UNIFORM CONVERGENCE IN VON NEUMANN'S ERGODIC THEOREM IN THE ABSENCE OF A SPECTRAL GAP

JONATHAN BEN-ARTZI AND BAPTISTE MORISSE

Abstract. Von Neumann's original proof of the ergodic theorem is revisited. A uniform convergence rate is established under the assumption that one can control the density of the spectrum of the underlying self-adjoint operator when restricted to suitable subspaces. Explicit rates are obtained when the bound is polynomial, with applications to the linear Schrödinger and wave equations. In particular, decay estimates for time-averages of solutions are shown.

MSC (2010): 37A30 (primary); 37A25, 37A10, 37C40, 35Q41, 35L05

Keywords: Von Neumann's ergodic theorem; convergence rates; density of states

Acknowledgements: The authors acknowledge support from Fellowship EP/N020154/1 of the Engineering and Physical Sciences Research Council (EPSRC). The authors are grateful to the anonymous referee who read the manuscript carefully and helped to significantly improve it.

Date: March 2, 2020.

Contents

1. Introduction 1
2. Proof of the main result 6
3. Examples 7
References 12

1. Introduction

In this note we obtain uniform rates of convergence in von Neumann's ergodic theorem for continuous time dynamical systems lacking a spectral gap. It is well-known that a spectral gap leads to the rate T^{-1}, however systems that lack a spectral gap are not as well understood. The key new ingredient is an estimate of the *Density of States* (DoS) of the generator near 0. Such estimates are readily available for differential operators, as is shown in the examples in Section 3.

Our setup is as follows. Let $\mathcal{H}$ be a separable Hilbert space and let $U_t : \mathcal{H} \to \mathcal{H}$ be a one-parameter group of unitary transformations. Let $H : D(H) \subset \mathcal{H} \to \mathcal{H}$ be its self-adjoint

generator: $U_t = e^{itH}$. We assume that H has continuous spectrum in a neighborhood of 0 (and 0 itself is often an eigenvalue) and show that a bound on the DoS near 0 leads to a uniform convergence rate on a suitable subspace $\mathcal{X} \subset \mathcal{H}$. We apply this to the linear Schrödinger and wave equations, to obtain various decay estimates depending on the space $\mathcal{X}$ (which is taken to either be a weighted-L^2 space, or the space $L^1 \cap L^2$) which consequently lead to global-in-time results, see (3.5), (3.8), (3.12), (3.14).

1.1. **Von Neumann's ergodic theorem.** Von Neumann's ergodic theorem [19] is a pillar of modern mathematics. Defining

$$P^T := \frac{1}{2T} \int_{-T}^{T} U_t \, \mathrm{d}t,$$

and

$$P := \text{orthogonal projection of } \mathcal{H} \text{ onto } \ker H,$$

it states

Theorem 1.1 (Ergodic theorem [19]). *For any $f \in \mathcal{H}$, $P^T f \to Pf$ as $T \to +\infty$.*

Sketch of proof. The original proof relies on Stone's theorem (and the spectral theorem, by proxy), i.e. the fact that U_t has a resolution of the identity $\{E(\lambda)\}_{\lambda \in \mathbb{R}}$ (for a precise definition see Definition 1.3 below) for which $U_t = \int_{\mathbb{R}} e^{it\lambda} \mathrm{d}E(\lambda)$. This leads to:

$$\begin{aligned}(P^T - P)f &= \frac{1}{2T} \int_{-T}^{T} U_t f \, \mathrm{d}t - Pf = \frac{1}{2T} \int_{-T}^{T} \int_{\mathbb{R}} e^{it\lambda} \, \mathrm{d}E(\lambda) f \, \mathrm{d}t - Pf \\ &= \frac{1}{2T} \int_{-T}^{T} \int_{\mathbb{R}\setminus\{0\}} e^{it\lambda} \, \mathrm{d}E(\lambda) f \, \mathrm{d}t = \int_{\mathbb{R}\setminus\{0\}} \frac{\sin T\lambda}{T\lambda} \, \mathrm{d}E(\lambda) f. \end{aligned} \tag{1.1}$$

This last expression tends to 0 as $T \to +\infty$. □

The strong convergence $P^T \to P$ can be improved to uniform convergence if H has a spectral gap:

Theorem 1.2 (Ergodic theorem: case of spectral gap). *Assume that there exists $\gamma > 0$ such that $\sigma(H) \subset (-\infty, -\gamma] \cup \{0\} \cup [\gamma, +\infty)$. Then*

$$\|P^T - P\|_{\mathcal{H} \to \mathcal{H}} \leq \gamma^{-1} T^{-1}, \qquad \forall T > 1. \tag{1.2}$$

Proof. See Remark 2.2 below. □

1.2. **The spectral theorem.** Since the spectral theorem and the resolution of the identity of self-adjoint operators play a central role in this paper, we recall some basic facts.

