2\}$, $f[ij]$ means the matrix having f in the (i, j)-position and 0's elsewhere. It is useful to see $\mathcal{A}(K)$ as a $C(K)$-module in the natural manner, namely by defining the product of a function $f \in C(K)$ and a matrix $(f_{ij}) \in \mathcal{A}(K)$ by $f(f_{ij}) := (ff_{ij})$. In this regarding, we straightforwardly realize that $\mathcal{A}(K)$ becomes in fact an algebra over $C(K)$, i.e., the operators of left and right multiplication by arbitrary elements of $\mathcal{A}(K)$ are $C(K)$-module homomorphisms. Moreover, the symbol $f[ij]$ can now be read as the product of the function $f \in C(K)$ and the matrix

$[ij] \in \mathcal{A}(K)$, where, for $i, j \in \{1, 2\}$, $[ij]$ stands for the matrix having the constant function equal to one in the (i, j)-position and 0's elsewhere.

For K as a above, we denote by $C(K, M_2(\mathbb{C}))$ the C^*-algebra of all continuous functions from K to $M_2(\mathbb{C})$. We will see $C(K, M_2(\mathbb{C}))$ as a $C(K)$-module in the natural manner. From now on, u will always stand for the element of $C(K)$ defined by $u(t) := t$ for every $t \in K$.

Proposition 2.3. *Let K be a compact subset of $[1, \infty[$ whose maximum element is greater than 1. Then η_K is a nonself-adjoint idempotent in $C(K, M_2(\mathbb{C}))$ satisfying*

$$\operatorname{sp}\big(C\big(K, M_2(\mathbb{C})\big), \sqrt{\eta_K^* \eta_K}\big) \setminus \{0\} = K,$$

and the mapping $\mathcal{F}$ from $\mathcal{A}(K)$ to $C(K, M_2(\mathbb{C}))$, defined by

$$\mathcal{F}\big((f_{ij})\big) := \sum_{i,j \in \{1,2\}} f_{ij} \eta_{ij}^K,$$

becomes a continuous $*$-homomorphism satisfying $\mathcal{F}(u[21]) = \eta_K$.

Proof. By the first part of Lemma 2.2, for $t \in K$, $\eta(t)$ is an idempotent in $M_2(\mathbb{C})$ satisfying $\|\eta(t)\| = t$, which implies

$$\operatorname{sp}\big(M_2(\mathbb{C}), \sqrt{\eta(t)^* \eta(t)}\big) \setminus \{0\} = \{t\}.$$

It follows that η_K is a nonself-adjoint idempotent of $C(K, M_2(\mathbb{C}))$ satisfying

$$\operatorname{sp}\big(C\big(K, M_2(\mathbb{C})\big), \sqrt{\eta_K^* \eta_K}\big) \setminus \{0\} = K.$$

On the other hand, the mapping

$$\mathcal{F} : \mathcal{A}(K) \to C\big(K, M_2(\mathbb{C})\big)$$

is a $*$-homomorphism if (and only if), for every $t \in K$, the composition of $\mathcal{F}$ with the valuation at t is a $*$-homomorphism from $\mathcal{A}(K)$ to $M_2(\mathbb{C})$. But this last fact follows from the definition of the operations on $\mathcal{A}(K)$, and the second part of Lemma 2.2. Finally, both the continuity of $\mathcal{F}$ (it is in fact contractive) and that $\mathcal{F}(u[21]) = \eta_K$ become obvious facts. □

Now, we invoke one of the main results in [1], namely the following.

Proposition 2.4. *Let A be a C^*-algebra, and let e be a nonself-adjoint idempotent in A. Then $K := \operatorname{sp}(A, \sqrt{e^* e}) \setminus \{0\}$ is a compact subset of $[1, \infty[$ whose maximum element (namely $\|e\|$) is grater than 1, and there exists a unique continuous $*$-homomorphism $F : \mathcal{A}(K) \to A$ such that $F(u[21]) = e$. Moreover, we have:*

1. *The closure in A of the range of F coincides with the C^*-subalgebra of A generated by e.*
2. *F is injective if and only if either 1 does not belong to K or 1 is an accumulation point of K.*
3. *If 1 is an isolated point of K, then $\ker(F)$ consists precisely of those matrices $(f_{ij}) \in \mathcal{A}(K)$ which vanish at every $t \in K \setminus \{1\}$ and satisfy*

$$f_{11}(1) + f_{12}(1) + f_{21}(1) + f_{22}(1) = 0.$$

As an immediate consequence of Propositions 2.3 and 2.4, we obtain the following.

Corollary 2.5. *Let K be a compact subset of $[1,\infty[$ whose maximum element is greater than 1, and let $\mathcal{F}:\mathcal{A}(K)\to C(K,M_2(\mathbb{C}))$ be the $*$-homomorphism given by Proposition 2.3. Then we have:*

1. *The closure in $C(K,M_2(\mathbb{C}))$ of the range of $\mathcal{F}$ coincides with the C^*-subalgebra of $C(K,M_2(\mathbb{C}))$ generated by η_K.*
2. *$\mathcal{F}$ is injective if and only if either 1 does not belong to K or 1 is an accumulation point of K.*
3. *If 1 is an isolated point of K, then $\ker(\mathcal{F})$ consists precisely of those matrices $(f_{ij})\in\mathcal{A}(K)$ which vanish at every $t\in K\setminus\{1\}$ and satisfy*

$$f_{11}(1)+f_{12}(1)+f_{21}(1)+f_{22}(1)=0.$$

Lemma 2.6. *Let X be a complex normed space, let Ω be a Hausdorff compact topological space, and let f be a function from Ω to $\mathbb{C}$ such that there are continuous mappings $\alpha,\beta:\Omega\to X$ satisfying $\beta(t)\neq 0$ and $\alpha(t)=f(t)\beta(t)$ for every $t\in\Omega$. Then f is continuous.*

Proof. Put $M:=\max\{\|\alpha(t)\|:\ t\in\Omega\}$ and $m:=\min\{\|\beta(t)\|:\ t\in\Omega\}$. Then we have $m>0$, and hence $|f(t)|\leqslant m^{-1}M$ for every $t\in\Omega$, so that f is bounded. Let t be in Ω, and let $\{t_\lambda\}$ be a net in Ω converging to t. Take a cluster point z of the net $\{f(t_\lambda)\}$ in $\mathbb{C}$. Then $(z,\alpha(t))$ is a cluster point of the net $\{(f(t_\lambda),\alpha(t_\lambda))\}$ in $\mathbb{C}\times X$, and therefore we have $\alpha(t)=z\beta(t)$, which implies (since $\beta(t)\neq 0$) $z=f(t)$. In this way we have shown that $f(t)$ is the unique cluster point of $\{f(t_\lambda)\}$ in $\mathbb{C}$. Since $\{f(t_\lambda)\}$ is bounded, we deduce that $\{f(t_\lambda)\}$ converges to $f(t)$. □

Lemma 2.7. *Let K be a compact subset of $]1,\infty[$. Then the $*$-homomorphism $\mathcal{F}:\mathcal{A}(K)\to C(K,M_2(\mathbb{C}))$, given by Proposition 2.3, is surjective. As a consequence, $C(K,M_2(\mathbb{C}))$ is generated by η_K as a C^*-algebra.*

Proof. Let us fix $t\in K$. By Lemma 2.2, the linear hull of

$$\big\{\eta_{ij}(t):\ i,j\in\{1,2\}\big\}$$

is a $*$-invariant subalgebra of $M_2(\mathbb{C})$. Moreover, since $t\in\,]1,\infty[$, such a subalgebra is not commutative (indeed, $\eta_{12}(t)$ does not commute with $\eta_{21}(t)$). If follows that such a subalgebra is the whole algebra $M_2(\mathbb{C})$, and, consequently, that $\{\eta_{ij}(t):\ i,j\in\{1,2\}\}$ becomes a basis of $M_2(\mathbb{C})$.

Let α be in $C(K,M_2(\mathbb{C}))$. It follows from the above that, for each $t\in K$, there are complex numbers $f_{11}(t)$, $f_{12}(t)$, $f_{21}(t)$, $f_{22}(t)$ uniquely determined by the condition

$$\alpha(t)=f_{11}(t)\eta_{11}(t)+f_{12}(t)\eta_{12}(t)+f_{21}(t)\eta_{21}(t)+f_{22}(t)\eta_{22}(t). \tag{2.1}$$

Moreover, applying again Lemma 2.2, for every $t\in K$ we have:

$$\begin{aligned}
\eta_{11}(t)\alpha(t)\eta_{11}(t)&=\big(f_{11}(t)+t^{-1}f_{12}(t)+t^{-1}f_{21}(t)+t^{-2}f_{22}(t)\big)\eta_{11}(t),\\
\eta_{12}(t)\alpha(t)\eta_{12}(t)&=\big(t^{-1}f_{11}(t)+t^{-2}f_{12}(t)+f_{21}(t)+t^{-1}f_{22}(t)\big)\eta_{12}(t),\\
\eta_{21}(t)\alpha(t)\eta_{21}(t)&=\big(t^{-1}f_{11}(t)+f_{12}(t)+t^{-2}f_{21}(t)+t^{-1}f_{22}(t)\big)\eta_{21}(t),\\
\eta_{22}(t)\alpha(t)\eta_{22}(t)&=\big(t^{-2}f_{11}(t)+t^{-1}f_{12}(t)+t^{-1}f_{21}(t)+f_{22}(t)\big)\eta_{22}(t).
\end{aligned}$$

Since, for $i, j \in \{1,2\}$, $\eta_{ij}^K \alpha \eta_{ij}^K$ and η_{ij}^K are continuous functions on K, and $\eta_{ij}(t) \neq 0$ for every $t \in K$, it follows from Lemma 2.6 that the mappings

$$\begin{aligned}
&t \to f_{11}(t) + t^{-1} f_{12}(t) + t^{-1} f_{21}(t) + t^{-2} f_{22}(t),\\
&t \to t^{-1} f_{11}(t) + t^{-2} f_{12}(t) + f_{21}(t) + t^{-1} f_{22}(t),\\
&t \to t^{-1} f_{11}(t) + f_{12}(t) + t^{-2} f_{21}(t) + t^{-1} f_{22}(t),\\
&t \to t^{-2} f_{11}(t) + t^{-1} f_{12}(t) + t^{-1} f_{21}(t) + f_{22}(t)
\end{aligned}$$

from K to $\mathbb{C}$ are continuous. Since, for $t \in K$ we have

$$\begin{vmatrix} 1 & t^{-1} & t^{-1} & t^{-2} \\ t^{-1} & t^{-2} & 1 & t^{-1} \\ t^{-1} & 1 & t^{-2} & t^{-1} \\ t^{-2} & t^{-1} & t^{-1} & 1 \end{vmatrix} = t^{-8} \begin{vmatrix} t^2 & t & t & 1 \\ t & 1 & t^2 & t \\ t & t^2 & 1 & t \\ 1 & t & t & t^2 \end{vmatrix} = -t^{-8}\left(t^2-1\right)^4 \neq 0,$$

we deduce that, for all $i, j \in \{1,2\}$, the function $f_{ij} : t \to f_{ij}(t)$ from K to $\mathbb{C}$ is continuous. Therefore, we can consider the element (f_{ij}) of $\mathcal{A}(K)$, which, in view of (2.1), satisfies $\mathcal{F}((f_{ij})) = \alpha$. Since α is arbitrary in $C(K, M_2(\mathbb{C}))$, the surjectivity of $\mathcal{F}$ is proved. Now, it follows from assertion (1) in Corollary 2.5 that $C(K, M_2(\mathbb{C}))$ is generated by η_K as a C^*-algebra. $\square$

Now we are ready to prove the main result in this section.

Theorem 2.8. *Let A be a C^*-algebra, and let e be a nonself-adjoint idempotent in A. Put $K := \mathrm{sp}(A, \sqrt{e^*e}) \setminus \{0\}$ (which, in view of Proposition 2.4, is a compact subset of $[1, \infty[$ whose maximum element is greater than 1), and assume that 1 does not belong to K. Then the C^*-subalgebra of A generated by e is $*$-isomorphic to $C(K, M_2(\mathbb{C}))$. More precisely, we have*

(1) *There exists a unique $*$-homomorphism $\Phi : C(K, M_2(\mathbb{C})) \to A$ such that $\Phi(\eta_K) = e$.*
(2) *Such a $*$-homomorphism is isometric, and its range coincides with the C^*-subalgebra of A generated by e.*

Proof. Let $\mathcal{F} : \mathcal{A}(K) \to C(K, M_2(\mathbb{C}))$ and $F : \mathcal{A}(K) \to A$ be the $*$-homomorphisms given by Propositions 2.3 and 2.4, respectively. By assertion (2) in Corollary 2.5 (respectively, Proposition 2.3) $\mathcal{F}$ (respectively, F) is injective. On the other hand, by the first conclusion in Lemma 2.7, $\mathcal{F}$ is surjective. It follows that $\Phi := F \circ \mathcal{F}^{-1}$ is an injective $*$-homomorphism from $C(K, M_2(\mathbb{C}))$ to A satisfying $\Phi(\eta_K) = e$. As any injective $*$-homomorphism between C^*-algebras, Φ is isometric, and hence has closed range. Now, that Φ is the unique $*$-homomorphism from $C(K, M_2(\mathbb{C}))$ to A satisfying $\Phi(\eta_K) = e$, as well as that the range of Φ coincides with the C^*-subalgebra of A generated by e, follows from the fact (given also by Lemma 2.7) that $C(K, M_2(\mathbb{C}))$ is generated by η_K as a C^*-algebra. $\square$

3. The case of C^*-algebras: the second theorem

Let A be an associative complex algebra. The quasi-product $a \circ b$ of two elements a, b of A is defined by $a \circ b := ab - a - b$. An element $a \in A$ is said to be quasi-invertible in A if there exists $b \in A$ satisfying $a \circ b = b \circ a = 0$. It is well known and easy to see that the element $a \in A$

is quasi-invertible in A if and only if $1-a$ is invertible in A_1, if and only if there exists a unique element $b \in A$ satisfying $a \circ b = 0$.

Lemma 3.1. *Let K be a compact subset of $[1,\infty[$ whose maximum element is greater than 1, and let $\mathcal{F}: \mathcal{A}(K) \to C(K, M_2(\mathbb{C}))$ be the $*$-homomorphism given by Proposition 2.3. Then an element $x \in \mathcal{A}(K)$ is quasi-invertible in $\mathcal{A}(K)$ if and only if $\mathcal{F}(x)$ is quasi-invertible in $C(K, M_2(\mathbb{C}))$.*

Proof. Let $x = (f_{ij})$ be in $\mathcal{A}(K)$. We claim that x is quasi-invertible in $\mathcal{A}(K)$ if and only if $\lambda_x(t) \neq 0$ for every $t \in K$, where

$$\lambda_x(t) := \frac{t^2-1}{t^2}\big(f_{11}(t)f_{22}(t) - f_{12}(t)f_{21}(t)\big) - \frac{1}{t}\big(f_{12}(t) + f_{21}(t)\big) - f_{11}(t) - f_{22}(t) + 1.$$

Assume that x is quasi-invertible in $\mathcal{A}(K)$. Let us fix $t \in K$, and identify complex-valued continuous functions on $\{t\}$ with complex numbers. Then, since the restriction mapping $\mathcal{A}(K) \to \mathcal{A}(\{t\})$ is a homomorphism, $(f_{ij}(t))$ is a quasi-invertible element of $\mathcal{A}(\{t\})$, and hence there are complex numbers $g_{11}(t), g_{12}(t), g_{21}(t), g_{22}(t)$ uniquely determined by the condition $(f_{ij}(t)) \circ (g_{ij}(t)) = 0$. This means that the linear system in the indeterminates $x_{11}, x_{12}, x_{21}, x_{22} \in \mathbb{C}$

$$\begin{cases} \big(f_{11}(t) + t^{-1}f_{12}(t) - 1\big)x_{11} + \big(f_{12}(t) + t^{-1}f_{11}(t)\big)x_{21} = f_{11}(t), \\ \big(f_{11}(t) + t^{-1}f_{12}(t) - 1\big)x_{12} + \big(f_{12}(t) + t^{-1}f_{11}(t)\big)x_{22} = f_{12}(t), \\ \big(f_{21}(t) + t^{-1}f_{22}(t)\big)x_{11} + \big(f_{22}(t) + t^{-1}f_{21}(t) - 1\big)x_{21} = f_{21}(t), \\ \big(f_{21}(t) + t^{-1}f_{22}(t)\big)x_{12} + \big(f_{22}(t) + t^{-1}f_{21}(t) - 1\big)x_{22} = f_{22}(t) \end{cases} \tag{3.1}$$

has a unique solution (namely $x_{ij} = g_{ij}(t)$), and hence that the principal determinant of the system (by the way, equal to $\lambda_x(t)^2$) is non-zero. Conversely, assume that $\lambda_x(t) \neq 0$ for every $t \in K$. Then, for each $t \in K$, the system (3.1) has a unique solution $x_{ij} = g_{ij}(t)$, and, since the function $t \to \lambda_x(t)$ from K to $\mathbb{C}$ is continuous, the functions $g_{ij}: t \to g_{ij}(t)$ from K to $\mathbb{C}$ are continuous. Then we easily realize that $y := (g_{ij}) \in \mathcal{A}(K)$ is the unique element of $\mathcal{A}(K)$ satisfying $x \circ y = 0$, which implies that x is quasi-invertible in $\mathcal{A}(K)$. Now, the claim is proved.

On the other hand, $\mathcal{F}(x)$ is quasi-invertible in $C(K, M_2(\mathbb{C}))$ if and only if $1 - \mathcal{F}(x)$ is invertible in $C(K, M_2(\mathbb{C}))$, if and only if $1 - \mathcal{F}(x)(t)$ is invertible in $M_2(\mathbb{C})$ for every $t \in K$, if and only if $\det(1 - \mathcal{F}(x)(t)) \neq 0$ for every $t \in K$, where $\det(\cdot)$ means determinant. But, for $t \in K$, a straightforward but tedious computation shows that $\det(1 - \mathcal{F}(x)(t)) = \lambda_x(t)$. Therefore, $\mathcal{F}(x)$ is quasi-invertible in $C(K, M_2(\mathbb{C}))$ if and only if $\lambda_x(t) \neq 0$. By invoking the claim proved in the preceding paragraph, the result follows. □

Let K be a compact subset of $[1,\infty[$ with $1 \in K$, and let p be a self-adjoint idempotent in $M_2(\mathbb{C})$, different from 0 and 1. Then $\mathbb{C}p$ is a self-adjoint subalgebra of $M_2(\mathbb{C})$, and hence

$$C_p\big(K, M_2(\mathbb{C})\big) := \big\{\alpha \in C\big(K, M_2(\mathbb{C})\big) \colon \alpha(1) \in \mathbb{C}p\big\}$$

is a proper C^*-subalgebra of $C(K, M_2(\mathbb{C}))$. We note that, in the construction of the C^*-algebra $C_p(K, M_2(\mathbb{C}))$, the choice of the idempotent p is structurally irrelevant. Indeed, if, for $i \in \{1,2\}$, p_i is a self-adjoint idempotent in $M_2(\mathbb{C})$, different from 0 and 1, then there exists a norm-one element χ_i in the Hilbert space $\mathbb{C}^2$ such that p_i is the operator $\chi \to (\chi|\chi_i)\chi_i$ on $\mathbb{C}^2$, and hence, since there exists a unitary element $v \in M_2(\mathbb{C})$ with $v\chi_1 = \chi_2$ (by transitivity of Hilbert spaces), the mapping $\alpha \to v\alpha v^*$ becomes a $*$-automorphism of $C(K, M_2(\mathbb{C}))$ sending $C_{p_1}(K, M_2(\mathbb{C}))$ onto $C_{p_2}(K, M_2(\mathbb{C}))$. We also note that, if we take $p = \eta(1)$, then $C_p(K, M_2(\mathbb{C}))$ contains η_K.

Lemma 3.2. *Let K be a compact subset of* $[1, \infty[$ *with* $1 \in K$, *and whose maximum element is greater than* 1, *and let $\mathcal{F}: \mathcal{A}(K) \to C(K, M_2(\mathbb{C}))$ be the $*$-homomorphism given by Proposition* 2.3. *Then the closure in $C(K, M_2(\mathbb{C}))$ of the range of $\mathcal{F}$ coincides with $C_{\eta(1)}(K, M_2(\mathbb{C}))$. As a consequence, $C_{\eta(1)}(K, M_2(\mathbb{C}))$ is generated by η_K as a C^*-algebra.*

Proof. For $x = (f_{ij})$ in $\mathcal{A}(K)$, we have

$$\mathcal{F}(x)(1) = \big(f_{11}(1) + f_{12}(1) + f_{21}(1) + f_{22}(1)\big)\eta(1) \in \mathbb{C}\eta(1),$$

and therefore $\mathcal{F}(x)$ lies in $C_{\eta(1)}(K, M_2(\mathbb{C}))$. This shows that the range of $\mathcal{F}$ (say B) is contained in $C_{\eta(1)}(K, M_2(\mathbb{C}))$.

To continue our argument, it is useful to identify $C(K, M_2(\mathbb{C}))$ with $C(K) \otimes M_2(\mathbb{C})$ in the natural manner. Then we have:

$$2 \otimes \eta(1) = 1 \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = (1+u)^{-1} u\big(\eta_{11}^K + \eta_{12}^K + \eta_{21}^K + \eta_{22}^K\big) \in B, \tag{3.2}$$

$$\sqrt{u^2 - 1} \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = u\big(\eta_{21}^K - \eta_{12}^K\big) \in B, \tag{3.3}$$

$$\sqrt{u^2 - 1} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = u\big(\eta_{22}^K - \eta_{11}^K\big) \in B, \tag{3.4}$$

$$\big(u^2 - 1\big) \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = u^2\big(\eta_{22}^K + \eta_{11}^K\big) - u\big(\eta_{12}^K + \eta_{21}^K\big) \in B. \tag{3.5}$$

Now, keep in mind that B is a $C(K)$-submodule of $C(K, M_2(\mathbb{C}))$, and denote by $C_1(K)$ the closed ideal of $C(K)$ consisting of those complex-valued continuous functions on K vanishing at 1. It follows from (3.2) that

$$C(K) \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \subseteq B,$$

and, by invoking the Stone–Weierstrass theorem, it follows from (3.3)–(3.5) that

$$C_1(K) \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \subseteq \overline{B}, \qquad C_1(K) \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \subseteq \overline{B}, \qquad C_1(K) \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \subseteq \overline{B}.$$

Since

$$\left\{ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

is a basis of $M_2(\mathbb{C})$, we deduce that $C_1(K) \otimes M_2(\mathbb{C}) \subseteq \overline{B}$. Since

$$C_{\eta(1)}\big(K, M_2(\mathbb{C})\big) = \big[\mathbb{C} \otimes \eta(1)\big] \oplus \big[C_1(K) \otimes M_2(\mathbb{C})\big],$$

and $\mathbb{C} \otimes \eta(1) \subseteq B$ (by (3.2)), we obtain $C_{\eta(1)}(K, M_2(\mathbb{C})) \subseteq \overline{B}$. By invoking the first paragraph in the present proof, we have $C_{\eta(1)}(K, M_2(\mathbb{C})) = \overline{B}$.

Now, it follows from assertion (1) in Corollary 2.5 that $C_{\eta(1)}(K, M_2(\mathbb{C}))$ is generated by η_K as a C^*-algebra. $\square$

Now we are ready to prove the main result in this section.

Theorem 3.3. *Let A be a C^*-algebra, and let e be a nonself-adjoint idempotent in A. Put $K := \mathrm{sp}(A, \sqrt{e^*e}) \setminus \{0\}$ (which, in view of Proposition 2.4, is a compact subset of $[1, \infty[$ whose maximum element is greater than 1), and assume that 1 belongs to K. Then the C^*-subalgebra of A generated by e is $*$-isomorphic to $C_p(K, M_2(\mathbb{C}))$ for any self-adjoint idempotent $p \in M_2(\mathbb{C})$ different from 0 and 1. More precisely, we have*:

(1) *There exists a unique $*$-homomorphism $\Phi : C_{\eta(1)}(K, M_2(\mathbb{C})) \to A$ such that $\Phi(\eta_K) = e$.*
(2) *Such a $*$-homomorphism is isometric, and its range coincides with the C^*-subalgebra of A generated by e.*

Proof. For every element c in a complex Banach algebra C, put

$$r(C, c) := \max\bigl\{|\lambda|:\ \lambda \in \mathrm{sp}(C, c)\bigr\},$$

and note that, since

$$\{0\} \cup \mathrm{sp}(C, c) = \{0\} \cup \bigl\{\lambda \in \mathbb{C} \setminus \{0\}:\ \lambda^{-1}c \notin q\text{-inv}(C)\bigr\}$$

(where q-inv(C) stands for the set of all quasi-invertible elements of C), we have

$$r(C, c) = \max\bigl[\{0\} \cup \bigl\{|\lambda|:\ \lambda \in \mathbb{C} \setminus \{0\},\ \lambda^{-1}c \notin q\text{-inv}(C)\bigr\}\bigr]. \tag{3.6}$$

Now, let $\mathcal{F} : \mathcal{A}(K) \to C(K, M_2(\mathbb{C}))$ and $F : \mathcal{A}(K) \to A$ be the $*$-homomorphisms given by Propositions 2.3 and 2.4, respectively. Then, for $x \in \mathcal{A}(K)$ we have

$$\bigl\|F(x)\bigr\|^2 = r\bigl(A, F\bigl(x^*x\bigr)\bigr) \leqslant r\bigl(\mathcal{A}(K), x^*x\bigr),$$

and, by keeping in mind Lemma 3.1 and (3.6), we have also

$$r\bigl(\mathcal{A}(K), x^*x\bigr) = r\bigl(C\bigl(K, M_2(\mathbb{C})\bigr), \mathcal{F}\bigl(x^*x\bigr)\bigr) = \bigl\|\mathcal{F}(x)\bigr\|^2,$$

so that the inequality $\|F(x)\| \leqslant \|\mathcal{F}(x)\|$ holds. Therefore $\mathcal{F}(x) \to F(x)$ ($x \in \mathcal{A}(K)$) becomes a (well-defined) continuous $*$-homomorphism from the range of $\mathcal{F}$ to A. Then, by the first conclusion in Lemma 3.2, such a $*$-homomorphism extends by continuity to a $*$-homomorphism

$$\Phi : C_{\eta(1)}\bigl(K, M_2(\mathbb{C})\bigr) \to A$$

satisfying $\Phi \circ \mathcal{F} = F$, and hence $\Phi(\eta_K) = e$. Now, that Φ is the unique $*$-homomorphism from $C_{\eta(1)}(K, M_2(\mathbb{C}))$ to A satisfying $\Phi(\eta_K) = e$, as well as that the range of Φ coincides with the C^*-subalgebra of A generated by e, follows from the fact (given also by Lemma 3.2) that $C_{\eta(1)}(K, M_2(\mathbb{C}))$ is generated by η_K as a C^*-algebra.

To conclude the proof, it is enough to show that Φ is injective. Let α be in $\ker(\Phi)$. Then, by Lemma 3.2, there exists a sequence $x_n = (f_{ij}^n)$ in $\mathcal{A}(K)$ such that $\mathcal{F}(x_n) \to \alpha$. For $n \in \mathbb{N}$ and $i, j \in \{1, 2\}$, define $g_{ij}^n \in C(K)$ by

$$\begin{aligned}
g_{11}^n &:= f_{11}^n + u^{-1} f_{12}^n + u^{-1} f_{21}^n + u^{-2} f_{22}^n, \\
g_{12}^n &:= u^{-1} f_{11}^n + u^{-2} f_{12}^n + f_{21}^n + u^{-1} f_{22}^n, \\
g_{21}^n &:= u^{-1} f_{11}^n + f_{12}^n + u^{-2} f_{21}^n + u^{-1} f_{22}^n, \\
g_{22}^n &:= u^{-2} f_{11}^n + u^{-1} f_{12}^n + u^{-1} f_{21}^n + f_{22}^n.
\end{aligned}$$

Then we have $[ij]x_n[ij] = g_{ij}^n[ij]$. Now, since the restriction of F to $C(K)[ij]$ is an isometry (by the proof of Theorem 2.6 of [1]), we deduce

$$\begin{aligned}
\|g_{ij}^n\| &= \|g_{ij}^n[ij]\| = \|F(g_{ij}^n[ij])\| = \|F([ij]x_n[ij])\| = \|F([ij])F(x_n)F([ij])\| \\
&= \|F([ij])\Phi(\mathcal{F}(x_n))F([ij])\| \to \|F([ij])\Phi(\alpha)F([ij])\| = 0.
\end{aligned}$$

As a consequence, $g_{ij}^n(t) \to 0$ for every $t \in K$. Since for $t \in K \setminus \{1\}$, we have

$$\begin{vmatrix}
1 & t^{-1} & t^{-1} & t^{-2} \\
t^{-1} & t^{-2} & 1 & t^{-1} \\
t^{-1} & 1 & t^{-2} & t^{-1} \\
t^{-2} & t^{-1} & t^{-1} & 1
\end{vmatrix} = -t^{-8}(t^2 - 1)^4 \neq 0,$$

it follows from the definition of g_{ij}^n that $f_{ij}^n(t) \to 0$ for every $t \in K \setminus \{1\}$. Now, since for $t \in K \setminus \{1\}$ we have $\mathcal{F}(x_n)(t) \to \alpha(t)$ and

$$\mathcal{F}(x_n)(t) = \sum_{i,j \in \{1,2\}} f_{ij}^n(t)\eta_{ij}(t) \to 0,$$

for such a t we obtain $\alpha(t) = 0$. Therefore, if 1 is an accumulation point of K, then $\alpha = 0$, as desired. Assume that 1 is an isolated point of K. Then the function $\chi : K \to \mathbb{C}$, defined by $\chi(1) := 1$ and $\chi(t) := 0$ for $t \in K \setminus \{1\}$, is continuous, and, since there exists $\lambda \in \mathbb{C}$ such that $\alpha(1) = \lambda\eta(1)$, for such a λ we have $\alpha = \lambda\chi\eta_{21}^K = \mathcal{F}(\lambda\chi[21])$. Therefore

$$0 = \Phi(\alpha) = \Phi(\mathcal{F}(\lambda\chi[21])) = F(\lambda\chi[21]),$$

which, in view of assertion (3) in Proposition 2.4, implies $\lambda = 0$, and hence $\alpha = 0$. □

4. The case of C^*-algebras: some consequences

In this section, we combine Theorems 2.8 and 3.3 to derive some attractive consequences. We begin with an easy corollary to Theorem 3.3.

Corollary 4.1. *Let A be a C^*-algebra generated by a nonself-adjoint idempotent e, and put $K := \mathrm{sp}(A, \sqrt{e^*e}) \setminus \{0\}$. If 1 is an isolated point of the compact set K, then A is $*$-isomorphic to the C^*-algebra*

$$\mathbb{C} \times C\big(K \setminus \{1\}, M_2(\mathbb{C})\big).$$

Proof. If 1 belongs to K, then, for α in $C_{\eta(1)}(K, M_2(\mathbb{C}))$, there exists a unique complex number $\lambda(\alpha)$ such that $\alpha(1) = \lambda(\alpha)\eta(1)$, and the mapping

$$\alpha \to \big(\lambda(\alpha), \alpha_{|K\setminus\{1\}}\big)$$

becomes an injective $*$-homomorphism from $C_{\eta(1)}(K, M_2(\mathbb{C}))$ to

$$\mathbb{C} \times C^{\mathrm{b}}\big(K \setminus \{1\}, M_2(\mathbb{C})\big),$$

where $C^{\mathrm{b}}(K \setminus \{1\}, M_2(\mathbb{C}))$ stands for the C^*-algebra of all bounded continuous function from $K \setminus \{1\}$ to $M_2(\mathbb{C})$. Moreover, if 1 is in fact an isolated point of K, then we have that

$$C^{\mathrm{b}}\big(K \setminus \{1\}, M_2(\mathbb{C})\big) = C\big(K \setminus \{1\}, M_2(\mathbb{C})\big),$$

and that the above $*$-homomorphism is surjective. Finally, apply Theorem 3.3. □

Corollary 4.2. *Let A be a C^*-algebra generated by a nonself-adjoint idempotent e, and put $K := \mathrm{sp}(A, \sqrt{e^*e}) \setminus \{0\}$. Then A has a unit if and only if either 1 does not belong to K or 1 is an isolated point of K.*

Proof. In view of Theorems 2.8 and 3.3, and Corollary 4.1, it is enough to show that, if 1 is an accumulation point of K, then $C_{\eta(1)}(K, M_2(\mathbb{C}))$ has not a unit. Assume that 1 belongs to K. We claim that, given $t_0 \in K \setminus \{1\}$, the valuation at t_0 (as a mapping from $C_{\eta(1)}(K, M_2(\mathbb{C}))$ to $M_2(\mathbb{C})$) is surjective. Indeed, if $a = (\lambda_{ij})$ is an arbitrary element of $M_2(\mathbb{C})$, then, for $i, j \in \{1, 2\}$, there exists $f_{ij} \in C(K)$ such that $f_{ij}(1) = 0$ and $f_{ij}(t_0) = \lambda_{ij}$, and hence the element α of $C(K, M_2(\mathbb{C}))$, defined by $\alpha(t) := (f_{ij}(t))$ for every $t \in K$, lies in $C_{\eta(1)}(K, M_2(\mathbb{C}))$ and satisfies $\alpha(t_0) = a$. Assume in addition that $C_{\eta(1)}(K, M_2(\mathbb{C}))$ has a unit $\mathbf{1}$. Then, by the claim just proved, for every $t \in K \setminus \{1\}$, $\mathbf{1}(t)$ must be equal to the unit of $M_2(\mathbb{C})$. Now, if 1 is in fact an accumulation point of K, then $\mathbf{1}(1)$ is the unit of $M_2(\mathbb{C})$, which is not possible because $\mathbf{1}(1)$ is a complex multiple of $\eta(1)$. □

Corollary 4.3. *Let A be a C^*-algebra. Then A has a nonself-adjoint idempotent* (*if and*) *only if it contains* (*as a C^*-subalgebra*) *a copy of either $M_2(\mathbb{C})$ or $C_p([1, 2], M_2(\mathbb{C}))$ for any self-adjoint idempotent $p \in M_2(\mathbb{C})$ different from 0 and 1.*

Proof. Assume that A has a nonself-adjoint idempotent e, and put $K := \mathrm{sp}(A, \sqrt{e^*e}) \setminus \{0\}$. We may suppose that A is generated by e. If 1 does not belong to K, then, by Theorem 2.8, A contains a copy of $M_2(\mathbb{C})$. Assume that 1 belongs to K, and that K is disconnected. Take a clopen proper subset U of K with $1 \in U$. Then, arguing as in the proof of Corollary 4.1, we realize that A is $*$-isomorphic to $C_p(U, M_2(\mathbb{C})) \times C(K \setminus U, M_2(\mathbb{C}))$, for some self-adjoint idempotent $p \in M_2(\mathbb{C})$ different from 0 and 1, and hence it contains a copy of $M_2(\mathbb{C})$. Finally, assume that 1 belongs to K, and that K is connected. Then we have $K = [1, \|e\|]$, and therefore, by Theorem 3.3, A is isomorphic to $C_p([1, \|e\|], M_2(\mathbb{C}))$, for some p as above. But, taking a homeomorphism ϕ from $[1, \|e\|]$ onto $[1, 2]$ with $\phi(1) = 1$, ϕ induces a $*$-isomorphism from $C([1, \|e\|], M_2(\mathbb{C}))$ onto $C([1, 2], M_2(\mathbb{C}))$ sending $C_p([1, \|e\|], M_2(\mathbb{C}))$ onto $C_p([1, 2], M_2(\mathbb{C}))$. $\square$

We remark that $C_p([1, 2], M_2(\mathbb{C}))$ does not contain any copy of $M_2(\mathbb{C})$. To realize this, we argue by contradiction, and hence we assume that $C_p([1, 2], M_2(\mathbb{C}))$ contains a copy (say B) of $M_2(\mathbb{C})$. For $\alpha \in C_p([1, 2], M_2(\mathbb{C}))$, let $\lambda(\alpha)$ stand for the unique complex number satisfying $\alpha(1) = \lambda(\alpha)p$. Then, since $\lambda : C_p([1, 2], M_2(\mathbb{C})) \to \mathbb{C}$ is a homomorphism, by the simplicity of B we have $\lambda(B) = 0$. Therefore B is contained in the ideal (say M) of $C([1, 2], M_2(\mathbb{C}))$ consisting of those continuous functions from $[1, 2]$ to $M_2(\mathbb{C})$ vanishing at 1. Now, since (clearly) M has no non-zero idempotent, and the unit of B is a non-zero idempotent of M, the contradiction is clear.

Remark 4.4. In relation to Corollary 4.3 above, it is worth mentioning that a C^*-algebra contains a nonself-adjoint idempotent if and only if it contains a non-central self-adjoint idempotent [1]. By the way, the "only if" part of the result in [1] just quoted follows easily from Corollary 4.3, whereas the "if part" is a consequence of Proposition 4.5 immediately below.

In relation to Proposition 4.5 immediately below, we note that nonself-adjoint idempotents in a C^*-algebra are non-central.

Proposition 4.5. *Let A be a C^*-algebra containing a non-central idempotent e. Then there exists a continuous mapping $r \to e_r$ from $[1, \infty[$ to the set of idempotents of A satisfying $e_{\|e\|} = e$ and $\|e_r\| = r$ for every $r \in [1, \infty[$.*

Proof. First assume that e is not self-adjoint. Then, by Theorems 2.8 and 3.3, we may assume that A is of the form $C(K, M_2(\mathbb{C}))$ or $C_{\eta(1)}(K, M_2(\mathbb{C}))$, where, in the first case, K is a compact subset of $]1, \infty[$ and, in the second case, K is a compact subset of $[1, \infty[$ whose maximum element is greater than 1 and such that $1 \in K$. In any case, put $\rho := \max K > 1$. Let r be in $[1, \infty[$, and let e_r denote the element of $C(K, M_2(\mathbb{C}))$ defined by

$$e_r(t) := \eta\left(1 + \frac{(r-1)(t-1)}{\rho - 1}\right)$$

for every $t \in K$. Noticing that, in the case that 1 belongs to K, e_r lies in $C_{\eta(1)}(K, M_2(\mathbb{C}))$, it turns out that, in any case e_r is an element of A. Moreover, keeping in mind Lemma 2.2, we easily realize that e_r is an idempotent, and that $\|e_r\| = r$. On the other hand, since $\|\eta_K\| = \rho$, we have $e_{\|\eta_K\|} = \eta_K$. Now it only remains to show that the mapping $r \to e_r$ is continuous. Fix $r \in [1, \infty[$ and $\varepsilon > 0$, and take $\delta > 0$ such that $\|\eta(s) - \eta(r)\| < \varepsilon$ whenever s is in $[1, \infty[$ with $|s - r| < \delta$. Then, for $s \in [1, \infty[$ with $|s - r| < \delta$, we have for every $t \in K$,

$$\left|\left[1 + \frac{(s-1)(t-1)}{\rho - 1}\right] - \left[1 + \frac{(r-1)(t-1)}{\rho - 1}\right]\right| = \frac{|s-r|(t-1)}{\rho - 1} \leqslant |s - r| < \delta,$$

so $\|e_s(t) - e_r(t)\| < \varepsilon$ for every $t \in K$, and so $\|e_s - e_r\| \leqslant \varepsilon$.

Now assume that e is self-adjoint. Since e is non-central, we may choose a self-adjoint element $a \in A$ with $ea - ae \neq 0$. Then the mapping $D : A \to A$ defined by $D(b) := ba - ab$ for every

$b \in A$ becomes a continuous derivation satisfying $D(e) \neq 0$ and $D(b^*) = -D(b)^*$ for every $b \in A$. Therefore, for $s \in \mathbb{R}$, $\exp(sD)$ is a continuous automorphism of A satisfying

$$\bigl[\exp(sD)(b)\bigr]^* = \exp(-sD)\bigl(b^*\bigr)$$

for every $b \in A$, and consequently

$$g(s) := \exp(sD)(e)$$

is a non-zero idempotent in A, and we have

$$g(s)^* = g(-s). \tag{4.1}$$

Now, consider the continuous mapping $f : \mathbb{R} \to [1, \infty[$ defined by

$$f(s) := \bigl\|g(s)\bigr\|.$$

By (4.1), we have

$$f(-s) = f(s) \tag{4.2}$$

for every $s \in \mathbb{R}$. Let r, s be in $\mathbb{R}$. Then, keeping in mind (4.1), (4.2), and that $\exp(\frac{s-r}{2}D)$ is an automorphism of A, we have

$$\begin{aligned}
f\left(\frac{r+s}{2}\right)^2 &= \left\|g\left(\frac{r+s}{2}\right)\right\|^2 = \left\|g\left(\frac{r+s}{2}\right)^* g\left(\frac{r+s}{2}\right)\right\| = r\left(A, g\left(\frac{r+s}{2}\right)^* g\left(\frac{r+s}{2}\right)\right)\\
&= r\left(A, g\left(-\frac{r+s}{2}\right) g\left(\frac{r+s}{2}\right)\right)\\
&= r\left[A, \exp\left(\frac{s-r}{2}D\right)\left(g\left(-\frac{r+s}{2}\right) g\left(\frac{r+s}{2}\right)\right)\right]\\
&= r\left[A, \left[\exp\left(\frac{s-r}{2}D\right)\left(g\left(-\frac{r+s}{2}\right)\right)\right]\left[\exp\left(\frac{s-r}{2}D\right) g\left(\left(\frac{r+s}{2}\right)\right)\right]\right]\\
&= r\bigl(A, g(-r)g(s)\bigr) \leqslant \bigl\|g(-r)\bigr\|\bigl\|g(s)\bigr\| = f(-r)f(s) = f(r)f(s),
\end{aligned}$$

and therefore

$$f\left(\frac{r+s}{2}\right) \leqslant \sqrt{f(r)f(s)} \leqslant \frac{f(r)+f(s)}{2}.$$

In this way we have shown that f is convex. Assume that $f(r) = 1$ for some $r \in]0, \infty[$. Then, by (4.2) and the convexity of f, we have $f(s) = 1$ for every $s \in [-r, r]$. Therefore, for $s \in [-r, r]$, the idempotent $g(s)$ has norm equal to 1, so it is self-adjoint, and so, by (4.1) the equality $g(s) = g(-s)$ holds. Since g is differentiable at 0 with $g'(0) = D(e)$, the above implies $D(e) = 0$, which is a contradiction. Thus, $f(r) > 1$ for every $r \in]0, \infty[$. Now, let $0 < r < s$. Noticing that $f(0) = 1$ and that then, by the convexity of f, the mapping $t \to \frac{f(t)-1}{t}$ is increasing, we have

$$0 < f(r) - 1 < \frac{s}{r}\big(f(r) - 1\big) \leqslant f(s) - 1.$$

In this way, we have shown that $f_{|[0,\infty[}$ is strictly increasing and not bounded. As a consequence, the range of $f_{|[0,\infty[}$ is $[1, \infty[$, and the inverse mapping $h : [1, \infty[\to [0, \infty[$ is continuous. Now, for $r \in [1, \infty[$, let e_r be the idempotent of A defined by $e_r := g(h(r))$. Then, clearly, the mapping $r \to e_r$ is continuous, and we have $e_1 = e$. Moreover, by the definition of g and h, we have also that $\|e_r\| = f(h(r)) = r$ for every $r \in [1, \infty[$. □

We recall that partial isometries in a C^*-algebra A are defined as those elements $a \in A$ satisfying $aa^*a = a$.

Lemma 4.6. *Let A be a C^*-algebra, and let a be a partial isometry in A such that both a^*a and aa^* lie in the centre of A. Then a is normal.*

Proof. For $x, y \in A$, put $[x, y] := xy - yx$. Since a^*a and aa^* lie in the centre of A, we have $[a^*a, a] = 0$ and $[aa^*, a] = 0$, which reads as $a^*a^2 = a$ and $a^2a^* = a$, respectively. The two last equalities, together with the one $aa^*a = a$, and those obtained by taking adjoints, imply $[[a, a^*], a] = 0$. By [2, Proposition 18.13], we have $r(A, [a, a^*]) = 0$, and hence, since $[a, a^*]$ is self-adjoint, we actually have $[a, a^*] = 0$. □

Let A denote the C^*-algebra of all bounded linear operators on an infinite-dimensional complex Hilbert space H, let $b : H \to H$ be any non-surjective linear isometry, and put $a := b$ (respectively $a := b^*$). Then a is a non-normal partial isometry in A such that a^*a (respectively aa^*) lies in the centre of A.

Corollary 4.7. *Let A be a C^*-algebra. Then the following assertions are equivalent*:

(1) *A contains a non-central self-adjoint idempotent.*
(2) *There exists a non-normal partial isometry $a \in A$ such that a belongs to a^2Aa^2.*
(3) *A contains a non-normal partial isometry.*

Proof. (1) ⇒ (2). By the assumption (1), Remark 4.4, and Theorems 2.8 and 3.3, we may assume that A is of the form $C(K, M_2(\mathbb{C}))$ or $C_{\eta(1)}(K, M_2(\mathbb{C}))$, where, in the first case, K is a compact subset of $]1, \infty[$ and, in the second case, K is a compact subset of $[1, \infty[$ whose maximum element is greater than 1 and such that $1 \in K$. In any case, by Lemma 2.2, η_{21}^K is a non-normal partial isometry in A, and we have $\eta_{21}^K = (\eta_{21}^K)^2(u^2\eta_{12}^K)(\eta_{21}^K)^2$.

(2) ⇒ (3). This is clear.

(3) ⇒ (1). Let a be the partial isometry whose existence is assumed in (3). Then, keeping in mind that both a^*a and aa^* are self-adjoint idempotents, it follows from Lemma 4.6 that A contains a non-central self-adjoint idempotent. □

Put $A := M_2(\mathbb{C})$ and $a := \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. Then a is a non-normal partial isometry in A, which does not belong to a^2Aa^2.

5. The case of JB^*-algebras: the main results

Over fields of characteristic different from two, Jordan algebras are defined as those (possibly non-associative) commutative algebras satisfying the identity $(x \cdot y) \cdot x^2 = x \cdot (y \cdot x^2)$. For a and b in a Jordan algebra, we put $U_a(b) := 2a \cdot (a \cdot b) - a^2 \cdot b$. Let A be an associative algebra. Then A becomes a Jordan algebra under the Jordan product defined by

$$a \cdot b := \frac{1}{2}(ab + ba).$$

Moreover, for all $a, b \in A$ we have

$$U_a(b) := 2a \cdot (a \cdot b) - a^2 \cdot b = aba. \tag{5.1}$$

Jordan subalgebras of A are, by definition, those subspaces J of A satisfying $J \cdot J \subseteq J$.

Let K be a compact subset of $[1, \infty[$. Then the linear mapping $\Theta : \mathcal{A}(K) \to \mathcal{A}(K)$, determined by

$$\Theta\big(f[ij]\big) := f[ij] \quad \text{if } i \neq j, \qquad \Theta\big(f[11]\big) := f[22], \qquad \Theta\big(f[22]\big) := f[11]$$

for every $f \in C(K)$, becomes an isometric involutive $*$-antiautomorphism of $\mathcal{A}(K)$. Therefore, the set of fixed elements for Θ is a closed $*$-invariant Jordan subalgebra of $\mathcal{A}(K)$, and hence a Banach–Jordan $*$-algebra. Such a Banach–Jordan $*$-algebra will be denoted by $\mathcal{J}(K)$. Note that elements of $\mathcal{J}(K)$ are precisely those matrices $(f_{ij}) \in \mathcal{A}(K)$ satisfying $f_{11} = f_{22}$, or equivalently, those elements of $\mathcal{A}(K)$ of the form $f([11] + [22]) + g[12] + h[21]$ with $f, g, h \in C(K)$.

We take from [1] the following.

Lemma 5.1. *Let K be a compact subset of $[1, \infty[$. Then $\mathcal{J}(K)$ is generated by $u[21]$ as a Jordan–Banach $*$-algebra.*

JB^*-algebras are defined as those complex Banach–Jordan $*$-algebras J satisfying $\|U_a(a^*)\| = \|a\|^3$ for every $a \in J$. By keeping in mind (5.1), it is easy to realize that C^*-algebras are JB^*-algebras under their Jordan products.

The mapping

$$\theta : \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} \to \begin{pmatrix} \lambda_{22} & \lambda_{12} \\ \lambda_{21} & \lambda_{11} \end{pmatrix}$$

is an involutive $*$-antiautomorphism of $M_2(\mathbb{C})$. Therefore, the set of fixed elements for θ is a $*$-invariant Jordan subalgebra of the C^*-algebra $M_2(\mathbb{C})$, and hence a JB^*-algebra. Such a JB^*-algebra is called the three-dimensional spin factor, and is denoted by $\mathcal{C}_3$.

Let K be a compact subset of $[1, \infty[$. We denote by $C(K, \mathcal{C}_3)$ the JB^*-algebra of all continuous functions from K to $\mathcal{C}_3$. We will identify $C(K, \mathcal{C}_3)$ with the JB^*-subalgebra of $C(K, M_2(\mathbb{C}))$ consisting of those continuous functions from K to $M_2(\mathbb{C})$ whose range is contained in $\mathcal{C}_3$.

Lemma 5.2. *Let K be a compact subset of $[1, \infty[$ whose maximum element is greater than 1, let $\mathcal{F} : \mathcal{A}(K) \to C(K, M_2(\mathbb{C}))$ be the $*$-homomorphism given by Proposition 2.3, and let $\mathcal{G}$ denote the restriction to $\mathcal{J}(K)$ of $\mathcal{F}$. Then $\mathcal{G}$ is a $*$-homomorphism from $\mathcal{J}(K)$ to the JB^*-algebra* *underlying $C(K, M_2(\mathbb{C}))$, and the closure in $C(K, M_2(\mathbb{C}))$ of the range of $\mathcal{G}$ coincides with the JB^*-subalgebra of $C(K, \mathcal{C}_3)$ generated by η_K.*

Proof. Noticing that $\mathcal{G}(u[21]) = \eta_K$, and keeping in mind Lemma 5.1, it is enough to show that the range of $\mathcal{G}$ is contained in $C(K, \mathcal{C}_3)$. But this follows from the fact that η_K actually belongs to $C(K, \mathcal{C}_3)$, and a new application of Lemma 5.1. □

Lemma 5.3. *Let K be a compact subset of $]1, \infty[$. Then $C(K, \mathcal{C}_3)$ is generated by η_K as a JB^*-algebra.*

Proof. Identifying $C(K, M_2(\mathbb{C}))$ with $C(K) \otimes M_2(\mathbb{C})$ in the natural manner, the operator $\widehat{\theta} := 1 \otimes \theta$ becomes an involutive $*$-antiautomorphism of $C(K, M_2(\mathbb{C}))$, whose set of fixed points is precisely $C(K, \mathcal{C}_3)$. Moreover, since $\mathcal{A}(K)$ is generated by $u[21]$ as a Banach $*$-algebra (by [1, Lemma 2.5]), and $\mathcal{F}(\Theta(u[21])) = \widehat{\theta}(\mathcal{F}(u[21]))$, we have $\mathcal{F} \circ \Theta = \widehat{\theta} \circ \mathcal{F}$. On the other hand, by Lemma 2.7, $\mathcal{F} : \mathcal{A}(K) \to C(K, M_2(\mathbb{C}))$ is surjective. Since $\mathcal{J}(K)$ is the set of fixed points for Θ, and $C(K, \mathcal{C}_3)$ is the set of fixed points for $\widehat{\theta}$, and $\mathcal{G}$ is the restriction to $\mathcal{J}(K)$ of $\mathcal{F}$, it follows that $\mathcal{G}$ (as a mapping from $\mathcal{J}(K) \to C(K, \mathcal{C}_3)$) is surjective. Now, apply Lemma 5.2. □

We recall that a JB^*-triple is a complex Banach space X endowed with a continuous triple product $\{\cdot, \cdot, \cdot\} : X \times X \times X \to X$ which is linear and symmetric in the outer variables, and conjugate-linear in the middle variable, and satisfies:

(1) For all x in X, the mapping $y \to \{x, x, y\}$ from X to X is a hermitian operator on X and has non-negative spectrum.
(2) The main identity

$$\bigl\{a, b, \{x, y, z\}\bigr\} = \bigl\{\{a, b, x\}, y, z\bigr\} - \bigl\{x, \{b, a, y, \}, z\bigr\} + \bigl\{x, y, \{a, b, z\}\bigr\}$$

holds for all a, b, x, y, z in X.
(3) $\|\{x, x, x\}\| = \|x\|^3$ for every x in X.

Concerning condition (1) above, we also recall that a bounded linear operator T on a complex Banach space X is said to be hermitian if $\|\exp(irT)\| = 1$ for every r in $\mathbb{R}$. Examples of JB^*-triples are all C^*-algebras under the triple product $\{\cdot, \cdot, \cdot\}$ determined by $\{a, b, a\} := ab^*a$.

Let X be a JB^*-triple, and let x be in X. It is well known that there is a unique couple (K, ϕ), where K is a compact subset of $[0, \infty[$ with $0 \in K$, and ϕ is an isometric triple homomorphism from $C_0(K)$ to X, such that the range of ϕ coincides with the JB^*-subtriple of X generated by x, and $\phi(v) = x$, where v stands for the mapping $t \to t$ from K to $\mathbb{C}$ (see [8, 4.8], [9, 1.15], and [5]). The locally compact subset $K \setminus \{0\}$ of $]0, \infty[$ is called the triple spectrum of x, and will be denoted by $\sigma(x)$. We note that $\sigma(x)$ does not change when we replace X with any JB^*-subtriple of X containing x.

We take from [1] the following.

Lemma 5.4. *Let A be a C^*-algebra, and let a be in A such that $0 \in \mathrm{sp}(a^*a)$. Then we have $\sigma(a) = \mathrm{sp}(A, \sqrt{a^*a}) \setminus \{0\}$.*

As in the particular case of C^*-algebras, already commented, JB^*-algebras are JB^*-triples under the triple product $\{\cdot,\cdot,\cdot\}$ determined by $\{a,b,a\} := U_a(b^*)$ (see [3,13]). For later reference, we remark that, if a JB^*-algebra J has a unit $\mathbf{1}$, then for $a,b \in J$ we have

$$a \cdot b = \{a, \mathbf{1}, b\} \quad \text{and} \quad a^* = \{\mathbf{1}, a, \mathbf{1}\}. \tag{5.2}$$

Theorem 5.5. *Let J be a JB^*-algebra, and let e be a nonself-adjoint idempotent in J. Put $K := \sigma(e)$, and assume that* 1 *does not belong to K. Then K is a compact subset of* $]1,\infty[$, *and the JB^*-subalgebra of J generated by e is $*$-isomorphic to $C(K,\mathcal{C}_3)$. More precisely, we have*:

(1) *There exists a unique $*$-homomorphism $\Psi : C(K,\mathcal{C}_3) \to J$ such that $\Psi(\eta_K) = e$.*
(2) *Such a $*$-homomorphism is isometric, and its range coincides with the JB^*-subalgebra of J generated by e.*

Proof. Let J_e denote the JB^*-subalgebra of J generated by e. By [13] and [12], there exists a C^*-algebra A containing J_e as a JB^*-subalgebra. Therefore, by Lemma 5.4 and Proposition 2.4, $K := \sigma(e)$ is a compact subset of $]1,\infty[$. By Theorem 2.8, there exists an isometric $*$-homomorphism $\Phi : C(K, M_2(\mathbb{C})) \to A$ such that $\Phi(\eta_K) = e$. Let Ψ stands for the restriction of Φ to $C(K,\mathcal{C}_3)$. Then, clearly, Ψ is an isometric $*$-homomorphism from $C(K,\mathcal{C}_3)$ to the JB^*-algebra underlying A, which satisfies $\Psi(\eta_K) = e$. Noticing that the JB^*-subalgebras of A and J generated by e coincide, it follows from Lemma 5.3 that the range of Ψ is J_e. This last fact allows us to see Ψ as a $*$-homomorphisms from $C(K,\mathcal{C}_3)$ to J. That Ψ is the unique (automatically continuous [12]) $*$-homomorphism from $C(K,\mathcal{C}_3)$ to J with $\Psi(\eta_K) = e$ follows from a new application of Lemma 5.3. □

Let K be a compact subset of $[1,\infty[$ with $1 \in K$, and let p be a self-adjoint idempotent in $\mathcal{C}_3$, different from 0 and 1. Then

$$C_p(K,\mathcal{C}_3) := \big\{\alpha \in C(K,\mathcal{C}_3) \colon \alpha(1) \in \mathbb{C}p\big\}$$

is a proper JB^*-subalgebra of $C(K,\mathcal{C}_3)$. As in the case of the C^*-algebra $C_p(K, M_2(\mathbb{C}))$, the JB^*-algebra $C_p(K,\mathcal{C}_3)$ does not depend structurally on p. Indeed, if, for $i \in \{1,2\}$, p_i is a self-adjoint idempotent in $M_2(\mathbb{C})$, different from 0 and 1, then $\{p_i, 1-p_i\}$ is a "frame of tripotents" in the simple JB^*-triple underlying $\mathcal{C}_3$, and therefore, by [10, Theorem 5.9], there exists a triple automorphism ϕ of $\mathcal{C}_3$ satisfying $\phi(p_1) = p_2$ and $\phi(1-p_1) = 1-p_2$. This implies that $\phi(1) = 1$, and then, by (5.2), that ϕ is actually an algebra $*$-automorphism. Such a $*$-automorphism of $\mathcal{C}_3$ induces a $*$-automorphism of $C(K,\mathcal{C}_3)$ sending $C_{p_1}(K,\mathcal{C}_3)$ onto $C_{p_2}(K,\mathcal{C}_3)$.

Lemma 5.6. *Let K be a compact subset of* $[1,\infty[$ *with* $1 \in K$, *and whose maximum element is greater than* 1. *Then $C_{\eta(1)}(K,\mathcal{C}_3)$ is generated by η_K as a JB^*-algebra.*

Proof. Argue as in the proof of Lemma 5.3, invoking Lemma 3.2 instead of Lemma 2.7. □

By invoking Theorem 3.3 and Lemma 5.6 instead of Theorem 2.8 and Lemma 5.3, respectively, the proof of the following theorem is similar to that of Theorem 5.5, and hence is omitted.

Theorem 5.7. *Let J be a JB^*-algebra, and let e be a nonself-adjoint idempotent in J. Put $K := \sigma(e)$, and assume that* 1 *belongs to K. Then K is a compact subset of* $[1,\infty[$ *whose* *maximum element is greater than* 1*, and the JB^*-subalgebra of J generated by e is $*$-isomorphic to $C_p(K,\mathcal{C}_3)$ for any self-adjoint idempotent $p \in \mathcal{C}_3$ different from* 0 *and* 1*. More precisely, we have*:

(1) *There exists a unique $*$-homomorphism $\Psi : C_{\eta(1)}(K,\mathcal{C}_3) \to J$ such that $\Psi(\eta_K) = e$.*
(2) *Such a $*$-homomorphism is isometric, and its range coincides with the JB^*-subalgebra of J generated by e.*

6. The case of JB^*-algebras: some consequences

In this section, we deal with the main corollaries to Theorems 5.5 and 5.7.

Corollary 6.1. *Let J be a JB^*-algebra generated by a nonself-adjoint idempotent e, and put $K := \sigma(e)$. If* 1 *is an isolated point of the compact set K, then J is $*$-isomorphic to the JB^*-algebra*

$$\mathbb{C} \times C\big(K \setminus \{1\}, \mathcal{C}_3\big).$$

Proof. Argue as in the proof of Corollary 4.1, invoking Theorem 5.7 instead of Theorem 3.3. □

Corollary 6.2. *Let J be a JB^*-algebra generated by a nonself-adjoint idempotent e, and put $K := \sigma(e)$. Then J has a unit if and only if either* 1 *does not belong to K or* 1 *is an isolated point of K.*

Proof. Argue as in the proof of Corollary 4.2, invoking Theorems 5.5 and 5.7, and Corollary 6.1 instead of Theorems 2.8 and 3.3, and Corollary 4.1, respectively. □

Corollary 6.3. *Let J be a JB^*-algebra. Then J has a nonself-adjoint idempotent* (*if and*) *only if it contains* (*as a JB^*-subalgebra*) *a copy of either $\mathcal{C}_3$ or $C_p([1,2],\mathcal{C}_3)$ for any self-adjoint idempotent $p \in \mathcal{C}_3$ different from* 0 *and* 1.

Proof. Argue as in the proof of Corollary 4.3, invoking Theorems 5.5 and 5.7, and Corollary 6.1 instead of Theorems 2.8 and 3.3, and Corollary 4.1, respectively. □

Arguing as in the comment following Corollary 4.3, one can realize that the JB^*-algebra $C_p([1,2],\mathcal{C}_3)$ does not contain any copy of $\mathcal{C}_3$.

Let J be a Jordan algebra. For $a, b, c \in J$, we put

$$[a,b,c] := (a \cdot b) \cdot c - a \cdot (b \cdot c).$$

The centre of J is defined as the set of those elements $a \in J$ such that $[a, J, J] = 0$. It is well known and easy to see that central elements a of J satisfy $[J, J, a] = [J, a, J] = 0$.

Remark 6.4. In relation to Corollary 6.3 above, it is worth mentioning that a JB^*-algebra contains a nonself-adjoint idempotent if and only if it contains a non-central self-adjoint idempotent [1]. Actually, the "only if" part of the result in [1] just quoted follows easily from Corollary 6.3, whereas the "if part" is a consequence of Proposition 6.5 immediately below.

Proposition 6.5. *Let J be a JB^*-algebra containing a non-central idempotent e. Then there exists a continuous mapping $r \to e_r$ from $[1, \infty[$ to the set of idempotents of J satisfying $e_{\|e\|} = e$ and $\|e_r\| = r$ for every $r \in [1, \infty[$.*

Proof. First assume that e is not self-adjoint. Then, invoking Theorems 5.5 and 5.7 instead of Theorems 2.8 and 3.3, respectively, and keeping in mind that, for every $t \in [1, \infty[$, $\eta(t)$ lies in $\mathcal{C}_3$, the first part of the proof of Proposition 4.5 works verbatim.

Now assume that e is self-adjoint. Since e is non-central, we may apply Lemma 2.5.5 of [6] to find $c \in J$ such that $U_e(c) \neq e \cdot c$ or, equivalently, $[e, e, c] \neq 0$. Moreover, clearly, such an element c can be chosen self-adjoint. There is no loss of generality in assuming that J is generated by $\{e, c\}$ as a JB^*-algebra. Then, by [12], there exists a C^*-algebra A containing J as a JB^*-subalgebra. Put $a := i(ec - ce) \in A$, and consider the mapping $D : A \to A$ defined by $D(b) := ba - ab$ for every $b \in A$. Then D becomes a continuous derivation of A satisfying $D(b^*) = -D(b)^*$ for every $b \in A$ (since a is self-adjoint). Moreover, for every $b \in J$ we have

$$D(b) = 4i[e, b, c] \in J, \tag{6.1}$$

and consequently $D(e) \neq 0$. By the second part of the proof of Proposition 4.5, there exists a continuous function $h : [1, \infty[\to \mathbb{R}$ such that the continuous mapping $e \to e_r := \exp(h(r)D)(e)$, from $[1, \infty[$ to the set of idempotents of A, satisfies $e_1 = e$ and $\|e_r\| = r$ for every $r \in [1, \infty[$. Therefore, the proof is concluded by realizing that, for every $r \in [1, \infty[$, e_r lies in J. But this follows from the fact that, by (6.1), J is invariant under D. □

An element a in a JB^*-algebra J is said to be normal if the equality $[a, a, a^*] = 0$ is satisfied. In the case that the JB^*-algebra J is a JB^*-subalgebra of a given C^*-algebra A, the equality $[a, a, a^*] = 0$ in J reads in A as $[[a, a^*], a] = 0$, and hence, by arguing as in the conclusion of the proof of Lemma 4.4, it is equivalent to the usual normality in A, namely $[a, a^*] = 0$.

An element x in a JB^*-triple is said to be a tripotent if the equality $\{x, x, x\} = x$ holds. Thus, the tripotents in a C^*-algebra are precisely the partial isometries, and, more generally, the tripotents in a JB^*-algebra are precisely those elements a satisfying $U_a(a^*) = a$.

Corollary 6.6. *Let J be a JB^*-algebra. Then the following assertions are equivalent*:

(1) *J contains a non-central self-adjoint idempotent.*
(2) *There exists a non-normal tripotent $a \in J$ such that a belongs to $U_{a^2}(J)$.*

Proof. (1) ⇒ (2). By the assumption (1), Remark 6.4, and Theorems 5.5 and 5.7, we may assume that J is of the form $C(K, \mathcal{C}_3)$ or $C_{\eta(1)}(K, \mathcal{C}_3)$, where, in the first case, K is a compact subset of $]1, \infty[$ and, in the second case, K is a compact subset of $[1, \infty[$ whose maximum element is greater than 1 and such that $1 \in K$. In any case, by Lemma 2.2, η_{21}^K is a non-normal partial isometry in J, and we have $\eta_{21}^K = U_{(\eta_{21}^K)^2}(u^2\eta_{12}^K)$, with $u^2\eta_{12}^K \in J$.

(2) ⇒ (1). Assume that assertion (2) holds. We may suppose that J is generated by a as a JB^*-algebra. Since a belongs to $U_{a^2}(J)$, [11, Lemma 1] applies, giving the existence of an idempotent $e \in J$ such that

$$U_a(J) = U_e(J).$$

Note that, by [7, pp. 118, 119], $U_e(J)$ is a subalgebra of J, and that e is a unit for such a subalgebra. Assume that e is self-adjoint. Then $U_e(J)$ is a JB^*-subalgebra of J, and hence, since

$$a = U_a\bigl(a^*\bigr) \in U_a(J) = U_e(J),$$

and J is generated by a as a JB^*-algebra, we deduce that $U_a(J) = J$ and that e is a unit for J. It follows from [7, Theorem 13, p. 52] that there exists a unique element $b \in J$ (called the "inverse" of a) such that $a = U_a(b)$, and that such a b satisfies $[a, x, b] = 0$ for every $x \in J$. Therefore we have that $b = a^*$, and then that $[a, a, a^*] = 0$, contrarily to the assumption that a is not normal. In this way we have shown that the idempotent e is not self-adjoint, and the proof is concluded by applying Remark 6.4. □

Comparing Corollary 6.6 with Corollary 4.7, one is tempted to conjecture that the equivalent assertions (1) and (2) in Corollary 6.6 are also equivalent to the following:

(3) J contains a non-normal tripotent.

As a matter of fact, we have been unable to prove or disprove the conjecture just formulated. Actually, an eventual verification of such a conjecture would provide in particular an affirmative answer to the following unsolved question.

Problem 6.7. Let J be a JB^*-algebra containing a non-zero tripotent. Does J contain a non-zero self-adjoint idempotent?

We conclude the paper with an application to the theory of JB-algebras. JB-algebras are defined as those Banach–Jordan real algebras J satisfying $\|a\|^2 \leqslant \|a^2 + b^2\|$ for all $a, b \in J$. The basic reference for JB-algebras is [6]. By [6, Proposition 3.8.2], the self-adjoint part of every JB^*-algebra becomes a JB-algebra. In particular, the self-adjoint part of the three-dimensional (complex) spin factor $\mathcal{C}_3$ is a JB-algebra, which is called the three-dimensional real spin factor, and is denoted by $\mathcal{S}_3$. We denote by $C([1,2], \mathcal{S}_3)$ the JB-algebra of all continuous functions from $[1,2]$ to $\mathcal{S}_3$. Moreover, given an idempotent $p \in \mathcal{S}_3$ different from 0 and 1, we denote by $C_p([1,2], \mathcal{S}_3)$ the JB-subalgebra of $C([1,2], \mathcal{S}_3)$ consisting of all elements $\alpha \in C([1,2], \mathcal{S}_3)$ such that $\alpha(1)$ belongs to $\mathbb{R}p$.

Now, we have the following.

Corollary 6.8. *Let J be a JB-algebra. Then J has a non-central idempotent (if and) only if it contains (as a JB-subalgebra) a copy of either $\mathcal{S}_3$ or $C_p([1,2], \mathcal{S}_3)$ for any idempotent $p \in \mathcal{S}_3$ different from* 0 *and* 1.

Proof. By [12,13], there exists a JB^*-algebra whose self-adjoint part is equal to J. Now apply Remark 6.4 and Corollary 6.3. □

Acknowledgment

The authors are very grateful to A.M. Peralta for fruitful discussions on the matter of the paper.

