Journal of Computational and Applied Mathematics 235 (2011) 5024–5039

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A Donoho–Stark criterion for stable signal recovery in discrete wavelet subspaces

Laurent Gosse

IAC–CNR "Mauro Picone" (sezione di Bari), Via Amendola 122/I, 70126 Bari, Italy

ARTICLE INFO

Article history:
Received 19 July 2010
Received in revised form 2 March 2011

MSC:
47a52
47b32
65r20
65t60
94a11

Keywords:
Product of orthogonal projections
Hilbert–Schmidt operator
Geometric harmonics
Singular operator with closed range
Gradient algorithms

ABSTRACT

We derive a sufficient condition by means of which one can recover a scale-limited signal from the knowledge of a truncated version of it in a stable manner following the canvas introduced by Donoho and Stark (1989) [4]. The proof follows from simple computations involving the Zak transform, well-known in solid-state physics. Geometric harmonics (in the terminology of Coifman and Lafon (2006) [22]) for scale-limited subspaces of $L^2(\mathbb{R})$ are also displayed for several test-cases. Finally, some algorithms are studied for the treatment of zero-angle problems.

1. Introduction

1.1. Preliminaries

The problem of signal recovery and extrapolation can be formalized in the following way.

- A signal is modeled as a function s of the variable t (usually standing for time) belonging to a certain closed linear subspace V of a (separable) Hilbert space $\mathcal{H}$, in general $L^2(\mathbb{R})$.
- Only a fraction r of s is observed: there is a set T (not necessarily an interval) such that for every $t \in T, r(t) = 0$, expressing the fact that the corresponding information has been lost. If χ_A stands for the characteristic function of the set A, one can write $r = (1 - \chi_T)s$.
- Worse, the observations can be corrupted by a noise ν, which is nonetheless assumed to be small in $L^2(\mathbb{R})$. In this last case, one observes $\tilde{r} = r + \nu$.

One can define two orthogonal projections depending on "the hole" T and V:

$$\mathcal{P} : L^2(\mathbb{R}) \to V, \qquad \mathcal{Q} : f \in L^2(\mathbb{R}) \mapsto f\chi_T.$$

We clearly cannot assume that $\mathcal{P}$ is compact: since it is idempotent, consider B the unit ball of the closed linear subspace V. If $\mathcal{P}$ is a compact operator, then $B = \mathcal{P}(B)$ is compact, and this implies that V is finite-dimensional which is too restrictive.

E-mail address: l.gosse@ba.iac.cnr.it.

0377-0427/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2011.04.034

We now deduce a "truncation operator", the projection $\mathcal{T}$:

$$\mathcal{T} : L^2(\mathbb{R}) \to L^2(\mathbb{R}), \qquad f \mapsto (1 - \chi_T)f.$$

It is assumed that $\mathcal{QP}$ is a compact operator (but not necessarily $\mathcal{TP}$). As $s \in V$, $\mathcal{P}s = s$, the observations rewrite:

$$r = \mathcal{T}s = \mathcal{TP}s = (\mathrm{Id} - \mathcal{Q})\mathcal{P}s, \qquad \tilde{r} = \mathcal{TP}s + \nu. \tag{1.1}$$

However, one can observe that, thanks again to the assumption $\mathcal{P}s = s$, one has moreover:

$$r = \mathcal{TP}s = (\mathrm{Id} - \mathcal{Q})\mathcal{P}s = (\mathrm{Id} - \mathcal{QP})s, \qquad \tilde{r} = (\mathrm{Id} - \mathcal{QP})s + \nu. \tag{1.2}$$

Actually, from this simple calculation, one can make the following important remarks.

- $\mathcal{QP}$ is a compact operator on $\mathcal{H} = L^2(\mathbb{R})$ as long as $|T|$ is finite[1]; hence its range is not closed and zero is an accumulation point in its spectrum. Eigenvalues can also display a very sharp decay rate depending on the smoothness of the functions in V (see *e.g.* [1,2]).
- the operator $\mathrm{Id} - \mathcal{QP}$ defined on $L^2(\mathbb{R})$ is a Fredholm operator with closed range and finite-dimensional null-space; its restriction to V coincides with $\mathcal{TP} = (\mathrm{Id} - \mathcal{Q})\mathcal{P}$.