Definition 1.3 (Resolution of the identity). *Let $H : D(H) \subset \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator. Its associated* resolution of the identity *$\{E(\lambda)\}_{\lambda\in\mathbb{R}}$ is a family of projection operators in $\mathcal{H}$ with the property that, for each $\lambda \in \mathbb{R}$, the subspace $\mathcal{H}^\lambda = E(\lambda)\mathcal{H}$ is the largest closed subspace such that*

i. $\mathcal{H}^\lambda$ reduces H, *namely, $HE(\lambda)g = E(\lambda)Hg$ for every $g \in D(H)$. In particular, if $g \in D(H)$ then also $E(\lambda)g \in D(H)$.*

ii. $(Hu, u)_{\mathcal{H}} \leq \lambda(u, u)_{\mathcal{H}}$ *for every $u \in \mathcal{H}^\lambda \cap D(H)$.*

Now we are able to state the spectral theorem:

Theorem 1.4 (Spectral theorem). *Let $H : D(H) \subset \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator and let $\{E(\lambda)\}_{\lambda\in\mathbb{R}}$ be the associated resolution of the identity. Then $\{E(\lambda)\}_{\lambda\in\mathbb{R}}$ is unique, and the identity $H = \int_{\mathbb{R}} \lambda \,\mathrm{d}E(\lambda)$ holds.*

In addition to the above, it is useful to state the definition of the spectral measure:

Definition 1.5 (Spectral measure). *Given any $f, g \in \mathcal{H}$ the resolution of the identity defines a complex function of bounded variation on the real line, given by*

$$\mathbb{R} \ni \lambda \mapsto (E(\lambda)f, g)_{\mathcal{H}}.$$

This function gives rise to a complex measure (depending on f, g) called the spectral measure.

Definition 1.6 (Density of states). *Let $\mathcal{X} \subset \mathcal{H}$ be some closed subspace. We call the bilinear form*

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\,(E(\lambda)\cdot, \cdot)_{\mathcal{H}} : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$$

the density of states of H at λ on the subspace $\mathcal{X}$.

1.3. **Main results.** As mentioned above, we assume the opposite of a spectral gap: we assume that $\sigma(H)$ contains a neighborhood of 0. However, we do not want to have "too much" spectrum near 0. We make this precise as follows. Let $\{E(\lambda)\}_{\lambda\in\mathbb{R}}$ be the resolution of the identity of H. Our main assumption is:

Assumption A1. There exist

i. a Banach subspace $\mathcal{X} \subset \mathcal{H}$ which is dense in $\mathcal{H}$ in the topology of $\mathcal{H}$, is continuously embedded in $\mathcal{H}$ (and therefore the norm $\|\cdot\|_{\mathcal{X}}$ is stronger than the norm $\|\cdot\|_{\mathcal{H}}$),

ii. a real number $r \in (0,1)$ and a function $\psi : [-r,r] \to \mathbb{R}$ that is strictly positive a.e. on $I_r = [-r,r]$ such that the following bound of the DoS of H holds:

$$\left| \frac{\mathrm{d}}{\mathrm{d}\lambda} (E(\lambda)f,g)_{\mathcal{H}} \right| \leq \psi(\lambda)\|f\|_{\mathcal{X}}\|g\|_{\mathcal{X}}, \qquad \forall f,g \in \mathcal{X},\, \forall \lambda \in I_r \setminus \{0\}, \tag{1.3}$$

iii. a real number $q > 0$ such that $|\lambda|^{-q}\psi(\lambda) \in L^1(I_r)$.

Our main result, to be proved in Section 2, is:

Theorem 1.7. *Under Assumption A1, letting $\ell = \min\{q,2\}$, the following uniform rate in von Neumann's ergodic theorem holds:*

$$\|P^T - P\|_{\mathcal{X}\to\mathcal{H}} \leq \frac{C}{T^{\ell/2}}, \qquad \forall T > 1, \tag{1.4}$$

where C is an explicit constant that does not depend on T.

When $\psi(\lambda)$ is a power of λ we immediately have:

Corollary 1.8. *If ψ has the form $\psi(\lambda) = c|\lambda|^{p-1}$ for some $c,p > 0$, then* (1.4) *holds with $\ell = \min\{p-\varepsilon, 2\}$ for any $\varepsilon > 0$.*

Proof of Corollary 1.8. Considering Assumption 1, if $\psi(\lambda) = c|\lambda|^{p-1}$, in order for $|\lambda|^{-q}\psi(\lambda)$ to be integrable in a neighborhood of 0, we need $q < p$. The assertion follows from applying Theorem 1.7. □

Remark 1.9. We note that the best rate that one could expect is T^{-1}, as in the case of a spectral gap (see [4, Remark 3] for a detailed proof). Hence, in Corollary 1.8 the rate cannot be improved beyond T^{-1} even if $p > 2$.

Remark 1.10. As we show in the examples in Section 3, there is a delicate interplay between the choice of subspace $\mathcal{X} \subset \mathcal{H}$ and the bound $\psi(\lambda)$ one can obtain for the DoS, both appearing in (1.3):

- In the case of the linear Schrödinger equation in $\mathbb{R}^d$, taking $\mathcal{X} = L^{2,s}$ (a weighted L^2 space) leads to $\psi(\lambda) \sim \lambda^{-1/2}$ (cf. (3.4)), while taking $\mathcal{X} = L^1 \cap L^2$ leads to $\psi(\lambda) \sim |\lambda|^{\frac{d}{2}-1}$ (cf. (3.7))
- In the case of the linear wave equation in $\mathbb{R}^d$, taking $\mathcal{X} = L^{2,s}$ leads to $\psi(\lambda) \sim 1$ (cf. (3.10)), while taking $\mathcal{X} = L^1 \cap L^2$ leads to $\psi(\lambda) \sim |\lambda|^{d-1}$ (cf. (3.13)).

Remark 1.11. In view of [4] (see discussion of previous results and Theorem 1.12 below) these results are *nearly* optimal, and might in fact be optimal. Establishing whether or not this is the case is the subject of ongoing research.