References

[1] J. Becerra, A. Rodríguez, Nonself-adjoint idempotents in C^*- and JB^*-algebras, preprint.
[2] F.F. Bonsall, J. Duncan, Complete Normed Algebras, Springer-Verlag, Berlin, 1973.
[3] R.B. Braun, W. Kaup, H. Upmeier, A holomorphic characterization of Jordan C^*-algebras, Math. Z. 161 (1978) 277–290.
[4] J. Duncan, P.J. Taylor, Norm inequalities for C^*-algebras, Proc. Roy. Soc. Edinburgh Sect. A 75 (1975/1976) 119–129.
[5] Y. Friedman, B. Russo, Structure of the predual of a JBW^*-triple, J. Reine Angew. Math. 356 (1985) 67–89.
[6] H. Hanche-Olsen, E. Stormer, Jordan Operator Algebras, Monogr. Stud. Math., vol. 21, Pitman, 1984.
[7] N. Jacobson, Structure and Representations of Jordan Algebras, Amer. Math. Soc. Colloq. Publ., vol. 39, Amer. Math. Soc., Providence, RI, 1968.
[8] W. Kaup, Algebraic characterization of symmetric complex Banach manifolds, Math. Ann. 228 (1977) 39–64.
[9] W. Kaup, A Riemann mapping theorem for bounded symmetric domains in complex Banach spaces, Math. Z. 183 (1983) 503–529.
[10] O. Loos, Bounded Symmetric Domains and Jordan Pairs, Mathematical Lectures, Univ. of California at Irvine, Irvine, 1977.
[11] O. Loos, Fitting decomposition in Jordan systems, J. Algebra 136 (1991) 92–102.
[12] J.D.M. Wright, Jordan C^*-algebras, Michigan Math. J. 24 (1977) 291–302.
[13] M.A. Youngson, Non-unital Banach Jordan algebras and C^*-triple systems, Proc. Edinb. Math. Soc. 24 (1981) 19–31.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Functional Analysis 248 (2007) 128–151
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

Wavelet bases on a manifold

Abdellatif Jouini *, Mouna Kratou

Faculty of Sciences of Tunis, Department of Mathematics, Campus, 1060 Tunis, Tunisia

Received 22 July 2006; accepted 29 March 2007

Available online 7 May 2007

Communicated by D. Voiculescu

Abstract

Using the decomposition method, we present in this paper constructions of multiresolution analyses on a compact Riemannian manifold M of dimension n ($n \in \mathbb{N}$). These analyses are generated by a finite number of basic functions and are adapted to the study of the Sobolev spaces $H^1(M)$ and $H_0^1(M)$.

Keywords: Multiresolution analyses; Wavelet bases; Sobolev spaces; Extension operators

1. Introduction

The search for wavelet bases on a bounded domain has been an active field for many years, since the beginning of the nineties. The constructions are based on the decomposition method, introduced by Z. Ciesielski and T. Fiegel in 1982 [2,3] to construct spline bases of generalized Sobolev spaces $W_p^k(M)$ ($k \in \mathbb{Z}$ and $1 < p < \infty$) on a Riemannian manifold M. In 1992, we constructed biorthogonal wavelet bases on two-dimensional manifold Ω [8,10]. In 1997, the decomposition method was used by A. Cohen, W. Dahmen and R. Schneider [4–6] to construct biorthogonal wavelet bases $(\psi_\lambda, \widetilde{\psi}_\lambda)_{\lambda\in\nabla}$ of $\mathbb{L}^2(\Omega)$ where Ω is a bounded domain of $\mathbb{R}^d$ ($d \in \mathbb{N}$); these bases were shown to be bases of Sobolev spaces $H^s(\Omega)$ for $|s| < 3/2$. There are others constructions based on the decomposition method as well by A. Canuto and coworkers [1,15,13] and by R. Masson [12]. These bases are continuous but not differentiable and have never been implemented. Moreover, there is a slight difficulty in their presentation, due to notational burden and it is often unclear how to get other regularity Sobolev estimates than for $|s| < 3/2$. Recently,

* Corresponding author.
E-mail address: abdellatif.jouini@fst.rnu.tn (A. Jouini).

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2007.03.024

in 2003 [11], we constructed in an elementary way two multiresolution analyses on the L-shaped domain which are adapted to higher regularity analysis (namely, to the study of the Sobolev space H^k, $k \in \mathbb{Z}$).

In this paper, we use the preceding method called decomposition method to construct biorthogonal multiresolution analyses on a compact Riemannian manifold M of dimension n (or an open bounded set of $\mathbb{R}^n$), $n \in \mathbb{N}^*$. The central problem is to construct extension operators which are straightforward, relatively simple and are adapted to the scale. The present construction of biorthogonal analyses differs from the previous one in the sense that these analyses are generated by a finite number of simple basic functions and have better stability constants. They are also adapted to the study of the Sobolev space $H^1(M)$ (or $H_0^1(M)$). If we take as example the L-shaped domain, we obtain easily biorthogonal wavelet bases because the diffeomorphisms are translations and then the nodes are dyadic points.

The contents of this paper is as follows. In Section 2, we define and study the decomposition method of a bounded domain. We prove some results associated to the decomposition method as Decomposition and Extension lemmas. Next we prove a Regular lemma which is very useful in many equivalence norms.

In Section 3, we define biorthogonal multiresolution analyses on the real line and we prove that these analyses are adapted to the study of the Sobolev space $H^1(\mathbb{R})$.

Section 4 is devoted to constructions of biorthogonal wavelet bases on the interval $[0, 1]$. Using tensorialisation, we construct biorthogonal multiresolution analyses on the cube $I^n = [0, 1]^n$.

In the last section, we construct on M biorthogonal wavelet bases which are generated by a finite number of basic functions.

2. Decomposition method

In the following, we denote by M a compact Riemannian manifold of dimension n ($n \in \mathbb{N}^*$). We assume that the Dirichlet boundary ∂M is piecewise C^1. In this section, we define the decomposition method and we prove some associated results as Decomposition and Extension lemmas which are very important for our constructions. The following definitions are very important for constructions of wavelet bases.

Definition 2.1. A subset $Q \subseteq M$ is said to be a n-cube if there exists a diffeomorphism $\Phi : U \to R^n$, where $U = \operatorname{Int} U \supset Q$, such that $\Phi(Q) = I^n$.

The norm on $L^2(Q)$ can be introduced as follows. We fix a $\Phi : Q \to I^n$ as in the definition and let for $f \in L^2(Q)$,

$$\|f\|_{L^2(Q)} = \left\| f \circ \Phi^{-1} \right\|_{L^2(I^n)}.$$

This norm depends on the choice of Φ. Choosing another diffeomorphism Ψ of Q onto I^n, we obtain an equivalent norm.

Definition 2.2. M has a decomposition in n-cubes if we can write $\overline{M} = \bigcup_{i=1}^N Q_i$ such that the n-cubes Q_i satisfy the following properties:

(i) If we denote Φ_j a diffeomorphism from Q_j onto I^n, $\Phi_i \circ \Phi_\ell^{-1}$ is an affine application from $\Phi_\ell(Q_i \cap Q_\ell)$ into I^n.

(ii) $Q_i \cap Q_\ell$, $\ell \neq i$, is a closed segment or a face.
(iii) $Q_i \cap \partial M$ is a union of segments or faces of Q_i.
(iv) The set $Z_i = Q_i \cap (\bigcup_{\ell<i} Q_\ell)$ is a union of segments or faces of Q_i and the set $\tilde{Z}_i = Q_i \cap (\partial M \cup \bigcup_{\ell<i} Q_i)$ is a union of segments or faces of Q_i.

The existence of decomposition is proved by Z. Ciesielski and T. Fiegel [3]. There are many examples of decompositions of bounded domains (triangle, polygon, disc, L-shaped domain, sphere, ...) [8,9].

We define the measure dm on M by

$$\int f\,dm = \sum_{i=1}^{N} \iint_{I^n} f|_{Q_i} \circ \Phi_i^{-1}\,dx.$$

This measure is equivalent to the Lebesgue measure and define a scalar product on $L^2(M)$:

$$\langle f|g\rangle_M = \int f\bar{g}\,dm.$$

We shall construct wavelet bases which are located in time and frequency. To realize this object, we introduce the notions of j-dyadic point and j-spline.

Definition 2.3.

(i) For $j \geqslant 0$, a j-dyadic point is a point $m \in M$ such that for $i \in \{1,\dots,N\}$, $m \in Q_i$ and $2^j \Phi_i(m) \in \mathbb{Z}^n$. We denote by M_j the set of j-dyadic points of M.
(ii) For $j \geqslant 0$, A j-spline is a continuous function $f : M \to \mathbb{C}$ such that, for $i \in \{1,\dots,N\}$ and every dyadic cube $Q_{j,(k_1,k_2,\dots,k_n)} = [\frac{k_1}{2^j}, \frac{k_1+1}{2^j}] \times \cdots \times [\frac{k_n}{2^j}, \frac{k_n+1}{2^j}]$ contained in Q, $f \circ \Phi_i^{-1}|_{Q_{j,k_1,k_2}}$ is a polynomial of the form $(a_1 + b_1 x_1) \times \cdots \times (a_n + b_n x_n)$.

We give here the first associated result with decomposition.

Lemma 2.1 *(Decomposition lemma). The application*

$$H^1(M) \to \prod_{i=1}^{N} H^1(Q_i), \quad f \to (f|Q_i)_{1\leqslant i\leqslant N},$$

is an isomorphism from $H^1(M)$ onto the closed subspace H of $\prod_{i=1}^N H^1(Q_i)$ of functions $(f_i)_{1\leqslant i\leqslant N}$ such that $f_i|_{Q_i\cap Q_\ell} = f_\ell|_{Q_i\cap Q_\ell}$ in $L^2(Q_i \cap Q_\ell)$, where $Q_i \cap Q_\ell$ is equal to a closed face. In particular, the norms $\|f\|_{H^1(M)}$ and $\sum_{i=1}^N \|f|_{Q_i}\|_{H^1(Q_i)}$ are equivalent on $H^1(M)$.

Proof. This lemma is immediate, because the compatibility condition $f_i|_{Q_i\cap Q_\ell} = f_\ell|_{Q_i\cap Q_\ell}$ permits to derive $f = \sum_{i=1}^N f_i \chi_{Q_i}$ on every square separately and to deduce: $\frac{\partial}{\partial x_\ell} f = \sum_{i=1}^N (\frac{\partial f_i}{\partial x_\ell}) \chi_{Q_i}$ where χ_{Q_i} is the characteristic function of Q_i. $\square$

We introduce here the notion of extensions which are very useful to construct wavelet bases in our domain M from those constructed in the cube $I^n = [0,1]^n$.

Definition 2.4. For $i = 1, \dots, N$, we denote by $Z_i = Q_i \cap (\bigcup_{\ell<i} Q_\ell)$. A family of continuous operators $(E_i : L^2(Q_i) \to L^2(M))_{1\leqslant i\leqslant N}$ is called a family of extensions if:

(i) $(E_i f)|_{Q_i} = f$ where $f \in L^2(Q_i)$.
(ii) $\mathrm{supp}(E_i f) \subset \bigcup_{\ell\geqslant i} Q_\ell$.
(iii) For $f \in H^1(Q_i)$ and $f|_{Z_i} = 0$, $E_i f \in H^1(M)$ and the application E_i is continuous from $H^{1,Z_i}(Q_i) = \{f \in H^1(Q_i) \mid f|_{Z_i} = 0\}$ into $H^1(M)$.

A second associated result with decomposition is given by the following lemma.

Lemma 2.2 *(Extension lemma). Let (E_i) be a family of extensions. Then the application*

$$(f_i)_{1\leqslant i\leqslant N} \to \sum_{i=1}^{N} E_i f_i$$

is an isomorphism from $\sum_{i=1}^N L^2(Q_i)$ onto $L^2(M)$ and from $\sum_{i=1}^N H^{1,Z_i}(Q_i)$ onto $H^1(M)$.

Proof. This lemma is immediate because the inverse isomorphism is given by $f \to (f_i)_{1\leqslant i\leqslant N}$ where f_i is defined by $f_i = (f - \sum_{\ell<i} E_\ell f_\ell)|_{Q_i}$. $\square$

Remark. We have the same result for $H_0^1(M)$ if we replace in Definition 2.4 and Lemma 2.2 the spaces $H^1(M)$ by $H_0^1(M)$ and $H^{1,Z_i}(Q_i)$ by $H^{1,\tilde{Z}_i}(Q_i)$ where $\tilde{Z}_i = Q_i \cap (\partial M \cup \bigcup_{\ell<i} Q_\ell)$.

We conclude that, for the study of $H^1(M)$ or $H_0^1(M)$, we need to study the spaces $H^{1,Z}(Q) = \{f \in H^1(Q) \mid f_{/Z} = 0\}$ where Z is a union of faces of ∂Q. For $T \subset \{0,1\}$ and $I = [0,1]$, we denote by $H^{1,T}(I)$ the space of functions $f \in H^1(I)$ such that $f|_T = 0$. For $T_1 = \{t \in \{0,1\} \mid \{t\} \times [0,1] \subset Z\}$ and $T_2 = \{t \in \{0,1\} \mid [0,1] \times \{t\} \subset Z\}$, we have $H^{1,Z}(Q) = H^{1,T_1}(I) \hat{\otimes} H^{1,T_2}(I)$.

We need the following results established in [10] to prove many equivalence norms in Sobolev spaces $H^1(M)$ and $H_0^1(M)$.

Lemma 2.3.

(i) *If $g \in L^2(\mathbb{R}^2)$ with compact support then there exists a positive constant C such that, for every sequence $(\lambda_{k_1,k_2}) \in \ell^2(\mathbb{Z}^2)$, we have*:

$$\left\| \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} \lambda_{k_1,k_2} g(x-k_1, y-k_2) \right\|_{L^2(\mathbb{R}^2)} \leqslant C \left(\sum_{k_1} \sum_{k_2} |\lambda_{k_1,k_2}|^2 \right)^{1/2}. \tag{2.1}$$

(ii) *Moreover, if $|\xi_1|^\alpha \hat{g}(\xi_1,\xi_2) \in L^2(\mathbb{R}^2)$ where $0 < \alpha < 1$, then there exists a constant positive C' such that, for every sequence $(\lambda_{k_1,k_2}) \in \ell^2(\mathbb{Z}^2)$, we have*

$$\left\| |\xi_1|^\alpha \left(\sum_{k_1} \sum_{k_2} \lambda_{k_1,k_2} e^{-k_1\xi_1} e^{-ik_2\xi_2} \right) \hat{g}(\xi_1,\xi_2) \right\|_{L^2(\mathbb{R}^2)} \leqslant C' \left(\sum_{k_1} \sum_{k_2} |\lambda_{k_1,k_2}|^2 \right)^{1/2}. \tag{2.2}$$

Lemma 2.4 *(Regular lemma). Let $g \in L^2(\mathbb{R}^n)$ be such that:*

(i) *g has a compact support.*
(ii) *$\int g(x)\,dx = 0$.*
(iii) *$g \in H^\varepsilon(\mathbb{R}^n)$ for $\varepsilon > 0$.*

Let us denote by $g_{j,k}$ the function $g_{j,k}(x) = 2^{jn/2} g(2^j x - k)$. Then, for $j \in \mathbb{Z}$ and $k \in \mathbb{Z}^n$, there exists a positive constant C such that for every sequence $(\lambda_{j,k}) \in \ell^2(\mathbb{Z} \times \mathbb{Z}^n)$ and every function $f \in L^2(\mathbb{R}^n)$ we have

$$\Big\| \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^n} \lambda_{j,k} g_{j,k} \Big\|_{L^2(\mathbb{R}^n)} \leqslant C \Big(\sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^n} |\lambda_{j,k}|^2 \Big)^{1/2}, \tag{2.3}$$

and

$$\Big(\sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^n} \big| \langle f | g_{j,k} \rangle \big|^2 \Big)^{1/2} \leqslant C \| f \|_{L^2(\mathbb{R}^n)}. \tag{2.4}$$

Proof. We prove first this lemma for $n = 2$. We denote by

$$\Omega_j = \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} \lambda_{j,k_1,k_2} 2^j g\big(2^j x - k_1, 2^j y - k_2\big).$$

We consider $0 < \varepsilon < 1$. From Lemma 2.3 there exists a positive constant C such that

$$\big\| \big(|\xi|^\varepsilon + |\eta|^\varepsilon\big) \hat{\Omega}_j(\xi,\eta) \big\|_{L^2(\mathbb{R}^2)} \leqslant C 2^{j\varepsilon} \Big(\sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} |\lambda_{j,k_1,k_2}|^2 \Big)^{1/2}.$$

We have

$$\iint g\,dx\,dy = 0.$$

Then, we write

$$g = \frac{\partial}{\partial x_1} g_1 + \frac{\partial}{\partial x_2} g_2,$$

where g_i, $i = 1, 2$, belongs to $L^2(\mathbb{R}^2)$, g_i has compact support and $\xi_i \hat{g}_i(\xi_1, \xi_2) \in L^2(\mathbb{R}^2)$. We conclude from Lemma 2.3 applied to g_i, $i = 1, 2$ and $\alpha = 1 - \varepsilon$, that if we denote by

$$\Omega_{i,j} = \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} \lambda_{j,k_1,k_2} 2^j g_i\big(2^j x - k_1, 2^j y - k_2\big),$$

we have

$$\begin{aligned}\big\|\big(|\xi|+|\eta|\big)^{-\varepsilon}\hat{\Omega}_j(\xi,\eta)\big\|_2 &\leqslant 2^{-j}\big(\big\||\xi|^{1-\varepsilon}\hat{\Omega}_{1,j}(\xi,\eta)\big\|_2+\big\||\eta|^{1-\varepsilon}\hat{\Omega}_{2,j}(\xi,\eta)\big\|_2\big),\\ &\leqslant C2^{-j}2^{j(1-\varepsilon)}\bigg(\sum_{k_1}\sum_{k_2}|\lambda_{j,k_1,k_2}|^2\bigg)^{1/2}\\ &= C2^{-j\varepsilon}\bigg(\sum_{k_1}\sum_{k_2}|\lambda_{j,k_1,k_2}|^2\bigg)^{1/2},\end{aligned}$$

where C is a positive constant. Then, we obtain, for $j \geqslant \ell$:

$$\begin{aligned}\big|\langle\Omega_j|\Omega_\ell\rangle\big| &= \frac{1}{4\pi^2}\big\|\big(|\xi|+|\eta|\big)^{-\varepsilon}\hat{\Omega}_j\big\|_2\big\|\big(|\xi|+|\eta|\big)^{\varepsilon}\hat{\Omega}_\ell\big\|_2,\\ &\leqslant C2^{-\varepsilon|j-\ell|}\bigg(\sum_{k_1}\sum_{k_2}|\lambda_{j,k_1,k_2}|^2\bigg)^{1/2}\bigg(\sum_{k_1}\sum_{k_2}|\lambda_{\ell,k_1,k_2}|^2\bigg)^{1/2}.\end{aligned}$$

Thus there exists a positive constant C' such that

$$\bigg\|\sum_{j\in\mathbb{Z}}\Omega_j\bigg\|_2 \leqslant C'\bigg(\sum_j\sum_{k_1}\sum_{k_2}|\lambda_{j,k_1,k_2}|^2\bigg)^{1/2}.$$

We have the same proof for every $n \geqslant 2$. Then the Regular lemma is proved. □

3. Multiresolution analyses on the real line

In this section, we present multiresolution analyses defined on the real line. These analyses will be used in next sections. We denote φ and φ^* the functions defined by:

$$\hat{\varphi}(\xi) = \prod_{j=1}^{\infty} m_0\bigg(\frac{\xi}{2^j}\bigg), \quad \text{where } m_0(\xi) = \frac{1+\cos\xi}{2}; \tag{3.1}$$

$$\hat{\varphi}^*(\xi) = \prod_{j=1}^{\infty} m_0^*\bigg(\frac{\xi}{2^j}\bigg), \quad \text{where } m_0^*(\xi) = \bigg(\frac{1+\cos\xi}{2}\bigg)(2-\cos\xi), \tag{3.2}$$

where "ˆ" is the classical Fourier transform on $\mathbb{R}$ [12].

We recall here some main properties of these analyses.

$$\varphi(x) = \big(1-|x|\big)^+ = \sup\big(1-|x|,0\big); \tag{3.3}$$

$$\varphi \text{ and } \varphi^* \text{ are in } L^2(\mathbb{R}); \tag{3.4}$$

$$\text{there exists } \varepsilon > 0 \text{ such that } \varphi \in H^{1+\varepsilon}(\mathbb{R}); \tag{3.5}$$

$$\operatorname{supp}\varphi = [-1,1], \quad \varphi(x) = \varphi(-x), \quad \sum_{k\in\mathbb{Z}}\varphi(x-k) = 1, \quad \sum_{k\in\mathbb{Z}}k\varphi(x-k) = x; \tag{3.6}$$

$$\operatorname{supp}\varphi^* = [-2,2], \quad \varphi^*(x) = \varphi^*(-x), \quad \sum_{k\in\mathbb{Z}}\varphi^*(x-k) = 1, \quad \sum_{k\in\mathbb{Z}}k\varphi^*(x-k) = x; \tag{3.7}$$

$$\varphi(x)=\frac{1}{2}\varphi(2x+1)+\varphi(2x)+\frac{1}{2}\varphi(2x-1); \tag{3.8}$$

$$\varphi^*(x)=\frac{1}{2}\varphi^*(2x+2)+\frac{1}{2}\varphi^*(2x+1)+\frac{3}{2}\varphi^*(2x)+\frac{1}{2}\varphi^*(2x-1)-\frac{1}{4}\varphi^*(2x-2); \tag{3.9}$$

$$\langle\varphi_{j,k},\varphi^*_{j,q}\rangle_{L^2(\mathbb{R})}=\delta_{k,q}. \tag{3.10}$$

It is clear that the functions $\varphi_{j,k}$, $k\in\mathbb{Z}$, form a Riesz basis of $V_j(\mathbb{R})$ and the functions $\varphi^*_{j,k}$, $k\in\mathbb{Z}$, constitute a Riesz basis of its dual $V^*_j(\mathbb{R})$ such that $L^2(\mathbb{R})=V_j(\mathbb{R})\oplus(V^*_j(\mathbb{R}))^\perp$. The projector P_j from $L^2(\mathbb{R})$ into $V_j(\mathbb{R})$ parallel to $(V^*_j(\mathbb{R}))^\perp$ is given by

$$P_jf=\sum_{k\in\mathbb{Z}}\langle f,\varphi^*_{j,k}\rangle\varphi_{j,k}, \tag{3.11}$$

and satisfies classical relations $P_{j+1}\circ P_j=P_j\circ P_{j+1}=P_j$, such that $Q_j=P_{j+1}-P_j$ is a projector from $L^2(\mathbb{R})$ into $W_j(\mathbb{R})=\operatorname{Im}Q_j=V_{j+1}(\mathbb{R})\oplus(V_j(\mathbb{R}))^\perp$ parallel to $(W^*_j(\mathbb{R}))^\perp$, where $W^*_j(\mathbb{R})=(\operatorname{Ker}Q_j)^\perp=V^*_{j+1}(\mathbb{R})\oplus(V_j(\mathbb{R}))^\perp$. A basis of $W_j(\mathbb{R})$ is given by the functions $\psi_{j,k}(x)=2^{\frac{j}{2}}\psi(2^jx-k)$, $k\in\mathbb{Z}$, where the associated wavelet ψ is usually defined by $\widehat{\psi}(2\xi)=e^{-i\xi}\overline{m(\xi+\pi)}\widehat{\varphi}(\xi)$ and a basis of $W^*_j(\mathbb{R})$ is given by the functions $\psi^*_{j,k}(x)=2^{\frac{j}{2}}\psi^*(2^jx-k)$, $k\in\mathbb{Z}$, where $\widehat{\psi^*}(2\xi)=e^{-i\xi}\overline{m(\xi+\pi)}\widehat{\varphi^*}(\xi)$. Moreover, the two bases are in duality for the scalar product of $L^2(\mathbb{R})$ and we have:

$$\langle\psi(x),\psi^*(x-k)\rangle_{L^2(\mathbb{R})}=\delta_{k,0}; \tag{3.12}$$

$$Q_jf=\sum_{k\in\mathbb{Z}}\langle f,\psi^*_{j,k}\rangle\psi_{j,k}; \tag{3.13}$$

$$\operatorname{supp}\psi=[-1,2],\qquad \psi(x)=\psi(1-x),\qquad \int_{-\infty}^{+\infty}\psi(x)\,dx=\int_{-\infty}^{+\infty}x\psi(x)\,dx=0,$$

$$\langle\psi(x),\varphi^*(x-k)\rangle_{L^2(\mathbb{R})}=0\quad\text{for }k\in\mathbb{Z}; \tag{3.14}$$

$$\operatorname{supp}\psi^*=[-1,2],\qquad \psi^*(x)=\psi^*(1-x),\qquad \int_{-\infty}^{+\infty}\psi^*(x)\,dx=\int_{-\infty}^{+\infty}x\psi^*(x)\,dx=0,$$

$$\langle\psi^*(x),\varphi(x-k)\rangle_{L^2(\mathbb{R})}=0\quad\text{for }k\in\mathbb{Z}; \tag{3.15}$$

$$\langle\psi_{j,k},\psi^*_{j,q}\rangle_{L^2(\mathbb{R})}=\delta_{j,j'}\delta_{k,k'}. \tag{3.16}$$

We have $\varphi\in H^{1+\varepsilon}(\mathbb{R})$, then we can apply the derivation method described in [9] to obtain new multiresolution analyses. More precisely, we denote by

$$\hat{\omega}(\xi)=\prod_{j=1}^{\infty}\tilde{m}_0\left(\frac{\xi}{2^j}\right),\quad\text{where }\tilde{m}_0(\xi)=m_0(\xi)\frac{2e^{-\xi}}{1+e^{-i\xi}}, \tag{3.17}$$

and

$$\hat{\omega}^*(\xi)=\prod_{j=1}^{\infty}\tilde{m}_0^*\left(\frac{\xi}{2^j}\right), \quad \text{where } \tilde{m}_0^*(\xi)=m_0^*(\xi)\frac{1+e^{-i\xi}}{2}. \tag{3.18}$$

Then, we have:

$$\frac{d\varphi}{dx}=\omega(x+1)-\omega(x), \tag{3.19}$$

$$\omega^*(x)=\int_{x-1}^{x}\varphi^*(t)\,dt, \tag{3.20}$$

$$\left\langle\omega(x)\,\middle|\,\omega^*(x-k)\right\rangle=\delta_{k,0}. \tag{3.21}$$

The functions $\omega_{j,k}=2^{j/2}\omega(2^jx-k)$, $k\in\mathbb{Z}$, constitute a Riesz basis of the space $\tilde{V}_j(\mathbb{R})$ and the functions $\omega^*_{j,k}=2^{j/2}\omega^*(2^jx-k)$, $k\in\mathbb{Z}$, constitute a Riesz basis of the space $\tilde{V}_j^*(\mathbb{R})$. The fundamental properties of ω and ω^* are:

$$\omega(x)=\chi_{[0,1]}(x), \tag{3.22}$$

where $\chi_{[0,1]}$ is the characteristic function of the interval [0, 1];

$$\omega\in H^{\varepsilon}(\mathbb{R}) \quad \text{and} \quad \omega^*\in H^{1+\varepsilon}(\mathbb{R}); \tag{3.23}$$

$$\operatorname{supp}\omega=[0,1], \qquad \omega(x)=\omega(1-x), \qquad \sum_{k\in\mathbb{Z}}\omega(x-k)=1; \tag{3.24}$$

$$\begin{gathered}\operatorname{supp}\omega^*=[-2,3], \qquad \omega^*(x)=\omega^*(1-x), \qquad \sum_{k\in\mathbb{Z}}\omega^*(x-k)=1,\\ \sum_{k\in\mathbb{Z}}k\omega^*(x-k)=x-\frac{1}{2}, \qquad \sum_{k\in\mathbb{Z}}k^2\omega^*(x-k)=x^2-x+\frac{1}{6};\end{gathered} \tag{3.25}$$

$$\omega(x)=\omega(2x)+\omega(2x-1); \tag{3.26}$$

$$\begin{aligned}\omega^*(x)={}&-\frac{1}{8}\omega^*(2x+2)+\frac{1}{8}\omega^*(2x+1)+\omega^*(2x)+\omega^*(2x-1)\\ &+\frac{1}{8}\omega^*(2x-2)-\frac{1}{8}\omega^*(2x-2)-\frac{1}{8}\omega^*(2x-3).\end{aligned} \tag{3.27}$$

The projector $\tilde{P}_j$ from $L^2(\mathbb{R})$ into $\tilde{V}_j(\mathbb{R})$ parallel to $\tilde{V}_j^*(\mathbb{R})^{\perp}$ has the following properties:

$$\tilde{P}_jf=\sum_{k\in\mathbb{Z}}\langle f|\omega^*_{j,k}\rangle\omega_{j,k}; \tag{3.28}$$

$$\tilde{P}_j\circ\tilde{P}_{j+1}=\tilde{P}_j. \tag{3.29}$$

Let f be in $H^1(\mathbb{R})$. Then we have the commutation property

$$\frac{d}{dx}(P_jf)=\tilde{P}_j\left(\frac{df}{dx}\right). \tag{3.30}$$

We consider $\tilde{Q}_j = \tilde{P}_{j+1} - \tilde{P}_j$, $\tilde{W}_j(\mathbb{R}) = \operatorname{Im}\tilde{Q}_j - \tilde{V}_{j+1}(\mathbb{R}) \cap \tilde{V}_j^*(\mathbb{R})^\perp$ and $\tilde{W}_j^*(\mathbb{R}) = (\ker\tilde{Q}_j)^\perp = \tilde{V}_{j+1}^*(\mathbb{R}) \cap \tilde{V}_j(\mathbb{R})^\perp$. $\tilde{W}_j(\mathbb{R})$ and $\tilde{W}_j^*(\mathbb{R})$ have the dual bases $\chi_{j,k}$ and $\chi_{j,k}^*$, $k \in \mathbb{Z}$ where $\chi_{j,\ell} = 2^{j/2}\chi(2^j x - k)$ and $\chi_{j,\ell}^* = 2^{j/2}\chi^*(2^j x - k)$, the functions χ and χ^* are given by

$$\chi = \frac{d\psi}{dx}, \tag{3.31}$$

$$\chi^* = \int_x^{+\infty} \psi^*(t)\,dt, \tag{3.32}$$

and the projectors Q_j and $\tilde{Q}_j$ are defined by

$$\tilde{Q}_j f = \sum_{k\in\mathbb{Z}} \langle f \mid \chi_{j,k}^*\rangle \chi_{j,k}, \tag{3.33}$$

and

$$\frac{d}{dx} Q_j f = \tilde{Q}_j\left(\frac{df}{dx}\right) \quad \text{for } f \in H^1(\mathbb{R}). \tag{3.34}$$

We can now establish the main result of this section.

Proposition 3.1. *Let $j_0 \in \mathbb{Z}$. Then*:

(i) *The norms $\|f\|_2$ and $\|P_{j_0} f\|_2 + (\sum_{j\geqslant j_0} \|Q_j f\|_2^2)^{1/2}$ are equivalent on $L^2(\mathbb{R})$.*
(ii) *The norms $\|f\|_{H^1}$ and $\|P_{j_0} f\|_2 + (\sum_{j\geqslant j_0} 4^j \|Q_j f\|_2^2)^{1/2}$ are equivalent on $H^1(\mathbb{R})$.*

Proof. We have $\sum_{k\in\mathbb{Z}} \varphi(x-k) = 1$ and $\sum_{k\in\mathbb{Z}} k\varphi(x-k) = x$. Then $\bigcup_{j\in\mathbb{Z}} V_j(\mathbb{R})$ is dense in $H^1(\mathbb{R})$ and, more precisely, there exists a positive constant C such that for $f \in H^2(\mathbb{R})$, we have

$$\left\| f - \sum_{k\in\mathbb{Z}} f(k) 2^{-j/2} \varphi_{j,k} \right\|_2 \leqslant C 2^{-2j} \|f''\|_2,$$

and

$$\left\| f - \sum_{k\in\mathbb{Z}} f(k) 2^{-j/2} \varphi_{j,k} \right\|_{H^1(\mathbb{R})} \leqslant C 2^{-j} \|f''\|_2.$$

We write $f = P_{j_0} f + \sum_{j\geqslant j_0} Q_j f$ and, by using Lemma 2.4, we obtain that $\sum_{k\in\mathbb{Z}} |\langle f \mid \psi_{j,k}^*\rangle|^2$ and $\|Q_j f\|_2^2$ are equivalent. Then property (2.1) gives the control of f by the norm

$$C\left(\|P_{j_0} f\|_2^2 + \sum_{j\geqslant j_0} \|Q_j f\|_2^2\right)^{1/2}.$$

To prove the inverse control, we consider ψ^*as vaguelette and we use property (2.2). We have the same result for χ and χ^*. Then, we obtain

$$\left\|\frac{df}{dx}\right\|_2 \approx \left\|\tilde{P}_{j_0}\frac{df}{dx}\right\|_2 + \left(\sum_{j\geqslant 0}\left\|\tilde{Q}_j\frac{df}{dx}\right\|_2^2\right)^{1/2},$$

and we conclude that

$$\left\|\tilde{Q}_j\frac{df}{dx}\right\|_2 = \left\|\frac{d}{dx}Q_j f\right\|_2 \approx 2^j\|Q_j f\|_2.$$

Then, the Proposition 3.1 is proved. □

4. Multiresolution analyses on the interval [0, 1]

In this section, we study the construction of biorthogonal multiresolution analyses on the interval $I=[0,1]$ with a specific treatment for the boundaries conditions. These analyses will be used for the cases of the square and the manifold.

We denote by $V_j(I)$ and $V_j^*(I)$ the spaces of restrictions to I of elements of $V_j(\mathbb{R})$ and of $V_j^*(\mathbb{R})$ where $(V_j(\mathbb{R}), V_j^*(\mathbb{R}))$ the biorthogonal multiresolution analysis of $L^2(\mathbb{R})$ described in the previous sections. The functions $\varphi_{j,k}|_I, 0\leqslant k\leqslant 2^j$, form a Riesz basis of $V_j(I)$ and the functions $\varphi^*_{j,k}|_I$, $-1\leqslant k\leqslant 2^j+1$, form a Riesz basis of $V_j^*(I)$. However, $V_j(I)$ and $V_j^*(I)$ do not have the same dimension because the functions φ and φ^* do not have the same support (see (4.3.6) and (4.3.7) [7,14]). Then, the spaces $V_j(I)$ and $V_j^*(I)$ are not in duality. For this second point, we have the following results for the scalar product of $\varphi_{j,k}|_I$ and $\varphi^*_{j,\ell}|_I$ for $j\geqslant 2$:

- If $k\in\{1,\dots,2^j-1\}$, we have $\operatorname{supp}\varphi_{j,k}\subset I$ and then $\langle\varphi_{j,k}|_I|\varphi^*_{j,\ell}|_I\rangle=\delta_{k,\ell}$.
- For $k=0$, we have $\langle\varphi_{j,0}|_I|\varphi^*_{j,\ell}|_I\rangle=0$ if $\ell\geqslant 2$ because $\operatorname{supp}\varphi^*_{j,\ell}\subset [0,+\infty[$.
- For $\ell\in\{-1,0,1\}$, we have

$$\langle\varphi_{j,0}|_I\big|\varphi^*_{j,\ell}\big|_I\rangle = \langle\varphi|_{[0,+\infty[}|\varphi^*(x-\ell)\rangle,$$
$$\langle\varphi|_{[0,+\infty[}|\varphi^*\rangle=\frac{1}{2},$$
$$\langle\varphi|_{[0,+\infty[}|\varphi^*(x+1)+\varphi^*(x)+\varphi^*(x-1)\rangle=\int_0^1(1-x)x\,dx=\frac{1}{6},$$

then

$$\langle\varphi|_{[0,+\infty[}|\varphi^*(x+1)\rangle=-\frac{1}{2}\quad\text{and}\quad\langle\varphi|_{[0,+\infty[}|\varphi^*(x-1)\rangle=\frac{1}{12}.$$

We define

$$V_j^T(I)=\{f\in V_j(I)\mid f|_T=0\},$$

where $T \subset \{0,1\}$ and $j \geqslant 2$. It is clear that

$$V_j^T(I) \subset V_{j+1}^T(I).$$

We shall construct a subspace $V_j^{*T}(I)$ of $V_j^*(I)$ such that $V_j^{*T}(I) \subset V_{j+1}^{*T}(I)$ and $V_j^T(I)$ and $V_j^{*T}(I)$ are in duality for the scalar product on I. To realize this object, we define $V_j^{*T}(I)$ as the space of restrictions to I of elements of $V_j^*(\mathbb{R})$ such that their restriction to $]-\infty,0]$ is a polynomial of degree $\leqslant d(0)$ and their restriction to $[1,+\infty[$ is a polynomial of degree $\leqslant d(1)$ where $d(t)=0$ if $t \in T$, $d(t)=1$ if $\ell \notin T$. We have immediately

$$V_j^{*T}(I) \subset V_{j+1}^{*T}(I).$$

We denote by $D_j^T = \{\frac{k}{2^j} \mid 0 \leqslant k \leqslant 2^j, \frac{k}{2^j} \notin T\}$. Then, the functions $\varphi_{j,k}|_I$, $\frac{k}{2^j} \in D_j^T$, form a Riesz basis of $V_j^T(I)$. A basis of $V_j^{*T}(I)$ is given by the functions $\varphi_{j,k}^*$, $2 \leqslant k \leqslant 2^j-2$ which are extended by 0 on the complement of I, the functions $(\varphi_{j,-1}^* + \varphi_{j,0}^* + \varphi_{j,1}^*)|_I$ and $(\varphi_{j,2^j-1}^* + \varphi_{2,2^j+1}^*)|_I$ (the first one is extended by $2^{j/2}$ on $]-\infty,0]$ and by 0 on $[1,+\infty[$ and the second is extended by 0 on $]-\infty,0]$ and by $2^{j/2}$ on $[1,+\infty[$) and the functions

$$\left(-\varphi_{j,-1}^* + \varphi_{j,1}^*\right)\big|_I \quad \text{if } 0 \notin T$$

and

$$\left(\varphi_{j,2^j-1}^* - \varphi_{j,2^j+1}^*\right)\big|_I \quad \text{if } 1 \notin T$$

(the first one is extended by $2^{3j/2}x$ and 0 and the second is extended by 0 and $2^{3j/2}(1-x)$). By using Gram–Schmidt orthogonalization, we obtain biorthogonal bases of $V_j^T(I)$ and $V_j^{*T}(I)$. More precisely, we have the following result.

Lemma 4.1. *The spaces $V_j^T(I)$ and $V_j^{*T}(I)$ have dual bases $(\varphi_{j,d}^T)$ and $(\varphi_{(j,d)}^{*T})$, $d \in D_j^T$, where*

$$\varphi_{(j,d)}^T = 2^{j/2}\varphi_{[j,d]}^T\big(2^j(x-d)\big) \quad \textit{and} \quad \varphi_{(j,d)}^{*T} = 2^{j/2}\varphi_{(j,d)}^{*T} = 2^{j/2}\varphi_{[j,d]}^{*T}\big(2^j(x-d)\big),$$

and:

- *for* $0<d<1$, $\varphi_{[j,d]}^T = \varphi$;
- *if* $0 \notin T$, $\varphi_{[j,0]}^T = (\varphi(x) - \frac{1}{2}\varphi(x-1))|_{[0,+\infty[} = \beta(x)$;
- *if* $1 \notin T$, $\varphi_{[j,1]}^T = (\varphi(x) - \frac{1}{2}\varphi(x+1))|_{]-\infty,0[} = \beta(-x)$;
- *for* $2 \leqslant 2^j d \leqslant 2^j-2$, $\varphi_{[j,d]}^{*T} = \varphi^*$;
- $\varphi_{[j,\frac{1}{2^j}]}^{*T} = (\varphi^*(x+2) + \varphi^*(x+1) + \varphi^*(x))|_{[-1,+\infty[} = \alpha^*(x)$;
- *if* $0 \notin T$, $\varphi_{[j,1]}^{*T} = (3\varphi^*(x) + 6\varphi^*(x-1))|_{]-\infty,0[} = \beta^*(-x)$.

We conclude that the projector P_j^T from $L^2(I)$ into V_j^T parallel to $V_j^{*T}(I)^{\perp}$ is given by

$$P_j^T f = \sum_{d \in D_j^T} \langle f | \varphi^{*T}_{(j,d)} \rangle^T_{(j,d)}, \tag{4.1}$$

and satisfies $P_j^T \circ P_{j+1}^T = P_{j+1}^T \circ P_j^T = P_j^T$. The projector on $W_j^T(I) = V_{j+1}^T(I) \cap V_j^{*T}(I)^{\perp}$ parallel to $W_j^{*T}(I) = V_{j+1}^{*T}(I) \cap V_j^T(I)^{\perp}$ is given by $Q_j^T = P_{j+1}^T - P_j^T$.

We will now construct bases of $W_j^T(I)$ and $W_j^{*T}(I)$. We remark first that $\dim W_j^T(I) = \dim W_j^{*T}(I) = 2^j$. We denote by $\Delta_j^T = \{d \in D_{j+1}^T \mid d \notin D_j^T\}$. The space $W_j^T(I)$ contains the functions $\psi_{j,k}$, $1 \leqslant k \leqslant 2^j - 2$, and the space $W_j^{*T}(I)$ contains the functions $\psi^*_{j,k}$, $1 \leqslant k \leqslant 2^j - 2$. We have $2^j - 2$ functions in $W_j^T(I)$ (the same number in $W_j^{*T}(I)$). Then, we must construct two functions in every space. We denote by $A_j^T(I) = V_j^T(I) \oplus \mathit{Vect}\{\psi_{j,k} \mid 1 \leqslant k \leqslant 2^j - 2\}$. We see that, for $3 \leqslant k \leqslant 2^{j+1} - 3$, $\varphi_{j+1,k} \notin A_j^T(I)$, $\varphi_{j+1,0}|_{[0,1]} + \frac{1}{2}\varphi_{j+1,1} \in V_j^T([0,1])$ if $0 \notin T$ and $\varphi_{j+1,1} + 2\varphi_{j+1,2} + \varphi_{j+1,2,3} \in V_j^T([0,1])$. We have the same treatment for $\varphi_{j+1,2^{j+1}}|_{[0,1]}$ and $\varphi_{j+1,2^{j+1}-1}$. We conclude that $\varphi_{j+1,2}$ and $\varphi_{j+1,2^{j+1}-2}$ form a generating system of a supplement of $A_j^T(I)$ in $V_{j+1}^T(I)$. We have the same result for $4 \leqslant k \leqslant 2^{j+1} - 4$. Then, $\varphi^*_{j+1,k} \in A_j^{*T}(I)$ and we have

$$\begin{gathered}
\left(\varphi^*_{j+1,-1} + \varphi^*_{j+1,0} + \varphi^*_{j+1,1}\right)\big|_{[0,1]} + \varphi^*_{j+1,2} + \varphi^*_{j+1,2} + \varphi^*_{j+1,3} \in A_j^{*T}(I), \\
\left(\varphi^*_{j+1,-1} + \varphi^*_{j+1,11}\right)\big|_{[0,1]} + 2\varphi^*_{j+1,2} + 3\varphi^*_{j+1,2} + 3\varphi^*_{j+1,2} \in A_j^{*T}(I),
\end{gathered}$$

and finally

$$\varphi^*_{j+1,3} - \frac{1}{2}\varphi^*_{j+1,2} \in A_j^{*T}(I).$$

We conclude that $\varphi^*_{j+1,2}$ and $\varphi^*_{j+1,2,2^{j+1}-2}$ form a generating system of a supplement of $A_j^{*T}(I)$ in $V_{j+1}^{*T}(I)$. By using Gram–Schmidt orthogonalization, we obtain the following result.

Lemma 4.2. *The spaces $W_j^T(I)$ and $W_j^{*T}(I)$ have dual bases $\psi^T_{(j,d)}$ and $\psi^{*T}_{(j,d)}$, $d \in \Delta_j^T$, where $\psi^T_{(j,d)} = 2^{j/2}\psi^T_{[j,d]}(2^j(x-d))$, $\psi^{*T}_{(j,d)} = 2^{j/2}\psi^{*,T}_{[j,d]}(2^j(x-d))$ and*:

- *for* $3 \leqslant 2^{j+1}d \leqslant 2^{j+1} - 3$, $\psi^T_{[j,d]} = \psi(x + \frac{1}{2})$;
- $$\psi^T_{[j,\frac{1}{2^{j+1}}]} = \begin{cases} \frac{1}{2}\varphi(x+\frac{1}{2}) + \psi(x+\frac{1}{2}) - \frac{1}{2}\varphi(x-\frac{1}{2}) = \gamma(x) & \textit{if } 0 \in T, \\ (-\varphi(x+\frac{1}{2}) + \psi(x+\frac{1}{2}) + \frac{1}{4}\varphi(x-\frac{1}{2}))_{[-\frac{1}{2},+\infty]} = \delta(x) & \textit{if } 0 \notin T; \end{cases}$$
- $$\psi^T_{[j,1-\frac{1}{2^{j+1}}]} = \begin{cases} -\frac{1}{2}\varphi(x+\frac{1}{2}) + \psi(x+\frac{1}{2}) + \frac{1}{2}\varphi(x-\frac{1}{2}) = \gamma(-x) & \textit{if } 1 \in T, \\ (\frac{1}{4}\varphi(x+\frac{1}{2}) + \psi(x+\frac{1}{2}) - \varphi(x-\frac{1}{2}))|_{]-\infty,\frac{1}{2}[} = \delta(-x) & \textit{if } 1 \notin T; \end{cases}$$
- *for* $3 \leqslant 2^{j+1}d \leqslant 2^{j+1} - 3$, $\psi^{*T}_{[j,d]} = \psi^{*T}_{[j,d]} = \psi^*(x+\frac{1}{2})$;
- $$\psi^{*T}_{[j,\frac{1}{2^{j+1}}]} = \begin{cases} (\frac{1}{2}\psi^*(x+\frac{1}{2}) + \varphi^*(x+\frac{3}{2}) + \varphi^*(x+\frac{1}{2}))|_{[-\frac{1}{2},+\infty[} = \gamma^*(x) & \textit{if } 0 \in T, \\ (2\psi^*(x+\frac{1}{2}) - 8\varphi^*(x+\frac{1}{2}) + 2\varphi^*(x+\frac{1}{2}))|_{[-\frac{1}{2},+\infty[} = \delta^*(x) & \textit{if } 0 \notin T; \end{cases}$$

- $\psi^{*T}_{[j,1-\frac{1}{2^{j+1}}]} = \begin{cases} (\frac{1}{2}\psi^*(x+\frac{1}{2}) + \varphi^*(x-\frac{1}{2}) + \varphi^*(x-\frac{3}{2}))|_{]-\infty,\frac{1}{2}]} = \gamma(-x) & \textit{if } 1 \in T, \\ (2\psi^*(x+\frac{1}{2}) + 2\varphi^*(x+\frac{1}{2}) - 8\varphi^*(x+\frac{1}{2}))|_{[-\frac{1}{2},+\infty[} = \delta^*(-x) & \textit{if } 1 \notin T. \end{cases}$

We have $\langle \psi^{*T}_{(j,d)} | \psi^{T}_{(j,d')} \rangle = \delta_{d,d'}$. *Then, the projector* Q^T_j *is given by*

$$Q^T_j f = \sum_{d \in \Delta^T_j} \langle f \mid \psi^{*T}_{(j,d)} \rangle \psi^T_{(j,d)}. \tag{4.2}$$

Proposition 4.1. *Let* $j_0 \geqslant 2$. *Then, we have*:

(i) *the norms* $\|f\|_2$ *and* $\|P^T_{j_0} f\|_2 + (\sum_{j \geqslant j_0} \|Q^T_j f\|_2^2)^{1/2}$ *are equivalent on* $L^2(I)$;
(ii) *the norms* $\|f\|_{H^1}$ *and* $\|P^T_{j_0} f\|_2 + (\sum_{j \geqslant j_0} 4^j \|Q^T_j f\|_2^2)^{1/2}$ *are equivalent on* $H^{1,T}(I)$.

Proof. We recall that $\chi^* = \int_x^{+\infty} \psi^*(u)\,du$ (see (3.32)). We denote:

$$\Gamma^*(x) = \int_x^{+\infty} \gamma^*(u)\,du \quad \text{for } x \geqslant -\frac{1}{2}$$

and

$$\Delta^*(x) = \int_x^{+\infty} \delta^*(u)\,du \quad \text{for } x \geqslant -\frac{1}{2}.$$

Then, we have $\Delta^*(-\frac{1}{2}) = 0$ and $\Gamma^*(-\frac{1}{2}) \neq 0$. We obtain, for $f \in H^1(I)$:

$$\begin{aligned}
\langle f | \psi^{*T}_{(j,d)} \rangle &= \left\langle \frac{df}{dx} \middle| 2^{j/2} \chi^*\left(2^j(x-d) + \frac{1}{2}\right) \right\rangle 2^{-j} \quad \text{if } 3 \leqslant 2^{j+1} d \leqslant 2^{j+1} - 3, \\
\langle f | \psi^{*T}_{(j,\frac{1}{2^{j+1}})} \rangle &= -\Gamma\left(-\frac{1}{2}\right) f(0) 2^{-j} + \left\langle \frac{df}{dx} \middle| 2^{j/2} \Gamma^*\left(2^j x - \frac{1}{2}\right) \right\rangle 2^{-j} \quad \text{if } 0 \in T, \\
&= \left\langle \frac{df}{dx} \middle| 2^{j/2} \Delta^*\left(2^j x - \frac{1}{2}\right) \right\rangle 2^{-j} \quad \text{if } 0 \notin T; \\
\langle f | \psi^{*T}_{(j,1-\frac{1}{2^{j+1}})} \rangle &= -\Gamma\left(-\frac{1}{2}\right) f(1) 2^{-j} + \left\langle \frac{df}{dx} \middle| 2^{j/2} \Gamma^*\left(2^j(1-x) - \frac{1}{2}\right) \right\rangle 2^{-j} \quad \text{if } 1 \in T, \\
&= \left\langle \frac{df}{dx} \middle| 2^{j/2} \Delta^*\left(2^j(1-x) - \frac{1}{2}\right) \right\rangle 2^{-j} \quad \text{if } 1 \notin T.
\end{aligned}$$

If $f \in H^{1,T}(I) = \{f \in H^1(I) \mid f|_T = 0\}$, then $\Gamma(-\frac{1}{2}) f(t)$, $t \in T$, are equal to zero and we obtain

$$Q^T_j f = 2^{-j} \tilde{Q}^T_j \left(\frac{df}{dx}\right),$$

where

$$\tilde{Q}_j^T u = \sum_{d \in \Delta_j^T} \langle f | \chi_{(j,d)}^{*T} \rangle \psi_{(j,d)}^T.$$

We extend now the functions $\beta, \gamma, \delta, \alpha^*, \beta^*, \gamma^*, \delta^*, \Delta^*$ and Γ^* in $H^{1+\varepsilon}(\mathbb{R})$ or $H^{\varepsilon}(\mathbb{R})$ such that these functions have integrals equal to zero and have compact support. Finally, and by using Lemma 2.4, we obtain the norm equivalences. □

5. Multiresolution analyses on the cube

We shall use multiresolution analyses constructed on the previous section (on the interval) to construct those on the cube $Q = [0,1]^n = I^n$. Using Definition 2.3 and classical method in wavelet theory, we define $V_j(I^n)$ as the space of continuous functions on I^n such that their restriction to every dyadic cube $Q_{j,(k_1,k_2,\ldots,k_n)} = [\frac{k_1}{2^j}, \frac{k_1+1}{2^j}] \times \cdots \times [\frac{k_n}{2^j}, \frac{k_n+1}{2^j}]$ is a polynomial of the form $(a_1 + b_1 x_1) \times \cdots \times (a_n + b_n x_n)$. We have $V_j(I^n) = V_j(I) \otimes V_j(I) \otimes \cdots \otimes V_j(I)$ and a basis of $V_j(I^n)$ is given by the functions $\varphi_{(j,k_1,k_2,\ldots,k_n)} = (\varphi_{j,k_2})|_I \otimes (\varphi_{j,k_2})|_I \otimes \cdots \otimes (\varphi_{j,k_2})|_I$ $(0 \leqslant k_1 \leqslant 2^j, \ldots, 0 \leqslant k_n \leqslant 2^j)$. We have the following result.

Proposition 5.1. *Let $j_0 \geqslant 2$. Then, we have*

(i) $V_j(I^n) \subset V_{j+1}(I^n) \subset H^1(I^n)$.
(ii) $\forall j \geqslant 0$, $\forall f \in V_j(I^n)$, $\|f\|_{H^1(I^n)} \leqslant C 2^j \|f\|_{L^2(I^n)}$, *where C is a positive constant.*
(iii) *If $f_j \in V_j(I^n)$, $\|\sum_{j=0}^{\infty} f_j\|_{H^1(I^n)} \leqslant C(\sum_{j=0}^{\infty} 4^j \|f_j\|_{L^2(I^n)}^2)^{1/2}$, where C is a positive constant.*
(iv) *For $f \in H^2(I^n)$, there exists a positive constant C such that*:

$$\left\| f - \sum_{k_1=0}^{2^j} \cdots \sum_{k_n=0}^{2^j} 2^{-j} f\left(\frac{k_1}{2^j}, \ldots, \frac{k_n}{2^j}\right) \varphi_{(j,k_2,\ldots,k_n)} \right\|_2 \leqslant C\, 2^{-2j} \left(\sum_{i,j=1}^{n} \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|_2 \right), \tag{5.1}$$

$$\left\| f - \sum_{k_1=0}^{2^j} \cdots \sum_{k_n=0}^{2^j} 2^{-j} f\left(\frac{k_1}{2^j}, \ldots, \frac{k_n}{2^j}\right) \varphi_{(j,k_1,\ldots,k_n)} \right\|_{H^1(I^n)} \leqslant C 2^{-j} \left(\sum_{i,j=1}^{n} \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|_2 \right). \tag{5.2}$$

The $\bigcup_j V_j(I^n)$ is dense in $H^1(I^n)$ and $\bigcup_j (V_j(I^n) \cap H_0^1(I^n))$ is dense in $H_0^1(I^n)$.

Proof. The points (i) and (ii) are immediate consequences of the definition of $V_j(I)$ and Proposition 4.1. To prove (iii), we extend the function $f_j = \sum \ldots \sum 2^{-j} f_j(\frac{k_1}{2^j}, \ldots, \frac{k_n}{2^j})\varphi_{(j,k_1,\ldots,k_n)}$ by extending $\varphi_{(j,k_1,\ldots,k_n)}$ in $\varphi_{j,k_1} \otimes \cdots \otimes \varphi_{j,k_n}$. The obtained function $\bar{f}_j$ satisfies

$$\|\bar{f}_j\|_{L^2(\mathbb{R}^n)} \leqslant C\|f_j\|_{L^2(I^n)},$$

where C is a positive constant. Then, we have

$$\left\| \sum_0^\infty \bar{f}_i \right\|_{L^2} \leqslant \left(\sum 4^{-j} \right)^{1/2} \left(\sum 4^j \|\bar{f}_j\|_2^2 \right)^{1/2}.$$

To estimate $\|\frac{\partial}{\partial x} \sum_{j=0}^\infty \bar{f}_j\|_2$, we apply Lemma 2.4 to the function $\frac{\partial \varphi}{\partial x} \otimes \cdots \otimes \varphi$, and we have

$$\left\| \frac{\partial}{\partial x} \sum_j f_j \right\|_2 = \left\| \sum_j \sum_{k_1} \cdots \sum_{k_n} f\left(\frac{k_1}{2^j}, \ldots, \frac{k_n}{2^j}\right) \left(\frac{\partial \varphi}{\partial x} \otimes \cdots \otimes \varphi \right)_{j,k_1,\ldots,k_n} \right\|_2.$$

There exists a positive constant C such that

$$\left\| \frac{\partial}{\partial x} \sum_j f_j \right\|_2 \leqslant C \left(\sum_j \sum_{k_1} \cdots \sum_{k_n} \left| f\left(\frac{k_1}{2^j}, \ldots, \frac{k_n}{2^j}\right) \right|^2 \right)^{1/2}.$$

Thus

$$\left\| \frac{\partial}{\partial x} \sum_j f_j \right\|_2 \leqslant C' \left(\sum_j 4^j \|f_j\|_2^2 \right)^{1/2},$$

where C' is a positive constant. Then $\sum_j \bar{f}_j \in H^1(\mathbb{R}^n)$ and $\sum_j f_j \in H^1(I^n)$. To prove (5.1) and (5.2), we study first the case $j = 0$. The application $f \to If$ where If is the polynomial $(a_1 + b_1 x_1) \times \cdots \times (a_n + b_n x_n)$ equal to f on the points $(\varepsilon_1, \ldots, \varepsilon_n)$, $\varepsilon_i \in \{0, 1\}$, is continuous from $H^2(I^n)$ in $H^1(I^n)$. Then $f - If$ vanishes on affine polynomials and there exists a positive constant such that

$$\|f - If\|_{H^1(I^n)} \leqslant C \inf_{a_i} \|f - a_0 - a_1 x_1 - \cdots - a_n x_n\|_{H^2(I^n)}.$$

Thus

$$\|f - If\|_{H^1(I^n)} \leqslant C' \left(\sum_{i,j=1}^n \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|_2 \right),$$

where C' is a positive constant.

In the general case, we consider $A_{(j,k_1,\ldots,k_n)} : I^n \to Q_{j,(k_1,\ldots,k_n)}$, given by $(x_1, \ldots, x_n) \to (\frac{k_1 + x_1}{2^j}, \ldots, \frac{k_n + x_n}{2^j})$. For $\varepsilon = 0$ or 1, we have

$$\begin{aligned}
&\left\| f-\sum_{k_1}\sum_{k_2}2^{-j}f\left(\frac{k_1}{2^j},\dots,\frac{k_n}{2^j}\right)\varphi_{(j,k_1,k_2)}\right\|^2_{H^\varepsilon(\mathbb{R}^n)}\\
&\quad=\sum_{k_1}\sum_{k_2}\left\|f-I\left(f\circ A_{(j,k_1,\dots,k_n)}\circ A^{-1}_{(j,k_1,\dots,k_n)}\right)\right\|^2_{H^\varepsilon(Q_j,k_1,\dots,k_n)}\\
&\quad\leqslant 2^{\varepsilon j}2^{-nj}\sum_{k_1}\sum_{k_2}\|f\circ A_{(j,k_1,\dots,k_n)}\|^2_{H^\varepsilon(I^n)}\\
&\quad\leqslant C'2^{\varepsilon j}2^{-nj}\sum_{k_1,\dots,k_n}\sum_{|\alpha|=2}\left\|\frac{\partial^2}{\partial x_1^\alpha}(f\circ A_{(j,k_1,\dots,k_n)})\right\|^2_{L^2(I^n)}\\
&\quad\leqslant C''2^{\varepsilon j}2^{-nj}\sum_{k_1}\sum_{k_2}\sum_{|\alpha|=2}\left\|\frac{\partial^2 f}{\partial x_1^\alpha}\sum|\alpha|=2\right\|\frac{\partial^2 f}{\partial x^\alpha}\right\|^2_{L^2(I^n_j,k_1,\dots,k_n)},
\end{aligned}$$

where C', C'' positive constants. Proposition 5.1 is then proved. □

We study now the spaces $H^{1,Z}(I^n)=\{f\in H^1(I^n)\mid f|_Z=0\}$ where Z is a union of faces of ∂I^n. We have $H^{1,Z}(I^n)=H^{1,T_1}(I)\mathbin{\hat\otimes}H^{1,T_2}(I)\mathbin{\hat\otimes}\cdots\mathbin{\hat\otimes}H^{1,T_n}(I)$ ($T_i, i=1,2,\dots,n$, are defined in Section 4) and then we define biorthogonal multiresolution analyses on I^n by

$$V_j^Z(I^n)=V_j(I^n)\cap H^{1,Z}(I^n)=V_j^{T_1}(I)\otimes V_j^{T_2}(I)\otimes\cdots\otimes V_j^{T_n}(I),$$

and

$$V_j^{*Z}(I^n)=V_j^{*T_1}(I)\otimes V_j^{*T_2}(I)\otimes\cdots\otimes V_j^{*T_n}(I).$$

The spaces $V_j^Z(I^n)$ and $V_j^{*Z}(I^n)$ are in duality for the scalar product on I^n and the projector P_j^Z from $L^2(I^n)$ into $V_j^Z(I^n)$ parallel to $V_j^{*Z}(I^n)^\perp$ is given by

$$Q_j^Z=\sum_\varepsilon\theta_j^{T_1,\varepsilon_1}\otimes\theta_j^{T_2,\varepsilon_2}\otimes\cdots\otimes\theta_j^{T_n,\varepsilon_n},$$

where $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)\in\{0,1\}^n\mid\{(0,\dots,0)\}$ and $\theta_j^{T_i,0}=P_j^{T_i}$ and $\theta_j^{T_i,1}=Q_j^{T_i}$.

To describe bases of $V_j^Z(I^n)$, $V_j^{*Z}(I^n)$, $W_j^Z(I^n)=\operatorname{Im}Q_j^Z$ and $W_j^{*Z}(I^n)=(\ker Q_j^Z)^\perp$, we denote by $D_j^n=D_j\times\cdots\times D_j$ the set of dyadic points of I^n, $D_j^Z=\{d\in D_j^n\mid d\notin Z\}$, $\Delta_j^Z=\{d\in D_{j+1}^Z\mid d\notin D_j^Z\}$. The dual bases of V_j^Z and V_j^{*Z} are given by $\varphi^Z_{(j,(d_1,\dots,d_n))}=\varphi^{T_1}_{(j,d_1)}\otimes\cdots\otimes\varphi^{T_n}_{(j,d_n)}$ and $\varphi^{*Z}_{(j,(d_1,\dots,d_n))}=\varphi^{*T_1}_{(j,d_1)}\otimes\cdots\otimes\varphi^{*T_n}_{(j,d_n)}$ for $(d_1,\dots,d_n)\in D_j^Z$. In the same way, dual bases of W_j^Z and W_j^{*Z} are given by $\psi^Z_{(j,(d_1,\dots,d_n))}$ and $\psi^{*Z}_{(j,(d_1,\dots,d_n))}$ for $(d_1,\dots,d_n)\in\Delta_j^Z$ where

$$\psi^{Z,(\varepsilon_1,\dots,\varepsilon_n)}_{(j,(d_1,\dots,d_n))}=\eta^{\varepsilon_1}_{(j,d_1)}\otimes\cdots\otimes\eta^{\varepsilon_n}_{(j,d_n)},\tag{5.2.3}$$

where $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)\in\{0,1\}^n\mid\{(0,\dots,0)\}$ and $\eta^0_{(j,d_i)}=\varphi^{T_i}_{(j,d_2)}$ if $2^jd_i\in\mathbb{Z}$ and $\eta^1_{(j,d_i)}=\psi^{T_i}_{(j,d_2)}$ for other case. We have the same result for $\psi^{*Z}_{(j,(d_1,d_2))}$. By using again Lemma 2.4 and classical wavelet theory, we obtain the following result.

Proposition 5.2. *Let* $j_0 \geqslant 2$. *Then, we have*:

(i) *the norms* $\|f\|_2$ *and* $\|P^Z_{j_0} f\|_2 + (\sum_{j\geqslant j_0} \|Q^Z_j f\|_2^2)^{1/2}$ *are equivalent in* $L^2(I^n)$;
(ii) *the norms* $\|f\|_{H^1}$ *and* $\|P^Z_{j_0} f\|_2 + (\sum_{j\geqslant j_0} 4^j \|Q^Z_j f\|_2^2)^{1/2}$ *are equivalent in* $H^{1,Z}(I^n)$.

6. Multiresolution analyses on a manifold

We recall that we denote a n-dimensional compact manifold by M and we assume that the Dirichlet boundary ∂M is piecewise C^1. In this section, we define and study the space $V_j(M)$ and we prove that this space allows to approximate the functions of $H^1(M)$ by simple functions. Next, we construct on M biorthogonal wavelet bases which are generated by a finite number of basic functions and adapted to the study of the Sobolev spaces $H^1(M)$ and $H^1_0(M)$.

We define $V_j(M)$ as the space of j-splines on M. We denote by $\bar\varphi_{j,m}$, $m \in M_j$ (see Definition 2.3), the function defined as follows: if $m \in Q_i$, $\bar\varphi_{j,m}|_{Q_i} = \varphi_{(j,2^j\Phi_i(m))}$ (where $\varphi_{(j,k_1,\dots,k_n)} = \varphi_{j,k_1}|_I \otimes \cdots \otimes \varphi_{j,k_n}|_I$) and if $m \notin Q_i$, $\bar\varphi_{j,m}|_{Q_i} = 0$. We can then describe a basis of $V_j(M)$.

Lemma 6.1. *The functions* $\bar\varphi_{j,m}, m \in M_j$, *form a basis of* $V_j(M)$. *More precisely, we have* $\bar\varphi_{j,m}(m') = \delta_{m,m'}$ *for* $m, m' \in M_j$ *and, for* $f \in V_j(M)$, $f = \sum_{m\in M_j} 2^{-j} f(m)\bar\varphi_{j,m}$.

Proof. To prove that $\bar\varphi_{j,m} \in V_j(M)$, it is enough to prove that $\bar\varphi_{j,m}$ is continuous. The function $\bar\varphi_{j,m}|_{Q_i}$ is continuous for all i, it is enough to prove

$$(\bar\varphi_{j,m}|_{Q_i})|_{Q_i\cap Q_\ell} = (\bar\varphi_{j,m}|_{Q_\ell})|_{Q_i\cap Q_\ell}.$$

If $Q_i \cap Q_\ell = \emptyset$, which is true. If $Q_i \cap Q_\ell$ is a line segment, we obtain $\bar\varphi_{j,m}|_{Q_i}(m) = \bar\varphi_{j,m}|_{Q_\ell}(m)$=$2^j$. Finally if $Q_i \cap Q_\ell$ is a face of Q_i and Q_ℓ, we discuss two cases:

* if $m \notin Q_i \cap Q_\ell$ then $(\bar\varphi_{j,m}|_{Q_i})|_{Q_i\cap Q_\ell} = (\bar\varphi_{j,m}|_{Q_\ell})|_{Q_i\cap Q_\ell} = 0$;
* if $m \in Q_i \cap Q_\ell$ then $(\bar\varphi_{j,m}|_{Q_i})|_{Q_i\cap Q_\ell} \circ \Phi_i^{-1}$ is a continuous function, affine on every dyadic interval of $\Phi_i(Q_i \cap Q_\ell)$, equal to 2^j on $\Phi_i(m)$ and 0 on other dyadic points.

We have the same result for $(\bar\varphi_{j,m}|_{Q_\ell})|_{Q_i\cap Q_\ell} \circ \Phi_\ell^{-1}$. We conclude that

$$(\bar\varphi_{j,m}|_{Q_i})|_{Q_i\cap Q_\ell} \circ \Phi_\ell^{-1} = (\bar\varphi_{j,m}|_{Q_i})|_{Q_i\cap Q_\ell\circ\Phi_i^{-1}} \circ \Phi_i \circ \Phi_\ell^{-1}$$

equal to $(\bar\varphi_{j,m}|_{Q_\ell})|_{Q_i\cap Q_\ell} \circ \Phi_\ell^{-1}$, and then we have $(\bar\varphi_{j,m}|_{Q_i})|_{Q_i\cap Q_\ell} = (\bar\varphi_{j,m}|_{Q_\ell})|_{Q_i\cap Q_i}$. For $f \in V_j(M)$, we consider the application

$$\tilde f = f - \sum_{m\in M_j} 2^{-j} f(m)\bar\varphi_{j,m}.$$

We have $\tilde f \in V_j(M)$ and $\tilde f(m) = 0$ where m is a dyadic point. We conclude that $\tilde f \circ \Phi_i^{-1}$ is equal to zero on every n-cube $Q_{j,k_1,k_2,\dots \tilde f\circ\Phi_i^{-1}}$ and then $\tilde f = 0$. The Lemma 6.1 is proved. □

We can now establish the first result of this section. In fact, the spaces $V_j(M)$, $j \geqslant 0$, have the following properties.

Proposition 6.1.

(i) $(V_j(M))_{j\geqslant 0}$ *is an increasing sequence of subspaces of* $H^1(M)$.
(ii) *(Bernstein inequality) There exists a constant* $C>0$ *such that*

$$\forall j\geqslant 0,\ \forall f\in V_j(M),\quad \|f\|_{H^1(M)}\leqslant C2^j\|f\|_{L^2(M)}.$$

(iii) *If* $f_j\in V_j(M)$ *and if* $\sum_{j=0}^{\infty}4^j\|f_j\|^2_{L^2(M)}<+\infty$ *then* $\sum_{j=0}^{+\infty}f_j\in H^1(M)$ *and* $\|\sum_{j=0}^{\infty}f_j\|_{H^1(M)}\leqslant C(\sum_{j=0}^{\infty}4^j\|f_j\|^2_{L^2(M)})^{1/2}$ *for a constant* $C>0$.
(iv) $U_{j\geqslant 0}V_j(M)$ *is dense in* $H^1(M)$ *and* $\bigcup_{j\geqslant 0}(V_j(M)\cap H^1_0(M))$ *is dense in* $H^1_0(M)$.