The Fredholm alternative applied to $\mathrm{Id} - \mathcal{QP} : V \to L^2(\mathbb{R})$ ensures that its range, $\mathrm{ran}(\mathrm{Id} - \mathcal{QP})$, is closed in $L^2(\mathbb{R})$ and $\mathrm{ran}(\mathrm{Id} - \mathcal{QP}) = \ker(\mathrm{Id} - \mathcal{PQ})^{\perp}$. Moreover, the null-space $\ker(\mathrm{Id} - \mathcal{QP})$ is at most of finite dimension and in case $\ker(\mathrm{Id} - \mathcal{QP}) = \{0\}$, $\mathrm{ran}(\mathrm{Id} - \mathcal{QP}) = L^2(\mathbb{R})$, thus the Eqs. (1.2) are invertible. Hence one switches from the potentially ill-posed inverse problem of trying to solve directly the equation $\mathcal{TP}s = r$ (see *e.g.* [3]) to the stable one $(\mathrm{Id} - \mathcal{QP})s = r$. More precisely:

Theorem 1. *Let V and T be such that the operator norm $\|\mathcal{QP}\| < 1$: in the noise-free case, any $s \in V$ can be fully recovered from r, i.e. $\ker(\mathrm{Id} - \mathcal{QP}) = \{0\}$ and $(\mathrm{Id} - \mathcal{QP})^{-1}r = s$. In the noisy case, the stability estimate holds:*

$$\|s - (\mathrm{Id} - \mathcal{QP})^{-1}\tilde{r}\|_{L^2(\mathbb{R})} \leq \frac{\|\nu\|_{L^2(\mathbb{R})}}{1 - \|\mathcal{QP}\|}. \tag{1.3}$$

The proof can be found in [4] (Theorem 4) and [5] (Corollary 1). The estimate (1.3) shows that the noise is at most amplified by a factor $(1 - \|\mathcal{QP}\|)^{-1}$; it is henceforth a convenient strategy to rely on the Fredholm operator $\mathrm{Id} - \mathcal{QP}$ to perform signal recovery/extrapolation. However, for $s \notin V$, the solution of $(\mathrm{Id} - \mathcal{QP})s = r$ and $\mathcal{TP}s = r$ will clearly differ. As a consequence of Theorem 1, $(\mathrm{Id} - \mathcal{QP})^{-1}$ can be computed (at least, theoretically) via an iterative scheme, the so-called Neumann series:

$$(\mathrm{Id} - \mathcal{QP})^{-1} = \sum_{k=0}^{\infty} (\mathcal{QP})^k. \tag{1.4}$$

This is usually called the Gerchberg–Papoulis (GP) algorithm [6,7] or the Alternating Projections (AP) method. The practical performance can be improved by following the results in *e.g.* [8–10]. The so-called "Generalized Gerchberg–Papoulis" algorithm studied in [11–14] reduces to the Alternating Projections method with the choice of a given multi-resolution subspace $V = V_J$ of scale-limited functions, for some scale parameter $J \in \mathbb{N}$.

Corollary 1. *Under the hypotheses of Theorem 1, let $s^\ell = \sum_{k=0}^{\ell} (\mathcal{QP})^k r$. The following error estimate holds (linear convergence of Alternating Projections):*

$$\|s - s^\ell\|_{L^2(\mathbb{R})} \leq \|\mathcal{QP}\|^\ell \|s - r\|_{L^2(\mathbb{R})}. \tag{1.5}$$

Proof. In the present case, the approximation s^ℓ satisfies the relation $s^0 = r, s^{k+1} = r + \mathcal{QP}s^k$. Hence $s^{k+1} - s = (\mathrm{Id} - \mathcal{Q})s + \mathcal{QP}s^k - s = \mathcal{QP}(s^k - s)$; the result (1.5) follows. □

Similar algorithms are also widely used in the context of irregular sampling, see for instance [15–17], which addresses the related issue of reconstructing a function belonging to a subspace V starting from a collection of point-wise observations; these results can be seen as an extreme example of the large-sieve stability estimates proved in [18].