1.4. **Previous results.** The idea that the spectrum encodes information about the dynamical system goes all the way back to von Neumann and his original proof of the mean ergodic theorem [19]. Soon after, Riesz [16] provided an alternative proof which does not rely on the spectral theorem and is considered simpler. As a result, in most textbooks the spectral approach appears to have been lost.

Kachurovskii and coauthors have published extensively over the last 20 years on the topic of rates of convergence in ergodic theorems, by revisiting von Neumann's original ideas, relying on the spectral theorem [4, 6, 7, 12–14]. We also mention [1] where some of the techniques of [6] were simplified. The survey [10] provides a detailed overview of this sequence of results. Closest in spirit to our result is [4], where the authors show:

Theorem 1.12 ([4]). *For any function $f \in \mathcal{H}$ and $p \in [0, 2)$, the two following conditions are equivalent:*

i. $((E(\lambda) - E(-\lambda) - E(\{0\}))\, f, f)_{\mathcal{H}} \leq A\lambda^p$, *for all* $\lambda > 0$,

ii. $\|(P^T - P)f\|_{\mathcal{H}} \leq \frac{B}{T^{p/2}} \|f\|_{\mathcal{H}}$, *for all* $T > 0$,

where given $A > 0$ one can compute $B = B(A, p) > 0$ explicitly, and, conversely, given $B > 0$ one can compute $A = A(B, p) > 0$ explicitly (and these dependencies do not involve the function f).

Another important direction of research relates to the appearance of the *Fejér kernel* $\mathrm{sinc}^2 T\lambda = \frac{\sin^2 T\lambda}{T^2\lambda^2}$ in the expression for the square of $P^T - P$ (see (1.1)). This is addressed in [8, 9, 11], for example.

Finally, it is natural to compare our result to the well-known RAGE theorem (see e.g. [3]). It states that for any compact operator K and any $f \in \mathcal{H}$,

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T \|KU_t P_{\mathrm{ac}} f\|^2 \, \mathrm{d}t = 0,$$

where P_{ac} is the orthogonal projection onto the absolutely continuous subspace of H. Where Theorem 1.7 proves uniform convergence to the projection onto the kernel of the generator, the RAGE theorem proves a weak convergence to 0 of the time average of the evolution of the continuous part of the spectrum.

Organization of the paper. Section 2 is devoted to the proof of the main theorem, Theorem 1.7. In Section 3 we apply this to the linear Schrödinger and wave equations and obtain decay estimates for averages of solutions.

2. Proof of the main result

Proof of Theorem 1.7. Our starting point is the observation [15, V-§2.1] that if the bilinear form $\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(E(\lambda)\cdot,\cdot\right)_{\mathcal{H}} : \mathcal{X}\times\mathcal{X}\to\mathbb{C}$ is bounded at a given $\lambda\in\mathbb{R}$, then there exists a bounded operator $A(\lambda):\mathcal{X}\to\mathcal{X}^*$ satisfying

$$\langle A(\lambda)f,g\rangle = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(E(\lambda)f,g\right)_{\mathcal{H}}, \qquad \forall f,g\in\mathcal{X},$$

where $\langle\cdot,\cdot\rangle$ is the $(\mathcal{X}^*,\mathcal{X})$ dual space pairing. Moreover, the operator norm of $A(\lambda)$ shares the same bound as the bilinear form. Now, recalling von Neumann's proof as sketched in (1.1) above, we have

$$(P^T-P)f = \int_{\mathbb{R}\setminus\{0\}} \operatorname{sinc}(T\lambda)\,\mathrm{d}E(\lambda)f$$

where $\operatorname{sinc} x := \sin x/x$. We split this integral as follows:

$$\int_{\mathbb{R}\setminus\{0\}} \operatorname{sinc}(T\lambda)\,\mathrm{d}E(\lambda)f = \left(\int_{I_r\setminus\{0\}} + \int_{I_r^c}\right)\operatorname{sinc}(T\lambda)\,\mathrm{d}E(\lambda)f$$

where $I_r=[-r,r]$. We start by estimating the second integral ("high frequency" part), using the fact that $\operatorname{sinc}^2 x\le |x|^{-2}$ and that projections onto different spectral parameters are mutually orthogonal:

$$\begin{aligned}\left\|\int_{I_r^c}\operatorname{sinc}(T\lambda)\,\mathrm{d}E(\lambda)f\right\|_{\mathcal{H}}^2 &= \int_{I_r^c}\operatorname{sinc}^2(T\lambda)\,\mathrm{d}(E(\lambda)f,f)_{\mathcal{H}}\\ &\le \frac{1}{T^2r^2}\int_{\mathbb{R}}\mathrm{d}(E(\lambda)f,f)_{\mathcal{H}} = \frac{1}{T^2r^2}\|f\|_{\mathcal{H}}^2 \le \frac{1}{T^2r^2}\|f\|_{\mathcal{X}}^2.\end{aligned}$$