Proof. It is clear that $V_j(M)\subset V_{j+1}(M)$ and $V_j(M)\subset H^1(M)$ (because $f|_{Q_i}\in H^1(Q_i)$ if $f\in V_j(M)$ and the continuity of f gives the accord of $f|_{Q_i}$ and $f|_{Q_\ell}$ on $Q_i\cap Q_\ell$). The Decomposition lemma and Proposition 5.1 give the Bernstein inequality and the inequality $\|\sum_0^\infty f_j\|_{H^1(M)}\leqslant C(\sum_0^\infty 4^j\|f_j\|_2^2)^{1/2}$. In the same way, if $f\in H^1(M)$ and for $i=1,\dots,N$, $f|_{Q_i}\in H^2(Q_i)$, Proposition 5.1 proves that $\sum_{m\in M_j}f(m)2^{-j}\bar\varphi_{j,m}$ has a limit f in $H^1(M)$ when $j\to+\infty$. Then, the proposition is proved. □

We introduce now the following spaces:

$$V_j^{Z_i}(Q_i)=Vect\{\bar\varphi_{j,m}|_{Q_i}\mid m\in Q_i,\ m\notin Z_i\},\quad 1\leqslant i\leqslant N,$$

and the extension operators

$$E_{i,j}:V_j^{Z_i}(Q_i)\to V_j(M),\quad \bar\varphi_{j,m}|_{Q_i}\to\bar\varphi_{j,m},$$
$$E_i^*:L^2(Q_i)\to L^2(M),\quad f\to\bar f\quad \text{with } \bar f|_{Q_i}=f, \text{ and } \bar f|_{M\setminus Q_i}=0.$$

We remark that the application $f\to f\circ\Phi_i^{-1}$ is an isomorphism from $V_j^{Z_i}(Q_i)$ onto $V_j^{\Phi_i(Z_i)}(Q)$ (multiresolution analyses defined in the previous section). We consider the bases $\varphi_{(j,d)}^{\Phi_i(Z_i)}$ and $\varphi_{(j,d)}^{*\Phi_i(Z_i)}$ respectively of $V_j^{\Phi_i(Z_i)}(Q)$ and $V_j^{*\Phi_i(Z_i)}(Q)$, and the application on Q_i defined by $g\in L^2(Q)\to g\circ\Phi_i\in L^2(Q_i)$. We obtain a basis $\varphi_{j,\phi_i(m))}^{\Phi_i(Z_i)}\circ\Phi_i, m\in Q_i, m\notin Z_i$, of $V_j^{Z_i}(Q_i)$ and a basis $\varphi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)}\circ\Phi_i$ of the subspace $V_j^{*Z_i}(Q_i)$ of $L^2(Q_i)$. We denote by

$$V_j^*(M)=\bigoplus_{i=1}^N E_i^*V_j^{*Z_i}(Q_i).$$

Then, we have the following result.

Lemma 6.2.

(i) $V_j^*(M)\subset V_{j+1}^*(M)$.
(ii) *The spaces* $V_j^*(M)$ *and* $V_j(M)$ *are in duality for the scalar product on* M.

(iii) *The basis* $(\varphi_{i,m})_{m\in M_j}$ *of* $V_j(M)$ *and the dual basis* $(\varphi^*_{j,m})_{m\in M_j}$ *of* $V^*_j(M)$ *are given by*

$$\varphi_{j,m} = E_{i,j}\big(\varphi^{\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i\big) \quad \textit{for } m\in Q_i,\ m\notin Z_i, \tag{6.1}$$

$$\varphi^*_{j,m} = E^*_i\big(\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i\big) - \sum_{p\in\bigcup_{\ell<i}Q_\ell}\big\langle E^*_i\big(\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i|\varphi_{j,p}\big\rangle_M\varphi^*_{j,p}$$
$$\textit{for } m\in Q_i,\ m\in Z_i. \tag{6.2}$$

Proof. We have from the previous section $V^{*\Phi_i(Z_i)}_j(Q)\subset V^{*\Phi_i(Z_i)}_{j+1}(Q)$ then we obtain $V^*_j(M)\subset V^*_{j+1}(M)$. It is clear that $V_j(M)=\bigoplus_{i=1}^N E_{i,j}(V^{Z_i}_j(Q_i))$ and then the functions $\varphi_{j,m}$ form a basis of $V_j(M)$. The functions $\varphi^*_{j,m}$ constitute a basis of $V^*_j(M)$ because the matrix of bases $\varphi^*_{j,m}$ and $E^*_i(\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i)$ is superior triangular: $\varphi^*_{j,m}-E^*_i(\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i)\in\bigoplus_{\ell<i}E^*_\ell V^{*Z_\ell}_{j(Q_\ell)}$. To prove the biorthogonality, we see that if $m\in Q_i$, $m\notin Z_i$, then $\langle\varphi^*_{j,m}|\varphi_{j,p}\rangle=\delta_{p,m}$. In fact, we have three possible cases for $p\in Q_\ell$, $p\notin Z_\ell$:

- if $\ell>i$ then $\operatorname{supp}\varphi^*_{j,m}\subset\bigcup_{n\leqslant i}Q_n$ and $\operatorname{supp}\varphi_{j,p}\subset\bigcup_{n\geqslant\ell}Q_n$ and we have

$$\big\langle\varphi^*_{j,m}\big|\varphi_{j,p}\big\rangle_M=0;$$

- if $\ell<i$ then we have

$$\big\langle\varphi^*_{j,m}\big|\varphi_{j,p}\big\rangle=\big\langle E^*_i\big(\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i\big)\big|\phi_{j,p}\big\rangle-\sum_{r\in\bigcup_{n<i}Q_n}\big\langle E^*_i\big(\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i\big)\big|\varphi_{j,r}\big\rangle\delta_{r,p}=0;$$

- if $\ell=i$ then we have

$$\operatorname{supp}\big\{\varphi^*_{j,m}-\big(\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\circ\Phi_i\big)\big\}\subset\bigcup_{n<i}Q_n,$$
$$\operatorname{supp}\big\{\varphi_{j,p}-\big(\varphi^{\Phi_i(Z_i)}_{(j,\Phi_i(p))}\circ\Phi_i\big)\big\}\subset\bigcup_{n>i}Q_n,$$

and thus

$$\big\langle\varphi^*_{j,m}\big|\varphi_{j,p}\big\rangle_M=\iint_Q\varphi^{*\Phi_i(Z_i)}_{(j,\Phi_m(m))}\varphi^{\Phi_i(Z_i)}_{(j,\Phi_i(p))}\,dx\,dy=\delta_{m,p}.$$

This completes the proof of Lemma 6.2. □

The preceding lemma allows that the projector P_j from $L^2(M)$ into $V_j(M)$ parallel to $V^*_j(M)^\perp$ is given by

$$P_jf=\sum_{m\in M_j}\big\langle f\big|\varphi^*_{j,m}\big\rangle_M\varphi_{j,m}.$$

We have $P_j \circ P_{j+1} = P_{j+1} \circ P_j = P_j$ and then $Q_j = P_{j+1} - P_j$ is the projector from $L^2(M)$ into $W_j(M) = V_{j+1}(M) \cap V_j^*(M)^\perp$ parallel to $(W_j^*(M))^\perp = (V_{j+1}^*(M) \cap V_j(M)^\perp)^\perp$. The following result describes bases of $W_j(M)$ and $W_j^*(M)$.

Lemma 6.3. *We have a basis (ψ_m), $m \in M_{j+1} \backslash M_j$, of $W_j(M)$ and a dual basis (ψ_m^*) of $W_j^*(M)$ which are given for $m \in Q_i$, $m \notin Z_i$ by*

$$\psi_m = E_{i,j+1}\big(\psi_{(j,\Phi_i(m))}^{\Phi_i(Z_i)} \circ \Phi_i\big) - \sum_{p \in M_j} \big\langle E_{i,j+1}\big(\psi_{(j,\Phi_i(m))}^{\Phi_i(Z_i)} \circ \Phi_i \big| \varphi_{j,p}^*\big)\big\rangle_m \varphi_{j,p},$$

$$\psi_m^* = E_i^*\big(\psi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)} \circ \Phi_i\big) - \sum_{p \in M_j} \big\langle E_i^*\big(\psi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)} \circ \Phi_i \big| \varphi_{j,p}\big)\big\rangle_m \varphi_{j,p}^*$$
$$- \sum_{q \in \bigcup_{\ell < i} Q_\ell, q \in M_{j+1} - M_j} \big\langle E_i^*\big(\psi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)} \circ \Phi_i\big) \big| \psi_q \big\rangle_M \psi_q^*.$$

Proof. It is clear that $\psi_m \in W_j(M)$ and $\psi_m^* \in W_j^*(M)$. Moreover, we have $\dim W_j(M) = \operatorname{card} M_{j+1} - \operatorname{card} M_j = \dim W_j^*(M)$ (see Definition 2.3 of M_j). We must prove now that $\langle \psi_m^* | \psi_p \rangle = \delta_{m,p}$ for $m, p \in M_{j+1} \backslash M_j$. We follow the technique used in the proof on Lemma 6.2. We have for $m \in Q_i$, $m \notin Z_i$:

- $\operatorname{supp} E_{i,j+1}(\psi_{(j,\Phi_i(m))}^{\Phi_i(Z_i)} \circ \Phi_i) \subset \bigcup_{\ell \geqslant i} Q_\ell$, and $\langle E_{i,j+1}(\psi_{(j,\Phi_i(m))}^{\Phi_i(Z_i)} \circ \Phi_i) | \varphi_{j,p}^* \rangle_M \neq 0$ is possible only if $p \in Q_\ell$, $p \notin Z_\ell$ for $\ell \geqslant i$. Then, we obtain $\operatorname{supp} \psi_m \subset \bigcup_{\ell \geqslant i} Q_\ell$.
- $\langle E_i^*(\psi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)} \circ \Phi_i) | \varphi_{j,p} \rangle_M \neq 0$ is possible only if $p \in \bigcup_{\ell \leqslant i} Q_\ell$ and then we obtain $\operatorname{supp} \psi_m^* \subset \bigcup_{\ell \leqslant i} Q_\ell$.

We consider now ψ_p with $p \in Q_\ell$, $p \notin Z_\ell$. We have:

- If $\ell > i$, $\langle \psi_m^* | \psi_p \rangle_M = 0$ because $|\operatorname{supp} \psi_m^* \cap \operatorname{supp} \psi_p| = 0$.
- If $\ell < i$, then
$$\big\langle \psi_m^* | \psi_p \big\rangle_M = \big\langle E_i^*\big(\psi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)} \circ \Phi_i\big) \big| \psi_p \big\rangle - \sum_{q \in \bigcup_{n < i} Q_m} \big\langle E_i^*\big(\psi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)} \circ \Phi_i\big) \big| \psi_q \big\rangle \delta_{q,p} = 0.$$
- If $\ell = i$, then we have
$$\operatorname{supp}\big\{\psi_m^* - \psi * \hat{\Phi_i}(Z_i)_{(j,\Phi_i(m))} \circ \Phi_i\big\} \subset \bigcup_{n < i} Q_n,$$
$$\operatorname{supp}\big\{\psi_p - \psi_{(j,\Phi_i(m))}^{\Phi_i(Z_i)} \circ \Phi_i\big\} \subset \bigcup_{n > i} Q_n,$$

  and thus
$$\big\langle \psi_m^* \big| \psi_p \big\rangle = \iint_Q \psi_{(j,\Phi_i(m))}^{*\Phi_i(Z_i)} \psi_{(j,\Phi_i(m))}^{\Phi_i(Z_i)} \, dx\, dy = \delta_{m,p}.$$

The Lemma 6.3 is then proved. □

Remark. There is few terms different from zero in series described in Lemma 6.3 when m is near Z_i. This proves that ψ_m and ψ_m^* have compact support which can be controlled.

We shall describe now extensions which allow to construct multiresolution analyses on the manifold from those constructed on the cube.

Lemma 6.4. *The operators E_i, defined on $L^2(Q_i)$ by*

$$E_i f = \sum_{m\in M_2, m\in Q_i\setminus Z_i} \langle f|\varphi_{2,m}^*\rangle_M \varphi_{2,m} + \sum_{m\in(\bigcup M_j)\setminus M_2, m\in Q_i\setminus Z_i} \langle f|\psi_m^*\rangle_M \psi_m$$

define a family of extensions in the sense of Definition 2.2.

Proof. We have $\operatorname{supp} f = Q_i$, then we obtain

$$\langle f|\psi_m^*\rangle = \langle f\circ\Phi_i^{-1}\big|\psi^{*\Phi_i(Z_i)}_{(j,\Phi_i(m))}\rangle \quad \text{for } m\in M_{j+1}\setminus M_j$$

and

$$\sum_{m\in M_{j+1}\setminus M_j, m\in Q_i\setminus Z_i} \langle f|\psi_m^*\rangle_M \psi_m|_{Q_i} = Q_j^{\Phi_i(Z_i)}\big(f\circ\Phi_i^{-1}\big)\circ\Phi_i.$$

Proposition 5.2 gives

$$\|f\|_{L^2(Q_i)} \approx \Big(\sum_{m\in M_2, m\in Q_i\setminus Z_i} \big|\langle f|\varphi_{2,m}^*\rangle\big|^2\Big)^{1/2} + \Big(\sum_{m\in(\bigcup M_j)\setminus M_2, m\in Q_i\setminus Z_i} \big|\langle f|\psi_m^*\rangle\big|^2\Big)^{1/2},$$

$$\|f\|_{H^{1,Z_i}(Q_i)} \approx \Big(\sum_{m\in M_2, m\in Q_i\setminus Z_i} \big|\langle f|\varphi_{2,m}^*\rangle\big|^2\Big)^{1/2} + \Big(\sum_{m\in(\bigcup M_j)\setminus M_2, m\in Q_i\setminus Z_i} 4^j\big|\langle f|\psi_m^*\rangle\big|^2\Big)^{1/2}.$$

Then, we have:

- $(E_i f)|_{Q_i} = f$ because $f\circ\Phi_i^{-1} = (P_2^{\Phi_i(Z_i)} + \sum_2^\infty Q_j^{\Phi_i(Z_i)})(f\circ\Phi_i^{-1})$;
- $\operatorname{supp} E_i f \subset \bigcup_{\ell\geqslant i} Q_\ell$;
- $E_i f\in L^2(M)$ (This result is a consequence of Decomposition and Regular lemmas);
- If $f\in H^{1,Z_i}(Q_i)$, $E_i f\in H^1(M)$ because the operators $E_{i,j+1}$ satisfy for $f\in V_j^{Z_i}(Q_i)$ the relation

$$\|E_{i,j+1} f\|_{L^2(M)} \leqslant C\|f\|_{L^2(M)},$$

where C is a positive constant independent of j. Then, $E_i f = \sum_2^\infty f_j$, where $f_j\in V_j(M)$ with $\sum_{j=2}^\infty 4^j\|f_j\|_2^2 < +\infty$, thus Proposition 6.1 gives $E_i f\in H^1(M)$.

This completes the proof of the lemma. □

We establish now the main result of this paper.

Theorem 6.1. *There exists a family of functions* $(\psi_m)_{m\in\bigcup_{j\in\mathbb{N}} M_j}$ *such that*:

(i) *the functions* (ψ_m) *form an unconditional basis of* $L^2(M)$ *and* $H^1(M)$. *We denote by* ψ_m^* *the application*: $f \in L^2 \Rightarrow f = \sum_{m\in\bigcup_{j\geqslant 0} M_j} \psi_m^*(f)\psi_m$.

(ii) $\psi_m \in V_{j(m)}(M)$ *where* $j(m) = \inf\{j \mid j \geqslant 2, m \in M_j\}$.

(iii) $\operatorname{supp}\psi_m \subset \{x \in M \mid d(x,m) \leqslant C2^{-j(m)}\}$ *and* $d(\operatorname{supp} f, m) \geqslant C2^{-j(m)} \Rightarrow \psi_m^*(f) = 0$, *where C is a positive constant and* $d(x,y)$ *is the distance between the points x and y of M*.

(iv) *There exists a constant* $C \geqslant 1$ *such that, for every sequence* $(\lambda_m) \in \ell^2(\bigcup M_j)$, *we have*

$$\frac{1}{C}\biggl(\sum_{m\in\bigcup M_j} |\lambda_m|^2\biggr)^{1/2} \leqslant \biggl\|\sum_{m\in\bigcup M_j} \lambda_m\psi_m\biggr\|_{L^2(M)} \leqslant C\biggl(\sum_{m\in\bigcup M_j} |\lambda_m|^2\biggr)^{1/2}.$$

In particular, the norms $\|f\|_{L^2(M)}$ *and* $(\sum_{m\in\bigcup M_j} |\psi_m^*(f)|^2)^{1/2}$ *are equivalent on* $L^2(M)$.

(v) *There exists a constant* $C' \geqslant 1$ *such that, for every sequence* $(\lambda_m) \in \ell^2(\bigcup M_j)$, *we have*

$$\frac{1}{C'}\biggl(\sum_{m\in\bigcup M_j} 4^{j(m)}|\lambda_m|^2\biggr)^{1/2} \leqslant \biggl\|\sum_{m\in\bigcup M_j} \lambda_m\psi_m\biggr\|_{H^1(M)} \leqslant C'\biggl(\sum_{m\in\bigcup M_j} 4^{j(m)}|\lambda_m|^2\biggr)^{1/2}.$$

In particular, the norms $\|f\|_{H^1(M)}$ *and* $(\sum_{m\in\bigcup M_j} 4^{j(m)}|\psi_m^*(f)|^2)^{1/2}$ *are equivalent on* $H^1(M)$.

(vi) *The functions* ψ_m, $m \in M_j$, *form a basis of* $V_j(M)$ (*for* $j \geqslant 2$).

(vii) *The projector* $P_j f = \sum_{m\in M_j} \psi_m^*(f)\psi_m$ *is normalized by*

$$P_j f = \sum_{m\in M_j} \varphi_{j,m}^*(f)\varphi_{j,m},$$

where the functions $\varphi_{j,m}$ *and the linear forms* $\varphi_{j,m}^*$ *satisfy*

$$\text{the } \varphi_{j,m}, m \in M_j, \text{ form a basis of } V_j(M) \text{ and } \langle\varphi_{j,m}^*|\varphi_{j,m'}\rangle_M = \delta_{m,m'}, \tag{6.3}$$

$$\operatorname{supp}\varphi_{j,m} \subset \{x \in M \mid d(x,m) \leqslant C2^{-j}\} \quad \text{and}$$

$$d(\operatorname{supp} f, m) \geqslant C2^{-j} \quad \Rightarrow \quad \varphi_{j,m}^*(f) = 0, \tag{6.4}$$

there exists a constant $C'' \geqslant 1$ *such that, for every sequence* $(\lambda_m) \in \ell^2(M_j)$, *we have*

$$\frac{1}{C''}\biggl(\sum_{m\in M_j} |\lambda_m|^2\biggr)^{1/2} \leqslant \biggl\|\sum_{m\in\bigcup M_j} \lambda_m\varphi_{j,m}\biggr\|_2 \leqslant C''\biggl(\sum_{m\in M_j} |\lambda_m|^2\biggr)^{1/2}. \tag{6.5}$$

Proof. The properties (i)–(iii) are immediate consequences of lemmas described above. To prove (iv) and (v), we denote by $\psi_m = \varphi_{2,m}$ and $\psi_m^* = \varphi_{2,m}^*$, for $m \in M_2$. We obtain the inequality $\|\sum_m \lambda_m\psi_m\|_{L^2(M)} \leqslant C(\sum_m |\lambda_m|^2)^{1/2}$ by Decomposition and Vaguelette lemmas. To obtain the inverse inequality $\sum_m |\langle f|\psi_m^*\rangle_M|^2 \leqslant C\|f\|_2$, we use extensions E_i: $f_i = (f - \sum_{\ell<i} E_\ell f_\ell)|_{Q_i}$. Then, we have

$$f = \sum_1^N E_i f_i, \qquad \|f\|_2 \approx \sum \|f_i\|_2$$

and

$$\sum_m |\langle E_i f_i|\psi_m^*\rangle|_M^2 = \sum_{m\in Q_i\setminus Z_i} |\langle E_i f_i|\psi_m^*\rangle_M|^2 = \sum_{m\in Q_i\setminus Z_i} |\langle f_i|\psi_m^*\rangle_M|^2 \leqslant C\|f_i\|_{L^2(Q_i)}^2,$$

where C is a positive constant.

In the same way, the inequality $\|\sum \lambda_m \psi_m\|_{H^1} \leqslant C(\sum_n 4^{j(m)}|\lambda_m|^2)^{1/2}$ is deduced from (iv) and Proposition 6.1. The inverse inequality is obtained by using $f = \sum_1^N E_i f_i$ and $\|f\|_{H^1(M)} \approx \sum_{i=1}^N \|f_i\|_{H^1(Q_i)}$. Then, Theorem 6.1 is proved. □

Remark. We see that the functions ψ_m and $\varphi_{j,m}$ are obtained from a finite number of basic functions. We have the same result for ψ_m^* and $\varphi_{j,m}^*$.

We have the same result for the functions which vanish on the boundary of M.

Theorem 6.2. *There exists an unconditional basis* $(\tilde\psi_m)$, $m \in \bigcup_{j\in\mathbb{N}} M_j$, $m \notin \partial M$, *of* $L^2(M)$ *and* $H_0^1(M)$ *such that the applications* $\tilde\psi_m^*$ *of the points* (ii)–(vi) *of Theorem* 6.1 (*by changing in the point* (v) $H^1(M)$ *by* $H_0^1(M)$ *and in the point* (vi) $V_j(M)$ *by* $V_j(M) \cap H_0^1(M)$). *The projector* $\tilde P_j f = \sum_{m\in M_j, m\notin\partial M} \tilde\psi_m^*(f)\tilde\psi_m$ *is normalized by* $\tilde P_j f = \sum_{m\in M_j, m\notin\partial M} \tilde\varphi_{j,m}^*$ *where the functions* $\tilde\varphi_{j,m}$ *satisfy the point* (vii) *of Theorem* 6.1 (*by changing* $V_j(M)$ *by* $V_j(M) \cap H_0^1(M)$).

7. Conclusion

In this paper, we have constructed biorthogonal multiresolution analyses on a compact manifold M of dimension n by using the decomposition method introduced by Z. Ciesielski and T. Fiegel and extension and regular lemmas. The scaling functions and the associated wavelets have compact support and are obtained from a finite number of basic functions. These bases are easy to implement. Finally, these analyses are adapted to the study of the Sobolev spaces $H^1(M)$ and $H_0^1(M)$.

References

[1] C. Canuto, A. Tabacco, K. Urban, The wavelet element method, preprint, 1997.
[2] Z. Ciesielski, T. Figiel, Spline bases in classical function spaces on compact C^∞ manifolds, parts I and II, Studia Math. LXXVI (1983) 1–58, 95–136.
[3] Z. Ciesielski, T. Figiel, Spline approximation and Besov spaces on compact manifolds, Studia Math. LXXV (1982) 13–36.
[4] A. Cohen, W. Dahmen, R. De Vore, Multiscale decomposition on bounded domains, IGPM, technical report, 2003.
[5] W. Dahmen, R. Schneider, Composite wavelet bases for operator equations, Acta Numer. 9 (1997) 55–228.

[6] W. Dahmen, R. Schneider, Wavelets on manifolds, construction and domain decomposition, SIAM J. Math. Anal. 31 (1) (2000) 184–230.
[7] I. Daubechies, Orthonormal bases of wavelets with compact support, Comm. Pure Appl. Math. 42 (1988) 906–996.
[8] A. Jouini, Constructions de bases d'ondelettes sur les variétés, thesis, Orsay, 1993.
[9] A. Jouini, P.G. Lemarié-Rieusset, Ondelettes sur un ouvert borné du plan, prepublication 46, Orsay, 1992.
[10] A. Jouini, P.G. Lemarié-Rieusset, Analyses multirésolutions biorthogonales sur l'intervalle et applications, Ann. Inst. H. Poincaré Anal. Non Linéaire 10 (4) (1993) 453–476.
[11] A. Jouini, P.G. Lemarié-Rieusset, Wavelets on the L-shaped domain, prepublication 178, Evry, 2003.
[12] J.P. Kahane, P.G. Lemarié-Rieusset, Fourier Series and Wavelets, vol. 3, Gordon & Breach, 1995.
[13] R. Masson, Méthodes d'ondelettes en simulation numérique pour les problèmes elliptiques et point selle, thesis, Université Paris 6, Jan. 1999.
[14] Y. Meyer, Ondelettes sur l'intervalle, Rev. Mat. Iberoamericana 7 (1991) 115–134.
[15] W. Sweldens, The lifting scheme: A custom design construction of biorthogonal wavelets, Appl. Comput. Harmon. Anal. 3 (1996) 186–200.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 248 (2007) 152–174
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

Operator space structure on Feichtinger's Segal algebra

Nico Spronk [1]

Department of Pure Mathematics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

Received 4 August 2006; accepted 15 March 2007

Communicated by Paul Malliavin

In memory of my friend, Sean Crawford Andrew

Abstract

We extend the definition, from the class of abelian groups to a general locally compact group G, of Feichtinger's remarkable Segal algebra $\mathrm{S}_0(G)$. In order to obtain functorial properties for non-abelian groups, in particular a tensor product formula, we endow $\mathrm{S}_0(G)$ with an operator space structure. With this structure $\mathrm{S}_0(G)$ is simultaneously an *operator Segal algebra* of the Fourier algebra $\mathrm{A}(G)$, and of the group algebra $\mathrm{L}^1(G)$. We show that this operator space structure is consistent with the major functorial properties: (i) $\mathrm{S}_0(G) \hat{\otimes} \mathrm{S}_0(H) \cong \mathrm{S}_0(G \times H)$ completely isomorphically (operator projective tensor product), if H is another locally compact group; (ii) the restriction map $u \mapsto u|_H : \mathrm{S}_0(G) \to \mathrm{S}_0(H)$ is completely surjective, if H is a closed subgroup; and (iii) $\tau_N : \mathrm{S}_0(G) \to \mathrm{S}_0(G/N)$ is completely surjective, where N is a normal subgroup and $\tau_N u(sN) = \int_N u(sn)\,dn$. We also show that $\mathrm{S}_0(G)$ is an invariant for G when it is treated simultaneously as a pointwise algebra and a convolutive algebra.

Keywords: Fourier algebra; Segal algebra; Operator space

1. Introduction and notation

1.1. History

In [7], Feichtinger defined, for any abelian group G, a Segal algebra $\mathrm{S}_0(G)$ of $\mathrm{L}^1(G)$. This Segal algebra is the minimal Segal algebra in $\mathrm{L}^1(G)$ which is closed under pointwise multipli-

E-mail address: nspronk@math.uwaterloo.ca.
[1] The author's research supported by NSERC under grant No. 312515-05.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2007.03.028

cation by characters and on which multiplication by any character is an isometry. It is proved in [7], that the Fourier transform induces an isomorphism $\mathrm{S}_0(G) \cong \mathrm{S}_0(\hat{G})$ where $\hat{G}$ is the dual group. Thus, we also have that $\mathrm{S}_0(G)$ is a Segal algebra in the Fourier algebra $\mathrm{A}(G)$ ($\cong \mathrm{L}^1(\hat{G})$), i.e. a dense ideal of $\mathrm{A}(G)$ which has a norm under which it is a Banach $\mathrm{A}(G)$-module. In fact, it is the minimal Segal algebra in $\mathrm{A}(G)$ which is translation invariant and on which translations are isometries.

For a general, not necessarily abelian, locally compact group G, the Fourier algebra is defined by Eymard [5]. There are hints in [6] of how to define $\mathrm{S}_0(G)$, as a Segal algebra in $\mathrm{A}(G)$. We develop this fully. $\mathrm{S}_0(G)$ is also a Segal algebra in the classical sense, i.e. a Segal algebra of $\mathrm{L}^1(G)$. We also develop, for general locally compact groups, the functorial properties which Feichtinger proved for abelian groups [7, Theorem 7]. One of Feichtinger's results is a tensor product formula: if G and H are locally compact abelian groups, then there is a natural isomorphism $\mathrm{S}_0(G) \otimes^\gamma \mathrm{S}_0(H) \cong \mathrm{S}_0(G \times H)$ (projective tensor product). For non-abelian G and H we cannot expect that $\mathrm{A}(G) \otimes^\gamma \mathrm{A}(H) \cong \mathrm{A}(G \times H)$, in general, by [19]. The theory of operator spaces, and the associated operator tensor product, allows us to obtain a satisfactory result from [3]: $\mathrm{A}(G) \hat{\otimes} \mathrm{A}(H) \cong \mathrm{A}(G \times H)$. Thus we are motivated to find a natural operator space structure on $\mathrm{S}_0(G)$, for a general locally compact G, which allows us to recover a tensor product formula. Analogous to the fact that $\mathrm{S}_0(G)$ does not have a fixed natural norm, but rather a family of equivalent norms, we find that our operator space structure is determined only up to complete isomorphism. In order to deal with our operator space structure, we find it more natural to deal with certain "dual" type matrices $\mathrm{T}_n(\mathcal{V})$ over an operator space $\mathcal{V}$, as opposed to the usual matrices $\mathrm{M}_n(\mathcal{V})$, which forces us to summarise a coherent theory of these in Section 1.3.

To underscore the naturality of our operator space structure we examine the other two major functorial properties, restriction to a closed subgroup and averaging over a closed normal subgroup. We show that our operator space structure is natural in the sense that it gives complete surjections, onto the Feichtinger algebra of the closed subgroup in the case of restriction, and onto the Feichtinger algebra of the quotient group in the case of averaging. See Sections 3.2 and 3.4.

In Section 3.5 we discuss an isomorphism theorem, characterising $\mathrm{S}_0(G)$ as an invariant of G. This makes no use of our operator space structure. As we note, our result is not even specific to $\mathrm{S}_0(G)$, but can be applied to many spaces which are simultaneously Segal algebras in $\mathrm{L}^1(G)$ and in some regular Banach subalgebra $\mathcal{A}$ of functions on G, having Gelfand spectrum G.

The Segal algebra $\mathrm{S}_0(G)$ seems interesting in and of itself simply for its wealth of structure and functorial properties. However, in the abelian case, $\mathrm{S}_0(G)$ is a fundamental example of *Wiener amalgam spaces* and of *modulation spaces*, which appear to be of tremendous use in time–frequency analysis. See [8] and references therein. We hope our $\mathrm{S}_0(G)$, for non-abelian G, may prove as inspirational and useful.

1.2. Harmonic analysis

Let G be a locally compact group with fixed left Haar measure m. We will denote integration of a function f with respect to m variously by $\int_G f\,dm$ or $\int_G f(s)\,ds$. For any $1 \leqslant p \leqslant \infty$ we let $\mathrm{L}^p(G)$ be the usual L^p-space with respect to m. If $s \in G$ and f is a complex-valued function on G we denote the group action of left translation of s on f by $s * f(t) = f(s^{-1}t)$ for t in G. For any appropriate pair of functions f, g we denote $f * g = \int_G f(s) s * g\,ds$.

The *Fourier* and *Fourier–Stieltjes algebras*, $\mathrm{A}(G)$ and $\mathrm{B}(G)$ were defined by Eymard in [5]. We recall that $\mathrm{A}(G)$ consists of functions on G of the form $s \mapsto \langle \lambda(s) f | g \rangle = \bar{g} * \check{f}(s)$, where

$\lambda : G \to \mathcal{B}(\mathrm{L}^2(G))$ is the left regular representation where $\lambda(s)f(t) = f(s^{-1}t)$ for almost every t in G. We note that $\mathrm{A}(G)$ has Gelfand spectrum G, given by evaluation. We also remark that we have duality relations $\mathrm{A}(G)^* \cong \mathrm{VN}(G)$, where $\mathrm{VN}(G)$ is the von Neumann algebra generated by $\lambda(G)$, and $\mathrm{C}^*(G)^* \cong \mathrm{B}(G)$ where $\mathrm{C}^*(G)$ is the universal C*-algebra of G.

1.3. Operator spaces

Our main reference for operator spaces is [4]. An operator space is a complex vector space $\mathcal{V}$, equipped with a sequence of norms, one on each space of $n \times n$ matrices over $\mathcal{V}$, $\mathrm{M}_n(\mathcal{V})$, which satisfy Ruan's axioms; we call this an *operator space structure*. An operator space is complete if $\mathrm{M}_1(\mathcal{V})$ is complete, i.e. $\mathcal{V} = \mathrm{M}_1(\mathcal{V})$ is a Banach space. If $\mathcal{W}$ is another operator space and $T : \mathcal{V} \to \mathcal{W}$ is a linear map we let $T^{(n)} : \mathrm{M}_n(\mathcal{V}) \to \mathrm{M}_n(\mathcal{W})$ be the amplification given by $T^{(n)}[v_{ij}] = [Tv_{ij}]$. We say T is *completely bounded* if $\|T\|_{\mathrm{cb}} = \sup_{n\in\mathbb{N}} \|T^{(n)}\| < \infty$. Moreover we say T is a *complete contraction/isometry/quotient map* if each $T^{(n)}$ is a contraction/isometry/quotient map. The space of completely bounded linear maps from $\mathcal{V}$ to $\mathcal{W}$ is denoted $\mathcal{CB}(\mathcal{V}, \mathcal{W})$.

We make note of some basic operator space constructions. If $\mathcal{W}$ is a closed subspace of an operator space $\mathcal{V}$, then $\mathcal{W}$ inherits the operator space structure from $\mathcal{V}$. Moreover, the quotient space $\mathcal{V}/\mathcal{W}$ obtains the *quotient* operator space structure via isometric identifications $\mathrm{M}_n(\mathcal{V}/\mathcal{W}) \cong \mathrm{M}_n(\mathcal{V})/\mathrm{M}_n(\mathcal{W})$, i.e.

$$\big\|[v_{ij} + \mathcal{W}]\big\|_{\mathrm{M}_n(\mathcal{V}/\mathcal{W})} = \inf\big\{\big\|[v_{ij} + w_{ij}]\big\|_{\mathrm{M}_n(\mathcal{V})}\colon [w_{ij}] \in \mathrm{M}_n(\mathcal{W})\big\}.$$

If $\mathcal{V}$ and $\mathcal{W}$ are any two operator spaces, then the space $\mathcal{CB}(\mathcal{V}, \mathcal{W})$ obtains the *standard* operator space structure (see [1]) where we identify, for each n, the matrix $[S_{ij}]$ in $\mathrm{M}_n(\mathcal{CB}(\mathcal{V}, \mathcal{W}))$ with the operator $v \mapsto [S_{ij}v]$ in $\mathcal{CB}(\mathcal{V}, \mathrm{M}_n(\mathcal{W}))$. We will make extensive use of the *operator projective tensor product*, defined in [3]. If $\mathcal{V}$ and $\mathcal{W}$ are two complete operator spaces, let $\mathcal{V} \hat{\otimes} \mathcal{W}$ denote their operator projective tensor product. The algebraic tensor product of $\mathcal{V}$ and $\mathcal{W}$ forms a dense subspace of $\mathcal{V} \hat{\otimes} \mathcal{W}$ and we let $\mathcal{V} \otimes_\wedge \mathcal{W}$ denote this algebraic tensor product, given the operator space projective structure.

We note that for any operator space $\mathcal{V}$ that every continuous linear functional is automatically completely bounded, i.e. $\mathcal{V}^* = \mathcal{CB}(\mathcal{V}, \mathbb{C})$, and thus is an operator space with the standard operator space structure. For a locally compact group G, the space $\mathrm{B}(G) \cong \mathrm{C}^*(G)^*$ is always endowed with the standard operator space structure. $\mathrm{A}(G)$ obtains the same operator space structure as a subspace of $\mathrm{B}(G)$ as it does as the predual of $\mathrm{VN}(G)$, i.e. a subspace of the dual. $\mathrm{L}^1(G)$, as the predual of a commutative von Neumann algebra naturally admits the *maximal* operator space structure.

Let $\mathcal{A}$ be a Banach algebra, equipped with an operator space structure such that $\mathrm{M}_1(\mathcal{A}) = \mathcal{A}$ isometrically, and let $\mathcal{V}$ be a left $\mathcal{A}$-module which is also an operator space. For a in $\mathcal{A}$ we let $m_a : \mathcal{V} \to \mathcal{V}$ be the module action map, $M_a v = a \cdot v$. Then $\mathcal{V}$ is called a *completely bounded Banach $\mathcal{A}$-module* under any of the following equivalent assumptions:

(i) $\{M_a\colon a \in \mathcal{A}\} \subset \mathcal{CB}(\mathcal{V})$ and $a \mapsto M_a : \mathcal{A} \to \mathcal{CB}(\mathcal{V})$ is completely bounded,
(ii) there is $C > 0$ such that for any pair of matrices $[a_{ij}]$ in $\mathrm{M}_n(\mathcal{A})$ and $[v_{kl}]$ in $\mathrm{M}_m(\mathcal{V})$, we have $\|[a_{ij}v_{kl}]\|_{\mathrm{M}_{nm}(\mathcal{V})} \leqslant C\|[a_{ij}]\|_{\mathrm{M}_n(\mathcal{A})}\|[v_{kl}]\|_{\mathrm{M}_m(\mathcal{V})}$, and
(iii) the map $\mathcal{A} \otimes_\wedge \mathcal{V} \to \mathcal{V}$, given on elementary tensors by $a \otimes v \mapsto a \cdot v$ is completely bounded.

Similar definitions can be given with right and bimodules. We say $\mathcal{V}$ is a *completely contractive Banach $\mathcal{A}$-module* if the maps in (i) and (iii) above are complete contractions and in (ii) we can set $C = 1$. We call $\mathcal{A}$ a *completely bounded* (*contractive*) *Banach algebra*, if it is a completely bounded (contractive) module over itself. We note that $\mathrm{A}(G)$, $\mathrm{B}(G)$ and $\mathrm{L}^1(G)$ are all completely contractive Banach algebras with their standard operator space structures.

Operator Segal algebras were introduced in [9]. Let $\mathcal{A}$ be a completely contractive Banach algebra. An *operator Segal algebra* in $\mathcal{A}$ is a dense left ideal $\mathrm{S}\mathcal{A}$, equipped with a complete operator space structure such that:

(OSA1) $\mathrm{S}\mathcal{A}$ is a completely bounded Banach $\mathcal{A}$-module, and
(OSA2) the injection $\mathrm{S}\mathcal{A} \hookrightarrow \mathcal{A}$ is completely bounded.

We further call $\mathrm{S}\mathcal{A}$ a *contractive operator Segal algebra* in $\mathcal{A}$ if the maps above are complete contractions. However, by uniformly scaling the matricial norms of $\mathrm{S}\mathcal{A}$ with a fixed small enough constant, we may make any operator Segal algebra a contractive one, and we will not insist on doing so in the sequel. We note that $\mathrm{S}\mathcal{A}$ itself is a completely bounded Banach algebra.

We finish this section by outlining an approach to operator spaces and completely bounded maps which is dual to the traditional one. Let, for the remainder of the section, $\mathcal{V}$ and $\mathcal{W}$ be complete operator spaces.

We let for any n in $\mathbb{N}$, T_n denote the operator space of $n \times n$ matrices with the dual operator space structure, i.e., $\mathrm{T}_n \cong \mathrm{M}_n^*$ completely isometrically. We let $\mathrm{T}_n(\mathcal{V}) = \mathrm{T}_n \hat{\otimes} \mathcal{V}$, which we regard as matrices. If $S : \mathcal{V} \to \mathcal{W}$ is a completely bounded map, then we let $\mathrm{T}_n(S) : \mathrm{T}_n(\mathcal{V}) \to \mathrm{T}_n(\mathcal{W})$ denote the amplification, i.e. $\mathrm{T}_n(S) = \mathrm{id}_{\mathrm{T}_n} \otimes S$. We also let T_∞ denote the set of $\mathbb{N} \times \mathbb{N}$ matrices which may be identified as trace class operators on $\ell^2(\mathbb{N})$, which we endow with the usual predual operator space structure $\mathrm{T}_\infty \cong \mathcal{B}(\ell^2(\mathbb{N}))_*$. We define $\mathrm{T}_\infty(\mathcal{V})$ analogously as above, and also the operator $\mathrm{T}_\infty(S)$, when it is defined. We note the following elementary, but important fact.

Proposition 1.1. *If $S : \mathcal{V} \to \mathcal{W}$ is a linear map, then the following are equivalent*:

(i) *S is a complete contraction* (*respectively complete quotient map*),
(ii) *each $\mathrm{T}_n(S)$ is a contraction* (*respectively quotient map*), *and*
(iii) *$\mathrm{T}_\infty(S)$ is defined and is a contraction* (*respectively quotient map*).

Proof. (i) $\Leftrightarrow$ (ii). This is [4, 4.18], in light of the identification [4, (7.1.90)]—which shows that our definition of $\mathrm{T}_n(\mathcal{V})$ coincides with theirs. If S is a complete quotient, then each $\mathrm{T}_n(S)$ is a quotient map by the projectivity property of the operator projective tensor product; see [4, 7.1.7].

(ii) $\Leftrightarrow$ (iii) for contractions. We have for each n a completely isomorphic embedding $\mathrm{T}_n \hookrightarrow \mathrm{T}_\infty$, given by identifying elements of T_n with elements of T_∞ whose non-zero entries are only in the upper left $n \times n$ corner. Since $\mathrm{T}_\infty^* \cong \mathcal{B}(\ell^2(\mathbb{N}))$ is an injective operator space, we obtain completely isometric imbeddings $\mathrm{T}_n(\mathcal{V}) = \mathrm{T}_n \hat{\otimes} \mathcal{V} \hookrightarrow \mathrm{T}_\infty(\mathcal{V}) = \mathrm{T}_\infty \hat{\otimes} \mathcal{V}$; see the discussion [4, p. 130]. As $\bigcup_{n=1}^\infty \mathrm{T}_n$ is dense in T_∞, it follows that $\mathrm{T}_{\mathrm{fin}}(\mathcal{V}) = \bigcup_{n=1}^\infty \mathrm{T}_n(\mathcal{V})$ is dense in $\mathrm{T}_\infty(\mathcal{V})$. Now we can define $\mathrm{T}_{\mathrm{fin}}(S) : \mathrm{T}_{\mathrm{fin}}(\mathcal{V}) \to \mathrm{T}_{\mathrm{fin}}(\mathcal{W})$ in the obvious way—so $\mathrm{T}_{\mathrm{fin}}(S) = \mathrm{T}_\infty(S)|_{\mathrm{T}_\infty(\mathcal{V})}$, when the latter makes sense. We have that

$$\left\|\mathrm{T}_{\mathrm{fin}}(S)\right\| = \sup_{n \in \mathbb{N}} \left\|\mathrm{T}_n(S)\right\|.$$

Thus if (ii) is assumed, then $\mathrm{T}_{\mathrm{fin}}(S)$ is contractive, and thus $\mathrm{T}_\infty(S)$ is defined and contractive. Conversely, if (iii) is assumed than $\mathrm{T}_{\mathrm{fin}}(S)$ is contractive, whence (ii).

(ii) $\Leftrightarrow$ (iii) for quotient maps. Suppose that $\mathrm{T}_\infty(S)$ is a quotient map. By [4, 10.1.4], we have that $\mathrm{T}_\infty(\mathcal{V})^* \cong \mathrm{M}_\infty(\mathcal{V}^*)$, where $\mathrm{M}_\infty(\mathcal{V}^*)$ is the space of $\mathbb{N} \times \mathbb{N}$ matrices with entries in $\mathcal{V}^*$ whose finite submatrices are uniformly bounded in norm. Thus $\mathrm{T}_\infty(S)^* = (S^*)^{(\infty)} : \mathrm{M}_\infty(\mathcal{W}^*) \to \mathrm{M}_\infty(\mathcal{V}^*)$ is an isometry. It follows that S^* is a complete isometry, whence S is a complete quotient map by [4, 4.1.8]. $\square$

We say that a linear operator $S : \mathcal{V} \to \mathcal{W}$ is a *complete isomorphism* if it is completely bounded, bijective, and $S^{-1} : \mathcal{W} \to \mathcal{V}$ is completely bounded too. We say that $S : \mathcal{V} \to \mathcal{W}$ is a *complete surjection* if the induced map $\tilde{S} : \mathcal{V}/\ker S \to \mathcal{W}$, defined by $\tilde{S}q = S$ where $q : \mathcal{V} \to \mathcal{V}/\ker S$ is the quotient map, is a complete isomorphism.

Corollary 1.2.

(i) *Suppose S in $\mathcal{CB}(\mathcal{V}, \mathcal{W})$ is a bijection. Then S is a complete isomorphism if and only if $\mathrm{T}_\infty(S) : \mathrm{T}_\infty(\mathcal{V}) \to \mathrm{T}_\infty(\mathcal{W})$ is an isomorphism of Banach spaces.*

(ii) *Suppose S in $\mathcal{CB}(\mathcal{V}, \mathcal{W})$ is surjective. Then S is a complete surjection if and only if $\mathrm{T}_\infty(S) : \mathrm{T}_\infty(\mathcal{V}) \to \mathrm{T}_\infty(\mathcal{W})$ is surjective.*

Proof. (i) If $S^{-1} \in \mathcal{CB}(\mathcal{W}, \mathcal{V})$, then $\mathrm{T}_\infty(S^{-1}) = \mathrm{T}_\infty(S)^{-1}$. Conversely, if $T = \mathrm{T}_\infty(S)^{-1}$ is a bounded operator, we have that for w in $\mathrm{T}_{\mathrm{fin}}(\mathcal{W})$ that

$$\mathrm{T}_\infty(S)Tw = w = \mathrm{T}_\infty(S)\mathrm{T}_{\mathrm{fin}}\big(S^{-1}\big)w$$

so $T|_{\mathrm{T}_{\mathrm{fin}}(\mathcal{W})} = \mathrm{T}_{\mathrm{fin}}(S^{-1})|_{\mathrm{T}_{\mathrm{fin}}(\mathcal{W})}$. Thus $\mathrm{T}_{\mathrm{fin}}(S^{-1})$ is bounded, whence so too is $\mathrm{T}_\infty(S^{-1})$.

(ii) If S is surjective than $\tilde{S}$ is bijective. From above, if $\tilde{S}$ is a complete isomorphism, then $\mathrm{T}_\infty(\tilde{S})$ is an isomorphism of Banach spaces. It follows that $\mathrm{T}_\infty(S) = \mathrm{T}_\infty(\tilde{S})\mathrm{T}_\infty(q)$ is surjective. On the other hand, if $\mathrm{T}_\infty(S) = \mathrm{T}_\infty(\tilde{S})\mathrm{T}_\infty(q)$ is surjective then $\mathrm{T}_\infty(\tilde{S})$ is surjective. As it is already injective, and bounded as $\mathrm{T}_\infty(q)$ is a quotient map, we obtain that $\mathrm{T}_\infty(\tilde{S})$ is a bounded bijection, hence an isomorphism by the open mapping theorem. $\square$

We have, by [4, 7.1.6] that $\mathrm{T}_n \hat{\otimes} \mathcal{W}^* = \mathrm{T}_n(\mathcal{W}^*) \cong \mathrm{M}_n(\mathcal{W})^* \cong (\mathrm{M}_n \check{\otimes} \mathcal{W})*$, where the duality is given in tensor form by $\langle t \otimes f, m \otimes w\rangle = \mathrm{trace}(tm)f(w)$, for $t \in \mathrm{T}_n$, $f \in \mathcal{W}^*$, $m \in \mathrm{M}_n$ and $w \in \mathcal{W}$. Thus, in matrix form, this dual paring becomes

$$\big\langle [f_{ij}], [w_{ij}]\big\rangle = \sum_{i,j=1}^{n} f_{ij}(w_{ji})$$

for $[f_{ij}]$ in $\mathrm{T}_n(\mathcal{W}^*)$ and $[w_{ij}]$ in $\mathrm{M}_n(\mathcal{W})$. Thus the map $[S_{ij}]$ in $\mathcal{CB}(\mathcal{V}, \mathrm{M}_n(\mathcal{W}))$ has adjoint $[S_{ij}]^*$ in $\mathcal{CB}(\mathrm{T}_n(\mathcal{W}^*), \mathcal{V}^*)$ given by

$$\big\langle [S_{ij}]^*[f_{ij}], v\big\rangle = \big\langle [f_{ij}], [S_{ij}v]\big\rangle = \sum_{i,j=1}^{n} f_{ij}(S_{ji}v) = \sum_{i,j=1}^{n} S_{ji}^* f_{ij}(v)$$

for $[f_{ij}]$ in $\mathrm{T}_n(\mathcal{W}^*)$ and $v \in \mathcal{V}$.

Now if $[S_{ij}] \in \mathcal{CB}(\mathcal{V}, \mathrm{M}_n(\mathcal{W}))$, we have that $[S_{ij}^*] \in \mathcal{CB}(\mathcal{W}^*, \mathrm{M}_n(\mathcal{V}^*))$, with $\|[S_{ij}^*]\|_{\mathrm{cb}} = \|[S_{ij}]\|_{\mathrm{cb}}$, with proof similar to that of [4, 3.1.2]. Then $[S_{ij}^*]^* \in \mathcal{CB}(\mathrm{T}_n(\mathcal{V}^{**}), \mathcal{W}^{**})$ and we let

$$[S_{ij}^*]_* = [S_{ij}^*]^*|_{\mathrm{T}_n(\mathcal{V})} \in \mathcal{CB}(\mathrm{T}_n(\mathcal{V}), \mathcal{W}). \tag{1.1}$$

Thus if $[v_{ij}] \in \mathrm{T}_n(\mathcal{V})$, we have obtain $[S_{ij}^*]_*[v_{ij}] = \sum_{i,j=1}^n S_{ji} v_{ij}$. We have adjoint $[S_{ij}^*]_*{}^* = [S_{ij}^*]$, and hence

$$\|[S_{ij}^*]_*\|_{\mathrm{cb}} = \|[S_{ij}^*]\|_{\mathrm{cb}} = \|[S_{ij}]\|_{\mathrm{cb}}. \tag{1.2}$$

This equation will be useful in the sequel when we determine that $\mathrm{S}_0(G)$ is an operator Segal algebra in $\mathrm{A}(G)$.

1.4. Localisation

Let $\mathcal{A}$ be a semi-simple commutative Banach algebra with Gelfand spectrum X. Via the Gelfand transform, we regard $\mathcal{A}$ as a subalgebra of $\mathcal{C}_0(X)$, the algebra of continuous functions on X vanishing at infinity. We say that $\mathcal{A}$ is *regular* if for every pair (x, F), where $x \in X$ and F is a closed subset of X with $x \notin F$, we have that there is u in $\mathcal{A}$ such that $u(x) = 1$ and $u|_F = 0$. We note, below, that such an algebra admits local inverses.

Proposition 1.3. *If $\mathcal{A}$ is a regular Banach algebra with Gelfand spectrum X, $u \in \mathcal{A}$, and K is a compact subset of X on which u does not vanish, then there is u' in $\mathcal{A}$ such that $uu'|_K = 1$.*

Proof. By [14, (39.12)], $\mathcal{A}|_K = \{u|_K\colon u \in \mathcal{A}\}$ is a regular Banach algebra with Gelfand spectrum K. Then we may apply analytic functional calculus [14, 39.14] to see that $1/(u|_K) \in \mathcal{A}|_K$. Find v in $\mathcal{A}$ such that $v|_K = 1/(u|_K)$. □

Thus, we see that a regular Banach algebra $\mathcal{A}$ is a "standard function algebra on its spectrum," in the sense of Reiter [23,24]. Let $\mathcal{I}$ be an ideal in $\mathcal{A}$, which is not necessarily assumed to be closed. We define the *hull* of $\mathcal{I}$ to be the set $\mathrm{h}(\mathcal{I}) = \{x \in X\colon u(x) = 0 \text{ for each } u \in \mathcal{I}\}$. We let $\mathcal{A}_c = \{u \in \mathcal{A}\colon \operatorname{supp} u \text{ is compact}\}$. Thus we obtain the following localisation result, proved in [23, 2.1.4] (or [24, 2.1.14]), which we restate here, without proof, for convenient reference.

Corollary 1.4. *Let $\mathcal{A}$ be a regular Banach algebra and $\mathcal{I}$ be an ideal of $\mathcal{A}$. If $u \in \mathcal{A}_c$ with $\operatorname{supp} u \cap \mathrm{h}(\mathcal{I}) = \varnothing$, then $u \in \mathcal{I}$. In particular, if $\mathrm{h}(\mathcal{I}) = \varnothing$, then $\mathcal{A}_c \subset \mathcal{I}$.*

We say that an ideal $\mathcal{I}$ of a regular Banach algebra $\mathcal{A}$ has *compact support*, if $\operatorname{supp}\mathcal{I} = \overline{G \setminus \mathrm{h}(\mathcal{I})}$ is compact in the spectrum X. The following result can be applied in more general situations than we give, but is only required for $\mathrm{A}(G)$ where G is a locally compact group.

Corollary 1.5. *Let $\mathcal{I}$ and $\mathcal{J}$ each be non-zero ideals having compact support in $\mathrm{A}(G)$ with $\mathcal{I} \subset \mathcal{J}$. Then there are $t_1, \ldots, t_n$ in G and $u_1, \ldots, u_n$ in $\mathcal{I}$ such that*

$$\sum_{l=1}^n (t_l * u_l)v = v \quad \textit{for all } v \textit{ in } \mathcal{J}.$$

Proof. Let Q be any compact subset of $\operatorname{supp}\mathcal{I}$ having non-empty interior Q°. Find $t_1,\dots,t_n$ in G so $\bigcup_{l=1}^n t_l Q^\circ \supset \operatorname{supp}\mathcal{J}$. Then $\mathcal{I}' = \sum_{l=1}^n t_l * \mathcal{I}$ is an ideal in $\mathrm{A}(G)$ for which $\operatorname{supp}\mathcal{J} \subset (\operatorname{supp}\mathcal{I})^\circ$. By the regularity of $\mathrm{A}(G)$ and [14, (39.15)] (or see [5, (3.2)]), there is a function $u \in \mathrm{A}(G)$ such that $u|_{\operatorname{supp}\mathcal{J}} = 1$ and $\operatorname{supp} u \subset \bigcup_{l=1}^n t_l Q$. By Corollary 1.4, above, $u \in \mathcal{I}'$, hence $u = \sum_{l=1}^n t_l * u_l$, as desired. □

2. Construction of Feichtinger's Segal algebra

Let G be a locally compact group. In this section, we reconstruct Feichtinger's Segal algebra, which was done explicitly for abelian G in [7] and [24, §6.2] and, implicitly for general G in [6]. The construction we give below is superficially different than the one of Feichtinger, but conceptually more useful to our task.

Let $\mathcal{I}$ be a non-zero closed ideal in $\mathrm{A}(G)$ which has compact support, as defined in Section 1.4. We let $\ell^1(G)$ be the usual ℓ^1-space, indexed over G, which will be identified with the closed linear span of the Dirac measures $\{\delta_s\colon s \in G\}$. We provide $\ell^1(G)$ with the maximal operator space structure. We define

$$q_{\mathcal{I}} : \ell^1(G) \hat{\otimes} \mathcal{I} \to \mathrm{A}(G) \quad \text{by} \quad q_{\mathcal{I}}(x \otimes u) = x * u = \sum_{s \in G} \alpha_s s * u$$

where $x = \sum_{s \in G} \alpha_s \delta_s$ and $s * u(t) = u(s^{-1}t)$ for t in G. We let

$$\mathrm{S}_0(G) = \operatorname{ran} q_{\mathcal{I}} \cong \ell^1(G) \hat{\otimes} \mathcal{I} / \ker q_{\mathcal{I}}.$$

We make $\mathrm{S}_0(G)$ into an operator space by giving it the quotient operator space structure. Let us note the following description of $\mathrm{S}_0(G)$, which is exactly that of [7], when $n = 1$ and $\mathcal{I} = \mathrm{A}_K(G)$, where

$$\mathrm{A}_K(G) = \bigl\{u \in \mathrm{A}(G)\colon \operatorname{supp} u \subset K\bigr\}$$

for K a compact subset of G with non-empty interior. The matrices $\mathrm{T}_n(\mathcal{V})$, for an operator space $\mathcal{V}$, are described in Section 1.3 above.

Lemma 2.1. *We have for any $n = 1, 2, \dots, \infty$ that*

$$\mathrm{T}_n\bigl(\mathrm{S}_0(G)\bigr) = \left\{ \sum_{k=1}^{\infty} \bigl[s_k * u_{ij}^{(k)}\bigr]\colon s_k \in G, \bigl[u_{ij}^{(k)}\bigr] \in \mathrm{T}_n(\mathcal{I}) \text{ with } \sum_{i=1}^{\infty} \bigl\|\bigl[u_{ij}^{(k)}\bigr]\bigr\|_{\mathrm{T}_n(\mathrm{A})} < +\infty \right\}. \tag{2.1}$$

The norm on $\mathrm{T}_n(\mathrm{S}_0(G)) = \mathrm{T}_n(\operatorname{ran} q_{\mathcal{I}})$ is given by

$$\bigl\|[u_{ij}]\bigr\|_{\mathrm{T}_n(\operatorname{ran} q_{\mathcal{I}})} = \inf \left\{ \sum_{k=1}^{\infty} \bigl\|\bigl[u_{ij}^{(k)}\bigr]\bigr\|_{\mathrm{T}_n(\mathrm{A})}\colon [u_{ij}] = \sum_{k=1}^{\infty} \bigl[s_k * u_{ij}^{(k)}\bigr] \text{ as above} \right\}.$$

Proof. First, let $n = 1$. We observe that there is an isometric identification $\ell^1(G) \hat{\otimes} \mathcal{I} = \ell^1(G) \otimes^\gamma \mathcal{I}$, by virtue of the fact that $\ell^1(G)$ has maximal operator space structure; see [4, 8.2.4]. Now if $t \in \ell^1(G) \otimes^\gamma \mathcal{I}$ and $\varepsilon > 0$, we can write

$$t=\sum_{k=1}^{\infty}\left(\sum_{s\in G}\alpha_{ks}\delta_s\right)\otimes u^{(k)},\quad \text{where } \sum_{k=1}^{\infty}\left(\sum_{s\in G}|\alpha_{ks}|\right)\left\|u^{(k)}\right\|_{\mathrm{A}}<\|t\|_{\gamma}+\varepsilon.$$

We see the sum can easily be rearranged in the form $t=\sum_{s\in G}\delta_s\otimes(\sum_{k=1}^{\infty}\alpha_{ks}u^{(k)})$. (This is essentially the proof that $\ell^1(G)\otimes^{\gamma}\mathcal{I}$ is isometrically isomorphic to the $\mathcal{I}$-valued ℓ^1-space, $\ell^1(G;\mathcal{I})$.) Thus we see that equations for u in $\mathrm{S}_0(G)$ can be arranged as suggested in (2.1), and any such equation describes an element of $\mathrm{S}_0(G)=\operatorname{ran}q_{\mathcal{I}}$. The formula for the norm follows immediately from the fact that $\mathrm{S}_0(G)$ is a quotient of $\ell^1(G)\otimes^{\gamma}\mathcal{I}$.

Now if $n>1$, we make identifications

$$\mathrm{T}_n\big(\ell^1(G)\hat{\otimes}\mathcal{I}\big)\cong\mathrm{T}_n\hat{\otimes}\big(\ell^1(G)\hat{\otimes}\mathcal{I}\big)\cong\ell^1(G)\hat{\otimes}(\mathrm{T}_n\hat{\otimes}\mathcal{I})=\ell^1(G)\otimes^{\gamma}\mathrm{T}_n(\mathcal{I}).$$

The proof then follows the $n=1$ case, given above. □

We now state some of the remarkable properties of $\mathrm{S}_0(G)$.

Theorem 2.2.

(i) *The space* $\mathrm{S}_0(G)$ *is a (contractive) operator Segal algebra in* $\mathrm{A}(G)$.
(ii) $\mathrm{S}_0(G)$ *is the smallest Segal algebra* $\mathrm{SA}(G)$ *in* $\mathrm{A}(G)$ *which is closed under left translations and on which left translations are isometric, i.e.* $s*u\in\mathrm{SA}(G)$ *for each* s *in* G *and* u *in* $\mathrm{SA}(G)$ *with* $\|s*u\|_{\mathrm{SA}}=\|u\|_{\mathrm{SA}}$. *Moreover, for each* $u\in\mathrm{S}_0(G)$, $s\mapsto s*u:G\to\mathrm{S}_0(G)$ *is continuous, and for each* s *in* G, $u\mapsto s*u:\mathrm{S}_0(G)\to\mathrm{S}_0(G)$ *is a complete isometry.*
(iii) *For any closed ideals* $\mathcal{I},\mathcal{J}$ *of* $\mathrm{A}(G)$, *each having compact support, the operator space structure on* $\mathrm{S}_0(G)$ *qua* $\operatorname{ran}q_{\mathcal{I}}$, *or on* $\mathrm{S}_0(G)$ *qua* $\operatorname{ran}q_{\mathcal{J}}$, *are completely isomorphic.*

Proof. (i) As in the construction above, we fix a non-zero closed ideal in $\mathrm{A}(G)$ with compact support.

We note that it is easy to see, using Lemma 2.1 above, and Proposition 1.1, that $\mathrm{S}_0(G)$ imbeds completely contractively into $\mathrm{A}(G)$. Moreover, using the $n=1$ case of the lemma, it is easy to see that $\mathrm{S}_0(G)$ is a Banach $\mathrm{A}(G)$-module. Unfortunately, it is somewhat involved to show that $\mathrm{S}_0(G)$ is a completely contractive $\mathrm{A}(G)$-module.

Let for v in $\mathrm{A}(G)$, $M_v:\mathrm{A}(G)\to\mathrm{A}(G)$ be the multiplication map. Using (1.1) and (1.2), it suffices to show that for any $[v_{ij}]\in\mathrm{M}_n(\mathrm{A}(G))$ we have that

$$\big[M^*_{v_{ij}}\big]_*\in\mathcal{CB}\big(\mathrm{T}_n\big(\mathrm{S}_0(G)\big),\mathrm{S}_0(G)\big)\quad\text{with } \big\|\big[M^*_{v_{ij}}\big]_*\big\|_{\mathrm{cb}}\leqslant\big\|[v_{ij}]\big\|_{\mathrm{M}_n(\mathrm{A})}.$$

By Proposition 1.1 it suffices to show for each m that

$$\mathrm{T}_m\big(\big[M^*_{v_{ij}}\big]_*\big):\mathrm{T}_m\big(\mathrm{T}_n\big(\mathrm{S}_0(G)\big)\big)\to\mathrm{T}_m\big(\mathrm{S}_0(G)\big)$$

satisfies

$$\big\|\mathrm{T}_m\big(\big[M^*_{v_{ij}}\big]_*\big)\big\|\leqslant\big\|[v_{ij}]\big\|_{\mathrm{M}_n(\mathrm{A})}. \tag{2.2}$$

We note that there is a natural identification

$$\mathrm{T}_m\big(\mathrm{T}_n\big(\mathrm{S}_0(G)\big)\big)\cong\mathrm{T}_{nm}\big(\mathrm{S}_0(G)\big)$$

of which we take advantage. However, we will prefer, for computational convenience, to label elements of this space by doubly indexed matrices $[u_{ij,pq}]$, where $i,j=1,\dots,n$ and $p,q=1,\dots,m$.

By Lemma 2.1, each element $[u_{ij,pq}]$ of $\mathrm{T}_{nm}(\mathrm{S}_0(G))$ admits, for any $\varepsilon>0$, the form

$$[u_{ij,pq}]=\sum_{k=1}^{\infty}\bigl[s_k*u_{ij,pq}^{(k)}\bigr]=\mathrm{T}_n(q_{\mathcal{I}})\left(\sum_{k=1}^{\infty}\delta_{s_k}\otimes\bigl[u_{ij,pq}^{(k)}\bigr]\right)$$

where $\sum_{k=1}^{\infty}\|[u_{ij,pq}^{(k)}]\|_{\mathrm{T}_{nm}(\mathrm{A})}\leqslant\|[u_{ij}]\|_{\mathrm{T}_{nm}(\operatorname{ran}q_{\mathcal{I}})}+\varepsilon$. We see that

$$\begin{aligned}
\mathrm{T}_m\bigl(\bigl[M_{v_{ij}}^*\bigr]_*\bigr)[w_{ij,pq}]&=\sum_{k=1}^{\infty}\mathrm{T}_m\bigl(\bigl[M_{v_{ij}}^*\bigr]_*\bigr)\bigl(\bigl[s_k*u_{ij,pq}^{(k)}\bigr]\bigr)\\
&=\sum_{k=1}^{\infty}\left[s_k*\sum_{i,j=1}^{m}\bigl(s_k^{-1}*v_{ji}\bigr)u_{ij,pq}\right]\\
&=\mathrm{T}_m(q_{\mathcal{I}})\left(\sum_{k=1}^{\infty}\delta_{s_k}\otimes\left[\sum_{i,j=1}^{m}\bigl(s_k^{-1}*v_{ji}\bigr)u_{ij,pq}\right]\right)\\
&=\mathrm{T}_m(q_{\mathcal{I}})\left(\sum_{k=1}^{\infty}\delta_{s_k}\otimes\mathrm{T}_m\bigl(\bigl[M_{s_k^{-1}*v_{ij}}^*\bigr]_*\bigr)\bigl[u_{ij,pq}^{(k)}\bigr]\right).
\end{aligned}$$

Using Proposition 1.1, (1.2), and the fact that translation is a complete isometry on $\mathrm{A}(G)$, we thus obtain

$$\begin{aligned}
&\left\|\sum_{k=1}^{\infty}\delta_{s_k}\otimes\mathrm{T}_m\bigl(\bigl[M_{s_k^{-1}*v_{ij}}^*\bigr]_*\bigr)\bigl[u_{ij,pq}^{(k)}\bigr]\right\|_{\mathrm{T}_{nm}(\ell^1\hat{\otimes}\mathcal{I})}\\
&\quad\leqslant\sum_{k=1}^{\infty}\bigl\|\mathrm{T}_m\bigl(\bigl[M_{s_k^{-1}*v_{ij}}^*\bigr]_*\bigr)\bigr\|\,\bigl\|\bigl[u_{ij,pq}^{(k)}\bigr]\bigr\|_{\mathrm{T}_{nm}(\mathrm{A})}\\
&\quad=\sum_{k=1}^{\infty}\bigl\|\bigl[s_k^{-1}*v_{ij}\bigr]\bigr\|_{\mathrm{M}_m(\mathrm{A})}\bigl\|\bigl[u_{ij,pq}^{(k)}\bigr]\bigr\|_{\mathrm{T}_{nm}(\mathrm{A})}\\
&\quad\leqslant\bigl\|[v_{ij}]\bigr\|_{\mathrm{M}_n(\mathrm{A})}\bigl(\bigl\|[u_{ij,pq}]\bigr\|_{\mathrm{T}_{nm}(\operatorname{ran}q_{\mathcal{I}})}+\varepsilon\bigr).
\end{aligned}$$

Thus, since ε is arbitrary, we obtain that

$$\bigl\|\mathrm{T}_m\bigl(\bigl[M_{v_{ij}}^*\bigr]_*\bigr)[u_{ij,pq}]\bigr\|_{\mathrm{T}_m(\operatorname{ran}q_{\mathcal{I}})}\leqslant\bigl\|[v_{ij}]\bigr\|_{\mathrm{M}_n(\mathrm{A})}\bigl\|[u_{ij,pq}]\bigr\|_{\mathrm{T}_{nm}(\operatorname{ran}q_{\mathcal{I}})}$$

and hence we obtain (2.2).

(ii) It is immediate from the construction of $\mathrm{S}_0(G)$ that it is closed under left translations. We note that the action of G on $\mathrm{S}_0(G)$ is continuous, and is one of isometries, in fact complete isometries. This follows by a straightforward application of Lemma 2.1 and Proposition 1.1.

Any Segal algebra $\mathrm{SA}(G)$ in $\mathrm{A}(G)$ is an ideal with empty hull, and thus, by Corollary 1.4, necessarily contains $\mathrm{A}_c(G)$. Hence for any non-zero compactly supported ideal $\mathcal{I}$ of $\mathrm{A}(G)$, we obtain that $\mathcal{I} \subset \mathrm{SA}(G)$. Thus, if translations are pointwise continuous and isometric on $\mathrm{SA}(G)$, we see by Lemma 2.1, in the case $n = 1$, that $\mathrm{SA}(G) \supset \mathrm{S}_0(G)$.