[1] The finiteness hypothesis for the measure of T can be understood through the simple example of the "sliding bumps": let φ be a C^∞ function supported in $[-1, 1]$, and define the sequence $\varphi_n(t) = \varphi(t - n)$ which is bounded in *e.g.* any Sobolev space $H^s(\mathbb{R})$. Since it "escapes at infinity", it converges weakly to zero but clearly not strongly; one cannot hope to have $\mathcal{QP}$ compact for general unbounded T.

1.2. Objectives and outline of the paper

We are partly motivated by the question raised in the paper [11] (bottom of page 229): "*we assume that the observation of the signal f inside the interval $[-T, T]$ can uniquely determine the value of f up to $[-\Pi, \Pi]$ in the time domain. Given T, the Π value will depend on the regularity of the signal and the scale parameters J. The mathematical relationship between these parameters is still open*". To the best of our knowledge, it is still unanswered; in the present paper, we thus propose to establish that under a rather simple criterion based on the Hilbert–Schmidt operator norm of the composite of two orthogonal projections, [4], stable recovery is possible by means of the iterative techniques presented in [13].

This paper is therefore organized as follows. In Section 2, we recall the subspaces of $L^2(\mathbb{R})$ which will be useful in the paper, namely the Paley–Wiener space of band-limited functions and the multi-resolution analysis; technical results about the composition product of two orthogonal projections in Hilbert space are also recalled, including the characterization through the minimal canonical angle between subspaces. In Section 3, we derive our Donoho–Stark criterion for stable signal recovery by computing the Hilbert–Schmidt norm of the product of projections $\mathcal{P}\mathcal{Q}$ by taking advantage of the structure of Reproducing Kernel Hilbert spaces; some consequences are obtained by using abstract results from [19]; numerical simulations following original ideas from [20] are displayed in Section 3.4. In Section 4, we exploit the fact that a Fredholm operator has a closed range to study iterative algorithms for singular operators (see [2,21]) in the context of signal recovery with a zero-angle problem (that is, when $\|\mathcal{Q}\mathcal{P}\| = 1$); non-uniqueness is resolved by working with the "minimum-norm least squares" solution. Final conclusions are drawn in Section 5. Appendices A and B contain auxiliary results about a technical lemma and the Zak transform.

Signal recovery can be performed through the use of Slepian and Pollak's *Prolate Spheroidal Wave functions* (PSWF) but numerically this yields an unstable ill-posed inversion; several fixes have been proposed to stabilize this approach, like [3] and the more recent "geometric harmonics" by Coifman and Lafon [22]. Several generalizations are proposed in [23,24]. Extrapolation in discrete wavelet subspaces has been developed in a series of papers [13,14,11,12] by computing "wavelet geometric harmonics"; in [20], early computations show the different behavior when one passes from a band-limited scaling function to another one with compact support. One idea contained in [4] is to compute explicitly an operator norm in order to get a sufficient condition ensuring that (1.2) is invertible and the results of [5] can be applied; several developments have been published recently, see [25,26]. A very original and seemingly unknown paper conducting wavelet extrapolation and comparing it to the MNLS algorithms of [3] is [27]. The Donoho–Stark criterion is studied in the context of irregular sampling in [6], see also [17]. Some elements dealing with Compressed Sensing and the product of two orthogonal projections are given in [28] and also [29], especially Section 5.

2. Band-limited and scale-limited extrapolations

2.1. Paley–Wiener space and Multi-Resolution Analysis (MRA)

In the majority of applications (except for [23,26,24]), V stands for a space of band-limited or scale-limited functions. For any $f \in L^2(\mathbb{R})$, we normalize its Fourier–Plancherel transform $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ as

$$\forall \xi \in \mathbb{R}, \quad [\mathcal{F}f](\xi) = \hat{f}(\xi) = \int_{\mathbb{R}} f(t) \exp(-2i\pi t\xi) dt.$$