Now we turn to the first integral ("low frequency" part). We use the estimate (1.3) of the DoS, as well as the boundedness of $|x|^\ell\operatorname{sinc}^2 x$ for $\ell\in[0,2]$:

$$\begin{aligned}\left\|\int_{I_r\setminus\{0\}}\operatorname{sinc}(T\lambda)\,\mathrm{d}E(\lambda)f\right\|_{\mathcal{H}}^2 &= \int_{I_r\setminus\{0\}}\operatorname{sinc}^2(T\lambda)\,\mathrm{d}(E(\lambda)f,f)_{\mathcal{H}}\\ &= \int_{I_r}\operatorname{sinc}^2(T\lambda)\,\langle A(\lambda)f,f\rangle\,\mathrm{d}\lambda\\ &\le \left(\int_{I_r}\operatorname{sinc}^2(T\lambda)\psi(\lambda)\,\mathrm{d}\lambda\right)\|f\|_{\mathcal{X}}^2\\ &\le \sup_{\lambda\in I_r}\left(|\lambda|^\ell\operatorname{sinc}^2(T\lambda)\right)\left(\int_{I_r}|\lambda|^{-\ell}\psi(\lambda)\,\mathrm{d}\lambda\right)\|f\|_{\mathcal{X}}^2.\end{aligned}$$

Letting $\ell=\min\{q,2\}$, and using the fact that $|\lambda|^{-\ell}\le|\lambda|^{-q}$ for $|\lambda|\le r<1$, there holds

$$\begin{aligned}\left\|\int_{I_r\setminus\{0\}}\operatorname{sinc}(T\lambda)\,\mathrm{d}E(\lambda)f\right\|_{\mathcal{H}}^2 &\le \sup_{\lambda\in I_r}\left(|\lambda|^\ell\operatorname{sinc}^2(T\lambda)\right)\left(\int_{I_r}|\lambda|^{-q}\psi(\lambda)\,\mathrm{d}\lambda\right)\|f\|_{\mathcal{X}}^2\\ &= \frac{1}{T^\ell}\sup_{y\in I_{rT}}\left(|y|^\ell\operatorname{sinc}^2 y\right)\Psi_q(r)\|f\|_{\mathcal{X}}^2 \le \frac{1}{T^\ell}\Psi_q(r)\|f\|_{\mathcal{X}}^2\end{aligned}$$

where we have denoted $\Psi_q(r) := \int_{I_r} |\lambda|^{-q}\psi(\lambda)\,\mathrm{d}\lambda$. Altogether, both estimates lead to

$$\|(P^T - P)f\|_{\mathcal{H}}^2 \leq \frac{1}{T^\ell}\left(\Psi_q(r) + \frac{1}{T^{2-\ell}r^2}\right)\|f\|_{\mathcal{X}}^2 \tag{2.1}$$

which completes the proof. □

Remark 2.1 (The constant in (2.1)). The constant in (2.1) is indeed uniformly bounded independent of T:

$$C = \Psi_q(r) + \frac{1}{T^{2-\ell}r^2} \leq \Psi_q(r) + \frac{1}{r^2}, \qquad \forall T > 1.$$

In fact, it even decreases slowly with T and one could attempt to optimize it by letting r tend to 0 at an appropriate rate. However, there is nothing to be gained by doing this. This is simply an artefact due to the condition $|\lambda|^{-q}\psi(\lambda) \in L^1(I_r)$ in Assumption 1: this is an open condition, in the sense that one could always increase q slightly and this condition will still hold. Doing this will cause C to converge to a constant independent of T.

Remark 2.2 (Spectral gap). In the case of a spectral gap (1.2) immediately follows. Indeed, with gap of size γ in the above proof ψ is trivial and one has

$$\|(P^T - P)f\|_{\mathcal{H}} \leq \gamma^{-1}T^{-1}\|f\|_{\mathcal{H}}.$$

Note that in this case the subspace $\mathcal{X}$ is no longer needed.

3. Examples

3.1. **The Laplace operator.** Let $\varphi : [0, +\infty) \to [0, +\infty)$ be some continuous and strictly increasing function and define $H = \varphi(-\Delta)$ as a function of the Laplace operator acting in $\mathcal{H} = L^2(\mathbb{R}^d)$ with an appropriate domain for self-adjointness. Note that if $\varphi(x) = x$ is the identity, then $-iH$ is the generator of the *Schrödinger equation*:

$$\begin{cases} \partial_t f(t,x) = i\Delta f(t,x), & t \in \mathbb{R},\, x \in \mathbb{R}^d, \\ f(0,x) = f_0(x), & x \in \mathbb{R}^d. \end{cases}$$

Let $\{E(\lambda)\}_{\lambda\in\mathbb{R}}$ be the resolution of the identity of H. We use the fact that the Fourier transform is a unitary map relating $-\Delta$ to multiplication by $|\xi|^2$ in order to get:

$$(E(\lambda)f, g)_{\mathcal{H}} = \int_{\varphi(|\xi|^2)\leq\lambda} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}\,\mathrm{d}\xi, \qquad \lambda \geq 0. \tag{3.1}$$

Let us show how different choices of subspaces $\mathcal{X}$ in the DoS estimate (1.3) can give rise to different results.