(iii) It follows from (ii), above, that $\mathrm{S}_0(G)$ is independent if the choice of ideal $\mathcal{I}$. Let us suppose that $\mathcal{I}$ and $\mathcal{J}$ are two closed non-zero ideals of $\mathrm{A}(G)$ having compact supports. By replacing $\mathcal{I}$ by $\mathcal{I} \cap \mathcal{J}$ for an appropriate choice of s, if necessary, we may suppose $\mathcal{I} \subset \mathcal{J}$. Then the injection $\iota : \ell^1(G) \hat{\otimes} \mathcal{I} \hookrightarrow \ell^1(G) \hat{\otimes} \mathcal{J}$ is a complete contraction. It is clear that $q_{\mathcal{J}} \circ \iota = q_{\mathcal{I}}$, so ι induces a completely contractive map $\tilde{\iota} : \operatorname{ran} q_{\mathcal{I}} \to \operatorname{ran} q_{\mathcal{J}}$, which is the identity map. Thus, by Proposition 1.1, $\mathrm{T}_\infty(\tilde{\iota}) : \mathrm{T}_\infty(\operatorname{ran} q_{\mathcal{I}}) \to \mathrm{T}_\infty(\operatorname{ran} q_{\mathcal{J}})$ is a contraction. Let us see that $\mathrm{T}_\infty(\tilde{\iota})$ is surjective. Let $u = \sum_{l=1}^n t_l * u_l$ be as in Corollary 1.5. We note that if $[w_{ij}] \in \mathrm{T}_\infty(\operatorname{ran} q_{\mathcal{J}})$, has form $[w_{ij}] = \sum_{k=1}^\infty [s_k * w_{ij}^{(k)}]$ as in (2.1), then we have

$$
\begin{aligned}
[w_{ij}] &= \sum_{k=1}^{\infty} \left[s_k * \left(u w_{ij}^{(k)}\right)\right] = \sum_{k=1}^{\infty} \sum_{l=1}^{n} \left[s_k * \left(t_l * u_l w_{ij}^{(k)}\right)\right] \\
&= \mathrm{T}_\infty(q_{\mathcal{I}}) \left(\sum_{k=1}^{\infty} \sum_{l=1}^{n} \delta_{s_k t_l} \otimes \left[u_l\left(t_l^{-1} * w_{ij}^{(k)}\right)\right] \right) \qquad (2.3)
\end{aligned}
$$

which is an element of $\mathrm{T}_\infty(\operatorname{ran} q_{\mathcal{I}})$. Hence $\mathrm{T}_\infty(\tilde{\iota})$ is surjective, and thus, by the open mapping theorem, an isomorphism of Banach spaces. We then appeal to Corollary 1.2. □

Let us note that Theorem 2.2, above, holds under more general assumptions. Though the assumptions we give below seem less natural, they are indispensable for actually working with $\mathrm{S}_0(G)$. Let $\mathcal{I}$ be a non-zero compactly supported ideal in $\mathrm{A}(G)$, which is not necessarily closed, but comes equipped with an operator space structure by which it is a completely contractive Banach $\mathrm{A}(G)$-module, and the inclusion map $\mathcal{I} \hookrightarrow \mathrm{A}(G)$ is completely bounded. We define $q_{\mathcal{I}} : \ell^1(G) \hat{\otimes} \mathcal{I} \to \mathrm{A}(G)$, and its quotient space $\operatorname{ran} q_{\mathcal{I}}$, with its quotient operator space structure, as before.

Corollary 2.3. *With $\mathcal{I}$ as above,* $\operatorname{ran} q_{\mathcal{I}}$ *is an operator Segal algebra in* $\mathrm{A}(G)$ *which is completely isomorphic with* $\operatorname{ran} q_{\overline{\mathcal{I}}}$, *where $\overline{\mathcal{I}}$ is the closure of $\mathcal{I}$ in* $\mathrm{A}(G)$. *Hence* $\operatorname{ran} q_{\mathcal{I}} = \mathrm{S}_0(G)$, *completely isomorphically.*

Proof. We first note that the proofs of Lemma 2.1, and then of (i) and (ii) of the theorem above, can be applied verbatim; though we should note that the inclusion $\operatorname{ran} q_{\mathcal{I}} \hookrightarrow \mathrm{A}(G)$ is completely bounded, instead of completely contractive. Thus we see that $\operatorname{ran} q_{\mathcal{I}}$ is an operator Segal algebra in $\mathrm{S}_0(G)$ on which G acts continuously and isomorphically by translations. Hence $\operatorname{ran} q_{\mathcal{I}} = \operatorname{ran} q_{\overline{\mathcal{I}}}$. Moreover the proof of part (iii) can be applied up to seeing that $\mathrm{T}_\infty(\tilde{\iota}) : \mathrm{T}_\infty(\operatorname{ran} q_{\mathcal{I}}) \to \mathrm{T}_\infty(\operatorname{ran} q_{\overline{\mathcal{I}}})$ is bounded. To see obtain surjectivity, we note for any u in $\mathcal{I}$ that $M_u : \mathrm{A}(G) \to \mathcal{I}$ is completely bounded, and then (2.3), where $[w_{ij}] \in \mathrm{T}_\infty(\operatorname{ran} q_{\overline{\mathcal{I}}})$, has the form

$$
[w_{ij}] = \mathrm{T}_\infty(q_{\mathcal{I}}) \left(\sum_{k=1}^{\infty} \sum_{l=1}^{n} \delta_{s_k t_l} \otimes \mathrm{T}_\infty(M_{u_l})\left[t_l^{-1} * w_{ij}^{(k)}\right] \right)
$$

where each $\mathrm{T}_\infty(M_{u_l})[t_l^{-1} * w_{ij}^{(k)}] \in \mathrm{T}_\infty(\mathcal{I})$. □

We recall that a Segal algebra $\mathrm{S}^1(G)$ in $\mathrm{L}^1(G)$ is called *symmetric* if for any s in G and f in $\mathrm{S}^1(G)$ we have

$$\|s * f\|_{\mathrm{S}^1} = \|f\|_{\mathrm{S}^1} = \|f * s\|_{\mathrm{S}^1}$$

where $f * s(t) = \Delta(s)^{-1} f(ts^{-1})$ for almost every t in G. This is a necessary and sufficient condition to make $\mathrm{S}^1(G)$ a two-sided ideal in $\mathrm{L}^1(G)$. $\mathrm{S}^1(G)$ is called *pseudo-symmetric* if the anti-action $s \mapsto f * s$ is continuous on G for any fixed f in $\mathrm{S}^1(G)$. We note that this is equivalent to having the action of right translation

$$s \mapsto s \cdot f, \quad \text{where } s \cdot f(t) = f(ts) \text{ for almost every } t \tag{2.4}$$

continuous on G for any fixed f in $\mathrm{S}^1(G)$. We also recall, by well-known technique (see [16, pp. 26, 27], for example), that any Banach space $\mathcal{V}$ is an contractive essential left/right $\mathrm{L}^1(G)$-module if and only if there is a continuous action/anti-action of G on $\mathcal{V}$ by linear isometries. Furthermore, if $\mathcal{V}$ is an operator space, $\mathcal{V}$ is a completely contractive $\mathrm{L}^1(G)$-module, if and only if the associated action of G on $\mathcal{V}$ is one by complete isometries.

It will be useful, below, to recall the *Lebesgue–Fourier algebra*

$$\mathrm{LA}(G) = \mathrm{A}(G) \cap \mathrm{L}^1(G) \tag{2.5}$$

studied in [9,11,12]. This is simultaneously a Segal algebra in $\mathrm{A}(G)$ and in $\mathrm{L}^1(G)$. In [9] it was shown that $\mathrm{LA}(G)$ is also a contractive operator Segal algebra in either context.

Corollary 2.4.

(i) $\mathrm{S}_0(G)$ *is a pseudo-symmetric operator Segal algebra in* $\mathrm{L}^1(G)$. *It is symmetric only if G is unimodular.*

(ii) *Let $\mathcal{I}$ be a fixed ideal in* $\mathrm{A}(G)$ *satisfying the assumptions of Corollary* 2.3. *Let*

$$q'_{\mathcal{I}} : \mathrm{L}^1(G) \hat{\otimes} \mathcal{I} \to \mathrm{A}(G)$$

be given, on elementary tensors, by $q'_{\mathcal{I}}(f \otimes u) = f * u$. Then $\operatorname{ran} q'_{\mathcal{I}}$, *with its quotient operator space structure, is completely isomorphic to* $\mathrm{S}_0(G)$.

Proof. (i) If $\mathcal{I}$ is a closed compactly supported pointwise ideal in $\mathrm{LA}(G)$. Then $\mathcal{I}$ imbeds completely contractively into $\mathrm{L}^1(G)$, and is a completely contractive $\mathrm{A}(G)$-module. Thus, it follows Corollary 2.3 above, that $\operatorname{ran} q_{\mathcal{I}}$, with its quotient operator space structure, imbeds completely contractively into $\mathrm{L}^1(G)$. It then follows, part (ii) of Theorem 2.2, that $\mathrm{S}_0(G)$ is an operator Segal algebra in $\mathrm{L}^1(G)$; and the same proof of can be trivially adapted to see G acts continuously and completely isometrically on $\mathrm{S}_0(G)$ by right translation. It is clear that $s \cdot u = \Delta(s) u * s^{-1}$ for all u in $\mathrm{S}_0(G)$ and s in G, and hence $\mathrm{S}_0(G)$ is a symmetric Segal algebra in $\mathrm{L}^1(G)$, if and only if G is unimodular.

(ii) Let (e_U) be the bounded approximate identity for $\mathrm{L}^1(G)$ given by normalised indicator functions of relatively compact neighbourhoods of the identity, e. If $[u_{ij}] \in \mathrm{T}_\infty(\mathcal{I})$, then for s in G we have that

$$[s * u_{ij}] = \lim_{U \searrow e} [s * e_U * u_{ij}] \in \mathrm{T}_\infty\big(\operatorname{ran} q'_{\mathcal{I}}\big).$$

Thus we see from Lemma 2.1, that $\mathrm{S}_0(G) \subset \operatorname{ran} q'_{\mathcal{I}}$, completely boundedly. To obtain the converse inclusion, we note, similarly as in the proof of Lemma 2.1, that

$$\mathrm{T}_\infty\big(\mathrm{L}^1(G) \hat{\otimes} \mathcal{I}\big) \cong \mathrm{L}^1(G) \hat{\otimes} \mathrm{T}_\infty(\mathcal{I}) = \mathrm{L}^1(G) \otimes^\gamma \mathrm{T}_\infty(\mathcal{I}).$$

Hence every element of $\operatorname{ran} q'_{\mathcal{I}}$ is of the form

$$\sum_{k=1}^{\infty} \big[f_k * u_{ij}^{(k)}\big] \quad \text{where } \sum_{k=1}^{\infty} \|f_k\|_{\mathrm{L}^1} \big\|\big[u_{ij}^{(k)}\big]\big\|_{\mathrm{T}_\infty(\mathrm{A})} < +\infty. \tag{2.6}$$

Now if $[u_{ij}] \in \mathrm{T}_\infty(\mathcal{I})$ and $f \in \mathrm{L}^1(G)$, then

$$[f * u_{ij}] = \int_G f(s)[s * u_{ij}] \in \mathrm{T}_\infty\big(\mathrm{S}_0(G)\big)$$

since the integral may be realised as a Bochner integral in $\mathrm{T}_\infty(\mathrm{S}_0(G))$, by the continuity of $s \mapsto [s * u_{ij}] : G \to \mathrm{T}_\infty(\mathrm{S}_0(G))$. Thus (2.6) shows that $\operatorname{ran} q'_{\mathcal{I}} \subset \mathrm{S}_0(G)$, completely boundedly. □

The next result shows the only occasions for which we know that $\mathrm{S}_0(G) = \mathrm{LA}(G)$, and we conjecture these are all such occasions. For this result, we will consider $\mathrm{S}_0(G)$ as a Segal algebra in $\mathrm{A}(G)$.

Corollary 2.5. *If K is an open compact subgroup, with T a transversal for left cosets, then there is a natural completely isomorphic algebra homomorphism*

$$\mathrm{S}_0(G) \cong \ell^1(T) \hat{\otimes} \mathrm{A}(K)$$

where $\ell^1(T)$ has pointwise multiplication. In particular,

(i) $\mathrm{S}_0(G) = \ell^1(G)$*, completely isomorphically, if G is discrete, and*
(ii) $\mathrm{S}_0(G) = \mathrm{A}(G)$*, completely isomorphically, if G is compact.*

Proof. We first note that $\mathrm{A}(K) \cong \mathrm{A}_K(G)$, completely isometrically. Second, for s, t in G, $s * \mathrm{A}_K(G) = \mathrm{A}_{sK}(G)$ and $t * \mathrm{A}_K(G) = \mathrm{A}_{tK}(G)$ are either identical or disjoint, depending on whether $s^{-1}t \in K$ or not. Thus we can see that $q_{\mathrm{A}_K(G)}|_{\ell^1(T)\hat{\otimes}\mathrm{A}_K(G)} : \ell^1(T) \hat{\otimes} \mathrm{A}_K(G) \to \operatorname{ran} q_{\mathrm{A}_K(G)}$ is a complete isomorphism. Indeed, it follows from Lemma 2.1 that

$$\mathrm{T}_\infty(q_{\mathrm{A}_K(G)}|_{\ell^1(T)\hat{\otimes}\mathrm{A}_K(G)}) : \mathrm{T}_\infty\big(\ell^1(T) \hat{\otimes} \mathrm{A}(K)\big) \to \mathrm{T}_\infty\big(\mathrm{S}_0(G)\big)$$

is an bijection, hence an isomorphism. □

It is useful to observe the more general fact below.

Corollary 2.6. *If H is an open subgroup of G, with T a transversal for left cosets, then there is a natural completely isomorphic algebra homomorphism*

$$\mathrm{S}_0(G) \cong \ell^1(T) \hat{\otimes} \mathrm{S}_0(H)$$

where $\ell^1(T)$ has pointwise multiplication.

Proof. Let $\mathcal{I}$ be a non-empty closed compactly supported ideal of $\mathrm{A}(G)$ and $\operatorname{supp}\mathcal{I} \subset H$. Then we may consider $\mathcal{I}$ to be an ideal in $\mathrm{A}(H) \cong \mathrm{A}_H(G)$, and we have that

$$q_{\mathcal{I}}^H = q_{\mathcal{I}}|_{\ell^1(H)\hat{\otimes}\mathcal{I}} : \ell^1(H) \hat{\otimes} \mathcal{I} \to \mathrm{S}_0(H)$$

is a complete surjection. The bijection $(t,s) \mapsto ts : T \times H \to G$ induces an isomorphism

$$\ell^1(T) \hat{\otimes} \ell^1(H) = \ell^1(T) \otimes^\gamma \ell^1(H) \cong \ell^1(T \times H) \cong \ell^1(G).$$

Then we obtain the following commuting diagram:

$$\begin{tikzcd}
\ell^1(T) \hat{\otimes} \ell^1(H) \hat{\otimes} \mathcal{I} \arrow[r, "\cong"] \arrow[d, "\mathrm{id}\otimes q_{\mathcal{I}}^H"'] & \ell^1(G) \hat{\otimes} \mathcal{I} \arrow[d, "q_{\mathcal{I}}"] \\
\ell^1(T) \hat{\otimes} \mathrm{S}_0(H) \arrow[r, "\delta_t\otimes u \mapsto t*u"] & \mathrm{S}_0(G)
\end{tikzcd}$$

We obtain, as in the proof of the result above, that the bottom arrow represents a complete isomorphism. □

3. Functorial properties

3.1. Tensor products

Let us first note the primary motivation for desiring an operator space structure on $\mathrm{S}_0(G)$. This is an analogue of a result from [3] which states that

$$\mathrm{A}(G) \hat{\otimes} \mathrm{A}(H) \cong \mathrm{A}(G \times H)$$

completely isometrically, via the natural morphism which identifies $u \otimes v$ with the function $(s,t) \mapsto u(s)v(t)$. Alternatively we may view this as an analogue of the classical result that

$$\mathrm{L}^1(G) \hat{\otimes} \mathrm{L}^1(H) = \mathrm{L}^1(G) \otimes^\gamma \mathrm{L}^1(H) \cong \mathrm{L}^1(G \times H)$$

where the projective and operator projective tensor products agree since $\mathrm{L}^1(G)$ (or $\mathrm{L}^1(H)$) is a maximal operator space.

Theorem 3.1. *Let G and H be locally compact groups. Then there is a natural complete isomorphism* $\mathrm{S}_0(G) \hat{\otimes} \mathrm{S}_0(H) \cong \mathrm{S}_0(G \times H)$.

Proof. By [15, (7.3) and (7.4)], there are almost connected open subgroups G_0 of G, and H_0 of H. Let $\mathcal{I}$ and $\mathcal{J}$ be compactly supported ideals of $\mathrm{A}(G_0)$ and $\mathrm{A}(H_0)$, respectively.

The dual spaces $\mathrm{A}(G_0)^* \cong \mathrm{VN}(G_0)$ and $\mathrm{A}(H_0)^* \cong \mathrm{VN}(H_0)$ are injective von Neumann algebras and hence injective operator spaces; see [22, pp. 227–228], for example. Thus, by comments in [4, p. 130], the inclusion maps induce a complete isometry $\mathcal{I}\hat{\otimes}\mathcal{J} \hookrightarrow \mathrm{A}(G_0)\hat{\otimes}\mathrm{A}(H_0)$. Thus, via the completely isometric injections

$$\mathcal{I}\hat{\otimes}\mathcal{J} \hookrightarrow \mathrm{A}(G_0)\hat{\otimes}\mathrm{A}(H_0) \cong \mathrm{A}(G_0 \times H_0) \hookrightarrow \mathrm{A}(G \times H)$$

we may regard $\mathcal{I}\hat{\otimes}\mathcal{J}$ as a compactly supported closed ideal of $\mathrm{A}(G \times H)$. Now we have a completely isometric identification

$$J : \left(\ell^1(G) \hat{\otimes} \mathcal{I}\right) \hat{\otimes} \left(\ell^1(H) \hat{\otimes} \mathcal{J}\right) \to \ell^1(G \times H) \hat{\otimes} (\mathcal{I} \hat{\otimes} \mathcal{J}).$$

Using finite sums of elementary tensors we see that $q_{\mathcal{I}} \otimes q_{\mathcal{J}} = q_{\mathcal{I}\hat{\otimes}\mathcal{J}} \circ J$. Hence $\operatorname{ran}(q_{\mathcal{I}} \otimes q_{\mathcal{J}})$, with its quotient operator space structure, must be completely (isometrically) isomorphic to $\operatorname{ran} q_{\mathcal{I}\hat{\otimes}\mathcal{J}}$, with its quotient operator space structure. By projectivity of the operator projective tensor product, see [4, 7.1.7], we have that

$$\operatorname{ran}(q_{\mathcal{I}} \otimes q_{\mathcal{J}}) = \mathrm{S}_0(G) \hat{\otimes} \mathrm{S}_0(H).$$

By Corollary 2.3 we have that $\operatorname{ran} q_{\mathcal{I}\hat{\otimes}\mathcal{J}} = \mathrm{S}_0(G \times H)$. $\square$

If G and H are abelian, the following recovers one of the main results of Feichtinger [7].

Corollary 3.2. *If either G or H admits an open abelian subgroup, then we have a natural isomorphism* $\mathrm{S}_0(G) \otimes^\gamma \mathrm{S}_0(H) \cong \mathrm{S}_0(G \times H)$.

Proof. If either G or H is abelian, then the proof above can be followed almost verbatim, with the projective tensor product $\otimes^\gamma$ playing the role of the operator projective tensor product $\hat{\otimes}$. The reason we require extra hypotheses here is that they are sufficient (and almost necessary) to obtain that $\mathrm{A}(G) \otimes^\gamma \mathrm{A}(H) \cong \mathrm{A}(G \times H)$, isomorphically, as proved in [19], for example. Thus we obtain that $\mathcal{I} \otimes^\gamma \mathcal{J}$ can be realised as an ideal in $\mathrm{A}(G \times H)$.

If G, say, has an open abelian subgroup A, with transversal for left cosets T, then by Corollary 2.6 and the reasoning above, we obtain isomorphic identifications

$$\begin{aligned}\mathrm{S}_0(G) \otimes^\gamma \mathrm{S}_0(H) &\cong \ell^1(T) \otimes^\gamma \mathrm{S}_0(A) \otimes^\gamma \mathrm{S}_0(H)\\ &\cong \ell^1\left(T \times \{e_H\}\right) \otimes^\gamma \mathrm{S}_0(A \times H) \cong \mathrm{S}_0(G \times H)\end{aligned}$$

where we obtain the last identification by realising $T \times \{e_H\}$ as a transversal for the left cosets of $A \times H$ in $G \times H$. $\square$

We note that if G and H are both compact, neither having an open abelian subgroup, then the above result fails by Corollary 2.5(ii) and [19]. We conjecture that our operator space structure on $\mathrm{S}_0(G)$ is the maximal operator space structure exactly when G admits an open abelian subgroup. This would imply the result above. However, it is clear, only when G has a compact abelian open subgroup, that our operator space structure on $\mathrm{S}_0(G)$ is the maximal one. Indeed if G is abelian, then for an arbitrary closed ideal $\mathcal{I}$ of $\mathrm{A}(G)$, i.e. of $\mathrm{L}^1(\hat{G})$, it is not clear that the subspace operator space structure is the maximal one, whence we have no means to deduce that $\mathrm{S}_0(G) \cong \ell^1(G) \hat{\otimes} \mathcal{I}/\ker q_{\mathcal{I}}$ is a maximal operator space. For results on subspaces of maximal operator spaces, see [21], for example.

3.2. Restriction

We recall from [13,26] that if H is a closed subgroup of G, then the restriction map $u \mapsto u|_H : \mathrm{A}(G) \to \mathrm{A}(H)$ is a quotient map. In fact, it is a complete quotient map since its adjoint map is an injective $*$-homomorphism from $\mathrm{VN}(H)$ onto the von Neumann algebra generated by $\{\lambda(s)\colon s\in H\}$ in $\mathrm{VN}(G)$.

The following result is due to Feichtinger [7], in the abelian case. However, most of his techniques rely on commutativity of G, and cannot be adapted to show the general case, even with no considerations for the operator space structure.

Theorem 3.3. *If H is a closed subgroup in G, then the restriction map*

$$u\mapsto u|_H:\mathrm{S}_0(G)\to\mathrm{S}_0(H)$$

is completely surjective.

Proof. First we must verify that if $u\in\mathrm{S}_0(G)$, then $u|_H\in\mathrm{S}_0(H)$. Let T be a transversal for the right cosets of H. The bijection $(t,s)\mapsto st:T\times H\to G$ induces an isomorphism

$$\ell^1(T)\mathbin{\hat{\otimes}}\ell^1(H)=\ell^1(T)\otimes^\gamma\ell^1(H)\cong\ell^1(T\times H)\cong\ell^1(G).$$

Now let $\mathcal{I}=\mathrm{A}_K(G)$, where K is a compact neighbourhood of the identity in G. If $t\in T$, then

$$(tK)\cap H\subset s_t\big(K^{-1}K\cap H\big),\quad \text{for some } s_t \text{ in } H. \tag{3.1}$$

Indeed, if $(tK)\cap H\neq\varnothing$, then there is $k\in K$ so $tk\in H$, so $t\in Hk^{-1}\subset HK^{-1}$, and thus there is s_t in H so $t\in s_tK^{-1}$, whence $tk\in s_tK^{-1}K$. Now, as in the proof of Lemma 2.1, any u in $\mathrm{S}_0(G)$ can be written in the form

$$u=\sum_{t\in T}\sum_{s\in H}s*t*u_{st},\quad \text{where } \sum_{t\in T}\sum_{s\in H}\|u_{st}\|_{\mathrm{A}}<+\infty$$

and each $u_{st}\in\mathrm{A}_K(G)$. We then have that for each t in T, using (3.1), that

$$\sum_{s\in H}(s*t*u_{st})|_H=\sum_{s\in H}s*s_t*\big(\big(s_t^{-1}*t*u_{st}\big)\big|_H\big)$$

where, $(s_t^{-1}*t*u_{st})|_H\in\mathcal{K}=\mathrm{A}_{K^{-1}K\cap H}(H)$ and $\|(s_t^{-1}*t*u_{st})|_H\|_{\mathrm{A}}\leqslant\|u_{st}\|_{\mathrm{A}}$. It then follows that $u|_H$, being a $\|\cdot\|_{\operatorname{ran}q_{\mathcal{K}}}$-summable series of elements from $\mathrm{S}_0(H)$, is itself in $\mathrm{S}_0(H)$.

Now let us see that restriction is completely surjective. Let $\mathcal{I}$ be as above so that $(\operatorname{supp}\mathcal{I})^\circ\cap H\neq\varnothing$. Note that $\mathcal{I}|_H$, with the operator space structure given by its being a quotient of $\mathcal{I}$ via the restriction map, is a completely contractive $\mathrm{A}(H)$-module. Indeed, this follows from the fact that $\mathrm{A}(H)$ is a complete quotient of $\mathrm{A}(G)$. Since $\ell^1(H)$ is a (completely) complemented subspace of $\ell^1(G)$, we have that $\ell^1(H)\mathbin{\hat{\otimes}}\mathcal{I}$ is a closed subspace of $\ell^1(G)\mathbin{\hat{\otimes}}\mathcal{I}$. We have that the following diagram commutes:

$$\begin{array}{ccc}
\mathrm{T}_\infty(\ell^1(H)\mathbin{\hat{\otimes}}\mathcal{I}) & \xrightarrow{\mathrm{T}_\infty(\mathrm{id}\otimes(u\mapsto u|_H))} & \mathrm{T}_\infty(\ell^1(H)\mathbin{\hat{\otimes}}\mathcal{I}|_H)\\
{\scriptstyle \mathrm{T}_\infty(q_{\mathcal{I}}|_{\ell^1(H)\hat{\otimes}\mathcal{I}})}\downarrow & & \downarrow{\scriptstyle \mathrm{T}_\infty(q_{\mathcal{I}|_H})}\\
\mathrm{T}_\infty(\mathrm{S}_0(G)) & \xrightarrow{[u_{ij}]\mapsto[u_{ij}|_H]} & \mathrm{T}_\infty(\mathrm{S}_0(H))
\end{array}$$

where $q_{\mathcal{I}}|_H$ is a complete surjection by Corollary 2.3, so $T_\infty(q_{\mathcal{I}}|_H)$ is a surjection by Corollary 1.2. Thus the restriction map $u \mapsto u|_H : S_0(G) \to S_0(H)$ is completely surjective. □

3.3. *Multipliers on* $\overline{L^2(G)} \otimes^\gamma L^2(G)$

The aim of this section is to develop some techniques for use in the next section on the averaging operation. Let

$$T(G) = \overline{L^2(G)} \otimes^\gamma L^2(G)$$

where $\overline{L^2(G)}$ denotes the conjugate space of $L^2(G)$. We recall that $T(G)^* \cong \mathcal{B}(L^2(G))$ via the dual pairing $\langle \bar{f} \otimes g, T\rangle = \langle Tg|f\rangle$. Thus $T(G)$ is an operator space with the predual operator space structure.

We may regard $T(G)$ as a space of equivalence classes of functions on $G \times G$: if $\omega = \sum_{k=1}^\infty \bar{f}_k \otimes g_k$, where $\{f_k\}_{k=1}^\infty$ and $\{g_k\}_{k=1}^\infty$ are each summable sequences from $L^2(G)$, then $\omega(s,t) = \sum_{k=1}^\infty \overline{f_i(s)} g_i(t)$ for almost every s and almost every t in G. A function $w : G \times G \to \mathbb{C}$ is called a *multiplier* of $T(G)$ if for every ω in $T(G)$, $m_w\omega$, defined for almost every s and almost every t in G by $m_w\omega(s,t) = w(s,t)\omega(s,t)$, determines an element of $T(G)$. We let $MT(G)$ denote the space of all multipliers w such that $m_w : T(G) \to T(G)$ is a bounded map. We summarise below, some results from [25]. We note that there are some trivial differences between our notations used here, and those in [25]; the result is stated below to be consistent with our present notation. For a different perspective, we also refer the reader to [20].

Theorem 3.4.

(i) [25, Theorem 3.3] *For each* w *in* $MT(G)$, $m_w : T(G) \to T(G)$ *is a completely bounded map with* $\|m_w\|_{\mathcal{CB}(T(G))} = \|m_w\|_{\mathcal{B}(T(G))}$. *Thus the space* $MT(G) \cong \{m_w\colon\ w \in MT(G)\}$ *is a closed subalgebra of* $\mathcal{CB}(T(G))$, *and hence a completely contractive Banach algebra.*

(ii) [25, Corollary 4.3 and Theorem 5.3] *If* $u \in B(G)$, *the map* $\gamma u : G \times G \to \mathbb{C}$, *given by* $\gamma u(s,t) = u(st^{-1})$, *is an element of* $MT(G)$. *Moreover,* $\gamma : B(G) \to MT(G)$ *is a completely contractive homomorphism.*

We let $P_G : T(G) \to A(G)$ be given by

$$P_G(\bar{f} \otimes g) = \langle \lambda(\cdot)g|f\rangle = \bar{f} * \check{g} \tag{3.2}$$

where $\check{g}(s) = g(s^{-1})$ for almost every s. The adjoint map, $P_G^* : VN(G) \to \mathcal{B}(L^2(G))$, is the inclusion map, hence a complete isometry, whence P_G is a complete contraction. We note for ω in $T(G)$ that

$$P_G\omega(s) = \int_G \omega(t, s^{-1}t)\, dt$$

for each s in g. Thus if $u \in B(G)$, then

$$P_G(m_{\gamma u}\omega)(s) = \int_G u\big(t(s^{-1}t)^{-1}\big)\omega(t, s^{-1}t)\, dt = u(s)P_G\omega(s).$$

We now introduce a class of ideals in $\mathrm{A}(G)$ which will prove useful. Let K be a compact subset of G of positive measure. Let $\mathrm{T}(K) = \overline{\mathrm{L}^2(K)} \otimes^\gamma \mathrm{L}^2(K)$, where we regard $\mathrm{L}^2(K)$ as a subspace of $\mathrm{L}^2(G)$ in the natural way. We define

$$\mathcal{M}(K) = P_G\big(\mathrm{T}(K)\big)$$

and endow $\mathcal{M}(K)$ with the quotient operator space structure so it is completely isometrically isomorphic with $\mathrm{T}(K)/\ker(P_G|_{\mathrm{T}(K)})$. Clearly, $\mathcal{M}(K) \subset \mathrm{A}(G)$, and has as a dense subspace $\mathrm{span}\{\bar{f} * \check{g}\colon f, g \in \mathrm{L}^2(K)\}$.

Proposition 3.5. *The space $\mathcal{M}(K)$ is a completely contractive $\mathrm{B}(G)$-module. Thus it is an ideal in $\mathrm{A}(G)$ with $\mathrm{supp}\,\mathcal{M}(K) \subset K^{-1}K$, and equipped with an operator space structure by which it is a completely contractive $\mathrm{A}(G)$-module.*

We remark that there is no reason to suspect that $\mathcal{M}(K)$ is a closed ideal in $\mathrm{A}(G)$ for a general compact set K.

Proof. Since $\mathrm{L}^2(K)$ is a complemented subspace of $\mathrm{L}^2(G)$, $\mathrm{T}(K)$ identifies isometrically as a closed subspace of $\mathrm{T}(G)$. As such, $\mathrm{T}(K)$ is a $\mathrm{MT}(G)$-submodule of $\mathrm{T}(G)$. Also, for u in $\mathrm{B}(G)$ and ω in $\mathrm{T}(K)$ we have $uP_G\omega = P_G(m_{\gamma u}\omega)$, so $\mathcal{M}(K)$ is a $\mathrm{B}(G)$-module. We thus have that the following diagram commutes:

$$\begin{tikzcd}
\mathrm{B}(G) \otimes_\wedge \mathrm{T}(K) \arrow[r, "\gamma\otimes\mathrm{id}"] \arrow[d, "\mathrm{id}\otimes P_G|_{\mathrm{T}(K)}"'] & \mathrm{MT}(G) \hat{\otimes} \mathrm{T}(K) \arrow[r, "w\otimes\omega\mapsto m_w\omega"] & \mathrm{T}(K) \arrow[d, "P_G|_{\mathrm{T}(K)}"] \\
\mathrm{B}(G) \otimes_\wedge \mathcal{M}(K) \arrow[rr, "u\otimes v\mapsto uv"] & & \mathcal{M}(K)
\end{tikzcd}$$

Since $\mathrm{id} \otimes P_G|_{\mathrm{T}(K)}$ is a complete quotient map, and the maps $\gamma \otimes \mathrm{id}$, $w \otimes \omega \mapsto m_w\omega$ and $P_G|_{\mathrm{T}(K)}$ are complete contractions, $u \otimes v \mapsto uv$ must be a complete contraction too.

Thus it is clear that $\mathcal{M}(K)$ is an ideal in $\mathrm{A}(G)$ and a completely contractive $\mathrm{A}(G)$-module. It is straightforward to verify that $\mathrm{supp}\,\mathcal{M}(K) \subset K^{-1}K$. □

Let us note that we may obtain a weak version of a "tensor product factorisation" result of [7]. Let K be a compact subset of G of non-empty interior, and define $q_K^2 : \ell^1(G) \otimes^\gamma \mathrm{L}^2(K) \to \mathrm{L}^2(G)$ by $q_K^2(\delta_s \otimes f) = s * f$. Let $\mathrm{W}^2(G) = \mathrm{ran}\, q_K^2$ and norm it as the quotient space. It is straightforward to verify that for any other compact set K', having non-empty interior, that $\mathrm{ran}\, q_{K'}^2 = \mathrm{ran}\, q_K^2$, and that the quotient norms are equivalent. Let $P_G' : \mathrm{W}^2(G) \otimes^\gamma \mathrm{W}^2(G) \to \mathrm{A}(G)$ be given by $P_G'(f \otimes g) = f * \check{g}$. It can be checked, similarly as in the proof of the corollary above, that $\mathrm{ran}\, P_G'$, with its quotient norm, is a Segal algebra in $\mathrm{A}(G)$ on which G acts isometrically by left translations. Also, $\mathrm{ran}\, P_G' \subset \mathrm{S}_0(G)$, and thus, by Theorem 2.2(ii), we obtain $\mathrm{ran}\, P_G' = \mathrm{S}_0(G)$.

However, unlike in the commutative case, we do not have that either of the maps from $\mathrm{S}_0(G) \otimes^\gamma \mathrm{S}_0(G)$ to $\mathrm{S}_0(G)$, given on elementary tensors by $u \otimes v \mapsto u * v$ or $u \otimes v \mapsto u * \check{v}$, are surjective. Indeed, this fails for compact groups which do not admit an abelian subgroup of finite index, a fact which follows from [17, Proposition 2.5], in light of Corollary 2.5 and the fact that $v \mapsto \check{v}$ is an isometry on $\mathrm{A}(G)$. It would be interesting to know when either of the aforementioned maps, extended to $\mathrm{S}_0(G) \hat{\otimes} \mathrm{S}_0(G)$, surjects onto $\mathrm{S}_0(G)$. This fails in general. Recent work of the author, with B.E. Forrest and E. Samei [10], has shown that if G is compact, and hence $\mathrm{S}_0(G) = \mathrm{A}(G)$, then each such map is surjective only when G admits an abelian subgroup of finite index.

3.4. Averaging over a normal subgroup

Let N be a closed normal subgroup of G and $\tau_N : \mathrm{L}^1(G) \to \mathrm{L}^1(G/N)$ be given for f in $\mathrm{L}^1(G)$ and almost every $sN \in G/N$, accepting a mild abuse of notation, by

$$\tau_N(f)(sN) = \int_N f(sn)\,dn.$$

This operator is a complete quotient map as observed in [9]. It was shown in [7], for an abelian G, that $\tau_N(\mathrm{S}_0(G)) = \mathrm{S}_0(G/N)$. We obtain a generalisation of that result.

Theorem 3.6. *We have for any locally compact group G with closed normal subgroup N that $\tau_N(\mathrm{S}_0(G)) = \mathrm{S}_0(G/N)$, and $\tau_N : \mathrm{S}_0(G) \to \mathrm{S}_0(G/N)$ is a complete surjection.*

Proof. We divide the proof into three stages.

(I) $\tau_N(\mathcal{M}(K)) \subset \mathrm{A}(G/N)$ *and* $\tau_N : \mathcal{M}(K) \to \mathrm{A}(G/N)$ *is completely bounded.*

Let us first show that $\tau_N(\mathrm{L}^2(K)) \subset \mathrm{L}^2(G/N)$ and that $\tau_N : \mathrm{L}^2(K) \to \mathrm{L}^2(G/N)$ is bounded. Let $\varphi : G \to \mathbb{C}$ be a continuous function of compact support such that $\varphi|_K = 1$. Then for any f in $\mathrm{L}^2(K)$ we have, using Hölder's inequality and the Weyl integral formula, that

$$\begin{aligned}\|\tau_N(f)\|^2_{\mathrm{L}^2(G/N)} &= \int_{G/N} \left| \int_N \varphi(sn) f(sn)\,dn \right|^2 dsN \\ &\leqslant \int_{G/N} \int_N \left|\varphi(sn')\right|^2 dn' \int_N \left|f(sn)\right|^2 dn\,dsN \\ &\leqslant \sup_{s\in G} \tau_N\big(|\varphi|^2\big)(sN) \int_G \left|f(s)\right|^2 ds.\end{aligned}$$

We note that $\sup_{s\in G} \tau_N(|\varphi|^2)(sN) < \infty$ since $\tau_N(|\varphi|^2)$ is itself continuous and of compact support on G/N, as can be checked using the uniform continuity of $|\varphi|^2$. Thus

$$\|\tau_N\|_{\mathcal{B}(\mathrm{L}^2(K),\mathrm{L}^2(G/N))} \leqslant \sup_{s\in G} \tau_N\big(|\varphi|^2\big)(sN)^{1/2}.$$

It is shown in [18, p. 187] that for any compactly supported $f \in \mathrm{L}^1(G)$

$$\tau_N(\check{f}) = \big[\Delta_{G/N}\tau_N(\check{\Delta}_G f)\big]^\vee.$$

Now let $\theta_N : \mathrm{L}^2(K) \to \mathrm{L}^2(G/N)$ be given by

$$\theta_N(f) = \Delta_{G/N}\tau_N(\check{\Delta}_G f).$$

Then we have that θ_N is bounded with

$$\|\theta_N\|_{\mathcal{B}(\mathrm{L}^2(K),\mathrm{L}^2(G/N))} \leqslant \sup_{s\in K} \Delta_{G/N}(sN)\|\tau_N\|_{\mathcal{B}(\mathrm{L}^2(K),\mathrm{L}^2(G/N))} \sup_{t\in K} \frac{1}{\Delta_G(t)}.$$

Now if $f, g \in \mathrm{L}^2(K)$, then

$$\tau_N(\bar{f} * \check{g}) = \tau_N(\bar{f}) * \tau_N(\check{g}) = \overline{\tau_N(f)} * \left[\theta_N(g)\right]^{\vee} \in \mathrm{A}(G/N).$$

Hence it follows that $\tau_N(\mathcal{M}(K)) \subset \mathrm{A}(G/N)$.

We now wish to establish that $\tau_N : \mathcal{M}(K) \to \mathrm{A}(G/N)$ is completely bounded. As in [4, Section 3.4], we assign $\overline{\mathrm{L}^2(K)}$ the row space operator space structure, denoted $\overline{\mathrm{L}^2(K)}_{\mathrm{r}}$; and $\mathrm{L}^2(K)$ the column operator space structure space, denoted $\mathrm{L}^2(K)_{\mathrm{c}}$. Then we have a completely isometric equality

$$\mathrm{T}(G) = \overline{\mathrm{L}^2(K)}_{\mathrm{r}} \hat{\otimes} \mathrm{L}^2(K)_{\mathrm{c}}$$

by [4, 9.3.2 and 9.3.4]. Then, by [4, 3.4.1 and 7.1.3], we have that

$$\tau_N \otimes \theta_N : \overline{\mathrm{L}^2(K)}_{\mathrm{r}} \hat{\otimes} \mathrm{L}^2(K)_{\mathrm{c}} \to \overline{\mathrm{L}^2(G/N)}_{\mathrm{r}} \hat{\otimes} \mathrm{L}^2(G/N)_{\mathrm{c}}$$

is completely bounded, thus is a completely bounded map on $\mathrm{T}(K)$. Hence we see that the following diagram commutes:

$$\begin{tikzcd} \mathrm{T}(K) \arrow[r, "\tau_N\otimes\theta_N"] \arrow[d, "P_G|_{\mathrm{T}(K)}"'] & \mathrm{T}(G/N) \arrow[d, "P_{G/N}"] \\ \mathcal{M}(K) \arrow[r, "\tau_N"] & \mathrm{A}(G/N) \end{tikzcd}$$

where P_G is defined in (3.2), and $P_{G/N}$ is defined analogously. Since $P_G|_{\mathrm{T}(K)}$ is a complete quotient, and $\tau_N \otimes \theta_N$ and $P_{G/N}$ are completely bounded, $\tau_N : \mathcal{M}(K) \to \mathrm{A}(G/N)$ must be completely bounded too.

(II) $\tau_N(\mathcal{M}(K))$, *with its quotient operator space structure, i.e. naturally identified with the quotient space* $\mathcal{M}(K)/\ker(\tau_N|_{\mathcal{M}(K)})$, *is a completely contractive* $\mathrm{A}(G/N)$*-module.*

Let $\pi_N : G \to G/N$ denote the quotient map. By [5, (2.26)] the function $u \mapsto u \circ \pi_N$ defines a complete isometry from $\mathrm{B}(G/N)$ to $\mathrm{B}(G : N)$, the closed subspace of $\mathrm{B}(G)$ of functions which are constant on cosets of N. Now if $u \in \mathrm{B}(G/N)$ and $v \in \mathcal{M}(K)$, then for any s in G we have

$$u(sN)\tau_N(u)(sN) = \int_N u \circ \pi_N(sn)v(sn)\,dn = \tau_N(u \circ \pi_N v)(sN).$$

Hence $u\tau_N(v) = \tau(u \circ \pi_N v) \in \tau_N(\mathcal{M}(K))$, by Proposition 3.5.

We now establish that $\tau_N(\mathcal{M}(K))$ is a completely contractive $\mathrm{B}(G/N)$-module, hence a completely contractive $\mathrm{A}(G/N)$-module. Letting $\iota:\mathrm{B}(G:N)\to\mathrm{B}(G/N)$ be the inverse of $u\mapsto u\circ\pi_N$, we obtain the following commuting diagram:

$$\begin{tikzcd}
\mathrm{B}(G:N)\otimes_\wedge\mathcal{M}(K) \arrow[r, "u\otimes v\mapsto uv"] \arrow[d, "\iota\otimes\tau_N|_{\mathcal{M}(K)}"'] & \mathcal{M}(K) \arrow[d, "\tau_N|_{\mathcal{M}(K)}"] \\
\mathrm{B}(G/N)\otimes_\wedge\tau_N(\mathcal{M}(K)) \arrow[r, "\tilde{u}\otimes\tilde{v}\mapsto\tilde{u}\tilde{v}"] & \mathrm{A}(G/N)
\end{tikzcd}$$

Since $\iota\otimes\tau_N|_{\mathcal{M}(K)}$ is a complete contraction, and $u\otimes v\mapsto uv$ and $\tau_N|_{\mathcal{M}(K)}$ are complete contractions, we obtain that $\tilde{u}\otimes\tilde{v}\mapsto\tilde{u}\tilde{v}$ is thus a complete contraction too.

(III) *The finale.*

We recall from Corollary 2.4(ii), that the map $q'_{\mathcal{M}(K)}:\mathrm{L}^1(G)\hat{\otimes}\mathcal{M}(K)\to\mathrm{S}_0(G)$, given on elementary tensors by $q'_{\mathcal{M}(K)}f\otimes u=f*u$ is a complete surjection. Similarly, appealing also to (II) above, we have that $q'_{\tau_N(\mathcal{M}(K))}:\mathrm{L}^1(G/N)\hat{\otimes}\tau_N(\mathcal{M}(K))\to\mathrm{S}_0(G/N)$ is a complete surjection. It is then clear that the following diagram commutes:

$$\begin{tikzcd}
\mathrm{L}^1(G)\hat{\otimes}\mathcal{M}(K) \arrow[r, "\tau_N\otimes\tau_N"] \arrow[d, "q'_{\mathcal{M}(K)}"'] & \mathrm{L}^1(G/N)\hat{\otimes}\tau_N(\mathcal{M}(K)) \arrow[d, "q'_{\tau_N(\mathcal{M}(K))}"] \\
\mathrm{S}_0(G) \arrow[r, "\tau_N"] & \mathrm{L}^1(G/N)
\end{tikzcd}$$

In particular $\tau_N(\mathrm{S}_0(G))\subset\operatorname{ran}q'_{\tau_N(\mathcal{M}(K))}=\mathrm{S}_0(G/N)$. Moreover, since $q'_{\mathcal{M}(K)}$ and $\tau_N\otimes\tau_N$, as above, and $q'_{\tau_N(\mathcal{M}(K))}:\mathrm{L}^1(G/N)\hat{\otimes}\tau_N(\mathcal{M}(K))\to\mathrm{S}_0(G/N)$ are complete surjections, then $\tau_N:\mathrm{S}_0(G)\to\mathrm{S}_0(G/N)$ is completely bounded. Amplifying the diagram by T_∞, and appealing to Corollary 1.2, as in the end of the proof of Theorem 3.3, we see that $\tau_N:\mathrm{S}_0(G)\to\mathrm{S}_0(G/N)$ is completely surjective. $\square$

We remark that it follows from the above theorem that $\tau_N(\mathrm{A}_\mathrm{c}(G))\subset\mathrm{A}_\mathrm{c}(G/N)$. We note that it was proved in [9] that $\tau_N(\mathrm{LA}(G))=\mathrm{L}^1(G/N)$. We also note that it was shown by Lohoué [18] that $\tau_N:\mathrm{A}_K(G)\to\mathrm{A}(G/N)$ is a bounded map, with bound depending on K. This fact can be deduced from our result, but Lohoué's proof is much simpler, though it is not obvious how to adapt his proof to show that $\tau_N\in\mathcal{CB}(\mathrm{A}_K(G),\mathrm{A}(G/N))$.

3.5. An isomorphism theorem

Let G and H be locally compact groups. Wendel [28] proved that there is an isometric isomorphism between the convolution algebras $\mathrm{L}^1(G)$ and $\mathrm{L}^1(H)$ if and only if G and H are isomorphic topological groups [28]. Also, Walter [27] proved that $\mathrm{A}(G)$ and $\mathrm{A}(H)$ are isometrically isomorphic if and only if G and H are isomorphic topological groups. Since we lack fixed norms on our algebras $\mathrm{S}_0(G)$ and $\mathrm{S}_0(H)$, it is not reasonable to expect and "isometric isomorphism" theorem, in the spirit of Wendel's and Walter's theorems. In fact, if G and H are both discrete groups having the same cardinality, then Corollary 2.5 tells us there is a multiplicative isomorphism identifying $\mathrm{S}_0(G) \cong \mathrm{S}_0(H)$. Similarly, if G and H are finite abelian groups, then there is a convolutive isomorphism identifying $\mathrm{S}_0(G) \cong \mathrm{S}_0(H)$. To obtain a satisfactory result, we must simultaneously exploit the facts $\mathrm{S}_0(G)$ and $\mathrm{S}_0(H)$ are pointwise and convolutive algebras.

Theorem 3.7. *Let G and H be locally compact groups and $\Phi : \mathrm{S}_0(G) \to \mathrm{S}_0(H)$ be a bounded linear bijection which satisfies*

$$\Phi(uv) = \Phi u \Phi v \quad \textit{and} \quad \Phi(u * v) = \Phi u * \Phi v$$

for every u, v in $\mathrm{S}_0(G)$. There is a homeomorphic isomorphism $\alpha : G \to H$ such that

$$\Phi u = u \circ \alpha$$

for each u in $\mathrm{S}_0(G)$.

Proof. We recall that $\mathrm{A}(G)$ has Gelfand spectrum G, implemented by evaluation functionals. Since $\mathrm{S}_0(G)$ is a Segal algebra in $\mathrm{A}(G)$, it follows from [2, Theorem 2.1] that $\mathrm{S}_0(G)$ has Gelfand spectrum G too. The same holds for $\mathrm{S}_0(H)$. Thus we may define $\alpha : H \to G$ by letting for h in H, $\alpha(h)$ be the element of G which satisfies $u(\alpha(h)) = \Phi u(h)$ for each $u \in \mathrm{S}_0(G)$. Then α is continuous. Indeed, if not, we may find a net $h_i \to h$ in H and a neighbourhood U of $\alpha(h)$, such that $\alpha(h_i) \notin U$ for each i. Using regularity we may find u in $\mathrm{S}_0(G)$ such that $u(\alpha(h)) = 1$ and $\operatorname{supp} u \subset U$. But then we would obtain

$$\lim_i \Phi u(h_i) = \lim_i u\big(\alpha(h_i)\big) = 0 \neq 1 = u\big(\alpha(h)\big) = \Phi u(h)$$

which contradicts that Φu is continuous, in particular that $\Phi u \in \mathrm{S}_0(H)$. We may similarly obtain a continuous map $\beta : G \to H$ satisfying $v(\beta(s)) = \Phi^{-1} v(s)$ for all $s \in G$ and v in $\mathrm{S}_0(H)$. We clearly have that $\beta = \alpha^{-1}$, hence α is a homeomorphism.

It remains to see that α is a group homomorphism. Let $\mathcal{U}$ be a compact neighbourhood basis of the identity e_G in G. For each U in $\mathcal{U}$ find u_U in $\mathrm{S}_0(G)$ such that

$$\operatorname{supp} u_U \subset U \quad \text{and} \quad \int_G \big|u_U(s)\big|\, ds = 1.$$

Then (u_U) is a bounded approximate identity for $\mathrm{L}^1(G)$, hence a convolutive approximate unit for $\mathrm{S}_0(G)$. Since Φ is a surjective convolutive homomorphism, (Φu_U) is a convolutive approximate identity for $\mathrm{S}_0(H)$. Let h_1, h_2 in H and suppose that $\alpha(h_1)^{-1}\alpha(h_2) \neq \alpha(h_1^{-1}h_2)$. We could then find $v \in \mathrm{S}_0(G)$ such that

$$v\big(\alpha(h_1)^{-1}\alpha(h_2)\big) = 1 \quad \text{and} \quad v(s) = 0 \quad \text{for all } s \text{ in a neighbourhood of } \alpha\big(h_1^{-1}h_2\big).$$

Then we would have that

$$\begin{aligned} v\big(\alpha(h_1)^{-1}\alpha(h_2)\big) &= \alpha(h_1) * v\big(\alpha(h_2)\big) = \lim_U \alpha(h_1) * u_U * v\big(\alpha(h_2)\big) \\ &= \lim_U \Phi\big(\alpha(h_1) * u_U * v\big)(h_2) \end{aligned}$$

$$= \lim_U \Phi\big(\alpha(h_1) * u_U\big) * \Phi v(h_2)$$
$$= \lim_U \int_H u_U\big(\alpha(h_1)^{-1}\alpha(r)\big) v\big(\alpha\big(r^{-1}h_2\big)\big)\, dr$$
$$= 0$$

where we obtain the last equality from the fact that

$$\operatorname{supp}\big(\alpha(h_1) * u\big) \circ \alpha \subset \big\{r \in H\colon \alpha(h_1)^{-1}\alpha(r) \in U\big\} = \alpha^{-1}\big(\alpha(h_1)U\big)$$

and the supposition that $v(\alpha(r^{-1}h_2)) = 0$ for all r in $\alpha^{-1}(\alpha(h_1)U)$, for a suitably small choice of U. This contradicts that $v(\alpha(h_1)^{-1}\alpha(h_2)) = 1$, whence such a v cannot be chosen, and we thus conclude that $\alpha(h_1)^{-1}\alpha(h_2) = \alpha(h_1^{-1}h_2)$. Substituting $e_H = h_1$, we see that $\alpha(e_H) = e_G$ and then, substituting e_H for h_2, we obtain that $\alpha(h_1^{-1}) = \alpha(h_1)^{-1}$ for each h_1 in H. Thus α is a group homomorphism. $\square$

Our theorem above is not special to the class of algebras $\mathrm{S}_0(G)$. In fact it can be applied to any class of regular Banach algebras with spectrum G, each of which is a Segal algebra of $\mathrm{L}^1(G)$. Examples of such are $\mathrm{LA}(G)$ from (2.5) and the Wiener algebra $\mathrm{W}_0(G)$ as defined in [6].

Acknowledgments

The author is grateful to Ebrahim Samei and Antoine Derighetti, for indicating to him the article [18], and helping to clarify some of the results therein. Though [18] was not ostensibly used, in the end, it inspired the author to find the proof of the result of Section 3.4.

References

[1] D.P. Blecher, The standard dual of an operator space, Pacific Math. J. 153 (1) (1992) 15–30.
[2] J.T. Burnham, Closed ideals in subalgebras of Banach algebras, I, Proc. Amer. Math. Soc. 32 (1972) 551–555.
[3] E.G. Effros, Z.-J. Ruan, On approximation properties for operator spaces, Internat. J. Math. 1 (1990) 163–187.
[4] E.G. Effros, Z.-J. Ruan, Operator Spaces, London Math. Soc. Monogr. (N.S.), vol. 23, Oxford Univ. Press, New York, 2000.
[5] P. Eymard, L'algebre de Fourier d'un groupe localement compact, Bull. Soc. Math. France 92 (1964) 181–236.
[6] H.G. Feichtinger, A characterization of minimal homogeneous Banach spaces, Proc. Amer. Math. Soc. 81 (1981) 55–61.
[7] H.G. Feichtinger, On a new Segal algebra, Montash. Math. 92 (1981) 269–289.
[8] H.G. Feichtinger, K. Gröchening, Gabor frames and time–frequency analysis of distributions, J. Funct. Anal. 146 (1997) 464–495.
[9] B.E. Forrest, N. Spronk, P.J. Wood, Operator Segal algebras in Fourier algebras, Studia Math. 179 (2007) 277–295.
[10] B.E. Forrest, E. Samei, N. Spronk, Convolutions on compact groups and Fourier algebras of coset spaces, preprint.
[11] F. Ghahramani, A.T.M. Lau, Weak amenability of certain classes of Banach algebras without bounded approximate units, Math. Proc. Cambridge Philos. Soc. 133 (2002) 357–371.
[12] F. Ghahramani, A.T.M. Lau, Approximate weak amenability, derivations and Arens regularity of Segal algebras, Studia Math. 169 (2005) 189–205.
[13] C.S. Herz, Harmonic synthesis for subgroups, Ann. Inst. Fourier (Grenoble) 23 (3) (1973) 91–123.
[14] E. Hewitt, K.A. Ross, Abstract Harmonic Analysis II, Grundlehren Math. Wiss., vol. 152, Springer, Berlin, 1970.
[15] E. Hewitt, K.A. Ross, Abstract Harmonic Analysis I, Grundlehren Math. Wiss., vol. 115, Springer, New York, 1979.
[16] B.E. Johnson, Cohomology in Banach algebras, Mem. Amer. Math. Soc. 127 (1972).

[17] B.E. Johnson, Non-amenability of the Fourier algebra of a compact group, J. London Math. Soc. 50 (2) (1994) 361–374.
[18] N. Lohoué, Remarques sur les ensembles de synthèse des algèbras de groupe localement compact, J. Funct. Anal. 13 (1973) 185–194.
[19] V. Losert, On tensor products of the Fourier algebras, Arch. Math. 43 (1984) 370–372.
[20] M. Neufang, Z.-J. Ruan, N. Spronk, Completely isometric representations of $M_{\mathrm{cb}}A(G)$ and $UCB(\hat{G})$, Trans. Amer. Math. Soc. (2006), in press; see http://www.ams.org/cgi-bin/mstrack/accepted_papers/tran.
[21] T. Oikhberg, Subspaces of maximal operator spaces, Integral Equations Operator Theory 48 (2004) 81–102.
[22] A.L.T. Paterson, The class of locally compact groups G for which $C^*(G)$ is amenable, in: Harmonic Analysis, Luxembourg, 1987, Springer, Berlin, 1988, pp. 226–237.
[23] H. Reiter, Classical Harmonic Analysis and Locally Compact Groups, Oxford Math. Monogr., Clarendon, Oxford, 1968.
[24] H. Reiter, J.D. Stegeman, Classical Harmonic Analysis and Locally Compact Groups, London Math. Soc. (N.S.), vol. 22, Oxford Univ. Press, New York, 2000.
[25] N. Spronk, Measurable Schur multipliers and completely bounded multipliers of the Fourier algebras, Proc. London Math. Soc. 89 (2004) 161–192.
[26] M. Takesaki, N. Tatsuuma, Duality and subgroups, II, J. Funct. Anal. 11 (1972) 184–190.
[27] M.E. Walter, W^*-algebras and nonabelian harmonic analysis, J. Funct. Anal. 11 (1972) 17–38.
[28] J.G. Wendel, Left centralizers and isomorphisms of group algebras, Pacific J. Math. 2 (1952) 251–261.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 248 (2007) 175–201
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

Large deviations for stochastic nonlinear beam equations

Tusheng Zhang

School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, England, UK

Received 20 August 2006; accepted 30 March 2007

Communicated by Paul Malliavin

Abstract

We establish a large deviation principle for the solutions of stochastic partial differential equations for nonlinear vibration of elastic panels (also called stochastic nonlinear beam equations).

Keywords: Stochastic partial differential equations; Stochastic beam equations; Large deviations; Exponential martingales; Exponential integrability

1. Introduction

Consider a bounded open interval on the real line, say, $(0,1)$. Let $L^2 = L^2(0,1)$. Denote by $H_0^1 = H_0^1(0,1)$ and $H_0^2 = H_0^2(0,1)$ the Sobolev spaces of order one and two satisfying the homogeneous boundary conditions. Denote by H_0^{-k} the dual space of H_0^k. $(\cdot,\cdot)$ will denote the L^2-inner product and $\langle\cdot,\cdot\rangle$ denotes the dual pairing. The norms on L^2, H_0^k and H_0^{-k} will be denoted respectively by $\|\cdot\|$, $\|\cdot\|_k$ and $\|\cdot\|_{-k}$. Consider the linear operator

$$Au = \alpha\partial_x^2 u - \gamma\partial_x^4 u,$$

E-mail address: tzhang@maths.man.ac.uk (T. Zhang).

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2007.03.029

and the nonlinear operator

$$B(u) = \beta\left(\int_0^1 |\partial_x u|^2\,dx\right)\partial_x^2 u.$$

The mathematical model for the nonlinear penal vibration is governed by the following partial differential equation:

$$\begin{gathered}\partial_t^2 u_t = \left(\alpha + \beta\int_0^1 |\partial_y u_t|^2\,dy\right)\partial_x^2 u_t - \gamma\partial_x^4 u_t + F(\dot{u}_t, u_t),\\ u_t(0) = u_t(1) = 0, \qquad \partial_x u_t(0) = \partial_x u_t(1) = 0,\\ u_0(x) = \phi_0(x), \qquad \partial_t u_0(x) = \phi_1(x),\end{gathered}\tag{1}$$

where $\dot{u}_t$ denotes the derivative of u with respect to the variable t. A detailed study of the model can be found in the book by Dowell [14]. The equation was also proposed by Woinowsky-Krieger in [22] as a model for the transversal deflection of an extensible beam of natural length 1. An equation in two space variables similar to (1) was suggested in [10] as a model of nonlinear oscillations of a plate in a supersonic flow of gas. It has also been studied by many other people, see [2,4,15,16] and references therein.

Let $W_t, t \geqslant 0$, be a Wiener process taking values in a Hilbert space. Without loss of generality, we may assume that W_t is l^2-valued Wiener process which admits the following representation:

$$W_t = \sum_{k=1}^{\infty}\lambda_k\beta_t^k e_k,$$

where $\lambda_k, k \geqslant 1$, is a sequence of non-negative numbers such that $\sum_{k=1}^{\infty}\lambda_k^2 < \infty$, $\beta_t^k, k \geqslant 1$, is a sequence of independent standard Brownian motions and $\{e_k, k \geqslant 1\}$ is the canonical orthonormal basis of l^2.

Taking into account the random fluctuations, Chow and Menaldi [9] considered the stochastic nonlinear partial differential equation for vibration of elastic panels:

$$\begin{gathered}\partial_t^2 u_t^\varepsilon = \left(\alpha + \beta\int_0^1 \left|\partial_y u_t^\varepsilon\right|^2 dy\right)\partial_x^2 u_t^\varepsilon - \gamma\partial_x^4 u_t^\varepsilon + \varepsilon\sigma\left(u_t^\varepsilon\right)\dot{W}_t + F\left(\dot{u}_t^\varepsilon, u_t^\varepsilon\right),\\ u_t^\varepsilon(0) = u_t^\varepsilon(1) = 0, \qquad \partial_x u_t^\varepsilon(0) = \partial_x u_t^\varepsilon(1) = 0,\\ u_0^\varepsilon(x) = \phi_0(x) \in H_0^2, \qquad \partial_t u_0^\varepsilon(x) = \phi_1(x) \in L^2,\end{gathered}\tag{2}$$

where for every $u \in H_0^2$, $\sigma(u)$ stands for a map from l^2 into H_0^2 which will be specified later, and $F(\cdot,\cdot)$ denotes a map from $L^2 \times H_0^2$ into L^2. It is proved in [9] that under reasonable conditions on σ, (2) has a unique solution with the property:

$$u^\varepsilon \in C\left([0,T]; H_0^2\right) \quad \text{and} \quad \dot{u}^\varepsilon \in C\left([0,T]; L^2\right).$$

A general formulation of the equation in an abstract Hilbert space was later studied by Brzeźniak, Maslowski and Seidler [4], where existence, uniqueness and asymptotic stability of the solution were discussed.

The aim of this paper is to establish a large deviation principle (LDP) for the vector process $v_t^\varepsilon = (u_t^\varepsilon, \dot{u}_t^\varepsilon)$ on the product space $C([0,T]; H_0^2) \times C([0,T]; L^2)$ as $\varepsilon \to 0$.

The large deviation problem for stochastic partial differential equations (SPDEs) has been studied by many people, but mainly for stochastic parabolic equations. For example, an LDP for stochastic reaction equations with nonlinear reaction term was established by Cerrai and Röckner [6]. An LDP for stochastic Burgers'-type SPDEs was considered by Cardon-Weber [5]. A uniform LDP for parabolic SPDEs was proved by Chenal and Millet [7]. In [19], Rovira and Sanz-Sole proved an LDP for a class of nonlinear hyperbolic SPDEs. An LDP was obtained by Chow [8] for some parabolic SPDEs. An LDP for stochastic reaction equations was established by Sowers [20]. A small time large deviation principle for stochastic parabolic equations was obtained by the author [23]. For the general theory of large deviations, readers are referred to the monograph [12]. For SPDEs in general, we refer readers to [18].

Because of the different nature of nonlinearity for different types of equations, the large deviations for SPDEs has to be dealt with on individual bases. There are two main issues which distinguish the current work from the previous ones. The first is the cubic nonlinear term $B(u)$ in Eq. (2) and the second is the second-order differentiation in t (not like the parabolic cases). Note that even the existence and uniqueness of the solution of this kind of equation was newly established. Although the second-order (in t) Eq. (2) can also be written as a system of parabolic equations as it was done in [4], but by doing so the operator (differential) becomes degenerate. The properties of the corresponding semigroups are therefore not good enough for the large deviation estimates, not like the parabolic cases in the existing literature. To tackle the first issue, our idea is to prove that the probability that the energy of the solution is big is exponentially small. To this end, a remarkable result of Davis [11], Barlow and Yor in [3] on the moment estimates of martingales plays a key role. To treat the second-order differentiation in t, we fully exploit the energy equality proved by Chow, Menaldi [9] and Pardoux [17], and establish some exponential integrability of Hilbert space-valued martingales. To achieve this, some exponential martingales are specially constructed.

The rest of the paper is organized as follows. In Section 2, the precise result is stated. In Section 3, the skeleton equation is studied. It is proved that the solution is a continuous map from the level set into the space $C([0,T]; H_0^2) \times C([0,T]; L^2)$. Section 4 is devoted to the proof of the large deviation principle. The long proof is split into several lemmas for clarity.

We end this introduction with a remark.

Remark 1.1. The main result in this paper is stated in the setting of one space dimension. This is just for simplicity. Our approach works equally well in high space dimensions and also in the general setting formulated in [4].

Throughout the paper, the generic constants may be different from line to line. If it is essential, the dependence of a constant on the parameters will be written explicitly.

2. Statement of the main result

We now state the precise conditions on σ. Let $\sigma_k(\cdot), k \geqslant 1$, be a sequence of mappings from H_0^2 into H_0^2 and $F(\cdot,\cdot)$ a mapping from $L^2 \times H_0^2$ into L^2. Introduce

$$\text{(A.1)} \qquad \left|\operatorname{trace}\sigma(u)\right|^2 = \sum_{k=1}^{\infty} \lambda_k^2 \left\|\sigma_k(u)\right\|_2^2 \leqslant c\left(1 + \|u\|_2^2\right).$$

$$\text{(A.2)} \qquad \sum_{k=1}^{\infty} \left\|\sigma_k(u)\right\|^2 \quad \text{is bounded on bounded subsets of } H_0^2.$$

$$\text{(A.3)} \qquad \left|\operatorname{trace}\left(\sigma(u) - \sigma(v)\right)\right|^2 = \sum_{k=1}^{\infty} \lambda_k^2 \left\|\sigma_k(u) - \sigma_k(v)\right\|^2 \leqslant c\left(\|u - v\|^2\right).$$

$$\text{(A.4)} \qquad \left\|F(v, u)\right\| \leqslant c\left(1 + \|v\| + \|u\|_2\right).$$

$$\text{(A.5)} \qquad \left\|F(v_1, u_1) - F(v_2, u_2)\right\| \leqslant c\left(\|v_1 - v_2\| + \|u_1 - u_2\|_2\right).$$

Throughout this paper, we assume (A.1)–(A.5) are in place. The time interval we consider is fixed as $[0, T]$. We notice that the Cameron–Martin space $\mathcal{H}$ corresponding to the Wiener process W_t is given by

$$\mathcal{H} = \left\{ h_t = \sum_{k=1}^{\infty} \lambda_k h_t^k e_k;\ \sum_{k=1}^{\infty} \int_0^T \left(\dot{h}_s^k\right)^2 ds < \infty \right\}.$$

For $h \in \mathcal{H}$, let u_t^h denote the solution of the following deterministic PDE, the so called skeleton equation:

$$d\dot{u}_t^h = Au_t^h\,dt + B\left(u_t^h\right)dt + \sum_{k=1}^{\infty} \lambda_k \sigma_k\left(u_t^h\right)\dot{h}_t^k\,dt, + F\left(\dot{u}_t^h, u_t^h\right)dt,$$
$$u_0^h = \phi_0 \in H_0^2, \qquad \dot{u}_0^h = \phi_1 \in L^2. \tag{3}$$

For $h_t = \sum_{k=1}^{\infty} \lambda_k h_t^k e_k \in \mathcal{H} \subset C([0, T]; l^2)$, define

$$I(h) = \frac{1}{2} \sum_{k=1}^{\infty} \int_0^T \left(\dot{h}_t^k\right)^2 dt.$$

Set $I(h) = \infty$ if $h \in C([0, T]; l^2) \setminus \mathcal{H}$. Notice that $I(\cdot)$ is the rate function for the large deviations of the l^2-valued Brownian motion

$$W_t = \sum_{k=1}^{\infty} \lambda_k \beta_t^k e_k.$$

This is clear by considering the finite-dimensional version: $W_t^d = \sum_{k=1}^{d} \lambda_k \beta_t^k e_k$.

For $f \in C([0, T]; H_0^2) \times C([0, T]; L^2)$, introduce

$$\mathcal{L}_f = \left\{ h \in \mathcal{H};\ f(t) = \left(u_t^h, \dot{u}_t^h\right) \right\}.$$

Define

$$R(f) = \begin{cases} \inf_{h\in\mathcal{L}_f} I(h) & \text{if } \mathcal{L}_f \neq \emptyset, \\ +\infty & \text{if } \mathcal{L}_f = \emptyset. \end{cases}$$

Theorem 1. *Assume* (A.1)–(A.5). *Let μ_ε be the law of $(u^\varepsilon, \dot{u}^\varepsilon)$ on the product space $C([0,T]; H_0^2) \times C([0,T]; L^2)$. Then $\{\mu_\varepsilon, \varepsilon > 0\}$ satisfies a large deviation principle with rate function $R(f)$, i.e.,*

(1) *for every closed subset $C \subset C([0,T]; H_0^2) \times C([0,T]; L^2)$,*

$$\limsup_{\varepsilon\to 0} \varepsilon^2 \log \mu_\varepsilon(C) \leqslant -\inf_{f\in C} R(f); \tag{4}$$

(2) *for every open subset $G \subset C([0,T]; H_0^2) \times C([0,T]; L^2)$,*

$$\liminf_{\varepsilon\to 0} \varepsilon^2 \log \mu_\varepsilon(G) \geqslant -\inf_{f\in G} R(f). \tag{5}$$

3. The skeleton equation

The purpose of this section is to study the skeleton equation. For $h \in \mathcal{H}$, recall that u_t^h denote the solution of the following deterministic PDE, the so called skeleton equation:

$$d\dot{u}_t^h = Au_t^h\,dt + B\big(u_t^h\big)\,dt + \sum_{k=1}^{\infty} \lambda_k \sigma_k\big(u_t^h\big)\dot{h}_t^k\,dt + F\big(\dot{u}_t^h, u_t^h\big)\,dt,$$
$$u_0^h = \phi_0 \in H_0^2, \qquad \dot{u}_0^h = \phi_1 \in L^2. \tag{6}$$

For $a > 0$, we aim to show that the mapping $v^h = (u^h, \dot{u}^h)$ from $(\{h; I(h) \leqslant a\}, \|\cdot\|_\infty)$ into $C([0,T]; H_0^2) \times C([0,T]; H)$ is continuous, where $\|\cdot\|_\infty$ denotes the uniform norm on $C([0,T]; l^2)$.

Proposition 2. *The map: $v^h = (u_t^h, \dot{u}_t^h)$ from $(\{h; I(h) \leqslant a\}, \|\cdot\|_\infty)$ into $C([0,T]; H_0^2) \times C([0,T]; H)$ is continuous.*

Proof. Let $h^n \in \{h; I(h) \leqslant a\}$ with $\lim_{n\to\infty} \sup_{0\leqslant t\leqslant T} \|h_t^n - h_t\|_{l^2} = 0$. Define

$$e(t,u) = \frac{1}{2}\left\{\|\dot{u}_t\|^2 + \alpha\|\partial_x u_t\|^2 + \frac{\beta}{2}\|\partial_x u_t\|^4 + \gamma\left\|\partial_x^2 u_t\right\|^2\right\}. \tag{7}$$

By the energy equality proved in [9, Theorem 3.1] and (A.1), we have

$$e\big(t,u^h\big) = e\big(0,u^h\big) + \sum_{k=1}^{\infty} \lambda_k \int_0^t \big(\dot{u}_s^h, \sigma_k\big(u_s^h\big)\big)\dot{h}_s^k\,ds + \int_0^t \big(F\big(\dot{u}_s^h, u_s^h\big), \dot{u}_s^h\big)\,ds$$

$$
\begin{aligned}
&\leqslant e\big(0,u^h\big)+\int_0^t \big\|\dot u_s^h\big\|\Bigg(\sum_{k=1}^\infty \lambda_k^2\big\|\sigma_k\big(u_s^h\big)\big\|^2\Bigg)^{\frac12}\Bigg(\sum_{k=1}^\infty \big(\dot h_s^k\big)^2\Bigg)^{\frac12} ds\\
&\quad + c\int_0^t \big\|\dot u_s^h\big\|\big(1+\big\|\dot u_s^h\big\|+\big\|u_s^h\big\|_2\big)\,ds\\
&\leqslant e\big(0,u^h\big)+c\int_0^t \big\|\dot u_s^h\big\|\big(1+\big\|u_s^h\big\|_2\big)\Bigg(\sum_{k=1}^\infty \big(\dot h_s^k\big)^2\Bigg)^{\frac12} ds\\
&\quad + c_T + c\int_0^t e\big(s,u^h\big)\,ds\\
&\leqslant c_T + e\big(0,u^h\big)+c\int_0^t e\big(s,u^h\big)\Bigg(\Bigg(\sum_{k=1}^\infty \big(\dot h_s^k\big)^2\Bigg)^{\frac12}+1\Bigg)\,ds. \qquad (8)
\end{aligned}
$$

When h is fixed, it is easy to check that $\sup_{0\leqslant t\leqslant T} e(t,u^h)<\infty$. Thus, applying Gronwall's inequality, we get

$$\sup_{h\in\{h;I(h)\leqslant a\}}\ \sup_{0\leqslant t\leqslant T} e\big(t,u^h\big)=M<\infty. \qquad (9)$$

Observe that

$$
\begin{aligned}
||A\phi\|_{-2} &\leqslant (\gamma+\alpha)||\phi||_2, \quad \phi\in H_0^2, \qquad (10)\\
\big||B(\phi)-B(\psi)\big\| &\leqslant \beta\big\|\partial_x^2\phi\big\|\|\partial_x\phi+\partial_x\psi\|\|\partial_x\phi-\partial_x\psi\|\\
&\quad+\beta\|\partial_x\psi\|^2\big\|\partial_x^2\phi-\partial_x^2\psi\big\|, \quad \phi,\psi\in H_0^2. \qquad (11)
\end{aligned}
$$

Thus, (9) implies that there exist constants C_1 and C_2 such that

$$\sup_{h\in\{h;I(h)\leqslant a\}}\ \sup_{0\leqslant t\leqslant T}\big\|Au_t^h\big\|_{-2}\leqslant C_1, \qquad (12)$$

and

$$\big\|B\big(u_t^{h_1}\big)-B\big(u_t^{h_2}\big)\big\|\leqslant C_2\big\|u_t^{h_1}-u_t^{h_2}\big\|_2, \quad h_1,h_2\in\big\{h;I(h)\leqslant a\big\}. \qquad (13)$$

Regarded as an equation in H_0^{-2}, one has

$$\dot u_t^h=\phi_1+\int_0^t Au_s^h\,ds+\int_0^t B\big(u_s^h\big)\,ds+\sum_{k=1}^\infty\lambda_k\int_0^t\sigma_k\big(u_s^h\big)\dot h_s^k\,ds+\int_0^t F\big(\dot u_s^h,u_s^h\big)\,ds. \qquad (14)$$

By (9), (A.1) and (A.4), we have

$$\sum_{k=1}^{\infty}\lambda_k\int_s^t \|\sigma_k(u_l^h)\||\dot h_l^k|\,dl$$
$$\leqslant \int_s^t\left(\sum_{k=1}^{\infty}\lambda_k^2\|\sigma_k(u_l^h)\|^2\right)^{\frac12}\left(\sum_{k=1}^{\infty}(\dot h_l^k)^2\right)^{\frac12}dl$$
$$\leqslant c\int_s^t(1+\|u_l^h\|_2)\left(\sum_{k=1}^{\infty}(\dot h_l^k)^2\right)^{\frac12}dl\leqslant c\int_s^t e(l,u^h)\left(\sum_{k=1}^{\infty}(\dot h_l^k)^2\right)^{\frac12}dl$$
$$\leqslant C_{M,a}(t-s)^{\frac12}, \tag{15}$$

and

$$\int_s^t \|F(\dot u_l^h,u_l^h)\|\,dl\leqslant c\int_s^t\big(1+e(l,u^h)\big)\,dl\leqslant C_{M,a}(t-s), \tag{16}$$

for some constant $C_{M,a}$. Combining this with (12) and (13), we see that there exists a constant C_3 so that

$$\sup_{h\in\{h;I_d(h)\leqslant a\}}\|u_t^h-u_s^h\|_{-2}\leqslant C_3|t-s|^{\frac12}. \tag{17}$$

Introduce

$$e_L(t,v)=\frac12\left\{\|\dot v_t\|^2+\alpha\|\partial_x v_t\|^2+\gamma\left\|\partial_x^2 v_t\right\|^2\right\}.$$

Set $v_t^n=u_t^{h_n}-u_t^h$. Write

$$h_t^n=\sum_{k=1}^{\infty}\lambda_k h_t^{k,n}e_k.$$

Note that $\sup_{0\leqslant t\leqslant T}e(t,v)$ dominates the norm of $(v_t,\dot v_t)$ in the space of $C([0,T];H_0^2)\times C([0,T];H)$. Applying the energy inequality in [9, Lemma 3.1], we have

$$e_L(t,v^n)=\int_0^t\big(B(u_s^{h_n})-B(u_s^h),\dot v_s^n\big)\,ds+\int_0^t\big(F(\dot u_s^{h_n},u_s^{h_n})-F(\dot u_s^h,u_s^h),\dot v_s^n\big)\,ds$$
$$+\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot v_s^n,\sigma_k(u_s^{h_n})\dot h_s^{k,n}-\sigma_k(u_s^h)\dot h_s^k\big)\,ds. \tag{18}$$

In virtue of (13),

$$\big|\big(B(u_s^{h_n})-B(u_s^h),\dot v_s^n\big)\big|\leqslant\big\|B(u_s^{h_n})-B(u_s^h)\big\|\|\dot v_s^n\|\leqslant ce_L(s,v^n),$$

for some constant c. By the Lipschitz condition and the Sobolev imbedding,

$$\big|\big(F\big(\dot u_s^{h_n},u_s^{h_n}\big)-F\big(\dot u_s^{h},u_s^{h}\big),\dot v_s^n\big)\big|\leqslant c\big(\big\|\dot v_s^n\big\|^2+\|u_s^n-u_s^h\|_2^2\big)\leqslant ce_L\big(s,v^n\big). \tag{19}$$

Let $s_m=[ms]/m$. Write

$$\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot v_s^n,\sigma_k\big(u_s^{h_n}\big)\dot h_s^{k,n}-\sigma_k\big(u_s^h\big)\dot h_s^k\big)\,ds=C_t^1+C_t^2+C_t^3+C_t^4, \tag{20}$$

where

$$C_t^1=\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot v_s^n,\big(\sigma_k\big(u_s^{h_n}\big)-\sigma_k\big(u_s^h\big)\big)\dot h_s^{k,n}\big)\,ds,$$

$$C_t^2=\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot v_s^n-\dot v_{s_m}^n,\sigma_k\big(u_s^h\big)\big)\big(\dot h_s^{k,n}-\dot h_s^k\big)\,ds,$$

$$C_t^3=\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot v_{s_m}^n,\sigma_k\big(u_s^h\big)-\sigma_k\big(u_{s_m}^h\big)\big)\big(\dot h_s^{k,n}-\dot h_s^k\big)\,ds,$$

$$C_t^4=\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot v_{s_m}^n,\sigma_k\big(u_{s_m}^h\big)\big)\big(\dot h_s^{k,n}-\dot h_s^k\big)\,ds.$$

We now estimate each of the terms. Keeping (A.3) in mind, we have

$$\begin{aligned}\big|C_t^1\big|&\leqslant\int_0^t\big\|\dot v_s^n\big\|\left(\sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_s^{h_n}\big)-\sigma_k\big(u_s^h\big)\big\|^2\right)^{\frac12}\left(\sum_{k=1}^{\infty}\big(\dot h_s^{k,n}\big)^2\right)^{\frac12}ds\\&\leqslant c\int_0^t\big\|\dot v_s^n\big\|\big\|u_s^{h_n}-u_s^h\big\|_2\left(\sum_{k=1}^{\infty}\big(\dot h_s^{k,n}\big)^2\right)^{\frac12}ds\\&\leqslant c\int_0^t e_L\big(s,v^n\big)\left(\sum_{k=1}^{\infty}\big(\dot h_s^{k,n}\big)^2\right)^{\frac12}ds.\end{aligned} \tag{21}$$

In view of (17) and (A.1),

$$\big|C_t^2\big|\leqslant\int_0^t\big\|\dot v_s^n-\dot v_{s_m}^n\big\|_{-2}\left(\sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_s^h\big)\big\|_2^2\right)^{\frac12}\left(\sum_{k=1}^{\infty}\big(\dot h_s^{k,n}-\dot h_s^k\big)^2\right)^{\frac12}ds$$

$$
\begin{aligned}
&\leqslant c\frac{1}{\sqrt{m}}\int_0^t \left(1+\|u_s^h\|_2\right)\left(\sum_{k=1}^{\infty}\left(\dot h_s^{k,n}-\dot h_s^k\right)^2\right)^{\frac12} ds\\
&\leqslant c_{a,M}\frac{1}{\sqrt{m}}, \qquad (22)
\end{aligned}
$$

where M is defined as in (9). Since $\|\dot u_s^h\|$ is dominated by $\sqrt{e(s,u^h)}$, (9) implies that

$$
\sup_{h\in\{h;I(h)\leqslant a\}}\left\|u_t^h-u_s^h\right\|\leqslant C|t-s|. \qquad (23)
$$

This together with (A.3) implies that

$$
\begin{aligned}
\left|C_t^3\right| &\leqslant \int_0^t \left\|\dot v_{s_m}^n\right\|\left(\sum_{k=1}^{\infty}\lambda_k^2\left\|\sigma_k\left(u_s^h\right)-\sigma_k\left(u_{s_m}^h\right)\right\|^2\right)^{\frac12}\left(\sum_{k=1}^{\infty}\left(\dot h_s^{k,n}-\dot h_s^k\right)^2\right)^{\frac12} ds\\
&\leqslant c\int_0^t \left(\left\|u_s^h-u_{s_m}^h\right\|\right)\left(\sum_{k=1}^{\infty}\left(\dot h_s^{k,n}-\dot h_s^k\right)^2\right)^{\frac12} ds\\
&\leqslant c_{a,M}\frac{1}{m}. \qquad (24)
\end{aligned}
$$

Now

$$
\begin{aligned}
\left|C_t^4\right| &\leqslant \left|\sum_{k=1}^{\infty}\lambda_k\sum_{l=1}^{[mt]-1}\left(\dot v_{\frac{l}{m}}^n,\sigma_k\left(u_{\frac{l}{m}}^h\right)\right)\left(\left(h_{\frac{(l+1)}{m}}^{k,n}-h_{\frac{(l+1)}{m}}^k\right)-\left(h_{\frac{l}{m}}^{k,n}-h_{\frac{l}{m}}^k\right)\right)\right|\\
&\quad+\left|\sum_{k=1}^{\infty}\lambda_k\left(\dot v_{\frac{[mt]}{m}}^n,\sigma_k\left(u_{\frac{[mt]}{m}}^h\right)\right)\left(\left(h_t^{k,n}-h_t^k\right)-\left(h_{\frac{[mt]}{m}}^{k,n}-h_{\frac{[mt]}{m}}^k\right)\right)\right|\\
&\leqslant \sum_{l=1}^{[mt]-1}\left\|\dot v_{\frac{l}{m}}^n\right\|\left(\sum_{k=1}^{\infty}\left\|\sigma_k\left(u_{\frac{l}{m}}^h\right)\right\|^2\right)^{\frac12}\left(\sum_{k=1}^{\infty}\lambda_k^2\left(\left(h_{\frac{(l+1)}{m}}^{k,n}-h_{\frac{(l+1)}{m}}^k\right)-\left(h_{\frac{l}{m}}^{k,n}-h_{\frac{l}{m}}^k\right)\right)^2\right)^{\frac12}\\
&\quad+\left\|\dot v_{\frac{[mt]}{m}}^n\right\|\left(\sum_{k=1}^{\infty}\left\|\sigma_k\left(u_{\frac{[mt]}{m}}^h\right)\right\|^2\right)^{\frac12}\left(\sum_{k=1}^{\infty}\lambda_k^2\left(\left(h_t^{k,n}-h_t^k\right)-\left(h_{\frac{[mt]}{m}}^{k,n}-h_{\frac{[mt]}{m}}^k\right)\right)^2\right)^{\frac12}\\
&\leqslant c_{m,M}\sup_{0\leqslant t\leqslant T}\left\|h_t^n-h_t\right\|_{l^2}, \qquad (25)
\end{aligned}
$$

where we have used (9) and the assumption (A.2). Putting together (18)–(25) we arrive at

$$
e_L\left(t,v^n\right)\leqslant c\left(\frac{1}{\sqrt{m}}+c_{m,M}\sup_{0\leqslant t\leqslant T}\left\|h_t^n-h_t\right\|_{l^2}\right)+c\int_0^t e_L\left(s,v^n\right) ds
$$

$$+c\int_0^t e_L\big(s,v^n\big)\left(\sum_{k=1}^{\infty}\big(\dot{h}_s^{k,n}\big)^2\right)^{\frac{1}{2}} ds. \tag{26}$$

Applying the Gronwall's inequality, we get

$$\begin{aligned}\sup_{0\leqslant t\leqslant T} e_L\big(t,v^n\big) &\leqslant c\left(\frac{1}{\sqrt{m}}+c_{m,M}\sup_{0\leqslant t\leqslant T}\big\|h_t^n-h_t\big\|_{l^2}\right)\exp\left(cT+c\int_0^T\left(\sum_{k=1}^{\infty}\big(\dot{h}_s^{k,n}\big)^2\right)^{\frac{1}{2}} ds\right)\\ &\leqslant c_a\left(\frac{1}{\sqrt{m}}+c_{m,M}\sup_{0\leqslant t\leqslant T}\big\|h_t^n-h_t\big\|_{l^2}\right).\end{aligned} \tag{27}$$

Given $\varepsilon>0$. We first choose m such that $c_a\frac{1}{m}\leqslant\frac{\varepsilon}{2}$. Then for such a m, there exists N so that for $n\geqslant N$,

$$c_a c_{m,M}\sup_{0\leqslant t\leqslant T}\big\|h_t^n-h_t\big\|_{l^2}\leqslant\frac{\varepsilon}{2}. \tag{28}$$

Therefore, for $n\geqslant N$,

$$\sup_{0\leqslant t\leqslant T} e_L\big(t,u^{h^n}-u^h\big)\leqslant\varepsilon,$$

which finishes the proof of the proposition. □

Corollary 3. *The rate function $R(\cdot)$ defined in Section 2 is a good rate function, i.e., for every $a>0$, $\{g;\,R(g)\leqslant a\}$ is compact.*

Proof. Notice that

$$\big\{g;\ R(g)\leqslant a\big\}=\big\{\big(u^h,\dot{u}^h\big);\ I(h)\leqslant a\big\}.$$

So the corollary is a consequence of Proposition 2 and the fact that $\{h;\,I(h)\leqslant a\}$ is compact in $C([0,T];l^2)$. □

4. Large deviations

Consider

$$\begin{aligned} d\dot{u}_t^\varepsilon &= \left(\alpha+\beta\int_0^1\big|\partial_y u_t^\varepsilon\big|^2\,dy\right)\partial_x^2 u_t^\varepsilon\,dt-\gamma\partial_x^4 u_t^\varepsilon\,dt+\varepsilon\sum_{k=1}^{\infty}\lambda_k\sigma_k\big(u_t^\varepsilon\big)\,d\beta_t^k+\int_0^t F\big(\dot{u}_s^\varepsilon,u_s^\varepsilon\big)\,ds,\\ &u_t^\varepsilon(0)=u_t^\varepsilon(1)=0,\qquad \partial_x u_t^\varepsilon(0)=\partial_x u_t^\varepsilon(1)=0,\\ &u_0^\varepsilon(x)=\phi_0(x)\in H_0^2,\qquad \partial_t u_0^\varepsilon(x)=\phi_1(x)\in L^2.\end{aligned} \tag{29}$$

In this section, we will establish the large deviation principle. We first prepare a number of preliminary results. Let $e(t,u^\varepsilon)$ be defined as in (7) in Section 3.

Lemma 4. *It holds that*

$$\lim_{M\to\infty}\limsup_{\varepsilon\to 0}\varepsilon^2\log P\Big(\sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)>M\Big)=-\infty. \tag{30}$$

Proof. By the energy equality [9, (3.14)],

$$\begin{aligned} e\big(t,u^\varepsilon\big)&=e\big(0,u^\varepsilon\big)+\varepsilon\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot u_s^\varepsilon,\sigma_k\big(u_s^\varepsilon\big)\big)\,d\beta_s^k\\ &\quad+\frac12\varepsilon^2\int_0^t\sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_s^\varepsilon\big)\big\|^2\,ds+\int_0^t\big(\dot u_s^\varepsilon,F\big(\dot u_s^\varepsilon,u_s^\varepsilon\big)\big)\,ds. \end{aligned} \tag{31}$$

Recall that it is proved in [3,11] that there exists a universal constant c such that, for any $p\geqslant 2$ and any continuous martingale (M_t) with $M_0=0$, one has

$$\big\|M_t^*\big\|_p\leqslant cp^{\frac12}\big\|\langle M\rangle_t^{\frac12}\big\|_p, \tag{32}$$

where $M_t^*=\sup_{0\leqslant s\leqslant t}|M_s|$, and $\|\cdot\|_p$ stands for the $L^p(\Omega)$-norm. Using (32) and (A.1), we have for $p\geqslant 2$,

$$\begin{aligned} &\Big(E\Big[\sup_{0\leqslant t\leqslant l}e\big(t,u^\varepsilon\big)^p\Big]\Big)^{\frac1p}\\ &\quad\leqslant e\big(0,u^\varepsilon\big)+\frac12\varepsilon^2\left(E\left[\left|\int_0^l\sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_s^\varepsilon\big)\big\|^2\,ds\right|^p\right]\right)^{\frac1p}\\ &\qquad+\varepsilon\left(E\left[\sup_{0\leqslant t\leqslant l}\left|\sum_{k=1}^{\infty}\lambda_k\int_0^t\big(\dot u_s^\varepsilon,\sigma_k\big(u_s^\varepsilon\big)\big)\,d\beta_s^k\right|^p\right]\right)^{\frac1p}+\left(E\left[\left|\int_0^l\big|\big(\dot u_s^\varepsilon,F\big(\dot u_s^\varepsilon,u_s^\varepsilon\big)\big)\big|\,ds\right|^p\right]\right)^{\frac1p}\\ &\quad\leqslant e\big(0,u^\varepsilon\big)+c\frac12\varepsilon^2\left(E\left[\left|\int_0^l\big(1+\|u_s^\varepsilon\|_2^2\big)\,ds\right|^p\right]\right)^{\frac1p}\\ &\qquad+c\left(E\left[\left|\int_0^l\big(1+\|\dot u_s^\varepsilon\|^2+\|u_s^\varepsilon\|_2^2\big)\,ds\right|^p\right]\right)^{\frac1p}\\ &\qquad+\varepsilon cp^{\frac12}\left(E\left[\left|\int_0^l\sum_{k=1}^{\infty}\lambda_k^2\big(\dot u_s^\varepsilon,\sigma_k\big(u_s^\varepsilon\big)\big)^2\,ds\right|^{\frac p2}\right]\right)^{\frac1p}\\ &\quad\leqslant c_T+c\frac12\varepsilon^2\int_0^l\Big(1+\big(E\big[\|u_s^\varepsilon\|_2^{2p}\big]\big)^{\frac1p}\Big)\,ds+\varepsilon cp^{\frac12}\left(E\left[\left|\int_0^l\|\dot u_s^\varepsilon\|^2\big(1+\|u_s^\varepsilon\|_2^2\big)\,ds\right|^{\frac p2}\right]\right)^{\frac1p} \end{aligned}$$

$$+c\int_0^l \left(E\left[e\left(s,u^\varepsilon\right)^p\right]\right)^{\frac{1}{p}}\,ds$$

$$\leqslant c_T+\left(c\frac{1}{2}\varepsilon^2+c\right)\int_0^l \left(E\left[e\left(s,u^\varepsilon\right)^p\right]\right)^{\frac{1}{p}}\,ds+\varepsilon c p^{\frac{1}{2}}\left(\int_0^l \left(E\left[\left\|\dot u_s^\varepsilon\right\|^p\left(1+\left\|u_s^\varepsilon\right\|_2^2\right)^{\frac{p}{2}}\right]\right)^{\frac{2}{p}}\,ds\right)^{\frac{1}{2}}$$

$$\leqslant c_T+\left(c\frac{1}{2}\varepsilon^2+c\right)\int_0^l \left(E\left[e\left(s,u^\varepsilon\right)^p\right]\right)^{\frac{1}{p}}\,ds$$

$$+\varepsilon c p^{\frac{1}{2}}\left(\int_0^l \left(E\left[\frac{1}{2}\left\|\dot u_s^\varepsilon\right\|^{2p}+\frac{1}{2}\left(1+\left\|u_s^\varepsilon\right\|_2^2\right)^p\right]\right)^{\frac{2}{p}}\,ds\right)^{\frac{1}{2}}$$

$$\leqslant c_T+\left(c\frac{1}{2}\varepsilon^2+c\right)\int_0^l \left(E\left[e\left(s,u^\varepsilon\right)^p\right]\right)^{\frac{1}{p}}\,ds+\varepsilon c p^{\frac{1}{2}}\left(\int_0^l \left(E\left[e\left(s,u^\varepsilon\right)^p\right]\right)^{\frac{2}{p}}\,ds\right)^{\frac{1}{2}}, \tag{33}$$

where we have used the inequality

$$\left(E\left[\left|\int_0^l f_s\,ds\right|^m\right]\right)^{\frac{1}{m}}\leqslant\int_0^l \left(E\left[|f_s|^m\right]\right)^{\frac{1}{m}}\,ds$$

in several places for an appropriate m. Therefore,

$$\left(E\left[\sup_{0\leqslant t\leqslant l} e\left(t,u^\varepsilon\right)^p\right]\right)^{\frac{2}{p}}\leqslant c_T+\left(\varepsilon^2 cp+c\varepsilon^4+c^2\right)\int_0^l \left(E\left[e\left(s,u^\varepsilon\right)^p\right]\right)^{\frac{2}{p}}\,ds.$$

By Gronwall's inequality, there exist constants c_1 and c_2 so that

$$E\left[\sup_{0\leqslant t\leqslant T} e\left(t,u^\varepsilon\right)^p\right]\leqslant c_1^p e^{c_2\varepsilon^2p^2}. \tag{34}$$

This implies, by Chebyshev inequality,

$$P\left(\sup_{0\leqslant t\leqslant T} e\left(t,u^\varepsilon\right)>M\right)\leqslant M^{-p}c_1^p e^{c_2\varepsilon^2p^2}.$$

Letting $p=\frac{1}{\varepsilon^2}$, we get

$$\varepsilon^2\log P\left(\sup_{0\leqslant t\leqslant T} e\left(t,u^\varepsilon\right)>M\right)\leqslant\log\left(\frac{1}{M}\right)+\log(c_1)+c_2$$

which yields

$$\lim_{M\to\infty}\limsup_{\varepsilon\to 0}\varepsilon^2\log P\Big(\sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)>M\Big)=-\infty. \qquad \square \tag{35}$$

We need a result of exponential integrability for a Hilbert space-valued martingale.

Proposition 5. *Let $f_k(s), k\geqslant 1$, be a sequence of adapted L^2-valued stochastic processes. Assume that there exists a constant K such that*

$$\sum_{k=1}^{\infty}\big\|f_k(s)\big\|^2\leqslant K \quad \textit{almost surely for all } s\geqslant 0.$$

Define

$$M_t=\sum_{k=1}^{\infty}\int_0^t f_k(s)\,d\beta_s^k.$$

Then there exists a constant $\delta_0>0$ such that

$$\sup_{t\neq s}E\bigg[\exp\bigg(\delta_0\frac{\|M_t-M_s\|^2}{|t-s|}\bigg)\bigg]<\infty. \tag{36}$$

Proof. For simplicity, denote L^2 by H. Without loss of generality, we may assume $s=0$. Otherwise consider $Y_u=M_{s+u}-M_s$. For $g\in C^2(H)$, by Ito's formula,

$$\begin{aligned}\exp\big(g(M_t)\big)&=\exp\big(g(M_0)\big)+\sum_{k=1}^{\infty}\int_0^t\exp\big(g(M_s)\big)\big(g'(M_s),f_k(s)\big)\,d\beta_s^k\\&\quad+\frac{1}{2}\sum_{k=1}^{\infty}\int_0^t\exp\big(g(M_s)\big)\big(\big(g'(M_s)\otimes g'(M_s)+g''(M_s)\big)f_k(s),f_k(s)\big)\,ds.\end{aligned}\tag{37}$$

Put

$$h_s=\frac{1}{2}\sum_{k=1}^{\infty}\big(\big(g'(M_s)\otimes g'(M_s)+g''(M_s)\big)f_k(s),f_k(s)\big). \tag{38}$$

By integration by parts formula and (37), it is easy to verify that

$$N_t^g=\exp\bigg(g(M_t)-g(0)-\int_0^t h_s\,ds\bigg)$$

is a non-negative local martingale. Now, for $\lambda > 0$ (which will be specified later), let $g_\lambda(x) = (1+\lambda|x|_H^2)^{\frac{1}{2}}$. Then

$$g'_\lambda(x) = \lambda\big(1+\lambda|x|_H^2\big)^{-\frac{1}{2}}x,$$
$$g''_\lambda(x) = -\lambda^2\big(1+\lambda|x|_H^2\big)^{-\frac{3}{2}}x\otimes x + \lambda\big(1+\lambda|x|_H^2\big)^{-\frac{1}{2}}I_H,$$

where I_H stands for the identity operator. It is easy to see that

$$\sup_x\big|g'_\lambda(x)\big| \leqslant \lambda^{\frac{1}{2}}, \qquad \sup_x\big\|g''_\lambda(x)\big\| \leqslant \lambda^{\frac{1}{2}},$$

where $\|\cdot\|$ stands for the operator norm. Define h_s^λ as in (38) replacing g by g_λ. Then,

$$\big|h_s^\lambda\big| \leqslant \frac{1}{2}\lambda\sum_{k=1}^{\infty}\big\|f_k(s)\big\|^2 \leqslant \frac{1}{2}\lambda K. \tag{39}$$

For any $r > 0$ and every $\lambda > 0$, we have

$$\begin{aligned}
&P\left(\frac{\|M_t\|}{\sqrt{t}} > r\right)\\
&= P\left(g_\lambda(M_t) \geqslant \big(1+\lambda t r^2\big)^{\frac{1}{2}}\right)\\
&= P\left(g_\lambda(M_t) - g_\lambda(0) - \int_0^t h_s^\lambda\,ds + g_\lambda(0) + \int_0^t h_s^\lambda\,ds \geqslant \big(1+\lambda t r^2\big)^{\frac{1}{2}}\right)\\
&\leqslant P\left(g_\lambda(M_t) - g_\lambda(0) - \int_0^t h_s^\lambda\,ds \geqslant \big(1+\lambda t r^2\big)^{\frac{1}{2}} - g_\lambda(0) - \frac{1}{2}\lambda K t\right)\\
&\leqslant E\big[N_t^{g_\lambda}\big]\exp\left(-\big(1+\lambda t r^2\big)^{\frac{1}{2}} + g_\lambda(0) + \frac{1}{2}\lambda K t\right)\\
&\leqslant \exp\left(-\big(1+\lambda t r^2\big)^{\frac{1}{2}} + g_\lambda(0) + \frac{1}{2}\lambda K t\right),
\end{aligned} \tag{40}$$

where the fact $E[N_t^{g_\lambda}] \leqslant 1$ has been used. Choosing $\lambda = t^{-1}\delta r^2$, we get that

$$P\left(\frac{\|M_t\|}{\sqrt{t}} > r\right) \leqslant \exp\left(-\delta^{\frac{1}{2}}r^2 + 1 + \frac{1}{2}K\delta r^2\right) = \exp\left(-\delta\left(\frac{1}{\delta^{\frac{1}{2}}} - \frac{1}{2}K\right)r^2\right).$$

Take $\delta > 0$ small enough so that $\delta^* := \delta(1/\delta^{1/2} - \frac{1}{2}K) > 0$. We arrive at

$$P\left(\frac{\|M_t\|}{\sqrt{t}} > r\right) \leqslant \exp\big(-\delta^* r^2 + 1\big), \tag{41}$$

where δ^* is independent of t. We can now easily deduce (36). Fix $\delta_0 < \delta^*$ and let $\xi = \|M_t\|/\sqrt{t}$. We have

$$E\left[\exp\left(\delta_0 \frac{\|M_t\|^2}{|t|}\right)\right] = E\left[\exp\left(\delta_0 \xi^2\right)\right] = -\int_0^\infty \exp\left(\delta_0 r^2\right) dP(\xi > r)$$

$$\leqslant 1 + 2\delta_0 \int_0^\infty \exp\left(\delta_0 r^2\right) P(\xi > r) r\, dr$$

$$\leqslant 1 + 2\delta_0 \int_0^\infty \exp\left(\delta_0 r^2\right) \exp\left(-\delta^* r^2 + 1\right) r\, dr < +\infty \tag{42}$$

which completes the proof of the proposition. □

Lemma 6. *For every $\delta_1 > 0$, $M > 0$,*

$$\lim_{m\to\infty} \limsup_{\varepsilon\to 0} \varepsilon^2 \log P\Big(\sup_{|s-t|\leqslant \frac{1}{m}, s,t\leqslant T} \left\|\dot{u}_s^\varepsilon - \dot{u}_t^\varepsilon\right\|_{-2} \geqslant \delta_1,\ \sup_{0\leqslant t\leqslant T} e\left(t, u^\varepsilon\right) \leqslant M\Big) = -\infty. \tag{43}$$

Proof. In view of (10), (11) and (A.4), $\|Au_s^\varepsilon\|_{-2}$, $\|B(u_s^\varepsilon)\|$ and $\|F(\dot{u}_s^\varepsilon, u_s^\varepsilon)\|$ are uniformly bounded by a constant (depending only on M) on the set $\{\omega, t \leqslant T;\ \sup_{0\leqslant t\leqslant T} e(t, u^\varepsilon) \leqslant M\}$. Regarded as an equation in H_0^{-2},

$$\dot{u}_t^\varepsilon = \phi_1 + \int_0^t Au_s^\varepsilon\, ds + \int_0^t B\left(u_s^\varepsilon\right) ds + \varepsilon \sum_{k=1}^\infty \lambda_k \int_0^t \sigma_k\left(u_s^\varepsilon\right) d\beta_s^k + \int_0^t F\left(\dot{u}_s^\varepsilon, u_s^\varepsilon\right) ds. \tag{44}$$

Therefore, on $\{\omega;\ \sup_{0\leqslant t\leqslant T} e(t, u^\varepsilon) \leqslant M\}$,

$$\sup_{|s-t|\leqslant \frac{1}{m}, s,t\leqslant T} \left\|\dot{u}_t^\varepsilon - \dot{u}_s^\varepsilon\right\|_{-2} \leqslant c_M \frac{1}{m} + \varepsilon \sup_{|s-t|\leqslant \frac{1}{m}, s,t\leqslant T} \|N_t - N_s\|, \tag{45}$$

where

$$N_t = \sum_{k=1}^\infty \lambda_k \int_0^t \sigma_k\left(u_s^\varepsilon\right) d\beta_s^k.$$

Thus for sufficiently big m,

$$P\Big(\sup_{|s-t|\leqslant \frac{1}{m}, s,t\leqslant T} \left\|\dot{u}_s^\varepsilon - \dot{u}_t^\varepsilon\right\|_{-2} \geqslant \delta_1,\ \sup_{0\leqslant t\leqslant T} e\left(t, u^\varepsilon\right) \leqslant M\Big)$$

$$\leqslant P\left(\varepsilon \sup_{|s-t|\leqslant \frac{1}{m}, s,t\leqslant T} \|N_t - N_s\| \geqslant \frac{1}{2}\delta_1,\ \sup_{0\leqslant t\leqslant T} e\left(t, u^\varepsilon\right) \leqslant M\right).$$

So it remains to show

$$\lim_{m\to\infty}\limsup_{\varepsilon\to 0}\varepsilon^2\log P\left(\varepsilon\sup_{|s-t|\leqslant\frac{1}{m},s,t\leqslant T}\|N_t-N_s\|\geqslant\frac{1}{2}\delta_1,\ \sup_{0\leqslant t\leqslant T}e\big(t,u^\varepsilon\big)\leqslant M\right)=-\infty. \tag{46}$$

Notice that by (A.1) there exists a constant K_M such that

$$\Big\{\omega;\ \sup_{0\leqslant t\leqslant T}e\big(t,u^\varepsilon\big)\leqslant M\Big\}\subset\left\{\omega;\ \sup_{0\leqslant t\leqslant T}\sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_t^\varepsilon\big)\big\|^2\leqslant K_M\right\}. \tag{47}$$

Define

$$\tau=\inf\left\{s\geqslant 0;\ \sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_s^\varepsilon\big)\big\|^2>K_M\right\}.$$

It follows from (47) that

$$\Big\{\omega;\ \sup_{0\leqslant t\leqslant T}e\big(t,u^\varepsilon\big)\leqslant M\Big\}\subset\{\tau\geqslant T\}.$$

Therefore,

$$\begin{aligned}&P\left(\varepsilon\sup_{|s-t|\leqslant\frac{1}{m},s,t\leqslant T}\|N_t-N_s\|\geqslant\frac{1}{2}\delta_1,\ \sup_{0\leqslant t\leqslant T}e\big(t,u^\varepsilon\big)\leqslant M\right)\\&\quad\leqslant P\left(\varepsilon\sup_{|s-t|\leqslant\frac{1}{m},s,t\leqslant T}\|N_{t\wedge\tau}-N_{s\wedge\tau}\|\geqslant\frac{1}{2}\delta_1\right).\end{aligned}$$

So, to prove (46), we may drop the event $\{\sup_{0\leqslant t\leqslant T}e(t,u^\varepsilon)\leqslant M\}$ and assume in the rest of the proof that

$$\sup_{0\leqslant t\leqslant T}\sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_t^\varepsilon\big)\big\|^2\leqslant K_M. \tag{48}$$

Applying Proposition 5, there exists a constant $\lambda_M>0$ such that

$$\sup_{t\neq s,s,t\leqslant T}E\left[\exp\left(\lambda_M\frac{\|N_t-N_s\|^2}{|t-s|}\right)\right]<\infty. \tag{49}$$

Introduce

$$D=\int_0^T\int_0^T\exp\left(\lambda_M\frac{\|N_t-N_s\|^2}{|t-s|}\right)ds\,dt.$$

Then we have $E[D] < \infty$. Now by Garsia lemma (see [21]) we have

$$\|N_t - N_s\| \leqslant \frac{8}{\sqrt{\lambda_M}} \int_0^{|t-s|} \left(\log \frac{D}{u^2}\right)^{\frac{1}{2}} dp(u), \tag{50}$$

where $p(u) = u^{\frac{1}{2}}$. For any $\delta < \frac{1}{2}$, say $\delta = \frac{1}{4}$, (50) implies that there exists a constant c such that

$$\|N_t - N_s\| \leqslant \frac{8c}{\sqrt{\lambda_M}} (\sqrt{\log D} + 1)|t-s|^{\frac{1}{4}}.$$

Consequently,

$$\sup_{|s-t| \leqslant \frac{1}{m}, s,t \leqslant T} \|N_t - N_s\| \leqslant \frac{8c}{\sqrt{\lambda_M}} (\sqrt{\log D} + 1)\left(\frac{1}{m}\right)^{\frac{1}{4}}.$$

Therefore,

$$\begin{aligned} P\left(\varepsilon \sup_{|s-t| \leqslant \frac{1}{m}, s,t \leqslant T} \|N_t - N_s\| \geqslant \frac{1}{2}\delta_1\right) &\leqslant P\left(\sqrt{\log D} > \frac{1}{2}\frac{\delta_1}{\varepsilon}\frac{\sqrt{\lambda_M}}{8c}\frac{1}{(\frac{1}{m})^{\frac{1}{4}}} - 1\right) \\ &\leqslant P\left(D > \exp\left(\frac{1}{2}\frac{\delta_1}{\varepsilon}\frac{\sqrt{\lambda_M}}{8c}\frac{1}{(\frac{1}{m})^{\frac{1}{4}}} - 1\right)^2\right) \\ &\leqslant E[D]\exp\left\{-\left(\frac{1}{2}\frac{\delta_1}{\varepsilon}\frac{\sqrt{\lambda_M}}{8c}\frac{1}{(\frac{1}{m})^{\frac{1}{4}}} - 1\right)^2\right\}. \end{aligned}$$

This yields

$$\lim_{m\to\infty} \limsup_{\varepsilon\to 0} \varepsilon^2 P\left(\varepsilon \sup_{|s-t| \leqslant \frac{1}{m}, s,t \leqslant T} \|N_t - N_s\| \geqslant \frac{1}{2}\delta_1\right) = -\infty,$$

which completes the proof. □

Lemma 7. *Let $f_s, s \geqslant 0$ be a H_0^{-2}-valued adapted stochastic process such that*

$$\sup_{0 \leqslant s \leqslant T} \|f_s\|_{-2} \leqslant \delta_1.$$

Set

$$M_t = \sum_{k=1}^{\infty} \lambda_k \int_0^t \langle f_s, \sigma_k(u_s^\varepsilon)\rangle d\beta_s^k. \tag{51}$$

Then there exist positive constants $c_1 > 0$, $c_2 > 0$ and c_M such that for $\eta_1 > 0$,

$$P\Big(\sup_{0\leqslant t\leqslant T}|M_t| > \eta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big)\leqslant c_1\exp\bigg(-\frac{c_2\eta_1^2}{c_M T\delta_1^2}\bigg). \tag{52}$$

Proof. Notice that $M_t, t \geqslant 0$ is a martingale whose bracket satisfies

$$\begin{aligned}
\langle M\rangle_t &= \sum_{k=1}^{\infty}\lambda_k^2\int_0^t \big\langle f_s,\sigma_k\big(u_s^\varepsilon\big)\big\rangle^2\,ds\\
&\leqslant \int_0^t \|f_s\|_{-2}^2\sum_{k=1}^{\infty}\lambda_k^2\big\|\sigma_k\big(u_s^\varepsilon\big)\big\|_2^2\,ds\\
&\leqslant c\int_0^t \|f_s\|_{-2}^2\big(1+\big\|u_s^\varepsilon\big\|_2^2\big)\,ds\\
&\leqslant c\delta_1^2\int_0^t e\big(s,u^\varepsilon\big)\,ds.
\end{aligned} \tag{53}$$

Thus, on $\{\omega, \sup_{0\leqslant t\leqslant T} e(t,u^\varepsilon)\leqslant M\}$, $\langle M\rangle_t \leqslant c_M\delta_1^2 =: b$. By the martingale representation theorem, there exists a standard Brownian motion $B_s, s\geqslant 0$ such that $M_t = B_{\langle M\rangle_t}$. We have

$$\begin{aligned}
&\Big\{\omega,\ \sup_{0\leqslant t\leqslant T}|M_t| > \eta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big\}\\
&\quad=\Big\{\omega,\ \sup_{0\leqslant t\leqslant T}|B_{\langle M\rangle_t}| > \eta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big\}\\
&\quad\subset\Big\{\omega;\ \sup_{0\leqslant u\leqslant b}|B_u| > \eta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big\}\\
&\quad=\Big\{\omega;\ \sqrt{b}\sup_{0\leqslant u\leqslant 1}|\tilde B_u| > \eta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big\},
\end{aligned}$$

where $\tilde B$ is another Brownian motion by the scaling invariance property. It is well known that there exists a constant $c_2 > 0$ such that $c_1 := E[\exp(c_2\sup_{0\leqslant u\leqslant 1}|\tilde B_u|^2)] < \infty$. Thus,

$$\begin{aligned}
P\Big(\sup_{0\leqslant t\leqslant T}|M_t| > \eta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big) &\leqslant P\bigg(\sup_{0\leqslant u\leqslant 1}|\tilde B_u| > \frac{\eta_1}{\sqrt{b}}\bigg)\\
&\leqslant E\Big[\exp\Big(c_2\sup_{0\leqslant u\leqslant 1}|\tilde B_u|^2\Big)\Big]\exp\bigg(-\frac{c_2\eta_1^2}{b}\bigg)\\
&= c_1\exp\bigg(-\frac{c_2\eta_1^2}{c_M\delta_1^2}\bigg),
\end{aligned}$$

proving the lemma. □

Recall

$$e_L(t,v)=\frac{1}{2}\left\{\|\dot v_t\|^2+\alpha\|\partial_x v_t\|^2+\gamma\left\|\partial_x^2 v_t\right\|^2\right\}. \tag{54}$$

Theorem 8. *For every $\eta>0$, $R>0$, $h\in\mathcal{H}$, there exists $\delta>0$ such that*

$$\limsup_{\varepsilon\to 0}\varepsilon^2\log P\Big(\sup_{0\leqslant t\leqslant T}e_L\big(t,u^\varepsilon-u^h\big)\geqslant\eta,\ \sup_{0\leqslant t\leqslant T}\|\varepsilon W-h\|_\infty<\delta\Big)\leqslant -R. \tag{55}$$

Proof. As an equation in H_0^{-2}, we have

$$\begin{aligned}\dot u_t^\varepsilon-\dot u_t^h&=\int_0^t A\big(u_s^\varepsilon-u_s^h\big)\,ds+\int_0^t\big(B\big(u_s^\varepsilon\big)-B\big(u_s^h\big)\big)\,ds\\&\quad+\varepsilon\sum_{k=1}^\infty\lambda_k\int_0^t\sigma_k\big(u_s^\varepsilon\big)\,d\beta_s^k-\sum_{k=1}^\infty\lambda_k\int_0^t\sigma_k\big(u_s^h\big)\dot h_s^k\,ds\\&\quad+\int_0^t\big(F\big(\dot u_s^\varepsilon,u_s^\varepsilon\big)-F\big(\dot u_s^h,u_s^h\big)\big)\,ds.\end{aligned} \tag{56}$$

For simplicity, denote $v_t:=u_t^\varepsilon-u_t^h$. By the energy inequality (3.10) in [9], we have

$$\begin{aligned}e_L(t,v)&=\int_0^t\big(B\big(u_s^\varepsilon\big)-B\big(u_s^h\big),\dot v_s\big)\,ds+\frac{1}{2}\varepsilon^2\int_0^t\sum_{k=1}^\infty\lambda_k^2\big\|\sigma_k\big(u_s^\varepsilon\big)\big\|^2\,ds\\&\quad+\sum_{k=1}^\infty\lambda_k\int_0^t\big(\dot v_s,\varepsilon\sigma_k\big(u_s^\varepsilon\big)\big)\,d\beta_s^k-\sum_{k=1}^\infty\lambda_k\int_0^t\big(\dot v_s,\sigma_k\big(u_s^h\big)\big)\dot h_s^k\,ds\\&\quad+\int_0^t\big(F\big(\dot u_s^\varepsilon,u_s^\varepsilon\big)-F\big(\dot u_s^h,u_s^h\big),\dot v_s\big)\,ds.\end{aligned} \tag{57}$$

Let $M>\sup_{0\leqslant t\leqslant T}e(t,u^h)$. In view of (11) and (A.1), on the event $\{\omega,\sup_{0\leqslant t\leqslant T}e(t,u^\varepsilon)\leqslant M\}$,

$$\begin{aligned}&\big\|B\big(u_s^\varepsilon\big)-B\big(u_s^h\big)\big\|\leqslant C_M\big(\big\|u_s^\varepsilon-u_s^h\big\|_2\big),\\&\sum_{k=1}^\infty\lambda_k^2\big\|\sigma_k\big(u_s^\varepsilon\big)\big\|^2\leqslant C_M.\end{aligned}$$

By the Lipschitz condition, (A.5) and the Sobolev imbedding,

$$\big|\big(F\big(\dot u_s^\varepsilon,u_s^\varepsilon\big)-F\big(\dot u_s^h,u_s^h\big),\dot v_s\big)\big|\leqslant ce_L(s,v).$$

So it follows from (57) that on $\{\omega, \sup_{0\leqslant t\leqslant T} e(t,u^\varepsilon)\leqslant M\}$,

$$e_L(t,v)\leqslant C_M\int_0^t e_L(s,v)\,ds+\frac{1}{2}\varepsilon^2 c_M T+\left|M_t^\varepsilon\right|, \tag{58}$$

where

$$M_t^\varepsilon=\sum_{k=1}^\infty \lambda_k\int_0^t \left(\dot v_s, \varepsilon\sigma_k\left(u_s^\varepsilon\right)\right)d\beta_s^k-\sum_{k=1}^\infty \lambda_k\int_0^t \left(\dot v_s, \sigma_k\left(u_s^h\right)\right)\dot h_s^k\,ds. \tag{59}$$

Let $s_m=[ms]/m$ and write

$$M_t^\varepsilon=N_t^{1,m}+N_t^{2,m}+N_t^{3,m}+N_t^{4,m},$$

where

$$N_t^{1,m}=\sum_{k=1}^\infty \lambda_k\int_0^t \left(\dot v_s, \sigma_k\left(u_s^\varepsilon\right)-\sigma_k\left(u_s^h\right)\right)\dot h_s^k\,ds,$$

$$N_t^{2,m}=\sum_{k=1}^\infty \lambda_k\int_0^t \left(\dot v_s-\dot v_{s_m}, \sigma_k\left(u_s^\varepsilon\right)\right)\left(\varepsilon\,d\beta_s^k-\dot h_s^k\,ds\right),$$

$$N_t^{3,m}=\sum_{k=1}^\infty \lambda_k\int_0^t \left(\dot v_{s_m}, \sigma_k\left(u_s^\varepsilon\right)-\sigma_k\left(u_{s_m}^\varepsilon\right)\right)\left(\varepsilon\,d\beta_s^k-\dot h_s^k\,ds\right),$$

$$N_t^{4,m}=\sum_{k=1}^\infty \lambda_k\int_0^t \left(\dot v_{s_m}, \sigma_k\left(u_{s_m}^\varepsilon\right)\right)\left(\varepsilon\,d\beta_s^k-\dot h_s^k\,ds\right).$$

Now,

$$\begin{aligned}\left|N_t^{1,m}\right| &\leqslant \int_0^t \|\dot v_s\|\left(\sum_{k=1}^\infty \lambda_k^2\left\|\sigma_k\left(u_s^\varepsilon\right)-\sigma_k\left(u_s^h\right)\right\|^2\right)^{\frac12}\left(\sum_{k=1}^\infty\left(\dot h_s^k\right)^2\right)^{\frac12}ds\\ &\leqslant c\int_0^t \|\dot v_s\|\left(\left\|u_s^\varepsilon-u_s^h\right\|_2\right)\left(\sum_{k=1}^\infty\left(\dot h_s^k\right)^2\right)^{\frac12}ds\\ &\leqslant c\int_0^t e_L(s,v)\left(\sum_{k=1}^\infty\left(\dot h_s^k\right)^2\right)^{\frac12}ds. \end{aligned} \tag{60}$$

So we deduce from (58) that on $\{\omega, \sup_{0\leqslant t\leqslant T} e(t,u^\varepsilon)\leqslant M\}$,

$$e_L(t,v)\leqslant C_M\int_0^t e_L(s,v)\left(1+\left(\sum_{k=1}^{\infty}\left(\dot h_s^k\right)^2\right)^{\frac12}\right)ds+\frac12\varepsilon^2 c_M T+\sum_{i=2}^{4}\left|N_t^{i,m}\right|. \tag{61}$$

By Gronwall's inequality we obtain that on $\{\omega, \sup_{0\leqslant t\leqslant T} e(t,u^\varepsilon)\leqslant M\}$,

$$e_L(t,v)\leqslant e^{C_M T}\left(\frac12\varepsilon^2 c_M T+\sum_{i=2}^{4}\sup_{0\leqslant t\leqslant T}\left|N_t^{i,m}\right|\right)$$

which implies that there exists a constant $\varepsilon_1>0$ such that for $\varepsilon\leqslant\varepsilon_1$,

$$\begin{aligned}&P\Big(\sup_{0\leqslant t\leqslant T} e_L\big(t,u^\varepsilon-u^h\big)\geqslant\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M,\ \|\varepsilon W-h\|_\infty<\delta\Big)\\&\quad\leqslant P\left(\sum_{i=2}^{4}\sup_{0\leqslant t\leqslant T}\left|N_t^{i,m}\right|>\frac12 e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M,\ \|\varepsilon W-h\|_\infty<\delta\right).\end{aligned}\tag{62}$$

Now,

$$\begin{aligned}&P\left(\sum_{i=2}^{4}\sup_{0\leqslant t\leqslant T}\left|N_t^{i,m}\right|>\frac12 e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M,\ \sup_{0\leqslant t\leqslant T}\|\varepsilon W-h\|_\infty<\delta\right)\\&\quad\leqslant P\Big(\sup_{0\leqslant t\leqslant T}\left|N_t^{4,m}\right|>\frac16 e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M,\ \sup_{0\leqslant t\leqslant T}\|\varepsilon W-h\|_\infty<\delta\Big) \qquad (63)\\&\qquad+P\Big(\sup_{0\leqslant t\leqslant T}\left|N_t^{2,m}\right|>\frac16 e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big)\\&\qquad+P\Big(\sup_{0\leqslant t\leqslant T}\left|N_t^{3,m}\right|>\frac16 e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big).\end{aligned}\tag{64}$$

Furthermore, for $\delta_1>0$,

$$\begin{aligned}&P\Big(\sup_{0\leqslant t\leqslant T}\left|N_t^{2,m}\right|>\frac16 e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big)\\&\quad\leqslant P\Big(\sup_{0\leqslant t\leqslant T}\left|N_t^{2,m}\right|>\frac16 e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M,\ \sup_{|t-s|\leqslant\frac1m}\|\dot v_t-\dot v_s\|_{-2}\leqslant\delta_1\Big)\\&\qquad+P\Big(\sup_{|t-s|\leqslant\frac1m}\|\dot v_t-\dot v_s\|_{-2}>\delta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t,u^\varepsilon\big)\leqslant M\Big).\end{aligned}\tag{65}$$

By the Girsanov theorem, we know that $W_t - \frac{1}{\varepsilon} h_t$, $t \geqslant 0$ is a Wiener process under the probability measure P^* given by

$$\frac{dP^*}{dP}\bigg|_{\mathcal{F}_t} = \exp\left(\frac{1}{\varepsilon}\int_0^t \langle \dot{h}_s, dW_s\rangle - \frac{1}{2}\frac{1}{\varepsilon^2}\sum_{k=1}^{\infty}\lambda_k^4 \int_0^t \left(\dot{h}_s^k\right)^2 ds\right).$$

Through a change of measure and applying Lemma 7, we can show that there exist constants c_1, c_2 such that

$$\begin{aligned} &P\left(\sup_{0\leqslant t\leqslant T}\left|N_t^{2,m}\right| > \frac{1}{4}e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\left(t, u^\varepsilon\right) \leqslant M,\ \sup_{|t-s|\leqslant \frac{1}{m}} \|\dot{v}_t - \dot{v}_s\|_{-2} \leqslant \delta_1\right) \\ &\quad \leqslant c_1 \exp\left(-\frac{c_2\eta^2}{cMT\varepsilon^2\delta_1^2}\right). \end{aligned} \tag{66}$$

Notice that on $\{\omega; \sup_{0\leqslant t\leqslant T} e(t, u^\varepsilon) \leqslant M\}$,

$$\begin{aligned} \sum_{k=1}^{\infty}\lambda_k^2\left(\dot{v}_{s_m}, \sigma_k\left(u_s^\varepsilon\right) - \sigma_k\left(u_{s_m}^\varepsilon\right)\right)^2 &\leqslant \|\dot{v}_{s_m}\|^2 \sum_{k=1}^{\infty}\lambda_k^2\left\|\sigma_k\left(u_s^\varepsilon\right) - \sigma_k\left(u_{s_m}^\varepsilon\right)\right\|^2 \\ &\leqslant c\|\dot{v}_{s_m}\|^2\left(\left\|u_s^\varepsilon - u_{s_m}^\varepsilon\right\|^2\right) \\ &\leqslant \|\dot{v}_{s_m}\|^2\left(\left\|\int_{s_m}^{s} \dot{u}_l^\varepsilon\, dl\right\|^2\right) \leqslant C_M\left(\frac{1}{m}\right)^2. \end{aligned} \tag{67}$$

Using this and following the same proof of Lemma 7, we can show that

$$\begin{aligned} &P\left(\sup_{0\leqslant t\leqslant T}\left|N_t^{3,m}\right| > \frac{1}{6}e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\left(t, u^\varepsilon\right) \leqslant M\right) \\ &\quad \leqslant c_1 \exp\left(-\frac{c_2\eta^2}{cMT\varepsilon^2(\frac{1}{m})^2}\right). \end{aligned} \tag{68}$$

Now given $R > 0$, $\eta > 0$. According to Lemma 4, we can choose M large enough and $\varepsilon_2 > 0$ such that for $\varepsilon \leqslant \varepsilon_2$,

$$P\left(\sup_{0\leqslant t\leqslant T} e\left(t, u^\varepsilon\right) > M\right) \leqslant \exp\left(-\frac{R}{\varepsilon^2}\right). \tag{69}$$

Next, we choose δ_1 , according to (66), so that for $\varepsilon \leqslant \varepsilon_3$

$$\begin{aligned} &P\left(\sup_{0\leqslant t\leqslant T}\left|N_t^{2,m}\right| > \frac{1}{4}e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\left(t, u^\varepsilon\right) \leqslant M,\ \sup_{|t-s|\leqslant \frac{1}{m}} \|\dot{v}_t - \dot{v}_s\|_{-2} \leqslant \delta_1\right) \\ &\quad \leqslant \exp\left(-\frac{R}{\varepsilon^2}\right), \end{aligned} \tag{70}$$

where ε_3 is a positive number . For such a $\delta_1 > 0$, by Lemma 6 and (68) there exist an integer m and $\varepsilon_4 > 0$ so that for $\varepsilon \leqslant \varepsilon_4$,

$$P\Big(\sup_{|t-s|\leqslant\frac{1}{m}} \|\dot{v}_t - \dot{v}_s\|_{-2} > \delta_1,\ \sup_{0\leqslant t\leqslant T} e\big(t, u^\varepsilon\big) \leqslant M\Big) \leqslant \exp\Big(-\frac{R}{\varepsilon^2}\Big), \tag{71}$$

$$P\bigg(\sup_{0\leqslant t\leqslant T} \big|N_t^{3,m}\big| > \frac{1}{6}e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t, u^\varepsilon\big) \leqslant M\bigg) \leqslant \exp\Big(-\frac{R}{\varepsilon^2}\Big). \tag{72}$$

When such an m is fixed,

$$\begin{aligned} N_t^{4,m} &= \sum_{k=1}^{\infty}\lambda_k \sum_{l=0}^{[mt]-1} \big(\dot{v}_{\frac{l}{m}}, \sigma_k\big(u^\varepsilon_{\frac{l}{m}}\big)\big)\big[\big(\varepsilon\beta^k_{\frac{l+1}{m}} - h^k_{\frac{l+1}{m}}\big) - \big(\varepsilon\beta^k_{\frac{l}{m}} - h^k_{\frac{l}{m}}\big)\big] \\ &\quad + \sum_{k=1}^{\infty}\lambda_k\big(\dot{v}_{\frac{[mt]}{m}}, \sigma_k\big(u^\varepsilon_{\frac{[mt]}{m}}\big)\big)\big[\big(\varepsilon\beta^k_t - h^k_t\big) - \big(\varepsilon\beta^k_{\frac{[mt]}{m}} - h^k_{\frac{[mt]}{m}}\big)\big]. \end{aligned} \tag{73}$$

Therefore,

$$\begin{aligned} \big|N_t^{4,m}\big| &\leqslant \sum_{l=0}^{[mt]-1} \|\dot{v}_{\frac{l}{m}}\| \bigg(\sum_{k=1}^{\infty}\big\|\sigma_k\big(u^\varepsilon_{\frac{l}{m}}\big)\big\|^2\bigg)^{1/2} \bigg(\sum_{k=1}^{\infty}\lambda_k^2\big[\big(\varepsilon\beta^k_{\frac{l+1}{m}} - h^k_{\frac{l+1}{m}}\big) - \big(\varepsilon\beta^k_{\frac{l}{m}} - h^k_{\frac{l}{m}}\big)\big]^2\bigg)^{1/2} \\ &\quad + \big\|\dot{v}_{\frac{[mt]}{m}}\big\| \bigg(\sum_{k=1}^{\infty}\big\|\sigma_k\big(u^\varepsilon_{\frac{[mt]}{m}}\big)\big\|^2\bigg)^{1/2} \bigg(\sum_{k=1}^{\infty}\lambda_k^2\big[\big(\varepsilon\beta^k_t - h^k_t\big) - \big(\varepsilon\beta^k_{\frac{[mt]}{m}} - h^k_{\frac{[mt]}{m}}\big)\big]^2\bigg)^{1/2} \\ &\leqslant \sum_{l=0}^{[mt]-1} \big\|\dot{v}_{\frac{l}{m}}\big\| \bigg(\sum_{k=1}^{\infty}\big\|\sigma_k\big(u^\varepsilon_{\frac{l}{m}}\big)\big\|^2\bigg)^{1/2} \|\varepsilon W - h\|_\infty \\ &\quad + \big\|\dot{v}_{\frac{[mt]}{m}}\big\| \bigg(\sum_{k=1}^{\infty}\big\|\sigma_k\big(u^\varepsilon_{\frac{[mt]}{m}}\big)\big\|^2\bigg)^{1/2} \|\varepsilon W - h\|_\infty. \end{aligned} \tag{74}$$

By the assumption (A.2), we know that on the event $\{\omega, \sup_{0\leqslant t\leqslant T} e(t, u^\varepsilon) \leqslant M\}$,

$$\sup_{0\leqslant s\leqslant T}\sum_{k=1}^{\infty}\big\|\sigma_k\big(u_s^\varepsilon\big)\big\|^2 \quad\text{and}\quad \sup_{0\leqslant s\leqslant T}\|\dot{v}_s\|$$

are uniformly bounded by some constant c_M. Thus, we see from (74) that there exists $\delta > 0$ such that the event

$$\bigg\{\sup_{0\leqslant t\leqslant T}\big|N_t^{4,m}\big| > \frac{1}{6}e^{-C_M T}\eta,\ \sup_{0\leqslant t\leqslant T} e\big(t, u^\varepsilon\big) \leqslant M,\ \sup_{0\leqslant t\leqslant T}\|\varepsilon W - h\|_\infty < \delta\bigg\}$$

is empty. Combining this with (62) and (69)–(72) we obtain that for sufficiently small ε,

$$P\Big(\sup_{0\leqslant t\leqslant T} e_L\big(t, u^\varepsilon - u^h\big) \geqslant \eta, \sup_{0\leqslant t\leqslant T} \|\varepsilon W - h\| < \delta\Big) \leqslant 5\exp\bigg(-\frac{R}{\varepsilon^2}\bigg).$$

This completes the proof. $\square$

After we established the key results, Proposition 2 and Theorem 8, there exists now a well-known method (see [1,13]) to deduce the large deviation principle. For completeness, we include a proof.

Theorem 9. $\{\mu_\varepsilon, \varepsilon > 0\}$ *satisfies a large deviation principle with rate function* $R(f)$.

Proof. Denote by $d(\cdot,\cdot)$ the distance in the metric space $C([0,T]; H_0^2) \times C([0,T]; L^2)$. First, fix any closed subset $C \subset C([0,T]; H_0^2) \times C([0,T]; L^2)$ and choose $a < \inf_{g\in C} R(g)$. Define

$$K_a = \big\{g \in C\big([0,T]; H_0^2\big) \times C\big([0,T]; L^2\big) \bigm| R(g) \leqslant a\big\},$$
$$C_a = \big\{f \in \mathcal{H} \bigm| I(f) \leqslant a\big\}.$$

Then

$$K_a = \big\{\big(u^f, \dot{u}^f\big);\ f \in C_a\big\}, \quad K_a \cap C = \emptyset,$$

where $(u^f, \dot{u}^f)$ denotes the solution of the skeleton equation in Section 2 with h replaced by f. For any $g \in K_a$, there exists an open neighborhood of g, V_g such that

$$V_g \cap C = \emptyset.$$

One can choose $\rho_g > 0$, such that

$$G_g = \big\{h \in C\big([0,T]; H_0^2\big) \times C\big([0,T]; L^2\big), d(h,g) \leqslant \rho_g\big\} \subseteq V_g.$$

For any $f_g \in C_a$ such that $g = (u^{f_g}, \dot{u}^{f_g})$ and $R > a$, by Theorem 8 one can find two constants $\varepsilon_g > 0, \alpha_g > 0$ such that for any $\varepsilon < \varepsilon_g$

$$P\big(\varepsilon W \in F_g, \big(u^\varepsilon, \dot{u}^\varepsilon\big) \in G_g^C\big) \leqslant \exp\bigg(-\frac{R}{\varepsilon^2}\bigg),$$

where $F_g = \{f \in C([0,T], l^2);\ \|f - f_g\|_\infty < \alpha_g\}$. Therefore,

$$P\big(\varepsilon W \in F_g, \big(u^\varepsilon, \dot{u}^\varepsilon\big) \in V_g^C\big) \leqslant P\big(\varepsilon W \in F_g, \big(u^\varepsilon, \dot{u}^\varepsilon\big) \in G_g^C\big) \leqslant \exp\bigg(-\frac{R}{\varepsilon^2}\bigg). \tag{75}$$

Since $(F_g)_{g\in K_a}$ forms a cover for the compact set C_a of $C([0,T],l^2)$, there exist $g_1,\dots,g_l\in K_a$ such that

$$F=\bigcup_{i=1}^{l}F_{g_i}\supset C_a.$$

Therefore, we have

$$\begin{aligned}
P\big(\varepsilon W\in F,\big(u^\varepsilon,\dot u^\varepsilon\big)\in C\big)&=P\left(\bigcup_{i=1}^{l}\big[\{\varepsilon W\in F_{g_i}\}\cup\big\{\big(u^\varepsilon,\dot u^\varepsilon\big)\in C\big\}\big]\right)\\
&\leqslant P\left(\bigcup_{i=1}^{l}\big[\{\varepsilon W\in F_{g_i}\}\cup\big\{\big(u^\varepsilon,\dot u^\varepsilon\big)\in V_{g_i}^C\big\}\big]\right)\\
&\leqslant\sum_{i=1}^{l}P\big(\varepsilon W\in F_{g_i},\big(u^\varepsilon,\dot u^\varepsilon\big)\in V_{g_i}^C\big)\\
&\leqslant l\exp\left(-\frac{R}{\varepsilon^2}\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
\limsup_{\varepsilon\to0}\varepsilon^2\log\mu_\varepsilon(C)&\leqslant\limsup_{\varepsilon\to0}\varepsilon^2\log\big[P\big(\varepsilon W\in F^C\big)+P\big(\varepsilon W\in F,\big(u^\varepsilon,\dot u^\varepsilon\big)\in C\big)\big]\\
&\leqslant\Big(-\inf_{f\in F^c}I(f)\Big)\vee(-R)\\
&=-a,
\end{aligned}$$

where we have used the fact that εW satisfies a large deviation principle with rate function $I(\cdot)$. Let $a\to\inf_{g\in C}R(g)$ to get

$$\limsup_{\varepsilon\to0}\varepsilon^2\log\mu_\varepsilon(C)\leqslant-\inf_{g\in C}R(g).$$

Let $G\subset C([0,T];H_0^2)\times C([0,T];L^2)$ be an open set, take $g\in G$ with $R(g)<\infty$. Then there exists $f\in\mathcal{H}$ such that

$$g=\big(u^f,\dot u^f\big),\qquad R(g)=I(f).$$

Choose $\rho>0$, such that

$$\big\{h\in C\big([0,T];H_0^2\big)\times C\big([0,T];L^2\big)\;\big|\;d(h,g)\leqslant\rho\big\}\subset G.$$

For any $R>R(g)$, by Theorem 8, $\exists\alpha>0,\varepsilon_0>0$, such that for any $\varepsilon\in(0,\varepsilon_0)$,

$$P\big(d\big(\big(u^\varepsilon,\dot u^\varepsilon\big),\big(u^f,\dot u^f\big)\big)>\rho,\ \|\varepsilon W-f\|_\infty<\alpha\big)\leqslant\exp\left(-\frac{R}{\varepsilon^2}\right).$$

Therefore,

$$\begin{aligned}
P\big((u^\varepsilon,\dot u^\varepsilon)\in G\big) &\geqslant P\big(d\big((u^\varepsilon,\dot u^\varepsilon),g\big)\leqslant\rho\big)\\
&\geqslant P\big(d\big((u^\varepsilon,\dot u^\varepsilon),\big(u^f,\dot u^f\big)\big)\leqslant\rho,\ \|\varepsilon W-f\|_\infty<\alpha\big)\\
&\geqslant P\big(\|\varepsilon W-f\|_\infty<\alpha\big)-P\big(d\big((u^\varepsilon,\dot u^\varepsilon),\big(u^f,\dot u^f\big)\big)>\rho,\ \|\varepsilon W-f\|_\infty<\alpha\big)\\
&\geqslant P\big(\|\varepsilon W-f\|_\infty<\alpha\big)-\exp\left(-\frac{R}{\varepsilon^2}\right).
\end{aligned}$$

But

$$\begin{aligned}
\lim_{\varepsilon\to0}\varepsilon^2\log P\big(\|\varepsilon W-f\|_\infty<\alpha\big) &\geqslant -\inf\big\{I(\varphi),\ \|\varphi-f\|_\infty\leqslant\alpha\big\}\\
&\geqslant -I(f)
\end{aligned}$$

and since $R>R(g)$,

$$\lim_{\varepsilon\to0}\varepsilon^2\log P\big((u^\varepsilon,\dot u^\varepsilon)\in G\big)\geqslant -I(f)=-R(g).$$

Since g is the arbitrary,

$$\lim_{\varepsilon\to0}\varepsilon^2\log P\big((u^\varepsilon,\dot u^\varepsilon)\in G\big)\geqslant -\inf_{g\in G}I(g).$$

This completes the proof of the theorem. □

Acknowledgment

This work was initiated when the author visited the University of Oslo. He would like to thank Bernt Øksendal for his hospitality and for the stimulating discussions.