3.1.1. *Hilbertian subspace.* Differentiating (3.1) in λ we get

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\Big|_{\lambda=\lambda_0}(E(\lambda)f,g)_{\mathcal{H}} = \int_{|\xi|=\sqrt{\varphi^{-1}(\lambda_0)}} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}\left|\nabla\left(\varphi(|\xi|^2)\right)\right|^{-1}\mathrm{d}\sigma \tag{3.2}$$

where $\mathrm{d}\sigma$ is the Lebesgue (uniform) surface measure on the $d-1$-dimensional sphere of radius $\sqrt{\varphi^{-1}(\lambda_0)}$. The term $\left|\nabla\left(\varphi(|\xi|^2)\right)\right|^{-1} = \frac{1}{2|\xi|\varphi'(|\xi|^2)}$ comes from the coarea formula [5, Appendix C3]. An evaluation of the L^2 functions $\widehat{f}$ and $\widehat{g}$ on the hypersurface $\{|\xi| = \sqrt{\varphi^{-1}(\lambda_0)}\}$ only makes sense if they belong to any Sobolev space $H^s(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$ with $s > 1/2$ by the trace lemma[1]. The functions $\widehat{f}$ and $\widehat{g}$ belong to $H^s(\mathbb{R}^d)$ if and only if f and g belong to $L^{2,s}(\mathbb{R}^d)$, defined as

$$L^{2,s}(\mathbb{R}^d) := \left\{ f \in L^2(\mathbb{R}^d) \ : \ \|f\|^2_{L^{2,s}(\mathbb{R}^d)} := \int_{\mathbb{R}^d} |f(x)|^2(1+|x|^2)^s\,\mathrm{d}x < +\infty \right\}.$$

We therefore conclude that we can bound (3.2) using the $L^{2,s}$-norms of f and g, which are stronger than their $\mathcal{H}$-norms:

$$\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\Big|_{\lambda=\lambda_0}(E(\lambda)f,g)_{\mathcal{H}}\right| \le \frac{1}{2\sqrt{\varphi^{-1}(\lambda_0)}\varphi'(\varphi^{-1}(\lambda_0))}\|f\|_{L^{2,s}}\|g\|_{L^{2,s}}, \tag{3.3}$$

where we denote $L^{2,s}$ rather than $L^{2,s}(\mathbb{R}^d)$ for brevity. Hence, using the notation of (1.3), we have that

$$\mathcal{X} = L^{2,s}(\mathbb{R}^d)$$
$$\psi(\lambda) = \frac{1}{2\sqrt{\varphi^{-1}(\lambda)}\varphi'(\varphi^{-1}(\lambda))}.$$

In the case of the Schrödinger equation ($\varphi = \mathrm{id}$) we get

$$\psi(\lambda) = \frac{1}{2\sqrt{\lambda}} \tag{3.4}$$

and from Corollary 1.8 we get a convergence rate of $T^{-\ell/2}$ for all $\ell < \frac{1}{2}$. Moreover, since $-\Delta$ has no eigenvalues in this setting (and, in particular, a trivial kernel), we conclude that

$$\left\|\frac{1}{2T}\int_{-T}^{T} e^{-it\Delta}f_0\,\mathrm{d}t\right\|_{L^2} = \|P^T f_0\|_{L^2} \le CT^{-\ell/2}\|f_0\|_{L^{2,s}}, \qquad \forall \ell < \frac{1}{2}.$$

This also implies that

$$\|P^T f_0\|_{L^q_T L^2_x([0,\infty)\times\mathbb{R}^d)} \le C(q)\|f_0\|_{L^{2,s}_x(\mathbb{R}^d)}, \qquad \forall q > 4. \tag{3.5}$$

Restricting to any bounded domain $\Omega \subset \mathbb{R}^d$ we may take L^2 norms rather than weighted norms (the weight is uniformly bounded away from 0 and $+\infty$ in Ω) so we have

$$\|P^T f_0\|_{L^q_T L^2_x([0,\infty)\times\Omega)} \le C(q,\Omega)\|f_0\|_{L^2_x(\Omega)}, \qquad \forall q > 4.$$

[1]This is not entirely optimal, since we are not making use of the fact that this hypersurface is in fact a sphere.

3.1.2. *Non-Hilbertian subspace.* Considering (3.2) again, we may change variables so that the integration takes place on the unit sphere in $\mathbb{R}^d$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\Big|_{\lambda=\lambda_0}(E(\lambda)f,g)_{\mathcal{H}} &= \int_{|\xi|=\sqrt{\varphi^{-1}(\lambda_0)}} \widehat{f}(\xi)\overline{\widehat{g}(\xi)}\left|\nabla\left(\varphi(|\xi|^2)\right)\right|^{-1}\mathrm{d}\sigma \\ &= \sqrt{\varphi^{-1}(\lambda_0)}^{d-1}\int_{\mathbb{S}^{d-1}} \frac{\widehat{f}\left(\tau\sqrt{\varphi^{-1}(\lambda_0)}\right)\overline{\widehat{g}\left(\tau\sqrt{\varphi^{-1}(\lambda_0)}\right)}}{2\sqrt{\varphi^{-1}(\lambda_0)}\varphi'(\varphi^{-1}(\lambda_0))}\mathrm{d}\tau \\ &= \frac{\sqrt{\varphi^{-1}(\lambda_0)}^{d-2}}{2\varphi'(\varphi^{-1}(\lambda_0))}\int_{\mathbb{S}^{d-1}} \widehat{f}\left(\tau\sqrt{\varphi^{-1}(\lambda_0)}\right)\overline{\widehat{g}\left(\tau\sqrt{\varphi^{-1}(\lambda_0)}\right)}\mathrm{d}\tau\end{aligned}$$

where $\mathrm{d}\tau$ is the uniform measure on the unit sphere in $\mathbb{R}^d$. Another way to make sense of the restriction of L^2 functions to a hypersurface is if they are bounded, i.e. one can bound:

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\Big|_{\lambda=\lambda_0}(E(\lambda)f,g)_{\mathcal{H}}\right| &\le \frac{|\mathbb{S}^{d-1}|\sqrt{\varphi^{-1}(\lambda_0)}^{d-2}}{2\varphi'(\varphi^{-1}(\lambda_0))}\|\widehat{f}\|_{L^\infty(\mathbb{R}^d)}\|\widehat{g}\|_{L^\infty(\mathbb{R}^d)} \\ &\le \frac{|\mathbb{S}^{d-1}|\sqrt{\varphi^{-1}(\lambda_0)}^{d-2}}{2\varphi'(\varphi^{-1}(\lambda_0))}\|f\|_{L^1(\mathbb{R}^d)}\|g\|_{L^1(\mathbb{R}^d)}. \qquad (3.6)\end{aligned}$$

Thus we obtain

$$\mathcal{X} = L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d) \quad \text{with norm} \quad \|\cdot\|_{\mathcal{X}} := \|\cdot\|_{L^1(\mathbb{R}^d)} + \|\cdot\|_{L^2(\mathbb{R}^d)}$$

$$\psi(\lambda) = \frac{|\mathbb{S}^{d-1}|\sqrt{\varphi^{-1}(\lambda)}^{d-2}}{2\varphi'(\varphi^{-1}(\lambda))}.$$

We again consider the Schrödinger case ($\varphi = \mathrm{id}$) where we obtain

$$\psi(\lambda) = \frac{1}{2}|\mathbb{S}^{d-1}|\lambda^{\frac{d}{2}-1}. \qquad (3.7)$$

From Corollary 1.8 we get a convergence rate of $T^{-\ell/2}$ where $\ell = \min\{\frac{d}{2}-\varepsilon, 2\}$ for any $\varepsilon > 0$. In particular, for dimensions $d \ge 5$ we obtain the optimal rate of convergence of T^{-1}. For any $f_0 \in L^2 \cap L^1$, there holds

$$\left\|\frac{1}{2T}\int_{-T}^{T} e^{-it\Delta}f_0\,\mathrm{d}t\right\|_{L^2} = \|P^T f_0\|_{L^2} \le CT^{-\ell/2}\|f_0\|_{\mathcal{X}}, \quad \ell = \min\left\{\frac{d}{2}-\varepsilon, 2\right\}.$$

This leads to the global-in-time estimate for any $f_0 \in L^2 \cap L^1$:

$$\|P^T f_0\|_{L^q_T L^2_x([0,\infty)\times\mathbb{R}^d)} \le C(q)\|f_0\|_{\mathcal{X}}, \qquad \forall q > \max\left\{\frac{4}{d}, 1\right\}. \qquad (3.8)$$

Remark 3.1. It is natural to compare the estimates (3.5) and (3.8) with:

1) The well-known *Strichartz estimates*

$$\left\|e^{it\Delta/2}f_0\right\|_{L^q_t L^r_x(\mathbb{R}\times\mathbb{R}^d)} \le C(d,q,r)\|f_0\|_{L^2_x(\mathbb{R}^d)}$$

where $2 \le q, r \le \infty$, $\frac{2}{q} + \frac{d}{r} = \frac{d}{2}$ and $(q,r,d) \ne (2,\infty,2)$, see [17].

2) *Smoothing estimates*, such as

$$\left\||D_x|^{1/2}e^{it\Delta}f_0\right\|_{L^2_t L^{2,-s}_x(\mathbb{R}\times\mathbb{R}^d)} \le C(d)\|f_0\|_{L^2_x(\mathbb{R}^d)}$$

where $s > 1/2$, see [2]. A detailed comparison between these estimates is elusive at the present time, and is the subject of ongoing research.

3.2. **The wave operator.** We now consider the linear, homogeneous wave equation

$$\begin{cases} \partial_t^2 f(t,x) - \Delta f(t,x) = 0, & t \in \mathbb{R},\, x \in \mathbb{R}^d, \\ f(0,x) = f_0(x),\, \partial_t f(0,x) = g_0(x), & x \in \mathbb{R}^d. \end{cases}$$

We let $\mathcal{H} = L^2(\mathbb{R}^d)$ and consider the self-adjoint operator (with an appropriate domain) $H = -\Delta$. We first need to convert the above problem into a first order system. We follow a well-known procedure: define

$$f_\pm := \frac{1}{2}\left(\sqrt{H} f \pm i\partial_t f\right). \tag{3.9}$$

Then we compute

$$\partial_t f_\pm = \frac{1}{2}\left(\sqrt{H}\partial_t f \pm i\partial_t^2 f\right) = \frac{1}{2}\left(\sqrt{H}\partial_t f \mp iHf\right) = \frac{i}{2}\sqrt{H}\left(-i\partial_t f \mp \sqrt{H} f\right) = \mp i\sqrt{H} f_\pm.$$

It follows that the vector

$$F(t,x) := \begin{pmatrix} f_+(t,x) \\ f_-(t,x) \end{pmatrix}$$

satisfies the equation

$$F'(t) = -iKF \qquad \text{where} \qquad K = \begin{pmatrix} \sqrt{H} & 0 \\ 0 & -\sqrt{H} \end{pmatrix}.$$

Denoting $\{E_{\sqrt{H}}(\lambda)\}_{\lambda\in\mathbb{R}}$ and $\{E_K(\lambda)\}_{\lambda\in\mathbb{R}}$ the resolutions of the identity of $\sqrt{H}$ and K, respectively, we first observe that $E_{-\sqrt{H}}(\lambda) = I - E_{\sqrt{H}}(-\lambda)$ so that

$$E_K(\lambda) = E_{\sqrt{H}}(\lambda) \oplus (I - E_{\sqrt{H}}(-\lambda)), \qquad \forall \lambda \in \mathbb{R}.$$