References

[1] R. Azencott, Grandes déviations et applications, in: Ecole d'été de Probabilité de Saint-Flour VIII, 1978, in: Lecture Notes in Math., vol. 774, Springer-Verlag, Berlin, 1980, pp. 1–176.
[2] J.M. Ball, Initial-boundary value problems for an extensible beam, J. Math. Anal. Appl. 42 (1973) 61–90.
[3] M.T. Barlow, M. Yor, Semi-martingale inequalities via the Garsia–Rodemich–Rumsey lemma, and applications to local times, J. Funct. Anal. 49 (1982) 198–229.
[4] Z. Brzeźniak, B. Maslowski, J. Seidler, Stochastic beam equations, Probab. Theory Related Fields 132 (2005) 119–149.
[5] C. Cardon-Weber, Large deviations for a Burgers'-type SPDE, Stoch. Process. Appl. 84 (1999) 53–70.
[6] S. Cerrai, M. Röckner, Large deviations for stochastic reaction–diffusion systems with multiplicative noise and non-Lipschitz reaction term, Ann. Probab. 32 (1) (2004) 1100–1139.
[7] F. Chenal, A. Millet, Uniform large deviations for parabolic SPDEs and applications, Stoch. Process. Appl. 72 (1997) 161–186.
[8] P.L. Chow, Large deviation problem for some parabolic Ito equations, Comm. Pure. Appl. Math 45 (1) (1992) 97–120.
[9] P.L. Chow, J.L. Menaldi, Stochastic PDE for nonlinear vibration of elastic panels, Differential Integral Equations 12 (3) (1999) 419–432.
[10] I.D. Chueshov, Introduction to the Theory of Infinite-dimensional Dissipative Systems, Acta, Kharkiv, 2002.

[11] B. Davis, On the L^p-norms of stochastic integrals and other martingales, Duke Math. J. 43 (1976) 697–704.
[12] A. Dembo, A. Zeitouni, Large Deviations Techniques and Applications, Springer-Verlag, Berlin, 1998.
[13] H. Doss, P. Priouret, Petites perturbations de systémes dynamiques avec réflexion, in: Séminaire de Probabilités XVII, in: Lecture Notes in Math., vol. 986, Springer-Verlag, Berlin, 1983, pp. 351–370.
[14] E.H. Dowell, Aeroelasticity of Plates and Shells, Nordhoff, Leyden, 1975.
[15] W.E. Fitzgibbon, Global existence and boundedness of solutions to the extensible beam equation, SIAM J. Math. Anal. 13 (1982) 739–745.
[16] S. Kouémou Patcheu, On a global solution and asymptotic behaviour for the generalized damped extensible beam equation, J. Differential Equations 135 (1997) 299–314.
[17] E. Pardoux, Equations aux dérivées partielles stochastiques non lineaires montones, thesis d'Etat, Université Paris Sud, 1975.
[18] G.D. Prato, J. Zabczyk, Stochastic Equations in Infinite Dimensions, Cambridge Univ. Press, Cambridge, 1992.
[19] C. Rovira, M. Sanz-Solé, The law of the solution to a nonlinear hyperbolic SPDE, J. Theoret. Probab. 4 (9) (1996) 863–901.
[20] R.B. Sowers, Large deviations for a reaction–diffusion equation with non-Gaussian perturbations, Ann. Probab. 20 (1) (1992) 504–537.
[21] J.B. Walsh, An introduction to stochastic partial differential equations, in: École d'Été de Probabilités de Saint-Flour XIV, 1984, in: Lecture Notes in Math., vol. 1180, Springer-Verlag, Berlin, 1986, pp. 265–437.
[22] S. Woinowsky-Krieger, The effect of an axial force on the vibration of hinged bars, J. Appl. Mech. 17 (1950) 35–36.
[23] T.S. Zhang, On small time asymptotics of diffusions on Hilbert spaces, Ann. Probab. 28 (2) (2000) 537–557.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Functional Analysis 248 (2007) 202–224
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

The asymptotic lift of a completely positive map

William Arveson [1]

Department of Mathematics, University of California, Berkeley, CA 94720, USA

Received 7 September 2006; accepted 30 November 2006

Available online 9 January 2007

Communicated by G. Pisier

Abstract

Starting with a unit-preserving normal completely positive map $L: M \to M$ acting on a von Neumann algebra—or more generally a dual operator system—we show that there is a unique reversible system $\alpha: N \to N$ (i.e., a complete order automorphism α of a dual operator system N) that captures all of the asymptotic behavior of L, called the *asymptotic lift* of L. This provides a noncommutative generalization of the Frobenius theorems that describe the asymptotic behavior of the sequence of powers of a stochastic $n \times n$ matrix. In cases where M is a von Neumann algebra, the asymptotic lift is shown to be a W^*-dynamical system $(N, \mathbb{Z})$, and we identify $(N, \mathbb{Z})$ as the tail flow of the minimal dilation of L. We are also able to identify the Poisson boundary of L as the fixed algebra N^α. In general, we show the action of the asymptotic lift is trivial iff L is *slowly oscillating* in the sense that

$$\lim_{n\to\infty} \left\| \rho \circ L^{n+1} - \rho \circ L^n \right\| = 0, \quad \rho \in M_*.$$

Hence α is often a nontrivial automorphism of N. The asymptotic lift of a variety of examples is calculated.

Keywords: Completely positive map; von Neumann algebra; Asymptotics

1. Introduction

Throughout this paper we use the term UCP map to denote a normal unit-preserving completely positive map $L: M_1 \to M_2$ of one dual operator system into another. While we are

E-mail address: arveson@math.berkeley.edu.
[1] Supported by NSF grant DMS-0100487.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2006.11.014

primarily concerned with the dynamical properties of UCP maps $L : M \to M$ that act on a von Neumann algebra M, it is appropriate to broaden that category to include UCP self-maps of more general dual operator systems.

Stochastic $n \times n$ matrices $P = (p_{ij})$ describe the transition probabilities of n-state Markov chains. The asymptotic properties of the sequence of powers of the transition matrix govern the long-term statistical behavior of the process after initial transient fluctuations have died out [6, pp. 170–185]. A stochastic $n \times n$ matrix $P = (p_{ij})$ gives rise to a UCP map of the commutative von Neumann algebra $\mathbb{C}^n$ by way of

$$(Px)_i = \sum_{j=1}^{n} p_{ij} x_j, \quad 1 \leqslant i \leqslant n,\ x = (x_1, \dots, x_n) \in \mathbb{C}^n,$$

and the classical Perron–Frobenius theory provides an effective description of the asymptotic behavior of the sequence $P, P^2, P^3, \dots$ (see Section 9).

Recently, several emerging areas of mathematics have opened a vista of potential applications for noncommutative generalizations of the Frobenius theorems. The most obvious examples are quantum probability, noncommutative dynamics [2], and quantum computing [15]. Such considerations led us to initiate a study of "almost periodic" UCP maps on von Neumann algebras in [3]. Broadly speaking, those results allowed us to relate the asymptotic behavior of the powers of *certain* UCP maps on von Neumann algebras to the asymptotic behavior of $*$-automorphisms of other naturally associated von Neumann algebras.

However, the hypothesis of almost periodicity is very restrictive. For example, it forces the spectrum of the UCP map to have a discrete component consisting of eigenvalues on the unit circle. While this discrete spectrum appeared to be essential for the results of [3], it is missing in many important examples. Nevertheless, in *all* of the examples that we were able to penetrate, there was a "hidden" W^*-dynamical system that shared the same asymptotic behavior. That led us to suspect that a different formulation might be possible, in which the almost periodic hypothesis is eliminated entirely. This turned out to be true, but the new results require a complete reformulation in terms of UCP maps on dual operator systems.

The main results of this paper are Theorems 3.2, 5.1, 6.1, 7.1. While these general results are not as sharp as those of [3] when restricted to the case of almost periodic maps, we believe that is compensated for by the simplicity and full generality of the new setting. There are natural variations of all of the above results for one-parameter semigroups of normal completely positive maps that will be taken up elsewhere. We also note that a very recent paper of Størmer [21] complements the results of [3].

2. Reversible lifts of UCP maps

Recall that an *operator system* is a norm-closed self-adjoint linear subspace M of the algebra $\mathcal{B}(H)$ of all bounded operators on a Hilbert space H, such that the identity operator $\mathbf{1}$ belongs to M. A *dual* operator system is an operator system that is closed in the weak*-topology of $\mathcal{B}(H)$. We write M_* for the predual of such an operator system M, namely the norm-closed linear subspace of the dual of M consisting of all restrictions to M of normal linear functionals on $\mathcal{B}(H)$; and since M can be naturally identified with the dual of M_*, we refer to the M_*-topology as the weak*-topology of M. A *normal* linear map of dual operator systems is a linear map $L : M \to N$ that is continuous relative to the respective weak*-topologies. Of course, such a map is the adjoint of a unique bounded linear map of preduals, namely $\rho \in N_* \mapsto \rho \circ L \in M_*$.

While there is an abstract characterization of operator systems [5, Theorem 4.4], we shall not require the details of it here. Only rarely do we require a realization $M \subseteq \mathcal{B}(H)$ of M as a concrete dual operator system; but when we do, we require that the realization have the above properties—namely that M is a weak*-closed concrete operator system in $\mathcal{B}(H)$ and M_* consists of all restrictions of normal linear functionals of $\mathcal{B}(H)$. For the most part, our conventions and basic terminology follow those of the monographs [7] and [17].

We fix attention on the category whose objects are UCP self-maps

$$L : M \to M$$

acting on dual operator systems M, and whose morphisms are equivariant UCP maps. Thus, a homomorphism from $L_1 : M_1 \to M_1$ to $L_2 : M_2 \to M_2$ is a UCP map

$$E : M_1 \to M_2$$

satisfying $E \circ L_1 = L_2 \circ E$. In this case we say that L_1 is a *lifting* of L_2 through E, or simply a *lift* of L_2. By an *automorphism* of a dual operator system N we mean a UCP map $\alpha : N \to N$ having a UCP inverse $\alpha^{-1} : N \to N$.

Definition 2.1. A *reversible lift* of a UCP map $L : M \to M$ is a triple (N, α, E) consisting of an automorphism $\alpha : N \to N$ of another dual operator system N and a UCP map $E : N \to M$ satisfying $E \circ \alpha = L \circ E$.

A UCP map $L : M \to M$ has many reversible lifts, the simplest being the trivial lift $(\mathbb{C}, \mathrm{id}, \iota)$, where $\iota : \mathbb{C} \to M$ is the inclusion $\iota(\lambda) = \lambda \cdot \mathbf{1}_M$. We begin by pointing out that all reversible lifts must satisfy a system of asymptotic inequalities. In the usual way, $L : M \to M$ gives rise to a hierarchy of UCP maps $L_n : M^{(n)} \to M^{(n)}$, $n = 1, 2, \ldots$, in which $M^{(n)} = M_n(\mathbb{C}) \otimes M$ is the $n \times n$ matrix system over M and $L_n = \mathrm{id} \otimes L : M^{(n)} \to M^{(n)}$ is the naturally induced UCP map. Similarly, a reversible lifting (N, α, E) of $L : M \to M$ gives rise to a hierarchy of reversible liftings $(N^{(n)}, \alpha_n, E_n)$ of $L_n : M^{(n)} \to M^{(n)}$, one for every $n = 1, 2, \ldots$.

Proposition 2.2. *Let (N, α, E) be a reversible lift of a given UCP map $L : M \to M$. For every bounded linear functional ρ on M, the sequence of norms $\|\rho \circ L^k\|$ decreases with increasing $k = 1, 2, \ldots$, and we have*

$$\|\rho \circ E\| \leqslant \lim_{k \to \infty} \|\rho \circ L^k\|. \tag{2.1}$$

Moreover, the inequalities (2.1) *persist throughout the hierarchy of liftings of $L_n : M^{(n)} \to M^{(n)}$, as ρ ranges over the dual of $M^{(n)}$, $n \geqslant 1$.*

Proof. Every UCP map is a contraction, hence $\|\rho \circ L^{k+1}\| \leqslant \|\rho \circ L^k\|$ for every $k \geqslant 0$. Moreover, for fixed $x \in N$ and $k \geqslant 0$ we can write

$$E(x) = E\big(\alpha^k\big(\alpha^{-k}(x)\big)\big) = L^k\big(E\big(\alpha^{-k}(x)\big)\big).$$

Since $\|E \circ \alpha^{-k}\| \leqslant 1$, we conclude that for every $\rho \in M'$,

$$\|\rho \circ E\| = \sup_{\|x\|=1} \bigl|\rho\bigl(E(x)\bigr)\bigr| = \sup_{\|x\|=1} \bigl|\rho \circ L^k\bigl(E \circ \alpha^{-k}(x)\bigr)\bigr| \leqslant \bigl\|\rho \circ L^k\bigr\|,$$

and (2.1) follows after passing to the limit as $k \to \infty$. The same argument applies throughout the hierarchy of liftings $(N^{(n)}, \mathrm{id}_n \otimes \alpha, \mathrm{id}_n \otimes E)$, after one replaces ρ with a bounded linear functional on $M^{(n)}$, $n = 1, 2, \ldots$. □

Remark 2.3 *(Nondegeneracy).* Let (N, α, E) be a reversible lift of a UCP map $L : M \to M$. Since $E : N \to M$ is normal, it has a pre-adjoint $E_* : M_* \to N_*$, defined by $E_*(\rho) = \rho \circ E$, $\rho \in M_*$. Consider the range

$$E_*(M_*) = \{\rho \circ E\colon\ \rho \in M_*\}$$

of this map. $E_*(M_*)$ is a linear subspace of N_*, and note that it is invariant under the invertible isometry $\alpha_* \in \mathcal{B}(N_*)$:

$$E_*(M_*) \circ \alpha \subseteq E_*(M_*).$$

Indeed, that is immediate from equivariance and normality of L, since for $\rho \in M_*$ we have $\rho \circ E \circ \alpha = (\rho \circ L) \circ E \in M_* \circ E$. It follows that the sequence $E_*(M_*) \circ \alpha^n$, $n \in \mathbb{Z}$, defines a doubly infinite tower of subspaces of N_*

$$\cdots \subseteq E_*(M_*) \circ \alpha \subseteq E_*(M_*) \subseteq E_*(M_*) \circ \alpha^{-1} \subseteq E_*(M_*) \circ \alpha^{-2} \subseteq \cdots.$$

A straightforward application of the Hahn–Banach theorem shows that this tower is norm-dense in N_* if and only if for every $y \in N$ one has

$$E\bigl(\alpha^{-n}(y)\bigr) = 0, \quad n = 0, 1, 2, \ldots \quad \Rightarrow \quad y = 0. \tag{2.2}$$

A reversible lifting (N, α, E) of L is said to be *nondegenerate* if condition (2.2) is satisfied.

It is significant that when (2.2) fails, one can always replace (N, α, E) with a *nondegenerate* reversible lifting $(\tilde{N}, \tilde{\alpha}, \tilde{E})$ using the following device. We may assume that $M \subseteq \mathcal{B}(H)$ is realized as a concrete dual operator system. Let $\tilde{H} = \ell^2(\mathbb{Z}) \otimes H$ be the Hilbert space of all square-summable bilateral sequences from H and define a map $\theta : N \to \mathcal{B}(\tilde{H})$ by

$$\bigl(\theta(y)\xi\bigr)(n) = E\bigl(\alpha^{-n}(y)\bigr)\xi(n), \quad \xi \in \tilde{H},\ n \in \mathbb{Z}. \tag{2.3}$$

$\tilde{N}$ is defined as the weak*-closure of $\theta(N)$. The unitary shift defined on $\tilde{H}$ by $U\xi(n) = \xi(n-1)$, $n \in \mathbb{Z}$, implements a $*$-automorphism $\tilde{\alpha}(X) = UXU^*$ of $\mathcal{B}(\tilde{H})$ such that $\tilde{\alpha}(\tilde{N}) = \tilde{N}$; indeed, $\theta(N)$ is stable under shifts to the left or right because α is an automorphism of N. Note too that:

$$\bigl(U\theta(y)U^{-1}\xi\bigr)(n) = E\bigl(\alpha^{-n+1}(y)\bigr)\xi(n) = L\bigl(E\bigl(\alpha^{-n}(y)\bigr)\bigr)\xi(n), \quad \xi \in \tilde{H},\ n \in \mathbb{Z}.$$

It follows that the map $\tilde{E} : \mathcal{B}(\tilde{H}) \to \mathcal{B}(H)$ that compresses an operator matrix in $\mathcal{B}(\tilde{H})$ to its 00th component restricts to a UCP map $\tilde{E} : \tilde{N} \to M$ satisfying $\tilde{E} \circ \tilde{\alpha} = L \circ \tilde{E}$. Thus, $(\tilde{N}, \tilde{\alpha}, \tilde{E})$ is a reversible lifting of L. It is a homomorphic image of (N, α, E) in the sense that the UCP map

$\theta : N \to \tilde{N}$ satisfies $\tilde{\alpha} \circ \theta = \theta \circ \alpha$ and $\tilde{E} \circ \theta = E$ (see Section 4 for a discussion of the category of reversible liftings of L). Finally, by examining components in the obvious way, one verifies directly that $(\tilde{N}, \tilde{\alpha}, \tilde{E})$ is nondegenerate.

3. Asymptotic lifts of UCP maps

In this section we show by a direct construction that there is a reversible lifting with favorable asymptotic properties and that, after degeneracies have been eliminated, it is *unique* up to natural isomorphism.

Definition 3.1. Let $L : M \to M$ be a UCP map on a dual operator system. An *asymptotic lift* of L is a reversible lifting (N, α, E) of L that satisfies nondegeneracy (2.2), such that the inequalities (2.1) become equalities for *normal* linear functionals throughout the entire matrix hierarchy

$$\left\| \rho \circ (\mathrm{id}_n \otimes E) \right\| = \lim_{k \to \infty} \left\| \rho \circ (\mathrm{id}_n \otimes L)^k \right\|, \quad \rho \in M_*^{(n)},\ n = 1, 2, \ldots. \tag{3.1}$$

We come now to a basic result.

Theorem 3.2. *Every UCP map $L : M \to M$ of a dual operator system has an asymptotic lifting. If (N_1, α_1, E_1) and (N_2, α_2, E_2) are two asymptotic liftings for L, then there is a unique isomorphism of dual operator systems $\theta : N_1 \to N_2$ such that $\theta \circ \alpha_1 = \alpha_2 \circ \theta$ and $E_2 \circ \theta = E_1$.*

The existence assertion of Theorem 3.2 is proved by a direct construction involving inverse sequences, which are defined as follows:

Definition 3.3. Let $L : M \to M$ be a UCP map on a dual operator system. By an *inverse sequence* for L we mean a bilateral sequence $(x_n)_{n \in \mathbb{Z}}$ of elements of M satisfying $\sup_n \|x_n\| < \infty$, and

$$x_n = L(x_{n+1}), \quad n \in \mathbb{Z}. \tag{3.2}$$

The set of all inverse sequences for L is denoted $\mathcal{S}_L$, or more simply $\mathcal{S}$ when there is no cause for confusion.

Remark 3.4 *(Properties of inverse sequences).* The set $\mathcal{S}$ of all inverse sequences for $L : M \to M$ is a vector space that is closed under pointwise involution $(x_n) \mapsto (x_n^*)$, and it contains all "constant" scalar sequences of the form $(\ldots, \lambda \cdot \mathbf{1}, \lambda \cdot \mathbf{1}, \lambda \cdot \mathbf{1}, \ldots)$, $\lambda \in \mathbb{C}$. More generally, the constant sequences $(\ldots, a, a, a, \ldots) \in \mathcal{S}$ correspond bijectively with the space of fixed elements $\{a \in M\colon\ L(a) = a\}$. Notice too that $\mathcal{S}$ is stable under shifting to the right or left; if $(x_n)_{n \in \mathbb{Z}}$ belongs to $\mathcal{S}$ then so does $(x_{n+k})_{n \in \mathbb{Z}}$ for every $k = 0, \pm 1, \pm 2, \ldots$.

Every element x_k of an inverse sequence (x_n) determines all of its predecessors uniquely, since $x_{k-1} = L(x_k)$ and, more generally, $x_r = L^{k-r}(x_k)$ for all $r \leqslant k$. On the other hand, x_k does *not* determine x_{k+1} uniquely, since there can be many solutions z of the equation $L(z) = x_k$. If we fix a particular solution z and replace x_{k+1} with z in the $(k+1)$st spot, then it may not be possible to solve the equation $z = L(w)$ for $w \in M$; and *in that case there is no inverse sequence whose $(k+1)$st term is the replaced element z and whose kth term is x_k.*

More generally, given an element $a \in M$, the question of whether or not there is an inverse sequence (x_n) satisfying $x_0 = a$ can be subtle.

Proof of Theorem 3.2. *Existence*. In order to construct an asymptotic lifting for L, we realize $M \subseteq \mathcal{B}(H)$ as a concrete weak*-closed operator system. Let $\ell^2 \otimes H$ be the Hilbert space of all sequences $n \in \mathbb{Z} \mapsto \xi_n \in H$ satisfying $\sum_n \|\xi_n\|^2 < \infty$, with its usual inner product.

We can realize the space $\mathcal{S}$ of all inverse sequences for L as an operator subspace $N \subseteq \mathcal{B}(\ell^2 \otimes H)$ by identifying an inverse sequence $(x_n) \in \mathcal{S}$ with the diagonal operator $D = \operatorname{diag}(x_n)$ defined by

$$(D\xi)_n = x_n \xi_n, \quad n \in \mathbb{Z}.$$

The operator norm of $\operatorname{diag}(x_n)$ is given by $\|\operatorname{diag}(x_n)\| = \sup_{n \geqslant 0} \|x_n\|$. Since L is a normal map, the relations defined by (3.2) are weak*-closed, and therefore the space $N = \{\operatorname{diag}(x_n)\colon (x_n) \in \mathcal{S}\}$ is closed in the weak*-topology of $\mathcal{B}(\ell^2 \otimes H)$. Remark 3.4 implies that N is self-adjoint and contains the identity operator of $\mathcal{B}(\ell^2 \otimes H)$, hence N acquires the structure of a dual operator system.

Consider the right-shift $\alpha : N \to N$ of diagonal operators

$$\alpha\big(\operatorname{diag}(x_n)\big) = \operatorname{diag}(x_{n-1}).$$

Letting U be the unitary bilateral shift acting on $\ell^2 \otimes H$ by

$$(U\xi)_n = \xi_{n-1}, \quad n \in \mathbb{Z},\ \xi \in \ell^2 \otimes H,$$

one finds that the associated $*$-automorphism $X \mapsto UXU^*$ of $B(\ell^2 \otimes H)$ implements the action of α on N. Hence α is a complete automorphism of the concrete operator system N. The natural inclusion of H in $\ell^2(\mathbb{Z}) \otimes H$, in which a vector $\xi \in H$ is identified with $\delta_0 \otimes \xi \in \ell^2(\mathbb{Z}) \otimes H$, gives rise to a normal map $X \mapsto P_H X\restriction_H$ that restricts to a map $E : N \to M$ satisfying

$$E\big(\operatorname{diag}(x_n)\big) = x_0, \quad (x_n) \in \mathcal{S}.$$

One has

$$E \circ \alpha\big(\operatorname{diag}(x_n)\big) = E\big(\operatorname{diag}(x_{n-1})\big) = x_{-1} = L(x_0) = L \circ E\big(\operatorname{diag}(x_n)\big),$$

so that (N, α, E) becomes a reversible lift of L.

It remains to verify (2.2) and (3.1). For (2.2), suppose that $X = \operatorname{diag}(x_k)$ satisfies $E(\alpha^{-n}(X)) = x_n = 0$ for $n \geqslant 0$. Since the sequence (x_k) belongs to S, this implies that $x_j = L^{|j|}(x_0) = 0$ for negative j as well, hence (x_k) is the zero sequence.

For (3.1), it is enough to verify the system of nontrivial inequalities

$$\big\|\rho \circ (\mathrm{id}_n \otimes E)\big\| \geqslant \lim_{k\to\infty} \big\|\rho \circ (\mathrm{id}_n \otimes L)^k\big\|, \quad \rho \in M_*^{(n)},\ n = 1, 2, \ldots. \tag{3.3}$$

Consider first the case $n = 1$, and choose $\rho \in M_*$. For each $k = 1, 2, \ldots$ we can find an element $u_k \in M$ satisfying $\|u_k\| = 1$ and $|\rho(L^k(u_k))| = \|\rho \circ L^k\|$. Consider the triangular array $s^k = (s_0^k, \ldots, s_k^k)$, $k = 1, 2, \ldots$, of elements of M defined by

$$\big(s_0^k, \ldots, s_k^k\big) = \big(L^k(u_k), L^{k-1}(u_k), \ldots, L(u_k), u_k\big), \quad k = 1, 2, \ldots.$$

Each component of every one of these sequences belongs to ball M, and $s_0^k = x_0$ for every $k \geqslant 1$. Moreover, the jth component s_j^k of any one of them is obtained from the $(j+1)$st by applying L,

$$s_j^k = L\big(s_{j+1}^k\big), \quad j = 0, 1, 2, \dots, k-1. \tag{3.4}$$

Since the infinite Cartesian product ball $M \times$ ball $M \times \cdots$ is compact in its weak* product topology, there is a subnet $s^{k'}$ of the sequence s^k with the property that each of its components (note that each component is well defined for sufficiently large k') converges weak* to an element of ball M. Hence we can define a single infinite sequence $x_0, x_1, x_2, \ldots \in$ ball M by

$$x_j = \lim_{k'} s_j^{k'}, \quad j = 0, 1, 2, \dots.$$

Since L is a normal map, the relations (3.4) imply that $x_{j+1} = L(x_j)$, $j = 0, 1, 2, \ldots$. If we continue the sequence into negative integers by setting $x_k = L^{|k|}(x_0)$ for $k \leqslant -1$, the result is a sequence $(x_n) \in \mathcal{S}$ satisfying $\sup_n \|x_n\| \leqslant 1$ and $E(\operatorname{diag}(x_n)) = x_0$. Thus we conclude that

$$\|\rho \circ E\| \geqslant \big|\rho(x_0)\big| = \lim_{k'} \big|\rho\big(L^{k'}(u_{k'})\big)\big| = \lim_{k'} \big\|\rho \circ L^{k'}\big\| = \lim_{k \to \infty} \big\|\rho \circ L^k\big\|,$$

and (3.3) follows.

Notice that this argument can be repeated *verbatim* to establish (3.3) throughout the matrix hierarchy for $n \geqslant 2$, since the inverse sequences for $\mathrm{id}_n \otimes L$ are bilateral sequences $(\tilde{x}_k)$ whose components $\tilde{x}_k$ are $n \times n$ matrices in $M^{(n)}$ that satisfy $(\mathrm{id}_n \otimes L)(\tilde{x}_{k+1}) = \tilde{x}_k$, $k \in \mathbb{Z}$. We conclude that (N, α, E) is an asymptotic lift of L. $\square$

Turning now to the uniqueness issue for asymptotic lifts, we require the dual formulation of (3.1)—Lemma 3.6—the proof of which makes use of the following elementary result. Since we lack an appropriate reference, we sketch a proof of the latter for completeness.

Lemma 3.5. *Let X be a Banach space, let $K_1 \supseteq K_2 \supseteq \cdots$ be a decreasing sequence of nonempty weak*-compact convex subsets of the dual X' with intersection K_∞. Then for every weak*-continuous linear functional $\rho \in X''$,*

$$\sup\big\{\big|\rho(x)\big|: x \in K_n\big\} \downarrow \sup\big\{\big|\rho(x)\big|: x \in K_\infty\big\}, \quad \text{as } n \to \infty. \tag{3.5}$$

Proof. Fix ρ. The sequence of nonnegative numbers $\sup\{|\rho(x)|: x \in K_n\}$ obviously decreases with n, and its limit ℓ satisfies

$$\ell \geqslant \sup\big\{\big|\rho(x)\big|: x \in K_\infty\big\}.$$

To prove the opposite inequality choose, for every $n = 1, 2, \ldots$, an element $x_n \in K_n$ such that $|\rho(x_n)| = \sup\{|\rho(x)|: x \in K_n\}$. By compactness of the unit ball of X', there is a subnet $\{x_{n'}\}$ of $\{x_n\}$ that converges weak* to x_∞. The limit point x_∞ must belong to $K_\infty = \bigcap_n K_n$ because the K_n decrease with n, and since the numbers $|\rho(x_n)| = \sup\{|\rho(x)|: x \in K_n\}$ converge to ℓ, it follows from weak*-continuity of ρ that

$$\bigl|\rho(x_\infty)\bigr| = \lim_{n'}\bigl|\rho(x_{n'})\bigr| = \lim_{n\to\infty}\bigl|\rho(x_n)\bigr| = \ell.$$

Since $\ell = |\rho(x_\infty)| \leqslant \sup\{|\rho(x)|\colon\ x \in K_\infty\}$, the proof is complete. □

Lemma 3.6. *Let $L : M \to M$ be a UCP map on a dual operator system and let (N, α, E) be a reversible lift of L. The following are equivalent*:

(i) *Every $\rho \in M_*$ satisfies* (3.1)

$$\|\rho \circ E\| = \lim_{n\to\infty}\bigl\|\rho \circ L^n\bigr\|.$$

(ii) *Writing* ball X *for the closed unit ball of a normed space X, we have*

$$E(\operatorname{ball} N) = \bigcap_{n=0}^{\infty} L^n(\operatorname{ball} M). \tag{3.6}$$

Proof. (i) ⇒ (ii). Choose an element $y \in \operatorname{ball} N$. Then for every $n = 0, 1, 2, \ldots$ we can write

$$E(y) = E\bigl(\alpha^n \circ \alpha^{-n}(y)\bigr) = L^n E\bigl(\alpha^{-n}(y)\bigr) \in L^n(\operatorname{ball} M)$$

since $E(\alpha^{-n}(y)) \in \operatorname{ball} M$. Hence $E(\operatorname{ball} N) \subseteq \bigcap_n L^n(\operatorname{ball} M)$. For the opposite inclusion, note that both sides of $E(\operatorname{ball} N) \subseteq \bigcap_n L^n(\operatorname{ball} M)$ are circled weak*-compact convex subsets of M, so by a standard separation theorem it suffices to show that for every $\rho \in M_*$ one has

$$\sup\bigl\{\bigl|\rho(x)\bigr|\colon\ x \in E(\operatorname{ball} N)\bigr\} = \sup\Bigl\{\bigl|\rho(x)\bigr|\colon\ x \in \bigcap_n L^n(\operatorname{ball} M)\Bigr\}. \tag{3.7}$$

The left-hand side of (3.7) is $\|\rho \circ E\|$, while by Lemma 3.5, the right-hand side is

$$\lim_{n\to\infty}\sup\bigl\{\bigl|\rho(x)\bigr|\colon\ x \in L^n(\operatorname{ball} M)\bigr\} = \lim_{n\to\infty}\bigl\|\rho \circ L^n\bigr\|.$$

An application of the hypothesis (i) now gives (3.7).

(ii) ⇒ (i). Choose $\rho \in N_*$. For $n = 1, 2, \ldots$ let $K_n = L^n(\operatorname{ball} M)$, and set $K_\infty = \bigcap_n K_n$. Lemma 3.5 implies that $\|\rho \circ L^n\| = \sup\{|\rho(y)|\colon\ y \in K_n\}$ decreases to $\sup\{|\rho(y)|\colon\ y \in K_\infty\}$ as $n \uparrow \infty$, while by (3.6),

$$\|\rho \circ E\| = \sup\bigl\{\bigl|\rho(y)\bigr|\colon\ y \in E(\operatorname{ball} N)\bigr\} = \sup\bigl\{\bigl|\rho(y)\bigr|\colon\ y \in K_\infty\bigr\}.$$

Thus $\lim_n \|\rho \circ L^n\|$ is identified with $\|\rho \circ E\|$, and (i) follows. □

Lemma 3.7. *Let $L : M \to M$ be a UCP map of a dual operator system and let (N, α, E) be a reversible lift of L that satisfies*

$$\|\rho \circ E\| = \lim_{n\to\infty}\bigl\|\rho \circ L^n\bigr\|, \quad \rho \in M_*. \tag{3.8}$$

Let $K=\{z\in N\colon\ E(\alpha^k(z))=0,\ k\in\mathbb{Z}\}$. *Then for every* $y\in N$ *we have*

$$\sup_{k\in\mathbb{Z}}\bigl\|E\bigl(\alpha^k(y)\bigr)\bigr\|=\inf_{z\in K}\|y+z\|. \tag{3.9}$$

Proof. The inequality $\leqslant$ is apparent, since for $z\in K$ and $k\in\mathbb{Z}$ we have

$$\bigl\|E\bigl(\alpha^k(y)\bigr)\bigr\|=\bigl\|E\bigl(\alpha^k(y+z)\bigr)\bigr\|\leqslant\|y+z\|.$$

In order to prove $\geqslant$, we may assume that $\sup_k\|E(\alpha^k(y))\|=1$. For fixed $n=1,2,\dots$, note first that $E(\alpha^{-n}(y))\in L^p(\operatorname{ball}M)$ for every $p=1,2,\dots$. Indeed, we have $E(\alpha^{-n}(y))=L^p(E(\alpha^{-n-p}(y))$, and $\|E(\alpha^{-n-p}(y))\|\leqslant 1$. Lemma 3.6 implies that $\bigcap_p L^p(\operatorname{ball}M)=E(\operatorname{ball}N)$; and since α^n is an invertible isometry of N, we can find an element $y_n\in N$ satisfying $\|y_n\|\leqslant 1$ and $E(\alpha^{-n}(y_n))=E(\alpha^{-n}(y))$.

Note that y and y_n satisfy the following relations:

$$E\bigl(\alpha^{-k}(y_n)\bigr)=E\bigl(\alpha^{-k}(y)\bigr),\quad k\in\mathbb{Z},\ k\leqslant n. \tag{3.10}$$

Indeed, an application of L^{n-k} to both sides of $E(\alpha^{-n}(y_n))=E(\alpha^{-n}(y))$ leads to

$$E\bigl(\alpha^{-k}(y_n)\bigr)=L^{n-k}E\bigl(\alpha^{-n}(y_n)\bigr)=L^{n-k}E\bigl(\alpha^{-n}(y)\bigr)=E\bigl(\alpha^{-k}(y)\bigr).$$

By weak*-compactness of the unit ball of N, there is a subnet y_{n_α} of the sequence y_n that converges weak* to an element $y_\infty\in N$ such that $\|y_\infty\|\leqslant 1$. Since each map $z\mapsto E(\alpha^{-k}(z))$ is weak*-continuous on the unit ball of N, the equations (3.10) imply that $E(\alpha^{-k}(y_\infty))=E(\alpha^{-k}(y))$, $k\in\mathbb{Z}$. Hence $y_\infty-y$ belongs to K. It follows that

$$\sup_{k\in\mathbb{Z}}\bigl\|E\bigl(\alpha^k(y)\bigr)\bigr\|=1\geqslant\|y_\infty\|=\bigl\|y+(y_\infty-y)\bigr\|\geqslant\inf_{z\in K}\|y+z\|$$

proving the inequality $\geqslant$ of (3.9). $\square$

Proof of Theorem 3.2. *Uniqueness*. Let (N,α,E) be the asymptotic lift of L constructed in the above existence proof and let $(\tilde N,\tilde\alpha,\tilde E)$ be another one. Define a map $\theta: y\in\tilde N\to\theta(y)\in N$ as follows:

$$\theta(y)=\operatorname{diag}(x_n),\quad\text{where } x_n=\tilde E\bigl(\tilde\alpha^{-n}(y)\bigr),\ n\in\mathbb{Z}.$$

Obviously, θ is a UCP map of $\tilde N$ into N satisfying $\theta\circ\tilde\alpha=\alpha\circ\theta$. Lemma 3.7 implies that θ is an injective isometry. Indeed, applying Lemma 3.7 repeatedly to the sequence of maps $\mathrm{id}_n\otimes\theta:\tilde N^{(n)}\to N^{(n)}$, $n=1,2,\dots$, we find that θ is a complete isometry. We claim that $\theta(\tilde N)=N$. Since θ is a weak*-continuous isometry, its range is a weak*-closed subspace of N, so to prove $\theta(\tilde N)=N$ it suffices to show that $\theta(\tilde N)$ is weak*-dense in N. For that, let (x_k) be an inverse sequence satisfying $\sup_k\|x_k\|=1$. Then x_n belongs to $E(\operatorname{ball}N)$ for every $n\geqslant 1$; and by Lemma 3.6, $E(\operatorname{ball}N)=\tilde E(\operatorname{ball}\tilde N)$. Hence there is an element $y_n\in\operatorname{ball}\tilde N$ such that $x_n=\tilde E(\tilde\alpha^{-n}(y_n))$. As in the proof of (3.10), this implies $x_k=\tilde E(\tilde\alpha^{-k}(y_n))$ for all $k\leqslant n$. These equations imply that $\theta(y_1),\theta(y_2),\dots$ converges component-by-component to $\operatorname{diag}(x_k)$ in the weak*-topology. Therefore $\theta(y_n)\to\operatorname{diag}(x_k)$ (weak*) as $n\to\infty$.

Hence θ is an equivariant isomorphism of $\tilde{N}$ on N. Since the zeroth component of $\theta(y)$ is $\tilde{E}(y)$, we also have $E \circ \theta = \tilde{E}$.

Finally, we claim that the two requirements $\theta \circ \tilde{\alpha} = \alpha \circ \theta$ and $E \circ \theta = \tilde{E}$ serve to determine such a UCP map θ uniquely. Indeed, if $\theta_1, \theta_2 : \tilde{N} \to N$ are two equivariant UCP maps satisfying $E \circ \theta_1 = E \circ \theta_2 = \tilde{E}$, then for every $y \in \tilde{N}$, $n \in \mathbb{Z}$ and $k = 1, 2$, we have

$$E\bigl(\alpha^n\bigl(\theta_k(y)\bigr)\bigr) = E\bigl(\theta_k\bigl(\tilde{\alpha}^n(y)\bigr)\bigr) = \tilde{E}\bigl(\tilde{\alpha}^n(y)\bigr)$$

so that $E(\alpha^n(\theta_1(y) - \theta_2(y))) = 0$, $n \in \mathbb{Z}$. Since (N, α, E) is nondegenerate, this implies $\theta_1(y) - \theta_2(y) = 0$. $\square$

4. The hierarchy of reversible lifts of L

We have emphasized that the asymptotic lift of a UCP map $L : M \to M$ is characterized by a family of asymptotic formulas (3.1). We now show that it is possible to characterize the asymptotic lift of L in a way that makes no explicit reference to the asymptotic behavior of the sequence $L, L^2, L^3, \ldots$, but rather makes use of a natural ordering of the set of all (equivalence classes of) nondegenerate reversible liftings of L.

Throughout this brief section, $L : M \to M$ will be a fixed UCP map acting on a dual operator system M. We consider the category Rev_L whose objects are *nondegenerate* reversible liftings (N, α, E) of L; a homomorphism from (N_1, α_1, E_1) to (N_2, α_2, E_2) is by definition a UCP map $\theta : N_1 \to N_2$ that satisfies $\theta \circ \alpha_1 = \alpha_2 \circ \theta$ and $E_2 \circ \theta = E_1$. Significantly, a homomorphism in this category gives rise to an embedding of operator systems:

Proposition 4.1. *A homomorphism $\theta : (N_1, \alpha_1, E_1) \to (N_2, \alpha_2, E_2)$ of nondegenerate reversible lifts of L defines an injective map of N_1 to N_2.*

Proof. If $y \in N_1$ satisfies $\theta(y) = 0$, then for every $n \in \mathbb{Z}$ we have

$$E_1\bigl(\alpha_1^n(y)\bigr) = E_2\bigl(\theta\bigl(\alpha_1^n(y)\bigr)\bigr) = E_2\bigl(\alpha_2^n\bigl(\theta(y)\bigr)\bigr) = 0,$$

hence $y = 0$ by nondegeneracy (2.2) of (N_1, α_1, E_1). $\square$

When a homomorphism $\theta : (N_1, \alpha_1, E_1) \to (N_2, \alpha_2, E_2)$ exists, we write

$$(N_1, \alpha_1, E_1) \leqslant (N_2, \alpha_2, E_2).$$

There is an obvious notion of isomorphism in this category, namely that there should exist a map θ as above which has a UCP inverse $\theta^{-1} : N_2 \to N_1$. Obviously, both relations $\geqslant$ and $\leqslant$ hold between isomorphic elements of Rev_L. Conversely,

Proposition 4.2. *Any two nondegenerate reversible liftings (N_k, α_k, E_k) that satisfy $(N_1, \alpha_1, E_1) \leqslant (N_2, \alpha_2, E_2) \leqslant (N_1, \alpha_1, E_1)$ are isomorphic.*

Proof. By hypothesis, there are equivariant UCP maps $\theta : N_1 \to N_2$ and $\phi : N_2 \to N_1$ such that $E_2 \circ \theta = E_1$ and $E_1 \circ \phi = E_2$. Consider the composition $\phi \circ \theta : N_1 \to N_1$. We claim that $\phi \circ \theta$ is the identity map of N_1. Indeed, for every $y \in N_1$ and $n \in \mathbb{Z}$, one can write

$$E_1\big(\alpha_1^n\big((\phi\circ\theta)(y)\big)\big)=E_1\big(\alpha_1^n\circ\phi\circ\theta(y)\big)=E_1\big(\phi\circ\alpha_2^n\circ\theta(y)\big)$$
$$=E_2\big(\alpha_2^n\circ\theta(y)\big)=E_2\big(\theta\circ\alpha_1^n(y)\big)=E_1\big(\alpha_1^n(y)\big).$$

It follows that $E_1(\alpha_1^n(\phi\circ\theta(y)-y))=0$ for all $n\in\mathbb{Z}$. By the nondegeneracy hypothesis (2.2) we conclude that $\phi\circ\theta(y)=y$. By symmetry, $\theta\circ\phi$ is the identity map of N_2. Hence θ is an isomorphism. □

Proposition 4.2 implies that the isomorphism classes of Rev_L form a *bona fide* partially ordered set. There is a smallest element—the class of the trivial lift $(\mathbb{C},\mathrm{id},\iota)$, $\iota:\mathbb{C}\to M$ denoting the inclusion $\iota(\lambda)=\lambda\cdot\mathbf{1}_M$. The following result characterizes the class of an asymptotic lift as the largest element.

Proposition 4.3. *Let $(N_\infty,\alpha_\infty,E_\infty)$ be an asymptotic lift of L. Then every $(N,\alpha,E)\in\mathrm{Rev}_L$ satisfies $(N,\alpha,E)\leqslant(N_\infty,\alpha_\infty,E_\infty)$.*

Proof. As in the proof of Theorem 3.2, we can realize N_∞ as the space of all diagonal operators $N_\infty=\{\mathrm{diag}(x_n):\ (x_n)\in\mathcal{S}\}$, α_∞ as the shift automorphism, and E_∞ as the 0th component map. Let $\theta:N\to N_\infty$ be the UCP map defined in (2.3). We have already pointed out in Remark 2.3 that θ is a homomorphism from (N,α,E) to $(N_\infty,\alpha_\infty,E_\infty)$. □

5. UCP maps on von Neumann algebras

In this section we prove that the asymptotic lift of a UCP map acting on a von Neumann algebra is actually a W^*-dynamical system.

Theorem 5.1. *Let (N,α,E) be the asymptotic lift of a UCP map $L:M\to M$ that acts on a von Neumann algebra M. Then N is a von Neumann algebra and α is a $*$-automorphism of N.*

We will deduce Theorem 5.1 from the following proposition, which is normally used to establish the existence and uniqueness of the Poisson boundary of a noncommutative space of "harmonic functions." The noncommutative Poisson boundary is a far-reaching generalization of the fact that the space of bounded harmonic functions in the open unit disk D is isometrically isomorphic to the abelian von Neumann algebra $L^\infty(\partial D,\frac{d\theta}{2\pi})$ of bounded measurable functions on the boundary ∂D of D. We sketch a proof for the reader's convenience; more detail can be found in [4] and [12]. The result itself appears to have been first discovered in [8, Corollary 1.6].

Proposition 5.2. *Let $\Lambda:M\to M$ be a UCP map on a von Neumann algebra and let H_Λ be the operator system of all harmonic elements of M*

$$H_\Lambda=\big\{x\in M:\ \Lambda(x)=x\big\}.$$

There is a unique associative multiplication $x,y\in H_\Lambda\mapsto x\circ y\in H_\Lambda$ that turns H_Λ into a von Neumann algebra with predual $(H_\Lambda)_*$, on which the group $\mathrm{aut}\,H_\Lambda$ of automorphisms of the operator system structure of H_Λ acts naturally as the group of all $*$-automorphisms.

Sketch of proof. *Uniqueness*. Given two such multiplications, the identity map defines a complete order isomorphism of one C^*-algebra structure to the other. Hence it is a $*$-isomorphism, and the two multiplications agree.

Existence. We claim that there is a completely positive idempotent linear map $E : M \to M$ that has range H_Λ. Indeed, if one topologizes the set of bounded linear maps of M to itself with the topology of point-weak*-convergence, then the set of completely positive unital maps on M becomes a compact space. Let E be any limit point of the sequence of averages

$$A_n = \frac{1}{n}\bigl(\Lambda + \Lambda^2 + \cdots + \Lambda^n\bigr), \quad n = 1, 2, \ldots.$$

Using $\Lambda A_n = A_n \Lambda = \frac{n+1}{n} A_{n+1} - \frac{1}{n}\Lambda$, together with the straightforward estimate $\|A_n - A_{n+1}\| \leqslant \frac{2}{n+1}$, one finds that $\Lambda E = E\Lambda = E$, and from that follows $E^2 = E$ as well as $E(M) = H_\Lambda$.

Such an idempotent E allows one to introduce a Choi–Effros multiplication on H_Λ, $x \circ y = E(xy)$, $x, y \in H_\Lambda$, which makes H_Λ into a C^*-algebra [5, Theorem 3.1, Corollary 3.2]. Since H_Λ is weak*-closed, it is the dual of the Banach space $(H_\Lambda)_*$, and a theorem of Sakai [20, Theorem 1.16.7] implies that H_Λ is a von Neumann algebra with predual $(H_\Lambda)_*$. $\square$

Proof of Theorem 5.1. Let (N, α, E) be the asymptotic lift constructed in the proof of Theorem 3.2, in which $N = \{\mathrm{diag}(x_n)\colon (x_n) \in \mathcal{S}\}$, α is the bilateral shift automorphism $\alpha(X) = UXU^{-1}$, and $E(X) = X_{00}$, for $X \in \mathcal{B}(\tilde{H})$, $\tilde{H} = \ell^2(\mathbb{Z}) \otimes H$ being the Hilbert space of sequences introduced there. In order to prove the existence of a von Neumann algebra structure on N, we appeal to Proposition 5.2 as follows.

Consider the larger von Neumann algebra $\tilde{M} \supseteq N$ of all bounded diagonal operators $Y = \mathrm{diag}(y_n)$ with components $y_n \in M$, and let Λ be the map defined on $\tilde{M}$ by

$$\Lambda\bigl(\mathrm{diag}(y_n)\bigr) = \mathrm{diag}(z_n), \quad \text{where } z_n = L(y_{n+1}),\ n \in \mathbb{Z}.$$

Obviously, Λ is a UCP map of $\tilde{M}$ with the property $\Lambda(Y) = Y$ if and only if Y has the form $Y = \mathrm{diag}(x_n)$ with (x_n) an inverse sequence for L. Thus, N is the space of all Λ-fixed elements of $\tilde{M}$. An application of Proposition 5.2 completes the proof. $\square$

6. Nontriviality of the asymptotic dynamics

A W^*-dynamical system $(A, \mathbb{Z})$ is considered trivial if the automorphism of A that implements the $\mathbb{Z}$-action is the identity automorphism. The purpose of this section is to show that the asymptotic lift of a UCP map is frequently a nontrivial dynamical system.

Theorem 6.1. *For every UCP map $L : M \to M$ of a dual operator system, the following are equivalent.*

(i) *The asymptotic lift (N, α, E) of L satisfies $\alpha(y) = y$, $y \in N$.*
(ii) *Every operator x in $\bigcap_n L^n(\mathrm{ball}\, M)$ satisfies $L(x) = x$.*
(iii) *The semigroup $\{L^n\colon n \geqslant 0\}$ oscillates slowly in the sense that*

$$\lim_{n\to\infty} \bigl\|\rho \circ L^n - \rho \circ L^{n+1}\bigr\| = 0, \quad \rho \in M_*.$$

Proof. (i) $\Rightarrow$ (ii). We have $E(\operatorname{ball} N) = \bigcap_n L^n(\operatorname{ball} M)$ by Lemma 3.6, so every $x \in \bigcap_n L^n(\operatorname{ball} M)$ has the form $x = E(y)$ for some $y \in N$. Hence $L(x) = L(E(y)) = E(\alpha(y)) = E(y) = x$.

(ii) $\Rightarrow$ (iii). Fix $\rho \in M_*$. Another application of Lemma 3.6 implies that $E(\operatorname{ball} N) = \bigcap_n L^n(\operatorname{ball} M)$ is pointwise fixed by L, hence $\|(\rho - \rho \circ L) \circ E\| = 0$. Using (3.1), we obtain

$$\lim_{n\to\infty} \|\rho \circ L^n - \rho \circ L^{n+1}\| = \lim_{n\to\infty} \|(\rho - \rho \circ L) \circ L^n\| = \|(\rho - \rho \circ L) \circ E\|,$$

and the right-hand side is zero.

(iii) $\Rightarrow$ (i). The preceding formula implies that $\|(\rho - \rho \circ L) \circ E\| = 0$ for every $\rho \in M_*$, hence $E = L \circ E$. So for every $n \geqslant 1$ and every $y \in N$,

$$E\big(\alpha^{-n}\big(y - \alpha(y)\big)\big) = E\big(\alpha^{-n}(y)\big) - E\big(\alpha^{-n+1}(y)\big) = E\big(\alpha^{-n}(y)\big) - L\big(E\big(\alpha^{-n}(y)\big)\big) = 0$$

and $\alpha(y) = y$ follows from nondegeneracy (2.2). $\square$

Remark 6.2 *(Matrix algebras).* Many UCP maps acting on matrix algebras are associated with nontrivial W^*-dynamical systems, because of the following observation: a UCP map L of $M_n = M_n(\mathbb{C})$ satisfies property (iii) of Theorem 6.1 $\Leftrightarrow$ $\sigma(L) \cap \mathbb{T} = \{1\}$. Indeed, to prove $\Rightarrow$ contrapositively, suppose there is a point λ in the spectrum of L that satisfies $\lambda \neq 1 = |\lambda|$. Choose an operator $x \in M_n$ satisfying $\|x\| = 1$ and $L(x) = \lambda x$, and choose $\rho \in M_*$ such that $\rho(x) = 1$. Then for all $n \geqslant 0$ we have

$$\begin{aligned}\|\rho \circ L^{n+1} - \rho \circ L^n\| &\geqslant \big|\rho\big(L^{n+1}(x)\big) - \rho\big(L^n(x)\big)\big| = \big|\lambda^{n+1}\rho(x) - \lambda^n \rho(x)\big| \\ &= |\lambda - 1|.\end{aligned}$$

Hence $\{L^n\}$ does not oscillate slowly.

Conversely, if $\sigma(L) \cap \mathbb{T} = \{1\}$, then since points of $\sigma(L) \cap \mathbb{T}$ are associated with *simple* eigenvectors of L, the spectrum of the restriction L_0 of L to the range of $\mathrm{id} - L$ is contained in the open unit disk $\{z\colon |z| < 1\}$, hence $\|L_0^n\| \to 0$ as $n \to \infty$. It follows that $L^{n+1} - L^n$ tends to zero (in norm, say) as $n \to \infty$; and that obviously implies condition (iii) of Theorem 6.1. We remark that the asymptotic behavior of $\|T^{n+1} - T^n\|$ for contractions T on Banach spaces has been much-studied; see [14] and references therein.

Elementary examples show that any finite subset of the unit circle that contains 1 and is stable under complex conjugation can occur as $\sigma(L) \cap \mathbb{T}$ for a UCP map L of a matrix algebra M_n (for appropriately large n). Hence there are many examples of UCP maps on finite-dimensional non-commutative von Neumann algebras whose asymptotic liftings are nontrivial finite-dimensional W^*-dynamical systems.

The asymptotic behavior of UCP maps on finite-dimensional algebras is discussed more completely in Section 9.

Remark 6.3 *(Automorphisms and endomorphisms).* Automorphisms are at the opposite extreme from slowly oscillating UCP maps. Indeed, if α is any automorphism of a dual operator system M, then α induces an isometry of the predual M_* via $\omega \mapsto \omega \circ \alpha$, and for every $\rho \in M_*$ and $n \in \mathbb{Z}$ we have

$$\|\rho \circ \alpha^{n+1} - \rho \circ \alpha^n\| = \|(\rho \circ \alpha - \rho) \circ \alpha^n\| = \|\rho \circ \alpha - \rho\|.$$

This formula obviously implies that the only slowly oscillating automorphism is the identity automorphism.

If α is merely an isometric UCP map on M and $M_\infty = \bigcap_n \alpha^n(M)$ is the "tail" operator system, then for every $\rho \in M_*$ we have

$$\left\|\rho \circ \alpha^{n+1} - \rho \circ \alpha^n\right\| = \left\|(\rho \circ \alpha - \rho) \circ \alpha^n\right\| = \left\|(\rho \circ \alpha - \rho)\restriction_{\alpha^n(M)}\right\|,$$

and the right-hand side decreases to $\|(\rho \circ \alpha - \rho)\restriction_{M_\infty}\|$ as $n \to \infty$ by Lemma 3.5. Hence α oscillates slowly iff it restricts to the identity map on M_∞.

7. Identification of the asymptotic dynamics

Let $L : M \to M$ be a UCP map acting on a von Neumann algebra M and let (N, α, E) be its asymptotic lift. We have seen that $(N, \mathbb{Z})$ is a W^*-dynamical system; indeed, the proof of Theorem 5.1 shows that the algebraic structure of N is that of the noncommutative Poisson boundary of an associated UCP map $\Lambda : \tilde{M} \to \tilde{M}$ on another von Neumann algebra.

In general, it can be a significant problem to find a concrete realization of the Poisson boundary, even when $M = L^\infty(X, \mu)$ is commutative (see [16]—this is called the *identification problem* in [13]). In the noncommutative case, there are only a few examples for which this problem has been effectively solved. Three of them are discussed in [12].

In this section we contribute to this circle of ideas by showing that the asymptotic lift of a UCP map L on a von Neumann algebra is isomorphic to the tail flow of the minimal dilation of L to a $*$-endomorphism whenever the minimal dilation has trivial kernel—which is automatic whenever L acts on a factor. We will clarify the precise relation between the asymptotic lift of L and the Poisson boundary of L in Section 8.

It will not be necessary to reiterate explicit details of the dilation theory of UCP maps acting on von Neumann algebras (see [2, Chapter 8]). Instead, we simply recall that every UCP map $L : M \to M$ acting on a von Neumann algebra can be dilated *minimally* to a normal unit-preserving $*$-endomorphism of a larger von Neumann algebra $\alpha : N \to N$ that contains $M = pNp$ as a corner. Any two minimal dilations of L are naturally isomorphic.

More generally, any unit-preserving normal isometric $*$-endomorphism $\alpha : N \to N$ of a von Neumann algebra N gives rise to a decreasing sequence of von Neumann subalgebras $N \supseteq \alpha(N) \supseteq \alpha^2(N) \supseteq \cdots$ whose intersection

$$A = \bigcap_{n \geqslant 0} \alpha^n(N)$$

is a von Neumann algebra with the property that the restriction of α to A is a $*$-*automorphism* of A. That W^*-dynamical system $(A, \mathbb{Z})$ (also written $(A, \alpha\restriction_A)$) is called the *tail flow* of the endomorphism $\alpha : N \to N$. The tail flow is clearly an interesting conjugacy invariant of the endomorphism α, but it has received little attention in the past.

The following result identifies the asymptotic lift of L as the tail flow of the minimal dilation of L whenever the minimal dilation has trivial kernel. Actually, we prove somewhat more, since the setting of Theorem 7.1 includes dilations that are not necessarily minimal.

Theorem 7.1. *Let $\alpha : N \to N$ be a unital normal isometric endomorphism of a von Neumann algebra N, let $p \in N$ be a projection in N satisfying $p \leqslant \alpha(p)$ and $\alpha^n(p) \uparrow \mathbf{1}_N$ as $n \uparrow \infty$. Let $M = pNp$ be the corresponding corner of N and let $L : M \to M$ be the UCP map defined by*

$$L(x) = p\alpha(x)p, \quad x \in M.$$

Then the asymptotic lift of $L : M \to M$ is isomorphic to $(A, \alpha\restriction_A, E)$, where $(A, \alpha\restriction_A)$ is the tail flow of α and $E : A \to M$ is the map $E(a) = pap$.

Proof. Obviously, $\alpha\restriction_A$ is an automorphism of the operator system structure of A, and E is a UCP map of A to $M = pNp$. We claim that the nondegeneracy requirement $E(\alpha^{-n}(a)) = 0,\ n \geqslant 1 \Rightarrow a = 0$ is satisfied. Indeed, fixing such an $a \in A$ and $n \geqslant 1$, we have

$$\alpha^n(p)a\alpha^n(p) = \alpha^n\big(p\alpha^{-n}(a)p\big) = \alpha^n\big(E\big(\alpha^{-n}(a)\big)\big) = 0, \quad n = 1, 2, \ldots.$$

Since $\alpha^n(p)$ converges strongly to $\mathbf{1}$ as $n \to \infty$, $\alpha^n(p)a\alpha^n(p)$ converges strongly to a, hence $a = 0$.

It remains to establish the formulas (3.1):

$$\lim_{k\to\infty} \big\|\rho \circ \big(\mathrm{id}_n \otimes L^k\big)\big\| = \big\|\rho \circ (\mathrm{id}_n \otimes E)\big\|, \tag{7.1}$$

for every $\rho \in (M_n \otimes M)_*$ and every $n = 1, 2, \ldots$. We first consider the case $n = 1$. Choose $\rho \in M_*$ and define $\bar\rho \in N_*$ by $\bar\rho(y) = \rho(pyp)$, $y \in N$. For every $k \geqslant 1$, we claim

$$\big\|\rho \circ L^k\big\| = \big\|\bar\rho \circ \alpha^k\big\| \quad \text{and} \quad \|\rho \circ E\| = \|\bar\rho\restriction_A\|. \tag{7.2}$$

Indeed, since $L^k(x) = p\alpha^k(x)p$ for all $x \in M$, it follows that for every $y \in N$ we have $\rho(L^k(pyp)) = \rho(p\alpha^k(pyp)p) = \rho(p\alpha^k(y)p) = \bar\rho(\alpha^k(y))$, and that identity clearly implies $\|\rho \circ L^k\| = \|\bar\rho \circ \alpha^k\|$. The second formula of (7.2) follows from the identity $\rho(E(a)) = \rho(pap) = \bar\rho(a)$ for $a \in A$.

If we now apply (3.5) to the decreasing sequence of weak*-compact sets $K_j = \alpha^j(\operatorname{ball} N)$, $j = 1, 2, \ldots$, having intersection $K_\infty = \operatorname{ball} A$, we find that $\|\bar\rho \circ \alpha^j\| = \sup\{|\rho(y)|\colon\ y \in K_j\}$ and $\|\bar\rho\restriction_A\| = \sup\{|\bar\rho(a)|\colon\ a \in \operatorname{ball} A\}$, so by Lemma 3.5 we conclude

$$\lim_{k\to\infty} \big\|\rho \circ L^k\big\| = \lim_{k\to\infty} \big\|\bar\rho \circ \alpha^k\big\| = \|\bar\rho\restriction_A\| = \|\rho \circ E\|,$$

proving (7.1).

This argument applies *verbatim* to establish (7.1) throughout the matrix hierarchy. Indeed, for $n \geqslant 2$, the hypotheses of Theorem 7.1 carry over to the corner $M^{(n)} = p_n N^{(n)} p_n$, where $p_n = \mathbf{1}_{M_n} \otimes p$, with α replaced by $\mathrm{id}_n \otimes \alpha$. We have $p_n \leqslant (\mathrm{id}_n \otimes \alpha)(p_n) \leqslant (\mathrm{id}_n \otimes \alpha^2)(p_n) \leqslant \cdots \uparrow \mathbf{1}_{N^{(n)}}$, and for a fixed $\rho \in M_*^{(n)}$ there are appropriate versions of the formulas (7.2). □

8. Identification of the Poisson boundary

In this section we show how the Poisson boundary can be characterized in terms of the asymptotic lift. Since the asymptotic lift can often be calculated explicitly—either directly as in

Section 9 or using the tools of dilation theory and Theorem 7.1—this result contributes to the identification problem for noncommutative Poisson boundaries as discussed in Section 7.

Let $L : M \to M$ be a UCP map acting on a von Neumann algebra, and consider the dual operator system of all noncommutative harmonic elements

$$H_L = \{x \in M:\ L(x) = x\}.$$

We have pointed out in Section 5 that H_L carries a unique von Neumann algebra structure, and that von Neumann algebra is called the Poisson boundary of the map $L : M \to M$. Given a $*$-automorphism of a von Neumann algebra N, we write $N^\alpha = \{y \in N:\ \alpha(y) = y\}$ for its fixed subalgebra.

Proposition 8.1. *Let $L : M \to M$ be a UCP map on a von Neumann algebra, let H_L be its Poisson boundary, and let (N, α, E) be its asymptotic lift. Then the restriction of E to the fixed algebra N^α implements an isomorphism of von Neumann algebras $N^\alpha \cong H_L$.*

Proof. The equivariance property $E \circ \alpha = L \circ E$, implies $E(N^\alpha) \subseteq H_L$. For the opposite inclusion, every element $a \in H_L$ gives rise to an inverse sequence $\bar{a} = (\ldots, a, a, a, \ldots)$ and, after realizing (N, α, E) concretely as in Section 3, we conclude that $a = E(\operatorname{diag} \bar{a}) \in E(N^\alpha)$. Finally, a straightforward application of formula (3.9) of Lemma 3.7 throughout the matrix hierarchy shows that $E\restriction_{N^\alpha}$ is a complete isometry. Since both N^α and H_L are von Neumann algebras, $E\restriction_{H_L}$ is a $*$-isomorphism.

9. Examples and concluding remarks

Theorem 7.1 identifies the asymptotic lift of a UCP map in terms of its minimal dilation. While one can often calculate properties of the minimal dilation in explicit terms (see [2, Chapter 8]), those computations can be cumbersome and sometimes difficult. On the other hand, given a specific UCP map, we have found that it is often easier to calculate its asymptotic lift directly in concrete terms. The purpose of this section is to illustrate that fact by carrying out calculations for some examples that require a variety of techniques.

9.1. Stochastic matrices

It is appropriate to begin with the classical commutative case having its origins in the theory of Markov chains. Let $P = (p_{ij})$ be an $n \times n$ matrix of nonnegative numbers satisfying $\sum_j p_{ij} = 1$ for every $i = 1, \ldots, n$. If we view the elements of the von Neumann algebra $M = \mathbb{C}^n$ as column vectors, then P gives rise to a UCP map on M by matrix multiplication. We now calculate the asymptotic lift of P, and we relate that to classical results of Frobenius [9,10], generalizing earlier results of Perron [18,19], on the asymptotic behavior of such matrices. To keep the discussion as simple as possible, we restrict attention to the case where P is *irreducible* in the sense that the only projections $e \in M$ satisfying $P(e) \leqslant e$ are $e = 0$ and $e = \mathbf{1}$ (but see Remark 9.4). We write $\sigma(P)$ for the spectrum of the linear operator $P : \mathbb{C}^n \to \mathbb{C}^n$; $\sigma(P)$ is a subset of the closed unit disk that contains the eigenvalue 1.

In this context, Theorem 2 of [11, see pp. 65–75] can be paraphrased as follows.

Theorem 9.1. *Let P be an irreducible stochastic $n \times n$ matrix. Then there is a k, $1 \leqslant k \leqslant n$, such that $\sigma(P) \cap \mathbb{T} = \{1, \zeta, \zeta^2, \ldots, \zeta^{k-1}\}$ is the set of all kth roots of unity, $\zeta = e^{2\pi i/k}$, each ζ^j being* *a simple eigenvalue. When $k > 1$, there is a permutation matrix U such that UPU^{-1} has cyclic form*

$$UPU^{-1} = \begin{pmatrix} 0 & C_0 & 0 & \dots & 0 \\ 0 & 0 & C_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & C_{k-2} \\ C_{k-1} & 0 & 0 & \dots & 0 \end{pmatrix} \tag{9.1}$$

in which the C_j are rectangular submatrices, and where the zero submatrices along the diagonal are all square.

Notice that we can eliminate the unitary permutation matrix U entirely by replacing P with a suitably relabeled stochastic matrix, and we do so. For each $j = 0, 1, \dots, k-1$, let e_j be the projection corresponding to the domain subspace of the block C_j. These projections are mutually orthogonal, have sum $\mathbf{1}$, and satisfy $P(e_j) \leqslant e_{j+1}$ (addition modulo k). Making use of $P(\mathbf{1}) = \mathbf{1}$, we find that in fact, $P(e_j) = e_{j+1}$ for every j. Let N be the linear span of $e_0, \dots, e_{k-1}$ and let α be the restriction of P to N. Clearly N is a $*$-subalgebra of M and $\alpha : N \to N$ is the $*$-automorphism associated with this cyclic permutation of the minimal projections $e_0, e_1, \dots, e_{k-1}$ of N.

Proposition 9.2. *The asymptotic lift of $P : M \to M$ is the triple (N, α, E), where $E : N \subseteq M$ is the inclusion map.*

Proof. For $k > 1$ as above, consider the UCP map $P^k : M \to M$. The spectrum of P^k consists of the eigenvalue 1, together with other spectral points in the open unit disk $\{|z| < 1\}$, the eigenvalue 1 having eigenspace N. Hence the sequence of powers $P^k, P^{2k}, P^{3k}, \dots$ converges to an idempotent linear map Q on M having range N. Clearly Q is a UCP projection onto N. For every $\rho \in M_*$, the norms $\|\rho \circ P^r\|$ decrease as r increases, hence

$$\lim_{r\to\infty} \|\rho \circ P^r\| = \lim_{m\to\infty} \|\rho \circ (P^k)^m\| = \|\rho \circ Q\| = \|\rho\restriction_N\| = \|\rho \circ E\|. \tag{9.2}$$

The same argument applies throughout the matrix hierarchy over M, and we conclude that the two criteria of Definition 3.1 are satisfied. Hence (N, α, E) is the asymptotic lift of P. □

Remark 9.3 *(Asymptotics of the powers of P).* The idempotent UCP map

$$Q = \lim_{m\to\infty} P^{mk}$$

exhibited in the proof of Proposition 9.2 provides a precise sense in which the asymptotic lift (N, α, E) contains all of the asymptotic information about the sequence of powers $P, P^2, P^3, \dots$. Indeed, we claim that there are positive constants c, r with $r < 1$ such that

$$\|P^n - \alpha^n Q\| \leqslant cr^n, \quad n = 1, 2, 3, \dots. \tag{9.3}$$

This follows from the fact that Q is an idempotent that commutes with P with the property that $\alpha = P\restriction_{N=\operatorname{ran} Q} f$ has spectrum in the unit circle and $P\restriction_{\operatorname{ran}(\mathrm{id}-Q)}$ has spectrum in $\{|z| < 1\}$.

Choosing a number $r < 1$ so that $\{|z| < r\}$ contains the spectrum of $P\restriction_{\mathrm{ran}(\mathrm{id}-Q)}$, the spectral radius formula of elementary Banach algebra theory implies that the sequence

$$r^{-n}\|P^n - \alpha^n Q\| = r^{-n}\|P^n - P^n Q\| = \left(r^{-1}\|P^n\restriction_{\mathrm{ran}(\mathrm{id}-Q)}\|^{1/n}\right)^n$$

tends to zero as $n \to \infty$, hence there is a $c > 0$ so that (9.3) is satisfied.

Remark 9.4 *(Reducibility and noncommutativity).* In the following subsection, we generalize the above result to the case of UCP maps acting on noncommutative finite-dimensional von Neumann algebras. In particular, the discussion of Section 9.2 applies equally to the reducible cases not covered in the statement of Theorem 9.1.

9.2. Finite-dimensional von Neumann algebras

One can compute the asymptotic lift of a UCP map on a finite-dimensional von Neumann algebra explicitly, using nothing but elementary methods along with the Choi–Effros multiplication. Since these results have significance for quantum computing [15], we sketch that calculation in some detail.

Let $L : M \to M$ be a UCP map on a finite-dimensional von Neumann algebra. For every point $\lambda \in \sigma(L) \cap \mathbb{T}$ let

$$N_\lambda = \{x \in M\colon\ L(x) = \lambda x\}$$

and let

$$N = \sum_{\lambda \in \sigma(L)\cap\mathbb{T}} N_\lambda$$

be the sum of these maximal eigenspaces. The identity operator belongs to N, and from the property $L(x^*) = L(x)^*$, $x \in M$, one deduces $N^* = N$. Hence N is an operator system such that the restriction $\alpha = L\restriction_N$ of L to N is a *diagonalizable* UCP map with spectrum $\sigma(L) \cap \mathbb{T}$. Indeed, it is not hard to show that α is a UCP automorphism of N.