As in the case of the Schrödinger equation, we may consider two cases:

3.2.1. *Hilbertian subspace.* Inserting $\varphi(H) = \sqrt{H}$ into the bound appearing in (3.3) one finds that $\psi(\lambda) = 1$ so that for $s > 1/2$,

$$\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(E_{\sqrt{H}}(\lambda) f, g\right)_{\mathcal{H}}\right| \le \|f\|_{L^{2,s}} \|g\|_{L^{2,s}}, \qquad \forall \lambda \in \mathbb{R}. \tag{3.10}$$

This implies that

$$\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(E_K(\lambda) F, G\right)_{\mathcal{H}\oplus\mathcal{H}}\right| \le \|F\|_{L^{2,s}\oplus L^{2,s}} \|G\|_{L^{2,s}\oplus L^{2,s}}, \qquad \forall \lambda \in \mathbb{R}.$$

A bound on the DoS of the form $\psi(\lambda) = 1$ leads to a convergence rate of $T^{-\ell/2}$ with $\ell < 1$ from Corollary 1.8. Therefore, noting that the kernel is empty:

$$\left\|\frac{1}{2T}\int_{-T}^{T} e^{-itK} F_0 \,\mathrm{d}t\right\|_{\mathcal{H}\oplus\mathcal{H}} \le CT^{-\ell/2} \|F_0\|_{L^{2,s}\oplus L^{2,s}}, \qquad \forall \ell < 1. \tag{3.11}$$

To obtain direct bounds for the average of the solution $f(t)$ of the wave equation, we use the identity $f = H^{-1/2}(f_+ + f_-)$ to write

$$\begin{aligned}\frac{1}{2T}\int_{-T}^{T} f(t)\,\mathrm{d}t &= \frac{1}{2T}\int_{-T}^{T} H^{-1/2}\left(f_+ + f_-\right)(t)\,\mathrm{d}t \\ &= \frac{1}{2T}\int_{-T}^{T} H^{-1/2}\Big(\left(e^{-itK}F_0\right)_1 + \left(e^{-itK}F_0\right)_2\Big)\,\mathrm{d}t.\end{aligned}$$

Estimate (3.11) leads then to

$$\begin{aligned}\left\|\frac{1}{2T}\int_{-T}^{T} f(t)\,\mathrm{d}t\right\|_{\mathcal{H}} &\leq CT^{-\ell/2}\left(\left\|H^{-1/2}\left(F_0\right)_1\right\|_{L^{2,s}} + \left\|H^{-1/2}\left(F_0\right)_2\right\|_{L^{2,s}}\right) \\ &\leq CT^{-\ell/2}\left(\|f_0\|_{L^{2,s}} + \left\|H^{-1/2}g_0\right\|_{L^{2,s}}\right),\end{aligned}$$

with $\ell < 1$ and where in the second inequality we have used (3.9):

$$2\left\|H^{-1/2}f_\pm(t=0)\right\|_{L^{2,s}} \leq \|f_0\|_{L^{2,s}} + \left\|H^{-1/2}g_0\right\|_{L^{2,s}}.$$

Denoting $P^T(f_0, g_0) := \frac{1}{2T}\int_{-T}^{T} f(t)\,\mathrm{d}t$ for brevity, we deduce the global-in-time estimate

$$\left\|P^T(f_0,g_0)\right\|_{L^q_T L^2_x([0,\infty)\times\mathbb{R}^d)} \leq C(q)\left(\|f_0\|_{L^{2,s}(\mathbb{R}^d)} + \left\|H^{-1/2}g_0\right\|_{L^{2,s}(\mathbb{R}^d)}\right), \qquad \forall q > 2. \tag{3.12}$$

3.2.2. *Non-Hilbertian subspace.* As before, taking the functional space to be $\mathcal{X} = L^1 \cap L^2$ (with norm $\|\cdot\|_{\mathcal{X}} = \|\cdot\|_{L^1} + \|\cdot\|_{L^2}$), (3.6) with $\varphi(H) = \sqrt{H}$ leads to

$$\psi(\lambda) = \frac{|\mathbb{S}^{d-1}|\sqrt{\varphi^{-1}(\lambda)}^{d-2}}{2\varphi'(\varphi^{-1}(\lambda))} = |\mathbb{S}^{d-1}||\lambda|^{d-1}$$

which then gives the bound

$$\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(E_{\sqrt{H}}(\lambda)f, g\right)_{\mathcal{H}}\right| \leq |\mathbb{S}^{d-1}||\lambda|^{d-1}\|f\|_{L^1}\|g\|_{L^1}, \qquad \forall\lambda\in\mathbb{R}. \tag{3.13}$$

Consequently, we have

$$\left|\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(E_K(\lambda)F, G\right)_{\mathcal{H}\oplus\mathcal{H}}\right| \leq |\mathbb{S}^{d-1}||\lambda|^{d-1}\|F\|_{L^1\oplus L^1}\|G\|_{L^1\oplus L^1}, \qquad \forall\lambda\in\mathbb{R}.$$