We digress momentarily to point out that in the classical setting of Frobenius' result for irreducible stochastic matrices P as formulated in Theorem 9.1, N coincides with the span of the projections $e_0, \ldots, e_{k-1}$ constructed above, the eigenvector associated with a kth root of unity $\lambda \in \sigma(P) \cap \mathbb{T}$ being given by

$$x_\lambda = e_0 + \bar{\lambda} e_1 + \bar{\lambda}^2 e_2 + \cdots + \bar{\lambda}^{k-1} e_{k-1}.$$

It follows that, in such commutative cases, N is closed under multiplication.

In the more general setting under discussion here, $\sigma(L) \cap \mathbb{T}$ need not consist of roots of unity and N need not be closed under multiplication. But in all cases N can be made into a von Neumann algebra. The most transparent proof of that fact uses the following observation of Kuperberg ([15], also see [3, Theorem 2.6]).

Lemma 9.5. *There is an increasing sequence of integers $n_1 < n_2 < \cdots$ such that $L^{n_1}, L^{n_2}, \ldots$ converges to an idempotent UCP map Q with the property $N = Q(M)$. In fact, Q is the unique idempotent limit point of the sequence of powers $L, L^2, L^3, \ldots$.*

One now uses the idempotent Q to introduce a Choi–Effros multiplication

$$x \circ y = Q(xy), \quad x, y \in N,$$

in N, thereby making it into a finite-dimensional von Neumann algebra. Hence (N, α) becomes a (typically noncommutative) W^*-dynamical system, and the inclusion $E : N \subseteq M$ becomes a UCP map satisfying $E \circ \alpha = L \circ E$.

Proposition 9.6. *The triple (N, α, E) is the asymptotic lift of $L : M \to M$.*

Sketch of proof. The inclusion of N in M is injective, hence the nondegeneracy condition (2.2) is satisfied. Thus we need only show that equality holds in the formulas (3.1), and that follows by a proof paralleling that of (9.2). Indeed, letting $n_1 < n_2 < \cdots$ be a sequence such that $L^{n_k} \to Q$ as in Lemma 9.5, one finds that for every bounded linear functional ρ on M,

$$\lim_{n\to\infty} \left\| \rho \circ L^n \right\| = \lim_{k\to\infty} \left\| \rho \circ L^{n_k} \right\| = \| \rho \circ Q \| = \| \rho \restriction_N \| = \| \rho \circ E \|.$$

Obviously, the same argument can be promoted throughout the matrix hierarchy, as one allows ρ to range over the dual of $M^{(n)}$, $n = 1, 2, \ldots$. □

Remark 9.7 *(Asymptotics of the powers of L).* Making use of the idempotent Q much as in Remark 9.3, one finds that (N, α, E) contains all asymptotic information about the sequence $L, L^2, L^3, \ldots$ because of the following precise estimate: There are positive constants c, r such that $r < 1$ and

$$\left\| L^n - \alpha^n Q \right\| \leqslant c r^n, \quad n = 1, 2, \ldots. \tag{9.4}$$

9.3. UCP maps on II_1 factors

We now calculate the asymptotic lifts of a family of nontrivial UCP maps acting on the hyperfinite II_1 factor R, the point being to show that the asymptotic lifts of these maps can be arbitrary $*$-automorphisms of R. There are many variations on these examples that exhibit a variety of phenomena in other von Neumann algebras. Here, we confine ourselves to the simplest nontrivial cases.

Let τ be the tracial state of R, and let P be any normal UCP map of R having τ as an absorbing state in the sense that for every normal state ρ of R, one has

$$\lim_{k\to\infty} \left\| \rho \circ P^k - \tau \right\| = 0. \tag{9.5}$$

Of course, the simplest such map is $P(x) = \tau(x)\mathbf{1}$; but we have less trivial examples in mind. For example, let $A_\theta = C^*(U, V)$ be the irrational rotation C^*-algebra, where U, V are unitaries satisfying $UV = e^{2\pi i\theta} VU$ and θ is an irrational number. Then for every number λ in the open unit interval (0, 1) one can show that there is a completely positive unit-preserving map P_λ that is defined uniquely on A_θ by its action on monomials:

$$P_\lambda\left(U^p V^q\right) = \lambda^{|p|+|q|} U^p V^q, \quad p, q \in \mathbb{Z}.$$

Moreover, since in every finite linear combination $x = \sum_{p,q} a_{p,q} U^p V^q$, the coefficients satisfy $|a_{p,q}| \leqslant \|x\|$, $p, q \in \mathbb{Z}$, a straightforward estimate leads to

$$\sup_{\|x\|\leqslant 1} \left\| P_\lambda^k(x) - \tau(x)\mathbf{1} \right\| \leqslant \sum_{(p,q)\neq(0,0)} \lambda^{(|p|+|q|)^k} \leqslant \lambda^{k-1} \sum_{(p,q)\neq(0,0)} \lambda^{|p|+|q|} = c\lambda^k,$$

for some positive constant c. We conclude that

$$\lim_{k\to\infty} \left\| P_\lambda^k - \tau(\cdot)\mathbf{1} \right\| = 0,$$

and in particular, (9.5) holds for every state ρ of $C^*(U, V)$.

The GNS construction applied to the tracial state of A_θ now provides a representation π of A_θ with the property that the weak closure of $\pi(A_\theta)$ is R, and there is a unique UCP map P on R such that $P(\pi(x)) = \pi(P_\lambda(x))$ for all $x \in A_\theta$. The preceding paragraph implies that the extended map $P : R \to R$ has the asserted property (9.5).

Similar examples of UCP maps acting on other II_1 factors can be constructed by replacing R with the group von Neumann algebras of discrete groups with infinite conjugacy classes (see [3, Proposition 4.4]).

Choose a UCP map $P : R \to R$ having property (9.5), choose an arbitrary $*$-automorphism α of R, and consider the UCP map $L = P \otimes \alpha$ defined uniquely on the spatial tensor product $R \otimes R$ by

$$L(x \otimes y) = P(x) \otimes \alpha(y), \quad x, y \in R.$$

Since $R \otimes R$ is isomorphic to R, one can think of L as a UCP map on R.

Proposition 9.8. *Let $E : R \to R \otimes R$ be the map $E(x) = \mathbf{1} \otimes x$, $x \in R$. The asymptotic lift of $L = P \otimes \alpha$ is the triple (R, α, E).*

Sketch of proof. E is clearly an injective UCP map of von Neumann algebras satisfying the equivariance condition $E \circ \alpha = L \circ E$. Thus it remains only to verify the formulas (3.1). For that, we will prove a stronger asymptotic relation. Let $Q : R \otimes R \to \mathbf{1} \otimes R$ be the τ-preserving conditional expectation

$$Q(x \otimes y) = \tau(x)y, \quad x, y \in R.$$

We claim that for all $\rho \in (R \otimes R)_*$, one has

$$\lim_{k\to\infty} \left\| \rho \circ L^k - \rho \circ (\mathrm{id_R} \otimes \alpha)^k \circ Q \right\| = 0. \tag{9.6}$$

Indeed, since $(R \otimes R)_*$ is the norm-closed linear span of functionals of the form $\rho_1 \otimes \rho_2$ with $\rho_k \in R_*$, obvious estimates show that it suffices to prove (9.6) for decomposable functionals of the form $\rho_1 \otimes \rho_2$, where $\rho_1, \rho_2 \in R_*$. Fix $\rho = \rho_1 \otimes \rho_2$ of this form. Using the decomposition

$$\big(L^k - (\mathrm{id}_R \otimes \alpha)^k \circ Q\big)(x \otimes y) = \big(P^k(x) - \tau(x)\mathbf{1}\big) \otimes \alpha^k(y),$$

we find that $\|\rho_1 \otimes \rho_2(L^k - (\mathrm{id}_R \otimes \alpha)^k \circ Q)\|$ decomposes into a product

$$\begin{aligned}\left\|(\rho_1 \otimes \rho_2)\big(\big(P^k - \tau(\cdot)\mathbf{1}\big) \otimes \alpha^k\big)\right\| &= \left\|\rho_1 \circ \big(P^k - \tau(\cdot)\mathbf{1}\big)\right\| \cdot \left\|\rho_2 \circ \alpha^k\right\| \\ &= \left\|\rho_1 \circ \big(P^k - \tau(\cdot)\mathbf{1}\big)\right\| \cdot \|\rho_2\| \\ &= \left\|\rho_1 \circ P^k - \tau\right\| \cdot \|\rho_2\|,\end{aligned}$$

which by (9.5), tends to zero as $k \to \infty$.

Note that (9.6) leads immediately to the case $n = 1$ of (3.1), since

$$\lim_{k\to\infty} \left\|\rho \circ L^k\right\| = \lim_{k\to\infty} \left\|\rho \circ \big(\mathrm{id}_R \otimes \alpha^k\big) \circ Q\right\| = \lim_{k\to\infty} \left\|\rho \circ Q \circ \alpha^k\right\| = \|\rho \circ Q\| = \|\rho \circ E\|.$$

With trivial changes, these arguments can be repeated throughout the matrix hierarchy, after one identifies $M_n \otimes (R \otimes R)$ with $(M_n \otimes R) \otimes R$. We omit those mind-numbing details. □

9.4. The CCR heat flow

The three asymptotic assertions (9.3), (9.4), (9.6) are much stronger than the requirements of (3.1), and one might ask if those stronger results can be established for the asymptotic lifts (N, α, E) of more general UCP maps on von Neumann algebras $L : M \to M$. In each of the preceding examples, it was possible to identify N with a dual operator subsystem $M_\infty \subseteq M$ (namely the range $E(N)$ of E), α with the restriction $\alpha_\infty = L\restriction_{M_\infty}$ of L to N, and E with the inclusion map $\iota : N \subseteq M$. There was also an idempotent completely positive projection Q mapping M onto M_∞, and together, these objects gave rise to the asymptotic relations

$$\lim_{k\to\infty} \left\|\rho \circ L^k - \rho \circ \alpha_\infty^k \circ Q\right\| = 0, \quad \rho \in M_*. \tag{9.7}$$

The relative strength of (9.7) and (3.1) is clearly seen if one reformulates the case $n = 1$ of (3.1) in this context as the following assertion:

$$\lim_{k\to\infty} \left|\left\|\rho \circ L^k\right\| - \left\|\rho \circ \alpha_\infty^k \circ Q\right\|\right| = \lim_{k\to\infty} \left|\left\|\rho \circ L^k\right\| - \|\rho \circ E\|\right| = 0$$

(see the proof of Proposition 9.8).

So it is natural to ask if the stronger relations (9.7) can be established more generally, at least in cases where $N = M_\infty \subseteq M$ is a subspace of M and $\alpha = \alpha_\infty$ is obtained by restricting L to M_∞. That is true, for example, in the more restricted context of [3]. The purpose of these remarks is to show that the answer is no in general, by describing examples with the two properties $M_\infty \subseteq M$ and $\alpha = L\restriction_{M_\infty}$, but for which there is no idempotent completely positive map Q satisfying (9.7).

The CCR heat flow is a semigroup of UCP maps $\{P_t\colon\ t \geqslant 0\}$ acting on the von Neumann algebra $M = \mathcal{B}(H)$ [1]. Consider the single UCP map $L = P_{t_0}$ for some fixed $t_0 > 0$. This map has the following two properties: (a) there is no normal state $\rho \in M_*$ satisfying $\rho \circ L = \rho$, and (b) for any two normal states $\rho_1, \rho_2 \in M_*$ one has

$$\lim_{k\to\infty} \left\|\rho_1 \circ L^k - \rho_2 \circ L^k\right\| = 0. \tag{9.8}$$

We claim first that the asymptotic lift of this L is the triple $(\mathbb{C}, \mathrm{id}, \iota)$, where ι is the inclusion of $\mathbb{C}$ in M, $\iota(\lambda) = \lambda \cdot \mathbf{1}_M$. Indeed, to sketch the proof of the key assertion—namely the case $n = 1$ of formula (3.1)—choose $\rho \in M_*$. We have $\|\rho \circ \iota\| = |\rho(\mathbf{1})|$, so that (3.1) reduces in this case to the formula

$$\lim_{k\to\infty} \left\|\rho \circ L^k\right\| = \left|\rho(\mathbf{1})\right|, \quad \rho \in M_*. \tag{9.9}$$

It is a simple exercise to show that (9.9) and (9.8) are in fact equivalent assertions about L—note for example that (9.8) is the special case of (9.9) in which $\rho(\mathbf{1}) = 0$—and of course the restriction of L to $\mathbb{C}\cdot\mathbf{1}$ is the identity map $\alpha_\infty = \mathrm{id}$. This argument promotes naturally throughout the matrix hierarchy over M, hence $(\mathbb{C}, \mathrm{id}, \iota)$ is the asymptotic lift of L.

We now show that one cannot obtain formulas like (9.7) for this example.

Proposition 9.9. *There is no completely positive projection Q of M on $\mathbb{C}\cdot\mathbf{1}$ that satisfies*

$$\lim_{k\to\infty} \left\|\rho \circ L^k - \rho \circ L^k \circ Q\right\| = 0, \quad \rho \in M_*. \tag{9.10}$$

Proof. Indeed, a completely positive idempotent Q with range $\mathbb{C}\cdot\mathbf{1}$ would have the form $Q(x) = \omega(x)\mathbf{1}$, $x \in M$, where ω is a state of M. For any normal state ρ we have $\rho \circ L^k \circ Q(x) = \rho(L^k(\omega(x)\mathbf{1})) = \omega(x)$, $x \in M$, hence in this case (9.10) makes the assertion

$$\lim_{k\to\infty} \left\|\rho \circ L^k - \omega\right\| = 0;$$

i.e., ω is an *absorbing state* for L. But an absorbing state ω for L is a *normal* state that satisfies $\omega \circ L = \omega$, contradicting property (a) above. $\square$

Remark 9.10 *(Further examples).* One can generalize this construction based on L. For example, let α be a $*$-automorphism of $\mathcal{B}(H)$ and consider the UCP map $L \otimes \alpha$ defined on $M = \mathcal{B}(H \otimes H)$. One can show that $L \otimes \alpha$ has asymptotic lift $(\mathcal{B}(H), \alpha, E)$ where $E : \mathcal{B}(H) \to M$ is the UCP map $E(x) = \mathbf{1} \otimes x$, much as in the proof of Proposition 9.8. With suitable choices of α (specifically, for any α that has a normal invariant state), one can also show that there is no completely positive projection $Q : M \to \mathbf{1} \otimes \mathcal{B}(H)$ that satisfies (9.10). Thus, even though a UCP map on a von Neumann algebra always has an asymptotic lift, *there are many examples for which one cannot expect precise asymptotic formulas such as* (9.7).

Acknowledgment

I want to thank Masaki Izumi for useful discussions about the properties of noncommutative Poisson boundaries during a pleasant visit to Kyoto in 2005. We will make use of the noncommutative Poisson boundary in the proof of Theorem 5.1; and we relate it to the fixed algebra of the asymptotic lift in Section 8.

References

[1] W. Arveson, The heat flow of the CCR algebra, Bull. London Math. Soc. 34 (2002) 73–83, arXiv: math.OA/0005250.

[2] W. Arveson, Noncommutative Dynamics and E-Semigroups, Monogr. Math., Springer-Verlag, New York, 2003.

[3] W. Arveson, Asymptotic stability I: Completely positive maps, Internat. J. Math. 15 (3) (2004) 289–312, arXiv: math.OA/0304488.

[4] W. Arveson, Noncommutative Poisson boundaries, unpublished. Available from http://math.berkeley.edu/~arveson/Dvi/290F04/22Sept.pdf, September 2004.
[5] M.-D. Choi, E. Effros, Injectivity and operator spaces, J. Funct. Anal. 24 (2) (1977) 156–209.
[6] J.L. Doob, Stochastic Processes, Publ. in Stat., Wiley, New York, 1953.
[7] E.G. Effros, Z.-J. Ruan, Operator Spaces, Oxford Univ. Press, Oxford, 2000.
[8] E.G. Effros, E. Størmer, Positive projections and Jordan structure in operator algebras, Math. Scand. 45 (1979) 127–138.
[9] G. Frobenius, Über Matrizen aus positiven Elementen, Sitzber. Akad. Wiss. Berlin, Phys. Math. Kl. (1909) 471–476.
[10] G. Frobenius, Über Matrizen aus nicht negativen Elementen, Sitzber. Akad. Wiss. Berlin, Phys. Math. Kl. (1912) 456–477.
[11] F.R. Gantmacher, Applications of the Theory of Matrices, Interscience, New York, 1959.
[12] M. Izumi, Non-commutative Poisson boundaries, in: M. Kotani, T. Shirai, T. Sunada (Eds.), Discrete Geometric Analysis, in: Contemp. Math., vol. 347, Amer. Math. Soc., 2004.
[13] V.A. Kaimanovich, Boundaries of invariant Markov operators: The identification problem, in: Discr. Geom. Anal., in: London Math. Soc. Lecture Note Ser., vol. 228, Cambridge Univ. Press, 1996, pp. 127–176.
[14] Y. Katznelson, L. Tzafiri, On power bounded operators, J. Funct. Anal. 68 (1986) 313–328.
[15] G. Kuperberg, The capacity of hybrid quantum memory, IEEE Trans. Inform. Theory 49 (6) (2003) 1465–1473, arXiv: quant-ph/0203105 v2.
[16] V.A. Kaimanovich, A.M. Vershik, Random walks on discrete groups: Boundary and entropy, Ann. Probab. 11 (3) (1983) 457–490.
[17] V. Paulsen, Completely Bounded Maps and Operator Algebras, Cambridge Univ. Press, Cambridge, 2002.
[18] O. Perron, Grundlagen für eine Theorie des jacobischen Kettenbruchalgorithmus, Math. Ann. 64 (1907) 1–76.
[19] O. Perron, Zur Theorie der Matrices, Math. Ann. 64 (1907) 248–263.
[20] S. Sakai, C^*-Algebras and W^*-Algebras, Classics Math., Springer-Verlag, New York, 1998. Reprinted from the 1971 edition.
[21] E. Størmer, Multiplicative properties of positive maps, preprint, 2006, pp. 1–8, arXiv: math.OA/0607090.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 248 (2007) 225–257
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

Optimal transport maps for Monge–Kantorovich problem on loop groups

Shizan Fang [a,*], Jinghai Shao [b]

[a] IMB UMR 5584, University of Bourgundy, Dijon, France
[b] School of Mathematics, Beijing Normal University, Beijing, China

Received 4 October 2006; accepted 19 November 2006

Available online 16 January 2007

Communicated by Paul Malliavin

Abstract

Let G be a compact Lie group, and consider the loop group $\mathcal{L}_eG := \{\ell \in C([0,1], G); \ell(0) = \ell(1) = e\}$. Let ν be the heat kernel measure at the time 1. For any density function F on $\mathcal{L}_eG$ such that $\mathrm{Ent}_\nu(F) < \infty$, we shall prove that there exists a unique optimal transportation map $\mathcal{T} : \mathcal{L}_eG \to \mathcal{L}_eG$ which pushes ν forward to $F\nu$.

Keywords: Wasserstein distance; Optimal transportation; Monge–Kantorovich problem; Loop groups; Riemannian distance; Heat measures

1. Introduction

The Monge–Kantorovich problem is to consider how to move one distribution of mass to another one as efficiently as possible. The efficiency is measured with respect to a cost function $c(x, y)$ specifying the transportation tariff per unit mass. For the quadratic cost $c(x, y) = |x - y|^2$ in the Euclidean space $\mathbb{R}^d$, the optimal transportation map was obtained by Brenier [3,4] (see also Rachev, Rüschendorf [18] and Knott, Smith [13]). On a compact Riemannian manifold, with respect to the square of the Riemannian distance d_M^2, such an optimal transportation was constructed by McCann [17]. For $c(x, y) = |x - y|$, see [6,10]. We shall not mention here all

* Corresponding author.
E-mail addresses: fang@u-bourgogne.fr (S. Fang), jinghai.shao@yahoo.com.cn (J. Shao).

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2006.11.016

contributions to this intensively studied topic, but refer to Villani's book [22] and Ambrosio's survey paper [1] for more details.

Recently the Monge–Kantorovich problem meets two direction of development: one explored essentially by K.T. Sturm [20,21], and Lott and Villani [15] to singular manifolds (of finite dimension) having length property such as cones or Alexandrov spaces; another one initiated by Feyel and Üstünel [11] to carry out the programme on infinite-dimensional settings: they resolved completely the case of Wiener space. The main goal of this work is to deal with curved infinite-dimensional spaces, namely loop groups.

The transportation cost inequality was first established by Talagrand [23] on Euclidean space $\mathbb{R}^n$ in 1996. Therefore, sometimes this inequality is also called Talagrand's inequality. This inequality concerns two important quantities: the Wasserstein distance between probability measures and the entropy of the dynamical system.

Firstly, let us recall the definition of the L^2-Wasserstein distance. Let μ, ν be two probability measures on a complete separable metric space (X, d). The L^2-Wasserstein distance between μ and ν is defined as follows:

$$W_2(\mu, \nu) = \inf_{\pi \in \mathcal{C}(\mu,\nu)} \left\{ \int_{X\times X} \frac{1}{2} d(x, y)^2 \,\mathrm{d}\pi(x, y) \right\}^{1/2}$$

where $\mathcal{C}(\mu, \nu)$ denotes the totality of the Borel probability measures on $X \times X$ with μ, ν as marginal laws, equivalently all couplings of μ and ν. In the definition of Wasserstein distance, $\frac{1}{2}d(x, y)^2$ acts as the cost function $c(x, y)$. We shall say that the transportation cost inequality or Talagrand's inequality holds for μ if there exists a constant $C > 0$ such that

$$W_2^2(f\mu, \mu) \leqslant C\,\mathrm{Ent}_\mu(f), \qquad f \geqslant 0,\ \int_X f \,\mathrm{d}\mu = 1,$$

holds where $\mathrm{Ent}_\mu(f) = \int_X f \log f \,\mathrm{d}\mu$. Through this inequality, we can get an upper estimate of the Wasserstein distance between two probability measures. On loop groups, we have established in [8,9,19] that the transportation cost inequality holds for the heat kernel measure with respect to several different distances including the Riemannian distance. In this paper, we guarantee the Wasserstein distance to be finite through the transportation cost inequality.

The main interest of this paper is to deal with the Monge–Kantorovich problem on loop groups. Our work will be based partly on the McCann's paper [17], where he treated ingeniously the singularity of the Riemannian distance; partly on the Feyel and Üstünel's paper [11], where the authors were the first to investigate the Monge–Kantorovich problem on an infinite-dimensional setting; the procedure of finite-dimensional approximating will be very useful to us.

Now let us explain more precisely the content of this paper. Let G be a connected compact Lie group and $\mathcal{G}$ its Lie algebra equipped with Ad_G-invariant inner product $\langle\,,\rangle$. Consider the based loop groups $\mathcal{L}_eG$:

$$\mathcal{L}_eG := \big\{\ell : [0, 1] \to G \text{ continuous; } \ell(0) = \ell(1) = e\big\},$$

where e is the unit element of G. Let $H_0(\mathcal{G})$ be the Cameron–Martin space of absolutely continuous curves h from $[0, 1]$ to $\mathcal{G}$ such that $h(0) = h(1) = 0$ and $|h|_{H_0}^2 = \int_0^1 |\dot{h}(\theta)|_{\mathcal{G}}^2 \,\mathrm{d}\theta < +\infty$.

$H_0(\mathcal{G})$, equipped with the Lie bracket: $[h_1, h_2](\theta) := [h_1(\theta), h_2(\theta)]$, plays the role of the Lie algebra of $\mathcal{L}_eG$. Consider a continuous curve $\gamma : [0, 1] \to \mathcal{L}_eG$; it is said to be admissible if there exists $z_t = \int_0^t z'_s \,\mathrm{d}s \in H_0(\mathcal{G})$ with $\int_0^1 |z'_s|^2_{H_0} \,\mathrm{d}s < +\infty$ such that for $\theta \in [0, 1]$,

$$d_t\gamma(t, \theta) = \gamma(t, \theta)z'_t(\theta)\,\mathrm{d}t, \qquad \gamma(0, \theta) = e. \tag{1.1}$$

For such a curve, we define

$$L(\gamma) = \left\{ \int_0^1 \left|z'_t\right|^2_{H_0} \mathrm{d}t \right\}^{1/2}.$$

Then the *Riemannian distance* d_L on $\mathcal{L}_eG$ is defined as

$$d_L(\ell_1, \ell_2) = \inf\bigl\{L(\gamma);\ \gamma \text{ admissible connecting } \mathbf{e} \text{ and } \ell_1^{-1}\ell_2\bigr\}, \tag{1.2}$$

where $\mathbf{e}$ denotes the identity loop. It is clear that d_L is left invariant: $d_L(\ell\ell_1, \ell\ell_2) = d_L(\ell_1, \ell_2)$, and it was proved in [9] that $(\ell_1, \ell_2) \mapsto d_L(\ell_1, \ell_2)$ is lower semi-continuous on $\mathcal{L}_eG \times \mathcal{L}_eG$.

Let

$$W_0(\mathcal{G}) = \bigl\{w : [0, 1] \to \mathcal{G} \text{ continuous};\ w(0) = w(1) = 0\bigr\}.$$

Then $(W_0(\mathcal{G}), H_0(\mathcal{G}))$ together with the Brownian bridge measure μ_0 on $W_0(\mathcal{G})$ is an abstract Wiener space. Let $x(t, \cdot)$ be a Brownian motion taking values on $W_0(\mathcal{G})$, with the covariance operator $\langle\, , \rangle_{H_0}$. For each $\theta \in [0, 1]$, we consider the s.d.e.

$$d_t g_x(t, \theta) = g_x(t, \theta) \circ d_t x(t, \theta), \qquad g_x(0, \theta) = e, \tag{1.3}$$

where d_t denotes the Stratonovich stochastic differential relative to the time t. It was proved in [16] that $(t, \theta) \mapsto g_x(t, \theta)$ admits a continuous version, that we denote by the same notation. Then we get a continuous stochastic process $t \mapsto g_x(t, \cdot)$ on $\mathcal{L}_eG$. Let ν denote the law of $x \mapsto g_x(1, \cdot)$ on $\mathcal{L}_eG$, which is called the heat kernel measure on $\mathcal{L}_eG$. Let $F : \mathcal{L}_eG \to \mathbb{R}_+$ be a nonnegative Borel function such that $\int_{\mathcal{L}_eG} F \,\mathrm{d}\nu = 1$. We introduce the Wasserstein distance $W_2(\nu, F\nu)$ between ν and $F\nu$:

$$W_2^2(\nu, F\nu) = \inf_{\pi \in \mathcal{C}(\nu, F\nu)} \left(\int_{\mathcal{L}_eG \times \mathcal{L}_eG} \frac{1}{2} d_L(\ell_1, \ell_2)^2 \pi(\mathrm{d}\ell_1, \mathrm{d}\ell_2) \right), \tag{1.4}$$

where $\mathcal{C}(\nu, F\nu)$ denotes the totality of probability measures on $\mathcal{L}_eG \times \mathcal{L}_eG$ having ν and $F\nu$ as marginal laws.

Theorem 1.1 *(Monge problem). Suppose that* $\mathrm{Ent}_\nu(F) < +\infty$. *Then there is a unique measurable map* $\mathcal{T} : \mathcal{L}_eG \to \mathcal{L}_eG$ *such that*

(i) $\mathcal{T}_\#\nu = F\nu$,
(ii) $W_2^2(\nu, F\nu) = \int_{\mathcal{L}_eG} \frac{1}{2} d_L(\ell, \mathcal{T}(\ell))^2\, \nu(\mathrm{d}\ell)$.

The paper is organized as follows. In Section 2, we present a new approach to McCann's optimal transportation map in the case of Lie groups; our method has the advantage to be generalized to infinite-dimensional settings. The proof of Theorem 1.1 will be done in Sections 3 and 4: we resolve first Kantorovich dual problem in Section 3 following [11], our objective will be then reached in Section 4, by developing McCann's observation on the upper differentiability of the Riemannian distance.

2. McCann's optimal transportation map on compact Lie groups

The exponential map defined by Riemannian geodesics on a Riemannian manifold plays a key role in McCann's construction of the optimal transportation. The resolution for geodesic equations meets a difficulty on infinite-dimensional settings; we refer to [14] for such kind of discussions on the case of Riemannian path spaces; however, the exponential map is not well defined. In this section, we shall develop an alternative approach of MaCann's result in the case of Lie groups.

Throughout this section, G will be a compact Lie group, and its Lie algebra $\mathcal{G}$ is endowed with an inner product $\langle\, ,\rangle$, which is *not* assumed to be Ad-invariant. The associated Riemannian distance d is defined as

$$d(x,y)=\inf\left\{L(\gamma):=\left(\int_0^1\left|\gamma(t)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)\right|_{\mathcal{G}}^2\mathrm{d}t\right)^{1/2}\right\},$$

where the infimum is taken over the set of all absolutely continuous curve connecting x and y.

Let $\psi : G \to \mathbb{R}\cup\{\pm\infty\}$be a measurable function. Consider

$$\psi^c(x):=\inf_{y\in G}\left\{\frac{1}{2}d^2(x,y)-\psi(y)\right\}. \tag{2.1}$$

If ψ is not bounded above, then ψ^c is identically equal to $-\infty$; $\psi^c=+\infty$ happens only if $\psi=-\infty$ everywhere; otherwise, ψ^c is a Lipschitz function; therefore it is differentiable out of a null set of G.

Given two probability measures μ and ν on G, then the Wasserstein distance $W_2(\mu,\nu)$ is defined as

$$W_2^2(\mu,\nu)=\inf\left\{\int_{G\times G}\frac{1}{2}d(x,y)^2\pi(\mathrm{d}x,\mathrm{d}y);\ \pi\in\mathcal{C}(\mu,\nu)\right\},$$

where $\mathcal{C}(\mu,\nu)$ denotes the set of couplings of μ and ν. According to the Kantorovich dual representation theorem (see [22, Chapter 1]), it holds

$$W_2^2(\mu,\nu)=\sup_{(\phi,\psi)\in\Phi_c}J(\phi,\psi), \tag{2.2}$$

where $J(\phi,\psi):=\int_G\phi\,\mathrm{d}\mu+\int_G\psi\,\mathrm{d}\nu$ and

$$\Phi_c:=\left\{\phi,\psi:G\to\mathbb{R}\text{ continuous};\ \phi(x)+\psi(y)\leqslant\frac{1}{2}d(x,y)^2\right\}. \tag{2.3}$$

Then it is known (see for example [17]) that there exists a couple of functions (ϕ, ψ) in Φ_c verifying $\psi = (\psi^c)^c$, $\phi = \psi^c$ such that

$$W_2^2(\mu, \nu) = J(\phi, \psi). \tag{2.4}$$

In particular, ϕ and ψ are Lipschitz continuous functions on G.

Proposition 2.1. *Let $(\phi, \psi) \in \Phi_c$ and $\phi = \psi^c$. If for some $x, y \in G$, ϕ is differentiable at x and*

$$\phi(x) + \psi(y) = \frac{1}{2} d(x, y)^2, \tag{2.5}$$

then y is uniquely determined by the gradient $\nabla\phi(x)$. More precisely, there exists a unique geodesic γ such that $\gamma(0) = e$, $\gamma(1) = x^{-1}y$ and

$$\int_0^1 \mathrm{Ad}^*_{\gamma_t^{-1}}\left(\gamma_t^{-1} \frac{\mathrm{d}}{\mathrm{d}t}\gamma_t\right) \mathrm{d}t = -\nabla\phi(x), \tag{2.6}$$

where Ad^*_g denotes the adjoint operator of Ad_g.

Proof. Let γ be minimizing geodesic such that $\gamma(0) = e$, $\gamma(1) = x^{-1}y$ and

$$L(\gamma) = \left(\int_0^1 \left|\gamma_t^{-1} \frac{\mathrm{d}}{\mathrm{d}t}\gamma_t\right|_{\mathcal{G}}^2 \mathrm{d}t\right)^{1/2} = d\left(e, x^{-1}y\right).$$

Set $z'_t = \gamma_t^{-1} \frac{\mathrm{d}}{\mathrm{d}t}\gamma_t$. Then

$$\mathrm{d}\gamma(t) = \gamma(t) z'_t \,\mathrm{d}t, \qquad \gamma(0) = e.$$

For any $a \in \mathcal{G}$, $\varepsilon > 0$, since $(\phi, \psi) \in \Phi_c$, we have

$$\phi\left(x e^{\varepsilon a}\right) + \psi(y) \leqslant \frac{1}{2} d\left(x e^{\varepsilon a}, y\right)^2.$$

Combining this with (2.5), we have

$$\phi\left(x e^{\varepsilon a}\right) - \phi(x) \leqslant \frac{1}{2} d\left(x e^{\varepsilon a}, y\right)^2 - \frac{1}{2} d(x, y)^2. \tag{2.7}$$

Let $\gamma_\varepsilon(t) = e^{-\varepsilon t a}\gamma(t)$, then $\gamma_\varepsilon(0) = e$, $\gamma_\varepsilon(1) = e^{-\varepsilon a} x^{-1} y$. We have

$$d_t \gamma_\varepsilon(t) = \gamma_\varepsilon(t)\left(z'_t - \varepsilon\, \mathrm{Ad}_{\gamma_t^{-1} e^{\varepsilon t a}}\, a\right) \mathrm{d}t.$$

It follows that

$$d\big(xe^{\varepsilon a},y\big)^2 \leqslant \int_0^1 \big|z_t' - \varepsilon\,\mathrm{Ad}_{\gamma_t^{-1}e^{\varepsilon ta}}\,a\big|_{\mathcal{G}}^2\,\mathrm{d}t$$

$$= \int_0^1 \big|z_t'\big|_{\mathcal{G}}^2\,\mathrm{d}t - 2\varepsilon\int_0^1 \big\langle z_t', \mathrm{Ad}_{\gamma_t^{-1}e^{\varepsilon ta}}\,a\big\rangle_{\mathcal{G}}\,\mathrm{d}t + \varepsilon^2\int_0^1 |\mathrm{Ad}_{\gamma_t^{-1}e^{\varepsilon ta}}\,a|_{\mathcal{G}}^2\,\mathrm{d}t$$

or

$$\frac{1}{2}d\big(xe^{\varepsilon a},y\big)^2 - \frac{1}{2}d(x,y)^2 \leqslant -\varepsilon\int_0^1 \big\langle z_t', \mathrm{Ad}_{\gamma_t^{-1}e^{\varepsilon ta}}\,a\big\rangle_{\mathcal{G}}\,\mathrm{d}t + \frac{1}{2}\varepsilon^2\int_0^1 |\mathrm{Ad}_{\gamma_t^{-1}e^{\varepsilon ta}}\,a|_{\mathcal{G}}^2\,\mathrm{d}t.$$

According to (2.7), we have, for each $\varepsilon > 0$,

$$\frac{\phi(xe^{\varepsilon a}) - \phi(x)}{\varepsilon} \leqslant -\int_0^1 \big\langle z_t', \mathrm{Ad}_{\gamma_t^{-1}e^{\varepsilon ta}}\,a\big\rangle_{\mathcal{G}}\,\mathrm{d}t + \frac{1}{2}\varepsilon\int_0^1 |\mathrm{Ad}_{\gamma_t^{-1}e^{\varepsilon ta}}\,a|_{\mathcal{G}}^2\,\mathrm{d}t.$$

Letting $\varepsilon \downarrow 0$ gives that

$$(D_a\phi)(x) \leqslant -\int_0^1 \big\langle z_t', \mathrm{Ad}_{\gamma_t^{-1}}\,a\big\rangle_{\mathcal{G}}\,\mathrm{d}t = \left\langle -\int_0^1 \mathrm{Ad}^*_{\gamma_t^{-1}}\,z_t'\,\mathrm{d}t, a\right\rangle_{\mathcal{G}}.$$

Let $\nabla\phi(x) \in \mathcal{G}$ be defined such that $(D_a\phi)(x) = \langle\nabla\phi(x), a\rangle_{\mathcal{G}}$. Since $a \in \mathcal{G}$ is arbitrary, we get the expression (2.6).

Set

$$V_t = \int_0^t \mathrm{Ad}^*_{\gamma_s^{-1}}\,z_s'\,\mathrm{d}s. \tag{2.8}$$

We shall prove that V_1 determines uniquely the minimizing geodesic $\{\gamma_t,\ 0 \leqslant t \leqslant 1\}$. Let $a \in \mathcal{G}$, $\varepsilon \in \mathbb{R}$ and $c \in C^2([0,1],\mathbb{R})$ such that $c(0) = c(1) = 0$. Consider $\gamma_\varepsilon(t) = e^{\varepsilon c(t)a}\gamma(t)$. Then γ_ε joins e and $x^{-1}y$. We have

$$\mathrm{d}\gamma_\varepsilon(t) = \gamma_\varepsilon(t)\big(z_t' + \varepsilon c'(t)\,\mathrm{Ad}_{\gamma_t^{-1}e^{-\varepsilon c(t)a}}\,a\big)\,\mathrm{d}t.$$

Then

$$L(\gamma_\varepsilon)^2 = \int_0^1 \big|z_t' + \varepsilon c'(t)\,\mathrm{Ad}_{\gamma_t^{-1}e^{-\varepsilon c(t)a}}\,a\big|_{\mathcal{G}}^2\,\mathrm{d}t.$$

Since $\varepsilon \mapsto L(\gamma_\varepsilon)^2$ realizes a minimum at $\varepsilon = 0$, we have

$$0 = \left\{\frac{\mathrm{d}}{\mathrm{d}\varepsilon} L(\gamma_\varepsilon)^2\right\}_{\varepsilon=0} = 2\int_0^1 \left\langle z_t', c'(t)\,\mathrm{Ad}_{\gamma^{-1}(t)}\, a\right\rangle_{\mathcal{G}} \mathrm{d}t,$$

or

$$\int_0^1 \left\langle \mathrm{Ad}^*_{\gamma^{-1}(t)}\, z_t', c'(t)a\right\rangle_{\mathcal{G}} \mathrm{d}t = 0.$$

Using the definition (2.8), the integration by parts yields to

$$\left\langle V_1, c'(1)a\right\rangle_{\mathcal{G}} = \int_0^1 \left\langle V_t, c''(t)a\right\rangle_{\mathcal{G}} \mathrm{d}t. \tag{2.9}$$

Now suppose that $(\tilde{\gamma}_t,\ 0 \leqslant t \leqslant 1)$ is another minimizing geodesic connecting e and $x^{-1}\tilde{y}$ such that $\tilde{V}_1 = V_1$. Then by (2.9), for each $c \in C^2([0,1],\mathbb{R})$ such that $c(0) = c(1) = 0$, we have

$$\int_0^1 \left\langle V_t, c''(t)a\right\rangle_{\mathcal{G}} \mathrm{d}t = \int_0^1 \left\langle \tilde{V}_t, c''(t)a\right\rangle_{\mathcal{G}} \mathrm{d}t.$$

Since the set of such function $t \mapsto c''(t)a$ is dense in $L^2([0,1],\mathcal{G})$, we deduce that $V_t = \tilde{V}_t$ almost everywhere. By continuity, $V_t = \tilde{V}_t$ for each $t \in [0,1]$. Namely,

$$\int_0^t \mathrm{Ad}^*_{\gamma_s^{-1}}\, z_s'\,\mathrm{d}s = \int_0^t \mathrm{Ad}^*_{\tilde{\gamma}_s^{-1}}\, \tilde{z}_s'\,\mathrm{d}s.$$

It follows that

$$K_t := \mathrm{Ad}^*_{\gamma_t^{-1}}\, z_t' = \mathrm{Ad}^*_{\tilde{\gamma}_t^{-1}}\, \tilde{z}_t'.$$

Note that γ and $\tilde{\gamma}$ both satisfy the differential equation

$$d_t\gamma(t) = \gamma(t) z_t'\,\mathrm{d}t = \gamma(t)\,\mathrm{Ad}^*_{\gamma_t} K_t \mathrm{d}t, \qquad \gamma(0) = e. \tag{2.10}$$

By the uniqueness of solution, $\tilde{\gamma}$ coincides with γ. In particular, $\tilde{y} = y$. $\square$

Theorem 2.2. *Let μ, ν be two probability measures on G. Suppose that μ is absolutely continuous with respect to the Haar measure on G. Then there exists a unique measurable mapping $\mathcal{T}: G \to G$ which pushes μ forward to ν and $(\mathrm{Id}\times\mathcal{T})_{\#}\mu$ is a unique coupling measure in $\mathcal{C}(\mu,\nu)$ which attains the Wasserstein distance $W_2(\mu,\nu)$.*

Proof. Let $(\phi, \psi) \in \Phi_c$ verifying $\psi = (\psi^c)^c$, $\phi = \psi^c$ such that (2.4) holds. By the Rademacher's theorem, ϕ is differentiable almost surely with respect to the Haar measure, so does also with respect to the measure μ. For a point $x \in G$ where ϕ is differentiable, the compactness of G yields the existence of $y \in G$ such that $\phi(x) = \frac{1}{2}d(x, y)^2 - \psi(y)$. By Proposition 2.1, such y is uniquely determined by $(\nabla\phi)(x)$. We shall denote it by

$$y = \mathcal{T}(x), \tag{2.11}$$

and prove that $\mathcal{T} : G \to G$ is a measurable mapping. Let $\{\beta_n,\ n \geqslant 1\} \subset C^\infty([0, 1], \mathbb{R})$ be an orthonormal basis of $H(\mathbb{R}) = \{f : [0, 1] \to \mathbb{R};\ f(0) = 0 \text{ and } \int_0^1 |\dot f(s)|^2 < +\infty\}$. Define

$$c_n(t) = \int_0^t \beta_n(s)\,\mathrm{d}s - t\int_0^1 \beta_n(s)\,\mathrm{d}s.$$

Let $\{e_1, \ldots, e_d\}$ be an orthonormal basis of $\mathcal{G}$. Then $\{\beta_n e_i,\ n \geqslant 1,\ i = 1, \ldots, d\}$ is an orthonormal basis of $H(\mathcal{G})$. Let $U_t = \int_0^t V_s\,\mathrm{d}s$. Then (2.9) can be rewritten as

$$\langle V_1, c_n'(1)e_i\rangle_{\mathcal{G}} = \int_0^1 \langle U_t', \beta_n'(t)e_i\rangle_{\mathcal{G}}\,\mathrm{d}t = \langle U, d_n e_i\rangle_{H(\mathcal{G})}.$$

It follows that

$$U = \sum_{n\geqslant 1}\sum_{i=1}^d \langle V_1, c_n'(1)e_i\rangle_{\mathcal{G}}\beta_n e_i,$$

converges in $H(\mathcal{G})$. A fortiori,

$$U_t = \sum_{n\geqslant 1}\sum_{i=1}^d \langle V_1, c_n'(1)e_i\rangle_{\mathcal{G}} \cdot \beta_n(t)e_i \tag{2.12}$$

converges in $\mathcal{G}$. Now replacing V_1 by $-\nabla\phi(x)$, we see that for each t, U_t is a measurable function of x so does V_t. The differential equation (2.10) can be rewritten in the form

$$\mathrm{d}\gamma(t) = \gamma(t)\,\mathrm{Ad}^*_{\gamma_t}\,\mathrm{d}V_t, \qquad \gamma(0) = e.$$

Therefore for each t, γ_t is a measurable mapping of x. In this way, we get the measurability of the mapping $\mathcal{T}(x) = x\gamma(1)$.

Take any $\Upsilon \in \mathcal{C}(\mu, \nu)$ which attains $W_2(\mu, \nu)$. Then

$$\int_{G\times G} \left(\frac{1}{2}d(x, y)^2 - \phi(x) - \psi(y)\right)\mathrm{d}\Upsilon(x, y) = 0.$$

Note that $\frac{1}{2}d(x,y)^2 - \phi(x) - \psi(y) \geqslant 0$, then Υ-a.s.

$$\frac{1}{2}d(x,y)^2 - \phi(x) - \psi(y) = 0.$$

By construction of $\mathcal{T}$, it holds that Υ-a.s.

$$y = \mathcal{T}(x).$$

Furthermore, this implies also the uniqueness of the mapping $\mathcal{T}$. Now for any bounded Borel function f on $G \times G$,

$$\int_{G\times G} f(x,y)\,\mathrm{d}\Upsilon(x,y) = \int_{G\times G} f\big(x,\mathcal{T}(x)\big)\,\mathrm{d}\Upsilon(x,y) = \int_G f\big(x,\mathcal{T}(x)\big)\,\mathrm{d}\mu(x).$$

Consequently, $\Upsilon = (\mathrm{Id}\times\mathcal{T})_{\#}\mu$. In particular, $\mathcal{T}_{\#}\mu = \nu$. □

Theorem 2.3. (McCann [17].) *The map $\mathcal{T}$ has the following explicit expression using the geodesic exponential map*:

$$\mathcal{T}(x) = \exp_x\big(-\nabla\phi(x)\big). \tag{2.13}$$

Proof. Let $x, y \in G$ and γ the minimizing geodesic connecting e and $x^{-1}y$. Consider $c_k(t) = \sin(k\pi t)$. Then (2.9) holds, or $\int_0^1 \langle \mathrm{Ad}^*_{\gamma_t^{-1}} z'_t, c'_k(t)a\rangle_{\mathcal{G}}\,\mathrm{d}t = 0$ where $a \in \mathcal{G}$. Therefore for any integer $k \geqslant 1$,

$$\int_0^1 \mathrm{Ad}^*_{\gamma_t^{-1}} z'_t \cos(k\pi t)\,\mathrm{d}t = 0,$$

from which we get that $t \to \mathrm{Ad}^*_{\gamma_t^{-1}} z'_t$ is a constant function over [0, 1]. In particular,

$$\gamma'(0) = z'(0) = \int_0^1 \mathrm{Ad}^*_{\gamma_s^{-1}} z'_s\,\mathrm{d}s.$$

Now combining (2.6) and the definition of geodesic exponential map, we get (2.13). □

3. Kantorovich dual problem

Let ν be the heat measure at the time 1 on $\mathcal{L}_eG$ defined by (1.3). The objective of this section is to resolve the Kantorovich dual problem for two probability measures ν and $F\nu$ on the loop group $\mathcal{L}_eG$. More precisely, let G be a connected compact Lie group, $\mathcal{G}$ its Lie algebra endowed with an Ad_G-invariant metric $\langle\,,\rangle_{\mathcal{G}}$. Recall that

$$\mathcal{L}_eG = \big\{\ell : [0,1] \to G \text{ continuous};\ \ell(0) = \ell(1) = e\big\}$$

and

$$H_0(\mathcal{G}) = \left\{ h:[0,1] \to \mathcal{G};\ h(0) = h(1) = 0,\ |h|_H^2 = \int_0^1 \left|\dot{h}(\theta)\right|_{\mathcal{G}}^2 \,\mathrm{d}\theta < +\infty \right\}.$$

For a cylindrical function $F : \mathcal{L}_e G \to \mathbb{R}$ in the form

$$F(\ell) = f\big(\ell(\theta_1), \ldots, \ell(\theta_n)\big), \quad f \in C^\infty\big(G^n\big),$$

and $h \in H_0(\mathcal{G})$, we define

$$(D_h F)(\ell) = \left\{ \frac{\mathrm{d}}{\mathrm{d}\varepsilon} F\big(\ell e^{\varepsilon h}\big) \right\}_{\varepsilon=0} = \sum_{i=1}^n \big\langle \partial_i f, \ell(\theta_i) h(\theta_i) \big\rangle_{T_{\ell(\theta_i)}G}, \tag{3.1}$$

where $\partial_i f$ denotes the ith partial derivative. The gradient operator $\nabla^{\mathcal{L}}$ on $\mathcal{L}_e G$ is defined as

$$\big(\nabla^{\mathcal{L}} F\big)(\ell) = \sum_{i=1}^n \ell^{-1}(\theta_i)(\partial_i f) G(\theta_i, \cdot), \tag{3.2}$$

where $G(\theta_i, \theta) := \theta_i \wedge \theta - \theta_i \theta$. Consider

$$\mathcal{E}(F, F) := \int_{\mathcal{L}_e G} \big|\nabla^{\mathcal{L}} F\big|_{H_0}^2 \,\mathrm{d}\nu.$$

By Driver [5], the integration by parts formula holds for the heat kernel measure ν, so that $\mathcal{E}(F, F)$ is closable. Let $D_1^2(\nu)$ be the domain of the associated Dirichlet form.

Now let $\mathcal{P} = \{0 < \theta_1 < \cdots < \theta_N < 1\}$ be a finite partition of $[0, 1]$. For any $h \in H_0(\mathcal{G})$, we define

$$\Pi_{\mathcal{P}} h = \sum_{i,j=1}^N G(\theta_i, \cdot) Q_{ij}^{\mathcal{P}} h(\theta_j), \tag{3.3}$$

where $(Q_{ij}^{\mathcal{P}})$ is the inverse matrix of $(G(\theta_i, \theta_j))_{ij}$. Note that $(\Pi_{\mathcal{P}} h)(\theta_i) = h(\theta_i)$ for $1 \leqslant i \leqslant N$. Set

$$H_{\mathcal{P}}(\mathcal{G}) = \big\{ \Pi_{\mathcal{P}} h;\ h \in H_0(\mathcal{G}) \big\}.$$

Then it is easy to see that $\langle \Pi_{\mathcal{P}} h, h \rangle_{H_0} = \langle \Pi_{\mathcal{P}} h, \Pi_{\mathcal{P}} h \rangle_{H_0}$, so that $\Pi_{\mathcal{P}}$ is the orthogonal projection from $H_0(\mathcal{G})$ onto $H_{\mathcal{P}}(\mathcal{G})$. Define $\Lambda_{\mathcal{P}} : H_0(\mathcal{G}) \to \mathcal{G}^{\mathcal{P}}$ by

$$\Lambda_{\mathcal{P}}(h) = \big(h(\theta_1), \ldots, h(\theta_N)\big). \tag{3.4}$$

Then by Driver [5], $\Lambda_{\mathcal{P}}$ is an isometric isomorphism from $H_{\mathcal{P}}(\mathcal{G})$ onto $\mathcal{G}^{\mathcal{P}}$, if $\mathcal{G}^{\mathcal{P}}$ is equipped with the metric

$$\langle a, b\rangle_{\mathcal{P}} = \sum_{i,j=1}^{N} Q_{ij}^{\mathcal{P}} \langle a_i, b_j\rangle_{\mathcal{G}}, \quad a = (a_1, \ldots, a_N),\ b = (b_1, \ldots, b_N) \in \mathcal{G}^N. \tag{3.5}$$

In the remainder of the paper, we use also the notation $\Lambda_{\mathcal{P}}$ to denote the projection from $\mathcal{L}_e G$ to $G^{\mathcal{P}}$:

$$\Lambda_{\mathcal{P}}(\ell) = \big(\ell(\theta_1), \ldots, \ell(\theta_N)\big).$$

Let $g_x(t,\cdot)$ be the continuous process on $\mathcal{L}_e G$ defined by the s.d.e. (1.3), we denote by $g_x^{\mathcal{P}}(t) = \Lambda_{\mathcal{P}}(g_x(t,\cdot)) = (g_x(t,\theta_1), \ldots, g_x(t,\theta_N))$ and $x_t^{\mathcal{P}} = \Lambda_{\mathcal{P}}(x(t,\cdot)) = (x(t,\theta_1), \ldots, x(t,\theta_N))$. Then $t \mapsto x_t^{\mathcal{P}}$ is a standard Brownian motion taking values in $(\mathcal{G}^{\mathcal{P}}, \langle\, ,\rangle_{\mathcal{P}})$ and

$$d g_x^{\mathcal{P}}(t) = g_x^{\mathcal{P}} \circ d x_t^{\mathcal{P}}, \qquad g_x^{\mathcal{P}}(0) = (e, \ldots, e). \tag{3.6}$$

Let $\nu_{\mathcal{P}} = (\Lambda_{\mathcal{P}})_{\#}\nu$. Then $\nu_{\mathcal{P}}$ is the law of $x \mapsto g_x^{\mathcal{P}}(1)$ on $G^{\mathcal{P}}$, which has a strictly positive density with respect to the Haar measure on $G^{\mathcal{P}}$.

Proposition 3.1. *Let $(\mathcal{P}_n)_{n\geqslant 1}$ be a sequence of partition of* $[0,1]$ *such that $\mathcal{P}_n \subset \mathcal{P}_{n+1}$ and $\bigcup_n \mathcal{P}_n$ is dense in* $[0,1]$. *Then for any $\ell, \ell' \in \mathcal{L}_e G$,*

$$d_n\big(\Lambda_n \ell, \Lambda_n \ell'\big) \uparrow d_L(\ell, \ell') \quad \text{as } n \to +\infty, \tag{3.7}$$

where $\Lambda_n = \Lambda_{\mathcal{P}_n}$ and d_n is the left invariant distance associated to $\langle\, ,\rangle_{\mathcal{P}_n}$ defined in (3.5).

Proof. We see first that if $\mathcal{P}_1 \subset \mathcal{P}_2$, then

$$d_{\mathcal{P}_1}\big(\Lambda_{\mathcal{P}_1}(\ell), \Lambda_{\mathcal{P}_1}(\ell')\big) \leqslant d_{\mathcal{P}_2}\big(\Lambda_{\mathcal{P}_2}(\ell), \Lambda_{\mathcal{P}_2}(\ell')\big). \tag{i}$$

In fact, let $\gamma : [0,1] \to G^{\mathcal{P}_2}$ be a minimizing geodesic such that $\gamma(0) = \Lambda_{\mathcal{P}_2}(\ell)$, $\gamma(1) = \Lambda_{\mathcal{P}_2}(\ell')$. Set $z_s' = \gamma^{-1}(s)\gamma'(s) \in \mathcal{G}^{\mathcal{P}_2}$. We have

$$d_{\mathcal{P}_2}^2\big(\Lambda_{\mathcal{P}_2}(\ell), \Lambda_{\mathcal{P}_2}(\ell')\big) = \int_0^1 \big|z_s'\big|_{\mathcal{P}_2}^2 \,\mathrm{d}s.$$

Let $\Lambda_{\mathcal{P}}^{-1} : \mathcal{G}^{\mathcal{P}} \to H_{\mathcal{P}}(\mathcal{G})$ be the inverse map of $\Lambda_{\mathcal{P}}$. Then

$$d_{\mathcal{P}_2}^2\big(\Lambda_{\mathcal{P}_2}(\ell), \Lambda_{\mathcal{P}_2}(\ell')\big) = \int_0^1 \big|\Lambda_{\mathcal{P}_2}^{-1}\big(z_s'\big)\big|_{H_0}^2 \,\mathrm{d}s.$$

Let $\tilde\gamma(s) = (\gamma(s,\theta))_{\theta\in\mathcal{P}_1}$ and $\tilde z_s = (z(s,\theta))_{\theta\in\mathcal{P}_1}$. We have $\tilde\gamma(0) = \Lambda_{\mathcal{P}_1}(\ell)$, $\tilde\gamma(1) = \Lambda_{\mathcal{P}_1}(\ell')$ and

$$d_{\mathcal{P}_1}^2\big(\Lambda_{\mathcal{P}_1}(\ell), \Lambda_{\mathcal{P}_1}(\ell')\big) \leqslant \int_0^1 \big|\tilde{z}_s'\big|_{\mathcal{P}_1}^2 \, \mathrm{d}s = \int_0^1 \big|\Pi_{\mathcal{P}_1}\big(\Lambda_{\mathcal{P}_2}^{-1}(z_s')\big)\big|_{H_0}^2 \, \mathrm{d}s$$
$$\leqslant \int_0^1 \big|\Lambda_{\mathcal{P}_2}^{-1}(z_s')\big|_{H_0}^2 \, \mathrm{d}s = d_{\mathcal{P}_2}^2\big(\Lambda_{\mathcal{P}_2}(\ell), \Lambda_{\mathcal{P}_2}(\ell')\big).$$

We get (i).

In the same way, we can prove that

$$d_{\mathcal{P}}\big(\Lambda_{\mathcal{P}}(\ell), \Lambda_{\mathcal{P}}(\ell')\big) \leqslant d_L(\ell, \ell'). \tag{ii}$$

Now we shall prove that

$$\sup_n d_n\big(\Lambda_n(\ell), \Lambda_n(\ell')\big) \geqslant d_L(\ell, \ell').$$

Suppose that $M_1 = \sup_n d_n(\Lambda_n(\ell), \Lambda_n(\ell')) < +\infty$. Then there exist geodesics γ_n in $C([0,1], G^{\mathcal{P}_n})$ such that $\gamma_n(0) = e \in G^{\mathcal{P}_n}$, $\gamma(1) = \Lambda_n(\ell^{-1}\ell')$, and $L(\gamma_n) = d_n(\Lambda_n(\ell), \Lambda_n(\ell'))$. Set

$$z_n'(s) = \gamma_n^{-1}(s)\gamma_n'(s) \in \mathcal{G}^{\mathcal{P}_n}$$

and

$$\tilde{z}_n'(s) = \Lambda_n^{-1}\big(z_n'(s)\big) \in H_{\mathcal{P}_n}(\mathcal{G}).$$

Then

$$\int_0^1 \big|\tilde{z}_n'(s)\big|_{H_0}^2 \, \mathrm{d}s = \int_0^1 \big|z_n'(s)\big|_{\mathcal{P}_n}^2 = d_n\big(\Lambda_n(\ell), \Lambda_n(\ell')\big) \leqslant M_1^2.$$

Therefore, $\{\tilde{z}_n;\ n \geqslant 1\}$ is contained in a bounded set of

$$H\big(H_0(\mathcal{G})\big) := \left\{ z_t = \int_0^t z_s' \, \mathrm{d}s;\ \int_0^1 \big|z_s'\big|_{H_0}^2 \, \mathrm{d}s < +\infty \right\}.$$

Up to a subsequence, $(\tilde{z}_n)_{n\geqslant 1}$ converges weakly to some $z \in H(H_0(\mathcal{G}))$ such that $\int_0^1 |z'(s)|_{H_0}^2 \, \mathrm{d}s \leqslant M_1^2$. Consider the following differential equation on $\mathcal{L}_e G$:

$$\mathrm{d}\tilde{\gamma}(t,\theta) = \tilde{\gamma}(t,\theta) z_t'(\theta) \, \mathrm{d}t, \qquad \tilde{\gamma}(0,\theta) = e;$$
$$\mathrm{d}\tilde{\gamma}_n(t,\theta) = \tilde{\gamma}_n(t,\theta) z_n'(t,\theta) \, \mathrm{d}t, \qquad \tilde{\gamma}_n(0,\theta) = e.$$

Then $\tilde{\gamma}_n$ converges to $\tilde{\gamma}$ uniformly on $[0,1]\times[0,1]$. In particular, $\tilde{\gamma}_n(1,\cdot)$ converges to $\tilde{\gamma}(1,\cdot)$. On the other hand, $\Lambda_n(\tilde{\gamma}_n(1,\cdot)) = \gamma_n(1) = \Lambda_n(\ell^{-1}\ell')$. This means that for each $\theta \in \mathcal{P}_{n_0}$, and $n \geqslant n_0$, $\tilde{\gamma}_n(1,\theta) = (\ell^{-1}\ell')(\theta)$. As $\bigcup_n \mathcal{P}_n$ is dense in $[0,1]$, we get that $\tilde{\gamma}(1,\cdot) = \ell^{-1}\ell'$, therefore,

$$d_L(\ell,\ell')^2 \leqslant \int_0^1 \left|z_s'\right|_{H_0}^2 \mathrm{d}s \leqslant M_1^2.$$

The proof of (3.7) is completed. □

Now let F be a non-negative Borel function on $\mathcal{L}_eG$ such that $\int_{\mathcal{L}_eG} F\,\mathrm{d}\nu = 1$. Set $\mu = F\nu$ and $\mu_{\mathcal{P}} = (\Lambda_{\mathcal{P}})_\#\mu$. Then $\mu_{\mathcal{P}} = F_{\mathcal{P}}\nu_{\mathcal{P}}$ with $F_{\mathcal{P}} \circ \Lambda_{\mathcal{P}} = \mathbb{E}^{\mathcal{B}_{\mathcal{P}}}(F)$, where $\mathcal{B}_{\mathcal{P}}$ is the sub σ-field on $\mathcal{L}_eG$ generated by $\Lambda_{\mathcal{P}}$.

Theorem 3.2. *We have*

$$\sup_{\mathcal{P}} W_2(\mu_{\mathcal{P}},\nu_{\mathcal{P}}) = W_2(\mu,\nu). \tag{3.8}$$

Proof. Let $\Upsilon \in \mathcal{C}(\mu,\nu)$. Define $\Upsilon_{\mathcal{P}} = (\Lambda_{\mathcal{P}} \times \Lambda_{\mathcal{P}})_\#\Upsilon$. Then $\Upsilon_{\mathcal{P}} \in \mathcal{C}(\mu_{\mathcal{P}},\nu_{\mathcal{P}})$. According to (3.7), we have

$$\begin{aligned} W_2^2(\mu_{\mathcal{P}},\nu_{\mathcal{P}}) &\leqslant \int_{\mathcal{L}_eG\times\mathcal{L}_eG} \frac{1}{2} d_{\mathcal{P}}^2(\Lambda_{\mathcal{P}}\ell_1,\Lambda_{\mathcal{P}}\ell_2)\Upsilon(\mathrm{d}\ell_1,\mathrm{d}\ell_2) \\ &\leqslant \int_{\mathcal{L}_eG\times\mathcal{L}_eG} \frac{1}{2} d_L^2(\ell_1,\ell_2)\Upsilon(\mathrm{d}\ell_1,\mathrm{d}\ell_2). \end{aligned}$$

Taking the infimum over $\Upsilon \in \mathcal{C}(\mu,\nu)$, we get $W_2(\mu_{\mathcal{P}},\nu_{\mathcal{P}}) \leqslant W_2(\mu,\nu)$.

Conversely, take $\varepsilon > 0$ small enough so that any two points $x, y \in G$ such that $d_G(x,y) < \varepsilon$ can be connected by a unique minimal geodesic in G, where d_G is the bi-invariant distance on G. For any partition $\mathcal{P} = \{0 < \theta_1 < \cdots < \theta_N < 1\}$, we set

$$U_{\mathcal{P}} = \left\{(y_1,\dots,y_N) \in G^{\mathcal{P}};\ d_G(y_n,y_{n-1}) < \varepsilon,\ i = 1,\dots,N+1\right\},$$

where $y_0 = y_{N+1} = e$. $U_{\mathcal{P}}$ is an open set of $G^{\mathcal{P}}$. For each $y = (y_1,\dots,y_N)$, we define $I_{\mathcal{P}}(y) \in \mathcal{L}_eG$ such that $I_{\mathcal{P}}(y)(\theta_i) = y_i$ and y_i, y_{i-1} are linked by the unique minimal geodesic. Then $I_{\mathcal{P}} : U_{\mathcal{P}} \to \mathcal{L}_eG$ is well defined and continuous.

Take a sequence of partition $\mathcal{P}_n$ of $[0,1]$ such that $\mathcal{P}_n \subset \mathcal{P}_{n+1}$ and $\bigcup_n \mathcal{P}_n$ is dense in $[0,1]$. Then we have

(i) $\quad \Lambda_{\mathcal{P}_n} \circ I_{\mathcal{P}_n} = \mathrm{Id}$ on $U_{\mathcal{P}_n}$, $\quad I_{\mathcal{P}_n} \circ \Lambda_{\mathcal{P}_n} \to \mathrm{Id}$ as $n \to +\infty$;

(ii) $\quad$ for each $\ell \in \mathcal{L}_eG$, $\quad \lim_{n\to\infty} \mathbf{1}_{U_{\mathcal{P}_n}}(\Lambda_{\mathcal{P}_n}\ell) = 1$.

Let $\Upsilon_n \in \mathcal{C}(\mu_{\mathcal{P}_n}, \nu_{\mathcal{P}_n})$ such that

$$W_2^2(\mu_{\mathcal{P}_n}, \nu_{\mathcal{P}_n}) = \int_{G^{\mathcal{P}_n} \times G^{\mathcal{P}_n}} \frac{1}{2} d_{\mathcal{P}_n}(x, y)^2 \Upsilon_n(\mathrm{d}x, \mathrm{d}y).$$

Define

$$\widetilde{\Upsilon}_n = (I_{\mathcal{P}_n} \times I_{\mathcal{P}_n})_{\#}\left(\frac{\mathbf{1}_{U_{\mathcal{P}_n} \times U_{\mathcal{P}_n}}}{\Upsilon_n(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})}\Upsilon_n\right).$$

According to (ii),

$$\lim_{n\to\infty} \nu_{\mathcal{P}_n}(U_{\mathcal{P}_n}) = \lim_{n\to\infty} \int_{\mathcal{L}_e G} \mathbf{1}_{U_{\mathcal{P}_n}}\left(\Lambda_{\mathcal{P}_n}(\ell)\right) \mathrm{d}\nu(\ell) = 1$$

and

$$\lim_{n\to\infty} \mu_{\mathcal{P}_n}(U_{\mathcal{P}_n}) = \lim_{n\to\infty} \int_{\mathcal{L}_e G} \mathbf{1}_{U_{\mathcal{P}_n}}\left(\Lambda_{\mathcal{P}_n}(\ell)\right) F \, \mathrm{d}\nu(\ell) = 1.$$

We have $(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})^c \subset (U_{\mathcal{P}_n}^c \times G^{\mathcal{P}_n}) \cup (G^{\mathcal{P}_n} \times U_{\mathcal{P}_n}^c)$. Then

$$\Upsilon_n\left((U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})^c\right) \leqslant \mu_{\mathcal{P}_n}\left(U_{\mathcal{P}_n}^c\right) + \nu_{\mathcal{P}_n}\left(U_{\mathcal{P}_n}^c\right) \to 0 \quad \text{as } n \to \infty.$$

Therefore,

$$\lim_{n\to\infty} \Upsilon_n(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n}) = 1.$$

Let $f : \mathcal{L}_e G \to \mathbb{R}$ be a bounded continuous function. Then

$$\begin{aligned}
&\left| \int_{G^{\mathcal{P}_n} \times G^{\mathcal{P}_n}} \frac{\mathbf{1}_{U_{\mathcal{P}_n} \times U_{\mathcal{P}_n}^c}(x, y)}{\Upsilon_n(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})} f\left(I_{\mathcal{P}_n}(x)\right) \mathrm{d}\Upsilon_n(x, y) \right| \\
&\quad \leqslant \frac{\|f\|_\infty}{\Upsilon_n(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})} \nu_{\mathcal{P}_n}\left(U_{\mathcal{P}_n}^c\right) \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{\mathcal{L}_e G \times \mathcal{L}_e G} f(\ell_1) \, \mathrm{d}\widetilde{\Upsilon}_n(\ell_1, \ell_2) &= \int_{G^{\mathcal{P}_n} \times G^{\mathcal{P}_n}} \frac{\mathbf{1}_{U_{\mathcal{P}_n} \times U_{\mathcal{P}_n}}(x, y)}{\Upsilon_n(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})} f\left(I_{\mathcal{P}_n}(x)\right) \mathrm{d}\Upsilon_n(x, y) \\
&= \int_{\mathcal{L}_e G} \frac{\mathbf{1}_{U_{\mathcal{P}_n}}(\Lambda_{\mathcal{P}_n}(\ell_1))}{\Upsilon_n(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})} f\left(I_{\mathcal{P}_n} \Lambda_{\mathcal{P}_n}(\ell_1)\right) \mathrm{d}\mu(\ell_1) \\
&\quad - \int_{G^{\mathcal{P}_n} \times G^{\mathcal{P}_n}} \frac{\mathbf{1}_{U_{\mathcal{P}_n} \times U_{\mathcal{P}_n}^c}(x, y)}{\Upsilon_n(U_{\mathcal{P}_n} \times U_{\mathcal{P}_n})} f\left(I_{\mathcal{P}_n}(x)\right) \mathrm{d}\Upsilon_n(x, y)
\end{aligned}$$

which converges to, as $n \to \infty$,

$$\int_{\mathcal{L}_e G} f(\ell)\,\mathrm{d}\mu(\ell).$$

If we denote by $\tilde{\mu}_n$ and $\tilde{\nu}_n$ the marginal laws of $\widetilde{\Upsilon}_n$, then $\tilde{\mu}_n$ converges weakly to μ, as well as $\tilde{\nu}_n$ converges weakly to ν. It follows that the family $\{\widetilde{\Upsilon}_n,\ n \geqslant 1\}$ is tight. Up to a subsequence, $\widetilde{\Upsilon}_n$ converges weakly to a probability measure Υ_0. Clearly, $\Upsilon_0 \in \mathcal{C}(\mu, \nu)$. Now following [24] and according to (3.7) and to (i), for $m \geqslant n$,

$$\begin{aligned}
&\int_{\mathcal{L}_e G\times\mathcal{L}_e G} \frac{1}{2} d^2_{\mathcal{P}_n}\big(\Lambda_{\mathcal{P}_n}(\ell_1), \Lambda_{\mathcal{P}_n}(\ell_2)\big)\,\mathrm{d}\widetilde{\Upsilon}_m(\mathrm{d}\ell_1, \mathrm{d}\ell_2) \\
&\quad\leqslant \int_{\mathcal{L}_e G\times\mathcal{L}_e G} \frac{1}{2} d^2_{\mathcal{P}_m}\big(\Lambda_{\mathcal{P}_m}(\ell_1), \Lambda_{\mathcal{P}_m}(\ell_2)\big)\widetilde{\Upsilon}_m(\mathrm{d}\ell_1, \mathrm{d}\ell_2) \\
&\quad\leqslant \frac{1}{\Upsilon_m(U_{\mathcal{P}_m} \times U_{\mathcal{P}_m})} \int_{G^{\mathcal{P}_m}\times G^{\mathcal{P}_m}} \frac{1}{2} d^2_{\mathcal{P}_m}(x, y)\Upsilon_m(\mathrm{d}x, \mathrm{d}y) \\
&\quad= \frac{1}{\Upsilon_m(U_{\mathcal{P}_m} \times U_{\mathcal{P}_m})} W_2^2(\mu_{\mathcal{P}_m}, \nu_{\mathcal{P}_m}) \\
&\quad\leqslant \frac{1}{\Upsilon_m(U_{\mathcal{P}_m} \times U_{\mathcal{P}_m})} \sup_{\mathcal{P}} W_2^2(\mu_{\mathcal{P}}, \nu_{\mathcal{P}}).
\end{aligned}$$

Letting $m \to +\infty$ yields

$$\int_{\mathcal{L}_e G\times\mathcal{L}_e G} \frac{1}{2} d^2_{\mathcal{P}_n}\big(\Lambda_{\mathcal{P}_n}(\ell_1), \Lambda_{\mathcal{P}_n}(\ell_2)\big)\Upsilon_0(\mathrm{d}\ell_1, \mathrm{d}\ell_2) \leqslant \sup_{\mathcal{P}} W_2^2(\mu_{\mathcal{P}}, \nu_{\mathcal{P}}).$$

Now letting $n \to +\infty$, by virtue of (3.7), we get

$$W_2^2(\mu, \nu) \leqslant \sup_{\mathcal{P}} W_2^2(\mu_{\mathcal{P}}, \nu_{\mathcal{P}}). \qquad \square$$

Theorem 3.3. *There exists a constant $C > 0$ such that*

$$W_2^2(F\nu, \nu) \leqslant C\,\mathrm{Ent}_\nu(F), \tag{3.9}$$

for any positive Borel function F on $\mathcal{L}_e G$ such that $\int_{\mathcal{L}_e G} F\,\mathrm{d}\nu = 1$.