This bound leads to a convergence rate of $T^{-\ell/2}$ where $\ell = \min\{d-\varepsilon, 2\}$ for any $\varepsilon > 0$ from Corollary 1.8. For $d \geq 3$ we obtain the optimal rate. Therefore, noting that the kernel is empty:

$$\left\|\frac{1}{2T}\int_{-T}^{T} e^{-itK}F_0\,\mathrm{d}t\right\|_{\mathcal{H}\oplus\mathcal{H}} \leq CT^{-\ell/2}\|F_0\|_{\mathcal{X}\oplus\mathcal{X}}, \qquad \ell = \min\{d-\varepsilon, 2\}.$$

Following the steps from the previous subsection, this leads to the global-in-time estimate

$$\left\|P^T(f_0,g_0)\right\|_{L^q_T L^2_x([0,\infty)\times\mathbb{R}^d)} \leq C(q)\left(\|f_0\|_{\mathcal{X}} + \left\|H^{-1/2}g_0\right\|_{\mathcal{X}}\right), \qquad \forall q > \max\left\{\frac{2}{d}, 1\right\}. \tag{3.14}$$

Remark 3.2. Here we compare our estimates to *Strichartz estimates* for the wave equation (see [18]):

$$\|f\|_{L_t^q L_x^p(\mathbb{R}\times\mathbb{R}^d)} \leq C(d,q,p,s)\left(\|f_0\|_{H^s(\mathbb{R}^d)} + \|g_0\|_{H^{s-1}(\mathbb{R}^d)}\right)$$

for triplets satisfying $2 \leq p, q \leq \infty$ and

$$\frac{1}{q} + \frac{d}{p} = \frac{d}{2} - s, \qquad \frac{2}{q} + \frac{d-1}{p} \leq \frac{d-1}{2}.$$

In particular, we can compare (3.12) with the Strichartz estimate for $(q,p,s) = (\infty, 2, 0)$:

$$\|f\|_{L_t^\infty L_x^2(\mathbb{R}\times\mathbb{R}^d)} \leq C(d)\left(\|f_0\|_{L_x^2(\mathbb{R}^d)} + \|g_0\|_{H_x^{-1}(\mathbb{R}^d)}\right).$$

References

[1] I. Assani and M. Lin. On the one-sided ergodic Hilbert transform. *Contemp. Math.*, 430:21–39, 2007.

[2] M. Ben-Artzi and S. Klainerman. Decay and regularity for the Schrödinger equation. *J. d'Analyse Mathématique*, 58(1):25–37, dec 1992.

[3] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon. *Schrödinger operators with application to quantum mechanics and global geometry*. Springer-Verlag, Berlin, 1987.

[4] N. A. Dzhulaii and A. G. Kachurovskii. Constants in the estimates of the rate of convergence in von Neumann's ergodic theorem with continuous time. *Sib. Math. J.*, 52(5):824–835, oct 2011.

[5] L. C. Evans. *Partial Differential Equations (Graduate Studies in Mathematics)*. American Mathematical Society, 2010.

[6] A. G. Kachurovskii. The rate of convergence in ergodic theorems. *Russ. Math. Surv.*, 51(4):653–703, aug 1996.

[7] A. G. Kachurovskii. On uniform convergence in the ergodic theorem. *J. Math. Sci.*, 95(5):2546–2551, jul 1999.

[8] A. G. Kachurovskii. Fejér integrals and the von Neumann ergodic theorem with continuous time. *Zap. Nauchn. Sem. S.-Peterburg. Otd. Mat. Inst. Steklov.*, 474:171–182, 2018.

[9] A. G. Kachurovskii and K. I. Knizhov. Deviations of Fejer Sums and Rates of Convergence in the von Neumann Ergodic Theorem. *Dokl. Math.*, 97(3):211–214, 2018.

[10] A. G. Kachurovskii and I. V. Podvigin. Estimates of the rate of convergence in the von Neumann and Birkhoff ergodic theorems. *Trans. Moscow Math. Soc.*, 77:1–53, 2016.

[11] A. G. Kachurovskii and I. V. Podvigin. Fejér Sums for Periodic Measures and the von Neumann Ergodic Theorem. *Dokl. Math.*, 98(1):344–347, 2018.

[12] A. G. Kachurovskii and A. V. Reshetenko. On the rate of convergence in von Neumann's ergodic theorem with continuous time. *Sb. Math.*, 201(4):493–500, 2010.

[13] A. G. Kachurovskii and V. V. Sedalishchev. On the constants in the estimates of the rate of convergence in von Neumann's ergodic theorem. *Math. Notes*, 87(5-6):720–727, jul 2010.

[14] A. G. Kachurovskii and V. V. Sedalishchev. Constants in estimates for the rates of convergence in von Neumann's and Birkhoff's ergodic theorems. *Sb. Math.*, 202(8):1105–1125, 2011.

[15] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, 1995.

[16] F. Riesz. Some Mean Ergodic Theorems. *J. London Math. Soc.*, s1-13(4):274–278, oct 1938.

[17] T. Tao. *Nonlinear Dispersive Equations: Local and Global Analysis*. 2006.

[18] D. Tataru. Strichartz estimates for second order hyperbolic operators with nonsmooth coefficients, II. *Am. J. Math.*, 123(3):385–423, 2001.

[19] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci.*, 18(2):70–82, 1932.

School of Mathematics, Cardiff University, Cardiff CF24 4AG, Wales, UK

E-mail address: Ben-ArtziJ@cardiff.ac.uk

E-mail address: MorisseB@cardiff.ac.uk