Proof. Let $\mathcal{P}$ be a finite partition of [0, 1]. Then it is known (cf. [2]) that

$$W_2^2(F_{\mathcal{P}}\nu_{\mathcal{P}}, \nu_{\mathcal{P}}) \leqslant C_{\mathcal{P}}\,\mathrm{Ent}_{\nu_{\mathcal{P}}}(F_{\mathcal{P}}), \tag{3.10}$$

where $C_{\mathcal{P}} = 2(e^{\|\mathrm{Ric}_{\mathcal{P}}\|} - 1)/\|\mathrm{Ric}\|_{\mathcal{P}}$ and $\mathrm{Ric}_{\mathcal{P}}$ is the Ricci tensor on $G^{\mathcal{P}}$. By [7], $\sup_{\mathcal{P}} \|\mathrm{Ric}_{\mathcal{P}}\| < +\infty$, it follows that $C := \sup_{\mathcal{P}} C_{\mathcal{P}} < +\infty$. Now

$$\mathrm{Ent}_{\nu_{\mathcal{P}}}(F_{\mathcal{P}}) = \int_{G^{\mathcal{P}}} F_{\mathcal{P}} \log F_{\mathcal{P}} \,\mathrm{d}\nu_{\mathcal{P}} = \int_{\mathcal{L}_e G} \mathbb{E}^{\mathcal{B}_{\mathcal{P}}}(F) \log \mathbb{E}^{\mathcal{B}_{\mathcal{P}}}(F) \,\mathrm{d}\nu \leqslant \mathrm{Ent}_{\nu}(F).$$

Hence (3.9) follows from (3.8) and (3.10). □

Theorem 3.4 *(Kantorovich Dual problem). Suppose that* $\mathrm{Ent}_\nu(F) < +\infty$*, then there exists a Borel function* $\phi \in \mathcal{D}_1^2(\nu)$ *and* ψ *in* $L^1(\mathcal{L}_e G, F\nu)$ *respectively such that*

(i) $$\phi(\ell) + \psi(\ell') \leqslant \frac{1}{2} d_L^2(\ell, \ell'), \quad \textit{for all } \ell, \ell' \in \mathcal{L}_e G, \quad \textit{and}$$

(ii) $$W_2^2(\nu, F\nu) = \int_{\mathcal{L}_e G} \phi \,\mathrm{d}\nu + \int_{\mathcal{L}_e G} \psi F \,\mathrm{d}\nu.$$

Proof. Take a sequence of finite partition $\mathcal{P}_n$ of $[0, 1]$ as above. On $G^{\mathcal{P}_n}$, according to (2.4), there exist two Lipschitz continuous functions ϕ_n, ψ_n on $G^{\mathcal{P}_n}$ such that

$$\phi_n(x) + \psi_n(y) \leqslant \frac{1}{2} d_{\mathcal{P}_n}(x, y)^2, \quad x, y \in G^{\mathcal{P}_n}, \tag{3.11}$$

$$\int \phi_n \,\mathrm{d}\nu_n + \int \psi_n \,\mathrm{d}\mu_n = W_2^2(\nu_n, \mu_n) = \int \frac{1}{2} |\nabla \phi_n|_{\mathcal{P}_n}^2 \,\mathrm{d}\nu_n. \tag{3.12}$$

Since $G^{\mathcal{P}_n}$ is compact, we have $\lambda_n = \int \phi_n \,\mathrm{d}\nu_n$ is finite. Let C be the constant in (3.9), then

$$\int_{G^{\mathcal{P}_n}} (\phi_n - \lambda_n)^2 \,\mathrm{d}\nu_n \leqslant C \int_{G^{\mathcal{P}_n}} |\nabla \phi_n|_{\mathcal{P}_n}^2 \,\mathrm{d}\nu_n \leqslant 2C^2 \,\mathrm{Ent}_\nu(F). \tag{3.13}$$

Define a new sequence $(\bar{\phi}_n, \bar{\psi}_n) = (\phi_n - \lambda_n, \psi_n + \lambda_n)$ which satisfies again (3.11) and (3.12). Let

$$\tilde{\phi}_n(\ell) = \bar{\phi}_n\big(\Lambda_n(\ell)\big), \qquad \tilde{\psi}_n(\ell) = \bar{\psi}_n\big(\Lambda_n(\ell)\big).$$

From a classical result in the Dirichlet form theory, we see that $\tilde{\phi}_n \in \mathcal{D}_1^2(\nu)$. By (3.1), we have, for $h \in H_0(\mathcal{G})$,

$$\big\langle \nabla \tilde{\phi}_n(\ell), h \big\rangle_{H_0} = (D_h \tilde{\phi}_n)(\ell) = (D_{\Lambda_n(h)} \bar{\phi}_n)\big(\Lambda_n(\ell)\big) = \big\langle (\nabla \bar{\phi}_n)\big(\Lambda_n(\ell)\big), \Lambda_n(h) \big\rangle_{\mathcal{P}_n}.$$

It follows that

$$|\nabla \tilde{\phi}_n|_{H_0}(\ell) = |\nabla \bar{\phi}_n|_{\mathcal{P}_n} \circ \Lambda_n(\ell).$$

Therefore by (3.13), $\{\tilde{\phi}_n\}$ is a bounded sequence in $\mathcal{D}_1^2(\nu)$. Up to a subsequence, $\tilde{\phi}_n$ converges weakly to a function $\hat{\phi} \in \mathcal{D}_1^2(\nu)$ such that $\int_{\mathcal{L}_eG} |\nabla\hat{\phi}|_{H_0}^2 \,\mathrm{d}\nu \leqslant C^2 \operatorname{Ent}_\nu(F)$. By Banach–Saks theorem, again up to a subsequence, the Cesaro mean $(\tilde{\phi}_1 + \cdots + \tilde{\phi}_n)/n$ converges to $\hat{\phi}$ in $\mathcal{D}_1^2(\nu)$. Let $\Upsilon_0 \in \mathcal{C}(F\nu, \nu)$ such that

$$W_2^2(F\nu, \nu) = \int\limits_{\mathcal{L}_eG\times\mathcal{L}_eG} \frac{1}{2} d_L(\ell_1, \ell_2)^2 \Upsilon_0(\ell_1, \ell_2).$$

Set

$$\begin{aligned} F_n(\ell_1, \ell_2) &= \frac{1}{2} d_L(\ell_1, \ell_2)^2 - \tilde{\phi}_n(\ell_1) - \tilde{\psi}_n(\ell_2) \\ &= \frac{1}{2} d_L(\ell_1, \ell_2)^2 - \bar{\phi}_n\big(\Lambda_n(\ell_1)\big) - \bar{\psi}_n\big(\Lambda_n(\ell_2)\big) \end{aligned}$$

which is positive by (3.7) and (3.11). Now using (3.8),

$$\begin{aligned} \int\limits_{\mathcal{L}_eG\times\mathcal{L}_eG} F_n(\ell_1, \ell_2)\Upsilon_0(\mathrm{d}\ell_1, \mathrm{d}\ell_2) &= W_2^2(F\nu, \nu) - \int\limits_{G^{\mathcal{P}_n}} \bar{\phi}_n \,\mathrm{d}\nu_n - \int\limits_{G^{\mathcal{P}_n}} \bar{\psi}_n \,\mathrm{d}\mu_n \\ &= W_2^2(F\nu, \nu) - W_2^2(\mu_n, \nu_n) \to 0, \end{aligned}$$

as $n \to +\infty$. This means that F_n converges to 0 in $L^1(\Upsilon_0)$. A fortiori, $(F_1 + \cdots + F_n)/n$ converges to 0. Therefore $(\tilde{\psi}_1 + \cdots + \tilde{\psi}_n)/n$ converges in $L^1(\Upsilon_0)$ to $\hat{\psi}(\ell_2) := \frac{1}{2} d_L(\ell_1, \ell_2)^2 - \hat{\phi}(\ell_1)$. Set

$$\phi_n' = \frac{\tilde{\phi}_1 + \cdots + \tilde{\phi}_n}{n}, \qquad \psi_n' = \frac{\tilde{\psi}_1 + \cdots + \tilde{\psi}_n}{n}.$$

Then we have

$$\phi_n'(\ell_1) + \psi_n'(\ell_2) \leqslant \frac{1}{2} d_L(\ell_1, \ell_2)^2, \quad \ell_1, \ell_2 \in \mathcal{L}_eG. \tag{3.14}$$

Again up to a subsequence, ϕ_n' converges Υ_0-a.s. to $\hat{\phi}$ and ψ_n' converges Υ_0-a.s. to $\hat{\psi}$; therefore, ϕ_n' converges to $\hat{\phi}$ ν-a.s. and ψ_n' converges to $\hat{\psi}$ μ-a.s. Now define

$$\phi = \varliminf_{n\to\infty} \phi_n', \qquad \psi = \varliminf_{n\to\infty} \psi_n'.$$

Then $\phi = \hat{\phi}$ ν-a.s. and $\phi = \hat{\psi}$ μ-a.s., and

$$\phi(\ell_1) + \psi(\ell_2) \leqslant \frac{1}{2} d_L(\ell_1, \ell_2)^2, \quad \ell_1, \ell_2 \in \mathcal{L}_eG. \tag{3.15}$$

We have

$$\int_{\mathcal{L}_e G} \phi \, \mathrm{d}\nu + \int_{\mathcal{L}_e G} \psi \, \mathrm{d}\mu = \int_{\mathcal{L}_e G \times \mathcal{L}_e G} \left(\hat{\phi}(\ell_1) + \frac{1}{2} d_L(\ell_1, \ell_2)^2 - \hat{\phi}(\ell_1) \right) \Upsilon_0(\mathrm{d}\ell_1, \mathrm{d}\ell_2)$$
$$= \int_{\mathcal{L}_e G \times \mathcal{L}_e G} \frac{1}{2} d_L(\ell_1, \ell_2)^2 \Upsilon_0(\mathrm{d}\ell_1, \mathrm{d}\ell_2) = W_2^2(\mu, \nu).$$

As a byproduct of the above equality, we have

$$\phi(\ell_1) + \psi(\ell_2) = \frac{1}{2} d_L(\ell_1, \ell_2)^2, \quad \Upsilon_0\text{-a.s.} \qquad \square \tag{3.16}$$

4. Proof of the main theorem

Let (ϕ, ψ) be the pair of Borel functions constructed in the proof of Theorem 3.4. We know that $\phi \in \mathcal{D}_1^2(\nu)$. According to [12, Lemma 4.14], for each $h \in H_0(\mathcal{G})$, there exists a full subset $\Omega_h \subset \mathcal{L}_e G$ w.r.t. ν such that for $\ell \in \Omega_h$, $t \mapsto \phi(\ell e^{th})$ is absolutely continuous and

$$\big\langle (\nabla \phi)(\ell), h \big\rangle_{H_0} = \left\{ \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \phi\big(\ell e^{\varepsilon h}\big) \right\}_{\varepsilon = 0}. \tag{4.1}$$

Let $D \subset H_0(\mathcal{G})$ be a dense countable subset. Then there is a full subset Ω such that for $\ell \in \Omega$, (4.1) holds for all $h \in D$. Now consider

$$\Theta = \big\{ (\ell_1, \ell_2) \in \mathcal{L}_e G \times \mathcal{L}_e G;\ \ell_1 \notin \Omega \big\}.$$

Let $\Upsilon_0 \in \mathcal{C}(\nu, \mu)$ which attains the Wasserstein distance $W_2(\nu, \mu)$. Then $\Upsilon_0(\Theta) = \nu(\Omega^c) = 0$. In the sequel, we consider $(\ell_1, \ell_2) \notin \Theta$ and satisfies the relation (3.16). Under the hypothesis $\mathrm{Ent}_\nu(F) < +\infty$, we can assume that $d_L(\ell_1, \ell_2) < +\infty$. By [9, Proposition 2.3(i)], there is a minimizing curve $\gamma : [0, 1] \to \mathcal{L}_e G$ such that $\gamma(0) = \mathbf{e}$, $\gamma(1) = \ell_1^{-1} \ell_2$ and

$$d_L(\ell_1, \ell_2)^2 = \int_0^1 \big| z'(t) \big|_{H_0}^2 \, \mathrm{d}t = L(\gamma)^2,$$

where

$$d_t \gamma(t, \theta) = \gamma(t, \theta) z_t'(\theta) \, \mathrm{d}t, \qquad \gamma(0, \theta) = e. \tag{4.2}$$

For $h \in H_0(\mathcal{G})$, define

$$\big(Q(\gamma, t) h \big)(\theta) = \mathrm{Ad}_{\gamma_t^{-1}(\theta)} h(\theta). \tag{4.3}$$

Proposition 4.1. *We have*

$$\big| Q(\gamma, t) h \big|_{H_0} \leqslant 2\big(1 + L(\gamma)\big) \cdot |h|_{H_0}. \tag{4.4}$$

Proof. By expression (4.3), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}\big(Q(\gamma,t)h\big) &= -\gamma_t^{-1}(\theta)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\gamma_t(\theta)\right)\gamma_t^{-1}(\theta)h(\theta)\gamma_t(\theta) \\ &\quad + \gamma_t^{-1}(\theta)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}h(\theta)\right)\gamma_t(\theta) + \gamma_t^{-1}(\theta)h(\theta)\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\gamma_t(\theta)\right) \\ &= \mathrm{Ad}_{\gamma_t^{-1}(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}h(\theta) - \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\gamma_t(\theta)\gamma_t^{-1}(\theta), h(\theta)\right]\right).\end{aligned}$$

To complete the proof, we need the following formula.

Lemma 4.2. *We have*

$$\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}\gamma_t^{-1}(\theta) = \int_0^t \mathrm{Ad}_{\gamma_s(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}z_s'(\theta)\right)\mathrm{d}s. \tag{4.5}$$

Proof. First suppose that $(t,\theta) \to z_t'(\theta)$ is smooth. By (4.2), we have $\gamma_t^{-1}\frac{\mathrm{d}}{\mathrm{d}\theta}\gamma_t(\theta) = z_t'(\theta)$. If we denote by $\frac{D^l}{\mathrm{d}\theta}$ the left covariant derivative and $\frac{D^r}{\mathrm{d}\theta}$ the right covariant derivative on G, then

$$\frac{D^r}{\mathrm{d}t}\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta} = \frac{D^l}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}t}\gamma_t(\theta).$$

Therefore,

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}\gamma_t^{-1}(\theta)\right)\right)\gamma_t(\theta) = \gamma_t(\theta)\frac{\mathrm{d}z_t'(\theta)}{\mathrm{d}\theta},$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}\gamma_t^{-1}(\theta)\right) = \mathrm{Ad}_{\gamma_t(\theta)}\left(\frac{\mathrm{d}z_t'(\theta)}{\mathrm{d}\theta}\right),$$

from which we get (4.5). The general case follows by density arguments. □

End of the proof of Proposition 4.1. Using the expression of $\frac{\mathrm{d}}{\mathrm{d}\theta}(Q(\gamma,t)h)$ and according to (4.5), we have

$$\begin{aligned}\big|Q(\gamma,t)h\big|_{H_0} &= \left(\int_0^1 \left|\mathrm{Ad}_{\gamma_t^{-1}(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}h(\theta) - \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\gamma_t(\theta)\gamma_t^{-1}(\theta), h(\theta)\right]\right)\right|_{\mathcal{G}}^2 \mathrm{d}\theta\right)^{\frac{1}{2}} \\ &\leqslant |h|_{H_0} + 2\left(\int_0^1 \left|\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}\gamma_t^{-1}(\theta)\right|_{\mathcal{G}}^2 \mathrm{d}\theta\right)^{\frac{1}{2}} \cdot \sup_\theta \big|h(\theta)\big|_{\mathcal{G}} \\ &\leqslant 2\left(1 + \left(\int_0^1\int_0^1 \left|\frac{\mathrm{d}}{\mathrm{d}\theta}z_s'(\theta)\right|_{\mathcal{G}}^2 \mathrm{d}\theta\,\mathrm{d}s\right)^{\frac{1}{2}}\right)|h|_{H_0},\end{aligned}$$

which is nothing but (4.4). □

Let $Q(\gamma,t)^*$ be the adjoint operator of $Q(\gamma,t)$ on $H_0(\mathcal{G})$.

Definition 4.3. Set

$$V_t = \int_0^t Q(\gamma,s)^* z_s' \,\mathrm{d}s. \tag{4.6}$$

Proposition 4.4. *We have*

$$V_1 = -(\nabla\phi)(\ell_1), \tag{4.7}$$

where ϕ is given in Theorem 3.4.

Proof. Let $h \in D$. Consider $\gamma_\varepsilon(t,\theta) := e^{-t\varepsilon h(\theta)}\gamma(t,\theta)$. Then γ_ε connects $\mathbf{e}$ and $e^{-\varepsilon h}(\ell_1^{-1}\ell_2)$. We have

$$\begin{aligned} d_t\gamma_\varepsilon(t,\theta) &= -\varepsilon h(\theta)e^{-t\varepsilon h(\theta)}\gamma(t,\theta)\,\mathrm{d}t + e^{-t\varepsilon h(\theta)} d_t\gamma(t,\theta) \\ &= \gamma_\varepsilon(t,\theta)\big(z_t'(\theta) - \varepsilon\,\mathrm{Ad}_{\gamma^{-1}(t,\theta)e^{t\varepsilon h(\theta)}}\big(h(\theta)\big)\big)\,\mathrm{d}t. \end{aligned}$$

Then

$$\begin{aligned} d_L\big(\ell_1 e^{\varepsilon h},\ell_2\big)^2 &\leqslant \int_0^1 \big|z_t' - \varepsilon Q(\gamma_\varepsilon,t)h\big|_{H_0}^2\,\mathrm{d}t \\ &= \int_0^1 \big|z_t'\big|_{H_0}^2\,\mathrm{d}t - 2\varepsilon\int_0^1 \big\langle z_t', Q(\gamma_\varepsilon,t)h\big\rangle_{H_0}\,\mathrm{d}t + \varepsilon^2\int_0^1 \big|Q(\gamma_\varepsilon,t)h\big|_{H_0}^2\,\mathrm{d}t. \end{aligned}$$

It follows that

$$\frac{1}{2}d_L^2\big(\ell_1 e^{\varepsilon h},\ell_2\big) - \frac{1}{2}d_L^2(\ell_1,\ell_2) \leqslant -\varepsilon\int_0^1 \big\langle z_t', Q(\gamma_\varepsilon,t)h\big\rangle_{H_0}\,\mathrm{d}t + \frac{1}{2}\varepsilon^2\int_0^1 \big|Q(\gamma_\varepsilon,t)h\big|_{H_0}^2\,\mathrm{d}t. \tag{4.8}$$

By (3.15) and (3.16), we have

$$\phi\big(\ell_1 e^{\varepsilon h}\big) - \phi(\ell_1) \leqslant \frac{1}{2}d_L^2\big(\ell_1 e^{\varepsilon h},\ell_2\big) - \frac{1}{2}d_L^2(\ell_1,\ell_2).$$

Using (4.1) and (4.8), we get

$$\big\langle\nabla\phi(\ell_1),h\big\rangle_{H_0} \leqslant -\int_0^1 \big\langle z_t', Q(\gamma,t)h\big\rangle_{H_0}\,\mathrm{d}t = -\left\langle \int_0^1 Q(\gamma,t)^* z_t'\,\mathrm{d}t, h\right\rangle_{H_0}$$

for each $h \in D$. Notice that D is dense in $H_0(\mathcal{G})$, hence

$$\nabla\phi(\ell_1) = -\int_0^1 Q(\gamma, t)^* z_t' \, \mathrm{d}t. \qquad \square$$

4.1. Construction of the optimal transportation $\mathcal{T}$

For further study, we need the explicit expression of $Q^*(\gamma, t)^{-1} = Q^*(\gamma^{-1}, t)$.

Lemma 4.5.

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\big(Q\big(\gamma^{-1}, t\big)^* k\big) = \mathrm{Ad}_{\gamma_t^{-1}(\theta)}\left(\frac{\mathrm{d}k}{\mathrm{d}\theta}\right) - \int_0^1 \frac{\mathrm{d}G_\sigma(\theta)}{\mathrm{d}\theta} \mathrm{Ad}_{\gamma_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma), \frac{\mathrm{d}k}{\mathrm{d}\sigma}\right] \mathrm{d}\sigma. \quad (4.9)$$

Proof. We have

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\big(Q\big(\gamma^{-1}, t\big)h\big) = \mathrm{Ad}_{\gamma_t(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}h(\theta) + \left[\gamma_t^{-1}(\theta)\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}, h(\theta)\right]\right).$$

Let $k \in H_0(\mathcal{G})$, then

$$\begin{aligned}
&\big\langle k, Q\big(\gamma^{-1}, t\big)h\big\rangle_{H_0} \\
&= \int_0^1 \left\langle \mathrm{Ad}_{\gamma_t^{-1}}\left(\frac{\mathrm{d}k}{\mathrm{d}\theta}\right), \frac{\mathrm{d}h}{\mathrm{d}\theta} + \left[\gamma_t^{-1}(\theta)\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}, h(\theta)\right]\right\rangle_{\mathcal{G}} \mathrm{d}\theta \\
&= \int_0^1 \left\langle \mathrm{Ad}_{\gamma_t^{-1}}\left(\frac{\mathrm{d}k}{\mathrm{d}\theta}\right), \frac{\mathrm{d}h}{\mathrm{d}\theta}\right\rangle_{\mathcal{G}} \mathrm{d}\theta - \int_0^1 \left\langle\left[\gamma_t^{-1}(\theta)\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}, \mathrm{Ad}_{\gamma_t^{-1}}\left(\frac{\mathrm{d}k}{\mathrm{d}\theta}\right)\right], h(\theta)\right\rangle_{\mathcal{G}} \mathrm{d}\theta.
\end{aligned}$$

Recall that G_θ is the function such that $h(\theta) = \int_0^1 \langle \frac{\mathrm{d}h(\sigma)}{\mathrm{d}\sigma}, \frac{\mathrm{d}G_\theta(\sigma)}{\mathrm{d}\sigma}\rangle \, \mathrm{d}\sigma$. By Fubini theorem, the last term in the above equality is equal to

$$-\int_0^1 \left\langle \frac{\mathrm{d}h(\sigma)}{\mathrm{d}\sigma}, \int_0^1 \frac{\mathrm{d}G_\theta(\sigma)}{\mathrm{d}\sigma}\left[\gamma_t^{-1}(\theta)\frac{\mathrm{d}\gamma_t(\theta)}{\mathrm{d}\theta}, \mathrm{Ad}_{\gamma_t^{-1}(\theta)}\left(\frac{\mathrm{d}k}{\mathrm{d}\theta}\right)\right]\mathrm{d}\theta\right\rangle_{\mathcal{G}} \mathrm{d}\sigma.$$

Therefore by changing notations, we get the expression for $\frac{\mathrm{d}}{\mathrm{d}\theta}(Q(\gamma^{-1}, t)^* k)$, which is equal to

$$\mathrm{Ad}_{\gamma_t^{-1}}\left(\frac{\mathrm{d}k}{\mathrm{d}\theta}\right) - \int_0^1 \frac{\mathrm{d}G_\sigma(\theta)}{\mathrm{d}\theta}\left[\gamma_t^{-1}(\sigma)\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}, \mathrm{Ad}_{\gamma_t^{-1}}\left(\frac{\mathrm{d}k}{\mathrm{d}\sigma}\right)\right]\mathrm{d}\sigma$$

or

$$\mathrm{Ad}_{\gamma_t^{-1}(\theta)}\left(\frac{\mathrm{d}k}{\mathrm{d}\theta}\right)-\int_0^1\frac{\mathrm{d}G_\sigma(\theta)}{\mathrm{d}\theta}\,\mathrm{Ad}_{\gamma_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma),\frac{\mathrm{d}k}{\mathrm{d}\sigma}\right]\mathrm{d}\sigma.$$

We get (4.9). □

In order to construct the Monge optimal transportation $\mathcal{T}(\ell)$, we shall inject the minimizing curve γ into an infinite-dimensional differential equation as in Section 2. To this end, we consider $\mathcal{P}_e^1(G)$, the path group over G of finite energy

$$\int_0^1\left|\gamma^{-1}(s)\dot\gamma(s)\right|_{\mathcal{G}}^2\,\mathrm{d}s<+\infty.$$

On $\mathcal{P}_e^1(G)$, we consider the distance $d_{\mathcal{P}}(\gamma_1,\gamma_2)$ defined by

$$d_{\mathcal{P}}(\gamma_1,\gamma_2)=\left(\int_0^1\left|v^{-1}(s)\dot v(s)\right|_{\mathcal{G}}^2\,\mathrm{d}s\right)^{1/2},\quad\text{where } v=\gamma_1^{-1}\gamma_2. \tag{4.10}$$

Then $d_{\mathcal{P}}$ is left invariant: $d_{\mathcal{P}}(\gamma\gamma_1,\gamma\gamma_2)=d_{\mathcal{P}}(\gamma_1,\gamma_2)$.

Proposition 4.6. *There exists a constant $C>0$ such that*

$$d_{\mathcal{P}}(\gamma_1\gamma,\gamma_2\gamma)\leqslant\big(Cd_{\mathcal{P}}(\gamma,\mathbf{e})+1\big)d_{\mathcal{P}}(\gamma_1,\gamma_2), \tag{4.11}$$

where $\mathbf{e}$ *denotes the identity path.*

Proof. Let $v=\gamma^{-1}(\gamma_1^{-1}\gamma_2)\gamma$. We have

$$\begin{aligned}\dot v&=-\gamma^{-1}\dot\gamma\gamma^{-1}\big(\gamma_1^{-1}\gamma_2\big)\gamma+\gamma^{-1}\overset{\cdot}{\overbrace{\big(\gamma_1^{-1}\gamma_2\big)}}\gamma+\gamma^{-1}\big(\gamma_1^{-1}\gamma_2\big)\dot\gamma,\\ \dot v v^{-1}&=-\gamma^{-1}\dot\gamma+\gamma^{-1}\overset{\cdot}{\overbrace{\big(\gamma_1^{-1}\gamma_2\big)}}\big(\gamma_1^{-1}\gamma_2\big)^{-1}\gamma+\gamma^{-1}\big(\gamma_1^{-1}\gamma_2\big)\dot\gamma\gamma^{-1}\big(\gamma_1^{-1}\gamma_2\big)^{-1}\gamma\\ &=\mathrm{Ad}_{\gamma^{-1}}\Big(\big(\gamma_1^{-1}\gamma_2\big)^{\cdot}\big(\gamma_1^{-1}\gamma_2\big)^{-1}+(\mathrm{Ad}_{\gamma_1^{-1}\gamma_2}-\mathrm{Id})\dot\gamma\gamma^{-1}\Big).\end{aligned}$$

It follows that

$$\begin{aligned}\left(\int_0^1\left|v^{-1}\dot v\right|_{\mathcal{G}}^2\,\mathrm{d}s\right)^{1/2}&=\left(\int_0^1\left|\dot v v^{-1}\right|_{\mathcal{G}}^2\,\mathrm{d}s\right)^{1/2}\\ &\leqslant\left(\int_0^1\left|\overset{\cdot}{\overbrace{\big(\gamma_1^{-1}\gamma_2\big)}}\big(\gamma_1^{-1}\gamma_2\big)^{-1}\right|_{\mathcal{G}}^2\,\mathrm{d}s\right)^{1/2}\\ &\quad+\left(\int_0^1\left|(\mathrm{Ad}_{(\gamma_1^{-1}\gamma_2)(s)}-\mathrm{Id})\dot\gamma\gamma^{-1}(s)\right|_{\mathcal{G}}^2\,\mathrm{d}s\right)^{1/2}.\end{aligned}$$

The first term on the right-hand side is $d_{\mathcal{P}}(\gamma_1, \gamma_2)$, while the second term is dominated by

$$C \sup_{0\leqslant s\leqslant 1} d_G\big(\gamma_1(s), \gamma_2(s)\big) \cdot d_G(\gamma, \mathbf{e}) \leqslant C d_{\mathcal{P}}(\gamma_1, \gamma_2) \cdot d_{\mathcal{P}}(\gamma, \mathbf{e}).$$

So we get (4.11). $\square$

Now let $\gamma : [0, 1] \to \mathcal{P}_e^1(G)$ be a continuous curve, where $\mathcal{P}_e^1(G)$ is endowed with the uniform topology. According to the expression (4.9), we define $A(\gamma, t)h \in H(\mathcal{G})$ by

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\big(A(\gamma, t)h\big) = \mathrm{Ad}_{\gamma_t^{-1}(\theta)}\left(\frac{\mathrm{d}h}{\mathrm{d}\theta}\right) - \int_0^1 \frac{\mathrm{d}G_\sigma(\theta)}{\mathrm{d}\theta} \mathrm{Ad}_{\gamma_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right]\mathrm{d}\sigma. \tag{4.12}$$

Proposition 4.7. *We have*

$$\big|A(\gamma, t)h\big|_H \leqslant \big(1 + 2d_{\mathcal{P}}(\gamma_t, \mathbf{e})\big)|h|_H. \tag{4.13}$$

Proof. Note that $|\frac{\mathrm{d}}{\mathrm{d}\theta} G_\sigma(\theta)| = |\mathbf{1}_{\{\theta<\sigma\}} - \sigma| \leqslant 1$. Then

$$\begin{aligned}
&\left|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta} G_\sigma(\theta)\, \mathrm{Ad}_{\gamma_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right]\mathrm{d}\sigma\right|_{\mathcal{G}} \\
&\quad\leqslant 2\int_0^1 \left|\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma)\right|_{\mathcal{G}} \left|\frac{\mathrm{d}h}{\mathrm{d}\sigma}\right|_{\mathcal{G}} \mathrm{d}\sigma \\
&\quad\leqslant 2\left(\int_0^1 \left|\gamma_t^{-1}(\sigma)\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\right|_{\mathcal{G}}^2 \mathrm{d}\sigma\right)^{1/2} \cdot |h|_H \\
&\quad= 2d_{\mathcal{P}}(\gamma_t, \mathbf{e})\, |h|_H.
\end{aligned}$$

Using (4.12), we get that

$$\big|A(\gamma, t)h\big|_H \leqslant |h|_H + 2d_{\mathcal{P}}(\gamma_t, \mathbf{e})|h|_H = \big(1 + 2d_{\mathcal{P}}(\gamma_t, \mathbf{e})\big)|h|_H. \qquad \square$$

Proposition 4.8. *We have*

$$\big\|A(\gamma, t) - A(\tilde{\gamma}, t)\big\|_{\mathrm{op}} \leqslant 2C\big(1 + d_{\mathcal{P}}(\gamma_t, \mathbf{e})\big)d_{\mathcal{P},\infty}(\gamma, \tilde{\gamma}) + 2d_{\mathcal{P}}(\gamma, \tilde{\gamma}), \tag{4.14}$$

where $d_{\mathcal{P},\infty}(\gamma_t, \tilde{\gamma}_t) := \sup_\theta d_G(\gamma_t(\theta), \tilde{\gamma}_t(\theta))$.

Proof. By (4.12),

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta}\big[A(\gamma,t)h - A(\tilde{\gamma},t)h\big] \\
&\quad = (\mathrm{Ad}_{\gamma_t^{-1}(\theta)} - \mathrm{Ad}_{\tilde{\gamma}_t^{-1}(\theta)})\left(\frac{\mathrm{d}h}{\mathrm{d}\theta}\right) - \int_0^1 \frac{\mathrm{d}G_\sigma(\theta)}{\mathrm{d}\theta}\left\{\mathrm{Ad}_{\gamma_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right]\right. \\
&\qquad \left. - \mathrm{Ad}_{\tilde{\gamma}_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\tilde{\gamma}_t(\sigma)}{\mathrm{d}\sigma}\tilde{\gamma}_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right]\right\}\mathrm{d}\sigma.
\end{aligned}$$

Since

$$\|\mathrm{Ad}_{\gamma_t^{-1}(\theta)} - \mathrm{Ad}_{\tilde{\gamma}_t^{-1}(\theta)}\| \leqslant C d_G\big(\gamma_t^{-1}(\theta), \tilde{\gamma}_t^{-1}(\theta)\big) \leqslant C \sup_\theta d_G\big(\gamma_t(\theta), \tilde{\gamma}(\theta)\big),$$

then

$$\left\|(\mathrm{Ad}_{\gamma_t^{-1}(\theta)} - \mathrm{Ad}_{\tilde{\gamma}_t^{-1}(\theta)})\frac{\mathrm{d}h}{\mathrm{d}\theta}\right\|_{L^2} \leqslant C d_{\mathcal{P},\infty}(\gamma_t, \tilde{\gamma}_t)|h|_H.$$

For estimating the second term, we consider $\gamma_t^{-1}\tilde{\gamma}_t$. Then

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\sigma}\big(\gamma_t^{-1}\tilde{\gamma}_t\big) &= -\gamma_t^{-1}\frac{\mathrm{d}\gamma_t}{\mathrm{d}\sigma}\gamma_t^{-1}\tilde{\gamma}_t + \gamma_t^{-1}\frac{\mathrm{d}\tilde{\gamma}_t}{\mathrm{d}\sigma} \\
&= \gamma_t^{-1}(\sigma)\left(-\frac{\mathrm{d}\gamma_t}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma) + \frac{\mathrm{d}\tilde{\gamma}_t}{\mathrm{d}\sigma}\tilde{\gamma}_t^{-1}(\sigma)\right)\tilde{\gamma}_t(\sigma)
\end{aligned}$$

and

$$\big(\gamma_t^{-1}(\sigma)\tilde{\gamma}_t(\sigma)\big)^{-1}\frac{\mathrm{d}}{\mathrm{d}\sigma}\big(\gamma_t^{-1}\tilde{\gamma}_t\big) = \mathrm{Ad}_{\tilde{\gamma}_t^{-1}(\sigma)}\left(\frac{\mathrm{d}\tilde{\gamma}_t}{\mathrm{d}\sigma}\tilde{\gamma}_t^{-1}(\sigma) - \frac{\mathrm{d}\gamma_t}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma)\right). \tag{4.15}$$

Therefore,

$$\begin{aligned}
M_t(\sigma) &:= \mathrm{Ad}_{\gamma_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\gamma_t(\sigma)}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right] - \mathrm{Ad}_{\tilde{\gamma}_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\tilde{\gamma}_t}{\mathrm{d}\sigma}\tilde{\gamma}_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right] \\
&= (\mathrm{Ad}_{\gamma_t^{-1}(\sigma)} - \mathrm{Ad}_{\tilde{\gamma}_t(\sigma)})\left[\frac{\mathrm{d}\gamma_t}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right] \\
&\quad + \mathrm{Ad}_{\tilde{\gamma}_t^{-1}(\sigma)}\left[\frac{\mathrm{d}\gamma_t}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma) - \frac{\mathrm{d}\tilde{\gamma}_t}{\mathrm{d}\sigma}\tilde{\gamma}_t^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right].
\end{aligned}$$

Then according to (4.15), we get

$$\big|M_t(\sigma)\big|_{\mathcal{G}} \leqslant 2C d_{\mathcal{P},\infty}(\gamma_t,\tilde{\gamma}_t)\left|\frac{\mathrm{d}\gamma_t}{\mathrm{d}\sigma}\gamma_t^{-1}(\sigma)\right|_{\mathcal{G}}\left|\frac{\mathrm{d}h}{\mathrm{d}\sigma}\right|_{\mathcal{G}} + 2\left|\big(\gamma_t^{-1}(\sigma)\tilde{\gamma}_t(\sigma)\big)^{-1}\frac{\mathrm{d}}{\mathrm{d}\sigma}\big(\gamma_t^{-1}(\sigma)\tilde{\gamma}_t(\sigma)\big)\right|_{\mathcal{G}}\left|\frac{\mathrm{d}h}{\mathrm{d}\sigma}\right|_{\mathcal{G}}.$$

Finally,

$$\int_0^1 \big|M_t(\sigma)\big|_{\mathcal{G}}\,\mathrm{d}\sigma \leqslant 2C d_{\mathcal{P},\infty}(\gamma_t,\tilde{\gamma}_t) d_{\mathcal{P}}(\gamma_t,\mathbf{e})|h|_H + 2d_{\mathcal{P}}(\gamma_t,\tilde{\gamma}_t)|h|_H.$$

Therefore,

$$\big\|A(\gamma,t)-A(\tilde{\gamma},t)\big\|_{\mathrm{op}} \leqslant 2C\big(1+d_{\mathcal{P}}(\gamma_t,\mathbf{e})\big)d_{\mathcal{P},\infty}(\gamma_t,\tilde{\gamma}_t) + 2d_{\mathcal{P}}(\gamma_t,\tilde{\gamma}_t). \qquad \square$$

Let $\gamma : [0,1] \to \mathcal{P}_e^1(G)$ defined by

$$d_t\gamma(t,\theta) = \gamma(t,\theta)z_t'(\theta)\,dt,$$

where $\int_0^1 |z_t'|_H^2\,\mathrm{d}t < +\infty$. Let $V_t = \int_0^t V_s'\,\mathrm{d}s$ with $\int_0^1 |V_t'|_H^2\,\mathrm{d}t < +\infty$. In what follows, suppose that γ resolve also the differential equation

$$d_t\gamma(t,\theta) = \gamma(t,\theta)\big(A(\gamma,t)V_t'\big)(\theta)\,\mathrm{d}t, \qquad \gamma(0,\theta)=e. \tag{4.16}$$

Proposition 4.9. *We have*

$$d_{\mathcal{P}}(\gamma_t,\mathbf{e}) \leqslant \sqrt{2}\,e^{4(\int_0^1 |V_s'|_H^2\,\mathrm{d}s)}\left(\int_0^1 \big|V_s'\big|_H^2\,\mathrm{d}s\right)^{1/2}. \tag{4.17}$$

In particular, $\sup_{0\leqslant t\leqslant 1} d_{\mathcal{P}}(\gamma_t,\mathbf{e}) < +\infty$.

Proof. By Lemma 4.2, we have

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\gamma_t(\theta)\gamma_t^{-1}(\theta) = \int_0^t \mathrm{Ad}_{\gamma_s(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}A(\gamma,s)V_s'\right)\mathrm{d}s. \tag{4.18}$$

Therefore,

$$d_{\mathcal{P}}(\gamma_t,\mathbf{e}) = \left\|\frac{\mathrm{d}\gamma_t}{\mathrm{d}\theta}\gamma_t^{-1}(\theta)\right\|_{L^2(\mathrm{d}\theta)} \leqslant \int_0^t\left(\int_0^1\left|\frac{\mathrm{d}}{\mathrm{d}\theta}A(\gamma,s)V_s'\right|_{\mathcal{G}}^2\mathrm{d}\theta\right)^{1/2}\mathrm{d}s,$$

which is smaller by (4.13) than

$$\int_0^t \bigl(1+2d_{\mathcal{P}}(\gamma_s,\mathbf{e})\bigr)\bigl|V_s'\bigr|_H\,\mathrm{d}s.$$

Hence

$$\begin{aligned}
d_{\mathcal{P}}^2(\gamma_t,\mathbf{e}) &\leqslant \left(\int_0^t \bigl(1+2d_{\mathcal{P}}(\gamma_s,\mathbf{e})\bigr)^2\,\mathrm{d}s\right)\left(\int_0^1 \bigl|V_s'\bigr|_H^2\,\mathrm{d}s\right)\\
&\leqslant 2\int_0^1 \bigl|V_s'\bigr|_H^2\,\mathrm{d}s + 8\left(\int_0^1 \bigl|V_s'\bigr|_H^2\,\mathrm{d}s\right)\int_0^t d_{\mathcal{P}}^2(\gamma_s,\mathbf{e})\,\mathrm{d}s.
\end{aligned}$$

Then the Gronwall's lemma yields

$$d_{\mathcal{P}}^2(\gamma_t,\mathbf{e}) \leqslant 2\left(\int_0^1 \bigl|V_s'\bigr|_H^2\,\mathrm{d}s\right)\cdot e^{8\int_0^1 |V_s'|_H^2\,\mathrm{d}s},$$

so we get (4.17). □

Theorem 4.10. *The differential equation* (4.16) *has at most one solution.*

Proof. Let γ_t, $\tilde{\gamma}_t$ be two solutions. Using (4.15) and (4.18),

$$\begin{aligned}
&\left|\bigl(\gamma_t^{-1}(\theta)\tilde{\gamma}_t(\theta)\bigr)^{-1}\frac{\mathrm{d}}{\mathrm{d}\theta}\bigl(\gamma_t^{-1}(\theta)\tilde{\gamma}_t(\theta)\bigr)\right|_{\mathcal{G}}\\
&\quad=\left|\int_0^t \left\{\mathrm{Ad}_{\gamma_s(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}A(\gamma,s)V_s'\right)-\mathrm{Ad}_{\tilde{\gamma}_s(\theta)}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}A(\tilde{\gamma},s)V_s'\right)\right\}\mathrm{d}s\right|_{\mathcal{G}}\\
&\quad\leqslant \int_0^t \|\mathrm{Ad}_{\gamma_s(\theta)}-\mathrm{Ad}_{\tilde{\gamma}_s(\theta)}\|\cdot\left|\frac{\mathrm{d}}{\mathrm{d}\theta}A(\gamma,s)V_s'\right|_{\mathcal{G}}\mathrm{d}s + \int_0^t\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\bigl(\bigl(A(\gamma,s)-A(\tilde{\gamma},s)\bigr)V_s'\bigr)\right|_{\mathcal{G}}\mathrm{d}s.
\end{aligned}$$

It follows that

$$d_{\mathcal{P}}(\gamma_t,\tilde{\gamma}_t)\leqslant \int_0^t C d_{\mathcal{P},\infty}(\gamma_s,\tilde{\gamma}_s)\bigl|A(\gamma,s)V_s'\bigr|_H\,\mathrm{d}s + \int_0^t \bigl|\bigl(A(\gamma,s)-A(\tilde{\gamma},s)\bigr)V_s'\bigr|_H\,\mathrm{d}s. \qquad (*)$$

According to (4.13), the first term is dominated by

$$C\int_0^t d_{\mathcal{P},\infty}(\gamma_s,\tilde{\gamma}_s)\big(1+2d_{\mathcal{P}}(\gamma_s,\mathbf{e})\big)\big|V_s'\big|_H\,\mathrm{d}s\leqslant K_1C\int_0^t d_{\mathcal{P},\infty}(\gamma_s,\tilde{\gamma}_s)\big|V_s'\big|_H\,\mathrm{d}s,$$

where $K_1:=1+2\sup_{0\leqslant t\leqslant 1}d_{\mathcal{P}}(\gamma_t,\mathbf{e})<+\infty$ by (4.17); according to (4.14), the second term in $(*)$ is dominated by

$$2C\int_0^t \big(1+d_{\mathcal{P}}(\gamma_s,\mathbf{e})\big)d_{\mathcal{P},\infty}(\gamma_s,\tilde{\gamma}_s)\big|V_s'\big|_H\,\mathrm{d}s+2\int_0^t d_{\mathcal{P}}(\gamma_s,\tilde{\gamma}_s)\big|V_s'\big|_H\,\mathrm{d}s.$$

Therefore,

$$\begin{aligned}d_{\mathcal{P}}^2(\gamma_t,\tilde{\gamma}_s)\leqslant{}&10K_1^2C^2\bigg(\int_0^t d_{\mathcal{P},\infty}(\gamma_s,\tilde{\gamma}_s)\,\mathrm{d}s\bigg)\int_0^1\big|V_s'\big|_H^2\,\mathrm{d}s\\&+8\bigg(\int_0^t d_{\mathcal{P}}^2(\gamma_s,\tilde{\gamma}_s)\,\mathrm{d}s\bigg)\int_0^1\big|V_s'\big|_H^2\,\mathrm{d}s.\end{aligned}$$

Since $d_{\mathcal{P},\infty}(\gamma_s,\tilde{\gamma}_s)\leqslant d_{\mathcal{P}}(\gamma_s,\tilde{\gamma}_s)$, we deduce that

$$d_{\mathcal{P}}^2(\gamma_t,\tilde{\gamma}_t)\leqslant\big(10K_1^2C^2+8\big)\bigg(\int_0^1\big|V_s'\big|_H^2\,\mathrm{d}s\bigg)\int_0^t d_{\mathcal{P}}^2(\gamma_s,\tilde{\gamma}_s)\,\mathrm{d}s.$$

Now the Gronwall's lemma yields that

$$d_{\mathcal{P}}(\gamma_t,\tilde{\gamma}_t)=0\quad\text{for all } t\in[0,1].\qquad\square$$

Let D be a countable dense subset of $H_0(\mathcal{G})$. For $h\in D$, $\varepsilon\in\mathbb{R}$ and $c\in C^2([0,1],\mathbb{R})$ such that $c(0)=c(1)=0$, consider $\gamma_\varepsilon(t,\theta):=e^{\varepsilon c(t)h(\theta)}\gamma(t,\theta)$. Then γ_ε is also a continuous curve connecting $\mathbf{e}$ and $\ell_1^{-1}\ell_2$.

$$\begin{aligned}d_t\gamma_\varepsilon(t,\theta)&=\varepsilon c'(t)h(\theta)e^{\varepsilon c(t)h(\theta)}\gamma(t,\theta)\,\mathrm{d}t+e^{\varepsilon c(t)h(\theta)}d_t\gamma(t,\theta)\\&=\gamma_\varepsilon(t,\theta)\big(z_t'(\theta)+\varepsilon c'(t)Q(\gamma_\varepsilon,t)h(\theta)\big),\end{aligned}$$

where $Q(\gamma,t)$ was defined in (4.3). Hence

$$\begin{aligned}L(\gamma_\varepsilon)^2&=\int_0^1\big|z_t'+\varepsilon c'(t)Q(\gamma_\varepsilon,t)h\big|_{H_0}^2\,\mathrm{d}t\\&=\int_0^1\big|z_t'\big|_{H_0}^2\,\mathrm{d}t+2\varepsilon\int_0^1\big\langle z_t',Q(\gamma_\varepsilon,t)h\big\rangle_{H_0}c'(t)\,\mathrm{d}t+\varepsilon^2\int_0^1\big|c'(t)Q(\gamma_\varepsilon,t)h\big|_{H_0}^2\,\mathrm{d}t.\end{aligned}$$

Since $\varepsilon \mapsto L(\gamma_\varepsilon)^2$ realizes a minimum at $\varepsilon = 0$, then

$$0 = \left\{ \frac{\mathrm{d}}{\mathrm{d}\varepsilon} L(\gamma_\varepsilon)^2 \right\}_{\varepsilon=0} = 2 \int_0^1 \langle z_t', Q(\gamma, t)h \rangle_{H_0} c'(t)\, \mathrm{d}t.$$

In what follows, we denote $V_t = \int_0^t Q(\gamma, s)^* z_s' \,\mathrm{d}s$. Applying the integration by parts formula, we get

$$\langle V_1, c'(1)h \rangle_{H_0} = \int_0^1 \langle V_t, c''(t)h \rangle_{H_0} \,\mathrm{d}t. \tag{4.19}$$

Now suppose that $\tilde{\gamma}_t$ is another curve connecting $\mathbf{e}$ with $\ell_1^{-1}\tilde{\ell}_2$ such that $\tilde{V}_1 = V_1$. Then by (4.19), we get

$$\int_0^1 \langle V_t - \tilde{V}_t, c''(t)h \rangle_{H_0} \,\mathrm{d}t = 0.$$

Since the set of such functions $t \mapsto c''(t)h$, for $h \in D$ is dense in $L^2([0,1], H_0(\mathcal{G}))$, we deduce that $V_t = \tilde{V}_t$ almost everywhere. By continuity, $V_t = \tilde{V}_t$ for each $t \in [0,1]$. Hence $Q(\gamma, t)^* z_t' = Q(\tilde{\gamma}, t)^* \tilde{z}_t'$. By (4.2),

$$\begin{aligned} d_t \gamma(t,\theta) &= \gamma(t,\theta) z_t'(\theta)\,\mathrm{d}t \\ &= \gamma(t,\theta)\big(A(\gamma,t)Q(\gamma,t)^* z_t'\big)(\theta)\,\mathrm{d}t \\ &= \gamma(t,\theta)\big(A(\gamma,t)V_t'\big)(\theta)\,\mathrm{d}t. \end{aligned}$$

According to Theorem 4.10, we have $\tilde{\gamma}(t,\theta) = \gamma(t,\theta)$ for $(t,\theta) \in [0,1] \times [0,1]$. In particular, $\tilde{\ell}_2 = \ell_2$. So ℓ_2 is uniquely determined by V_1. And Proposition 4.4 says that $V_1 = -\nabla\phi(\ell_1)$, so ℓ_2 is uniquely determined by $(\nabla\phi)(\ell_1)$. Denote

$$\ell_2 = \mathcal{T}(\ell_1).$$

4.2. Measurability of $\mathcal{T}$

In what follows, we shall establish the measurability of the map $\mathcal{T}$.

Theorem 4.11. *Let γ_t be the solution of* (4.16). *Then γ_t can be approximated by Euler scheme.*

Proof. Let $n \geqslant 1$. For $t \in [l2^{-n}, (l+1)2^{-n}[$, consider

$$\begin{aligned} &\gamma_n(t,\theta) = \gamma_n(l2^{-n},\theta)\exp\big\{(t - l2^{-n})2^n\big(A(\gamma_n, l2^{-n})(V_{(l+1)2^{-n}} - V_{l2^{-n}})\big)(\theta)\big\}, \\ &\gamma_n(0,\theta) = e, \end{aligned} \tag{4.20}$$

where

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta}\big(A\big(\gamma_n, l2^{-n}\big)h\big)\\
&\quad= \mathrm{Ad}_{\gamma_n^{-1}(l2^{-n},\theta)}\left(\frac{\mathrm{d}h}{\mathrm{d}\theta}\right)\\
&\qquad- \int_0^1 \frac{\mathrm{d}G_\sigma(\theta)}{\mathrm{d}\theta}\,\mathrm{Ad}_{\gamma_n^{-1}(l2^{-n},\sigma)}\left[\frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma_n\big(l2^{-n},\sigma\big)\gamma_n^{-1}\big(l2^{-n},\sigma\big), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\right]\mathrm{d}\sigma.
\end{aligned} \tag{4.21}$$

Set $V_n'(t) = 2^n(V_{(l+1)2^{-n}} - V_{l2^{-n}})$ for $t \in [l2^{-n}, (l+1)2^{-n}[$. Then (4.20) can be written in the form

$$\mathrm{d}\gamma_n(t,\theta) = \gamma_n(t,\theta)A(\gamma_n, t_n)V_n'(t_n)\,\mathrm{d}t, \qquad \gamma_n(0,\theta) = e, \tag{4.22}$$

where $t_n = [t2^n]2^{-n}$. Firstly, it is easy to see that

$$\lim_{n\to\infty}\int_0^1 \big|V_s' - V_n'(s)\big|_H^2\,\mathrm{d}s = 0. \tag{4.23}$$

Again by expression (4.15),

$$d_{\mathcal{P}}\big(\gamma_t, \gamma_n(t)\big) = \left(\int_0^1 \left|\frac{\mathrm{d}\gamma_t}{\mathrm{d}\theta}\gamma_t^{-1}(\theta) - \frac{\mathrm{d}\gamma_n(t,\theta)}{\mathrm{d}\theta}\gamma_n^{-1}(t,\theta)\right|_{\mathcal{G}}^2 \mathrm{d}\theta\right)^{1/2}.$$

Similarly, we have

$$\begin{aligned}
&\left|\frac{\mathrm{d}\gamma_t}{\mathrm{d}\theta}\gamma_t^{-1}(\theta) - \frac{\mathrm{d}\gamma_n(t,\theta)}{\mathrm{d}\theta}\gamma_n^{-1}(t,\theta)\right|_{\mathcal{G}}\\
&\quad\leqslant \int_0^t \left|\mathrm{Ad}_{\gamma_s(\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\big(A(\gamma,s)V_s'\big) - \mathrm{Ad}_{\gamma_n(s,\theta)}\frac{\mathrm{d}}{\mathrm{d}\theta}\big(A(\gamma_n,s_n)V_n(s_n)'\big)\right|_{\mathcal{G}}\mathrm{d}s\\
&\quad\leqslant \int_0^t \|\mathrm{Ad}_{\gamma_s(\theta)} - \mathrm{Ad}_{\gamma_n(s,\theta)}\|\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\big(A(\gamma,s)V_s'\big)\right|_{\mathcal{G}}\mathrm{d}s\\
&\qquad+ \int_0^t \left|\frac{\mathrm{d}}{\mathrm{d}\theta}\big(A(\gamma,s)V_s'\big) - \frac{\mathrm{d}}{\mathrm{d}\theta}\big(A(\gamma_n,s_n)V_n'(s_n)\big)\right|_{\mathcal{G}}\mathrm{d}s.
\end{aligned}$$

Then

$$d_{\mathcal{P}}\big(\gamma_t, \gamma_n(t)\big) \leqslant C \int_0^t d_{\mathcal{P},\infty}\big(\gamma_s, \gamma_n(s)\big)\big|A(\gamma, s)V_s'\big|_H \, \mathrm{d}s + \int_0^t \big|A(\gamma, s)V_s' - A(\gamma_n, s_n)V_n'(s_n)\big|_H \, \mathrm{d}s. \tag{4.24}$$

We have

$$A(\gamma, s)V_s' - A(\gamma_n, s_n)V_n'(s_n) = A(\gamma, s)\big(V_s' - V_n'(s_n)\big) + \big(A(\gamma, s) - A(\gamma, s_n)\big)V_n'(s_n) + \big(A(\gamma, s_n) - A(\gamma_n, s_n)\big)V_n'(s_n).$$

Then

$$\begin{aligned} &\big|A(\gamma, s)V_s' - A(\gamma_n, s_n)V_n'(s_n)\big|_H \\ &\quad \leqslant \big\|A(\gamma, s)\big\|_{\mathrm{op}}\big|V_s' - V_n'(s_n)\big|_H + \big\|A(\gamma, s) - A(\gamma, s_n)\big\|_{\mathrm{op}}\big|V_n'(s_n)\big|_H \\ &\qquad + \big\|A(\gamma, s_n) - A(\gamma_n, s_n)\big\|_{\mathrm{op}}\big|V_n'(s_n)\big|_H. \end{aligned} \tag{4.25}$$

Now for $s \in [l2^{-n}, (l+1)2^{-n}[$, take $h \in H$,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\theta}\big(A(\gamma, s)h\big) - \frac{\mathrm{d}}{\mathrm{d}\theta}\big(A\big(\gamma, l2^{-n}\big)h\big) \\ &\quad = (\mathrm{Ad}_{\gamma_s^{-1}(\theta)} - \mathrm{Ad}_{\gamma_{l2^{-n}}^{-1}(\theta)})\frac{\mathrm{d}h}{\mathrm{d}\theta} \\ &\qquad - \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta} G_\sigma(\theta)\bigg\{(\mathrm{Ad}_{\gamma_s^{-1}(\theta)} - \mathrm{Ad}_{\gamma_{l2^{-n}}^{-1}(\theta)})\bigg[\frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma_s(\sigma)\gamma_s^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\bigg]\bigg\} \, \mathrm{d}\sigma \\ &\qquad - \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta} G_\sigma(\theta) \, \mathrm{Ad}_{\gamma_{l2^{-n}}^{-1}(\theta)}\bigg[\frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma_s(\sigma)\gamma_s^{-1}(\sigma) - \frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma_{l2^{-n}}(\sigma)\gamma_{l2^{-n}}^{-1}(\sigma), \frac{\mathrm{d}h}{\mathrm{d}\sigma}\bigg] \, \mathrm{d}\sigma. \end{aligned}$$

It follows that for $s \in [l2^{-n}, (l+1)2^{-n}[$,

$$\big|A(\gamma, s)h - A\big(\gamma, l2^{-n}\big)h\big|_H \leqslant C d_{\mathcal{P},\infty}(\gamma_s, \gamma_{l2^{-n}})|h|_H + 2C d_{\mathcal{P},\infty}(\gamma_s, \gamma_{l2^{-n}}) d_{\mathcal{P}}(\gamma_s, \mathbf{e})|h|_H + R_n(s),$$

where

$$
\begin{aligned}
R_n(s) &= 2\int_0^1 \left|\frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma_s(\sigma)\gamma_s^{-1}(\sigma) - \frac{\mathrm{d}}{\mathrm{d}\sigma}\gamma_{l2^{-n}}(\sigma)\gamma_{l2^{-n}}^{-1}(\sigma)\right|_{\mathcal{G}} \left|\frac{\mathrm{d}}{\mathrm{d}\sigma}h\right|_{\mathcal{G}} \mathrm{d}\sigma \\
&\leqslant 2\int_0^1 \left(\int_{l2^{-n}}^s \left|\frac{\mathrm{d}}{\mathrm{d}\sigma}\big(A(\gamma,u)V_u'\big)\right|_{\mathcal{G}} \mathrm{d}u\right)\left|\frac{\mathrm{d}h}{\mathrm{d}\sigma}\right|_{\mathcal{G}} \mathrm{d}\sigma \\
&\leqslant 2\int_{l2^{-n}}^s \big|A(\gamma,u)V_u'\big|_H |h|_H \,\mathrm{d}u \\
&\leqslant 2K_1|h|_H \int_{l2^{-n}}^s \big|V_n'\big|_H \,\mathrm{d}u \\
&\leqslant 2K_1|h|_H \cdot 2^{-n/2}\left(\int_0^1 \big|V_s'\big|_H^2 \,\mathrm{d}s\right)^{1/2}.
\end{aligned}
$$

Since $\sup_s d_{\mathcal{P},\infty}(\gamma_s,\gamma_{s_n}) \to 0$ as $n \to +\infty$, there exists $\beta_n \to 0$ such that

$$
\sup_s \big\|A(\gamma,s) - A(\gamma,s_n)\big\|_{\mathrm{op}} \leqslant \beta_n \to 0.
$$

Returning back to (4.25), we get

$$
\begin{aligned}
\big|A(\gamma,s)V_s' - A(\gamma_n,s_n)V_n'(s_n)\big|_H \leqslant{}& K_1\big|V_s' - V_n'(s_n)\big|_H + 2CK_1 d_{\mathcal{P},\infty}\big(\gamma_{s_n},\gamma_n(s_n)\big)\big|V_n'(s_n)\big|_H \\
&+ 2d_{\mathcal{P}}\big(\gamma_{s_n},\gamma_n(s_n)\big)\big|V_n'(s_n)\big|_H + \beta_n\big|V_n'(s_n)\big|_H.
\end{aligned}
$$

By (4.24),

$$
\begin{aligned}
d_{\mathcal{P}}^2\big(\gamma_t,\gamma_n(t)\big) \leqslant{}& 2CK_1\left(\int_0^t d_{\mathcal{P},\infty}^2\big(\gamma_s,\gamma_n(s)\big)\,\mathrm{d}s\right)K_2 + 2K_1^2\left(\int_0^1 \big|V_s' - V_n'(s_n)\big|_H^2 \,\mathrm{d}s\right) \\
&+ 8C^2K_1^2\left(\int_0^t d_{\mathcal{P},\infty}^2\big(\gamma_{s_n},\gamma_n(s_n)\big)\,\mathrm{d}s\right)K_2 + 8\left(\int_0^t d_{\mathcal{P}}^2\big(\gamma_{s_n},\gamma_n(s_n)\big)\,\mathrm{d}s\right)K_2 \\
&+ \beta_n^2 K_2,
\end{aligned}
$$

where $K_2 = \int_0^1 |V_s'|_H^2 \,\mathrm{d}s$ and $\int_0^1 |V_n'(s)|_H^2 \,\mathrm{d}s \leqslant \int_0^1 |V_s'|_H^2 \,\mathrm{d}s = K_2$. It follows that

$$
\begin{aligned}
d_{\mathcal{P}}^2\big(\gamma_t,\gamma_n(t)\big) \leqslant{}& \beta_n^2 K_2 + 2K_1^2\left(\int_0^1 \big|V_s' - V_n'(s)\big|_H^2 \,\mathrm{d}s\right) + 2CK_1^2K_2\int_0^t d_{\mathcal{P}}^2\big(\gamma_s,\gamma_n(s)\big)\,\mathrm{d}s \\
&+ \big(8C^2K_1^2K_2 + 8K_2\big)\int_0^t d_{\mathcal{P}}^2\big(\gamma_{s_n},\gamma_n(s_n)\big)\,\mathrm{d}s
\end{aligned}
$$

from which we deduce that

$$\sup_{0\leqslant t\leqslant 1} d_{\mathcal{P}}^2\big(\gamma_t, \gamma_n(t)\big) \leqslant \alpha_n \cdot e^K \to 0 \quad \text{as } n \to +\infty. \qquad \square$$

Proof of Theorem 1.1. Just as we have done in the case of Lie group, we take $\{\beta_n,\ n \geqslant 1\}$ and $\{c_n,\ n \geqslant 1\}$ as in the proof of Theorem 2.2. Let $\{h_i,\ i \geqslant 1\} \subset D$ be a basis of $H_0(\mathcal{G})$. Let $U_t = \int_0^t V_s\,\mathrm{d}s$, and rewrite (4.19) as

$$\big\langle V_1, c_n'(1)h_i\big\rangle_{H_0} = \int_0^1 \big\langle U_t', \beta_n'(t)h_i\big\rangle_{H_0}\,\mathrm{d}t = \langle U, \beta_n h_i\rangle_{H(H_0(\mathcal{G}))}.$$

Therefore,

$$U = \sum_{n\geqslant 1}\sum_{i\geqslant 1}\big\langle V_1, c_n' h_i\big\rangle_{H_0} \beta_n h_i \in H\big(H_0(\mathcal{G})\big)$$

and

$$U_t = \sum_{n\geqslant 1}\sum_{i\geqslant 1}\big\langle V_1, c_n' h_i\big\rangle_{H_0} \beta_n(t) h_i \in H_0(\mathcal{G}).$$

Since $V_1 = -\nabla\phi(\ell_1)$ is measurable function of ℓ_1, we see for each $t \in [0, 1]$, U_t is measurable, so does V_t. By (4.20), we see that $\gamma_n(t)$ is measurable for each t. According to Theorem 4.11, $\gamma_n(t)$ converges to $\gamma(t)$, hence $\gamma(t)$ is measurable for each $t \in [0, 1]$. Therefore, $\mathcal{T}(\ell_1) = \ell_1\gamma(1)$ is a measurable function of ℓ_1 and Υ_0 is supported by the graph $\{(\ell_1, \mathcal{T}(\ell_1));\ \ell_1 \in \mathcal{L}_e G\}$. It follows that $\Upsilon_0 = (\mathrm{Id} \times \mathcal{T})_{\#}\nu$ and $\mathcal{T}_{\#}\nu = F\nu$. Furthermore, this result implies also the uniqueness of the mapping $\mathcal{T}$. The proof of Theorem 1.1 is complete now. $\square$

Acknowledgment

The authors are grateful to Professor Cédric Villani for his careful reading and useful suggestions.

References

[1] L. Ambrosio, Optimal transport maps in Monge–Kantorovich problem, in: Lecture Note in ICM, vol. III, 2002, pp. 131–140.
[2] S. Bobkov, I. Gentil, M. Ledoux, Hypercontractivity of Hamilton–Jacobi equations, J. Math. Pures Appl. 80 (2001) 669–696.
[3] Y. Brenier, Décomposition polaire et réarrangement monotone des champs de vecteurs, C. R. Acad. Sci. Paris Sér. I Math. 305 (1987) 805–808.
[4] Y. Brenier, Polar factorization and monotone rearrangement of vector valued functions, Comm. Pure Appl. Math. 44 (1991) 375–417.
[5] B. Driver, Integration by parts and quasi-invariance for heat measures on loop groups, J. Funct. Anal. 149 (1997) 470–547.
[6] L.C. Evans, W. Gangbo, Differential equations methods for the Monge–Kantorovich mass transfer problem, Mem. Amer. Math. Soc. 137 (1999) 1–66.
[7] S. Fang, Integration by parts for heat measures over loop groups, J. Math. Pures Appl. 78 (1999) 877–894.

[8] S. Fang, J. Shao, Transportation cost inequalities on path and loop groups, J. Funct. Anal. 218 (2005) 293–317.
[9] S. Fang, J. Shao, Distance riemannienne, théorème de Rademacher et inégalité de transport sur le groupe des lacets, C. R. Math. Acad. Sci. Paris 341 (2005) 445–450.
[10] M. Feldma, R.J. McCann, Monge's transport problem on a Riemannian manifold, Trans. Amer. Math. Soc. 354 (2002) 1667–1697.
[11] D. Feyel, A. Üstünel, Measure transport on Wiener space and the Girsanov theorem, C. R. Math. Acad. Sci. Paris 334 (2002) 1025–1028.
[12] M. Hino, J.A. Ramirez, Small time Gaussian behavior of symmetric diffusion semigroups, Ann. Probab. 31 (2003) 1254–1295.
[13] M. Knott, C.S. Smith, On the optimal mapping of distributions, J. Optim. Theory Appl. 43 (1) (1984) 39–49.
[14] X.D. Li, Existence uniqueness of geodesics on path spaces, J. Funct. Anal. 173 (1) (2000) 182–202.
[15] J. Lott, C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math., in press.
[16] P. Malliavin, Hypoellipticity in infinite dimension, in: M. Pinsky (Ed.), Diffusion Processes and Related Problems in Analysis, Birkhäuser, Boston, MA, pp. 17–33.
[17] R.J. McCann, Polar factorization of maps on Riemannian manifolds, Geom. Funct. Anal. 11 (3) (2001) 589–608.
[18] S.T. Rachev, L. Rüschendorf, Mass Transportation Problems, Probab. Appl., Springer-Verlag, New York, 1998.
[19] J. Shao, Hamilton–Jacobi semi-groups in infinite-dimensional spaces, Bull. Sci. Math., in press.
[20] K.T. Sturm, On the geometry of metric measure spaces. I, Acta Math. 196 (1) (2006) 65–131.
[21] K.T. Sturm, On the geometry of metric measure spaces. II, Acta Math. 196 (1) (2006) 133–177.
[22] C. Villani, Topics in Mass Transportation, Amer. Math. Soc., Providence, RI, 2003.
[23] M. Talagrand, Transportation cost for Gaussian and other product measures, Geom. Funct. Anal. 6 (1996) 587–600.
[24] F.Y. Wang, Probability distance inequalities on Riemannian manifolds and path spaces, J. Funct. Anal. 206 (2004) 167–190.