It follows that a function f is said to be band-limited as soon as there exists $\omega > 0$ such that $\hat{f}(\xi) = 0$ for any ξ with $|\xi| > \omega$. We can therefore introduce the Paley–Wiener space:

$$PW_\omega(\mathbb{R}) = \{f \in L^2(\mathbb{R}) \text{ such that } \hat{f}(\xi) = 0 \text{ for } |\xi| > \omega\}.$$

The Paley–Wiener theorem states that functions belonging to $PW_\omega(\mathbb{R})$ can be extended to the whole complex plane as entire functions of exponential type:

$$f \in PW_\omega(\mathbb{R}) \Rightarrow \forall z \in \mathbb{C}, \quad |f(z)| \le \sup_{t \in \mathbb{R}} |f(t)| \exp(\omega|\Im(z)|).$$

As a consequence of analytic continuation theory for functions of one complex variable, the knowledge of such a function restricted to any arbitrary interval of $\mathbb{R}$ allows us to deduce all its remaining values in $\mathbb{C}$. Thus *band-limited extrapolation* corresponds to the choice $V = PW_\omega(\mathbb{R})$. Next, we introduce briefly the concept of *Multi-Resolution Analysis* (MRA) (see *e.g.* [30] for details).

Definition 1. A sequence of nested subspaces V_j is called a *Multi-Resolution Analysis* of $L^2(\mathbb{R})$ if: $\{0\} \subset \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset L^2(\mathbb{R})$. Moreover, the following properties must hold:

- for all $f \in L^2(\mathbb{R})$, $\|\mathcal{P}_{V_j} f - f\|_{L^2} \to 0$ as $j \to +\infty$ also, $\mathcal{P}_{V_j} f \to 0$ as $j \to -\infty$;
- if $f(t) \in V_j$, then $f(t/2) \in V_{j-1}$ and for all $k \in \mathbb{Z}$, $f(t - 2^j k) \in V_j$;
- there exists a shift-invariant orthonormal base of V_0 given by the scaling function $\phi_n(t) = \phi(t - n)$ for $n \in \mathbb{Z}$.

In this definition, $\mathcal{P}_{V_j}$ stands for the orthogonal projector onto the subspace V_j. Intuitively, it asks for the V_j's to be linear subspaces of $L^2(\mathbb{R})$ with increasing temporal resolution: when j decreases, functions in V_j tend to become constants. Oppositely, when j increases, they are allowed to oscillate with high instantaneous frequency. The wavelet spaces W_j are defined as the orthogonal complement of V_j inside V_{j+1}, which means: for all $j \in \mathbb{Z}$, $V_{j+1} = V_j \oplus W_j$. From ϕ_n, the base of V_0, one can deduce a base of V_j by simple dilatation,

$$\phi_{j,n}(t) = \sqrt{2^j}\phi_n(2^j t) = \sqrt{2^j}\phi(2^j t - n). \tag{2.1}$$

Thus, the orthogonal projection of f onto the *scale-limited subspace* V_j reads

$$\mathcal{P}_{V_j} f = \sum_{n\in\mathbb{Z}} \langle f, \phi_{j,n}\rangle \phi_{j,n}, \quad \langle f, \phi_{j,n}\rangle = \int_{\mathbb{R}} f(t)\phi_{j,n}(t).\mathrm{d}t, \tag{2.2}$$

which is the best approximation of f in V_j in the least-squares sense.

2.2. Composite of two projections in Hilbert space and stable recovery

In all of the following, we shall use the following notation for the norm of any bounded operator $T : \mathcal{H} \to \mathcal{H}$, $\mathcal{H}$ being a separable Hilbert space,

$$\|T\| \doteq \sup_{f\in\mathcal{H}} \frac{\|Tf\|_{\mathcal{H}}}{\|f\|_{\mathcal{H}}}.$$

Moreover, $\ker T$ and $\mathrm{ran}(T)$ will stand for its null-space and its range, respectively. Very general results about the structure of the composition of two orthogonal projections in Hilbert space are given in [31].

Lemma 1. *Let $\mathcal{H}$ be a Hilbert space and P_A, P_B be two orthogonal projections onto A, B which are two closed linear subspaces of $\mathcal{H}$. Then there holds:*

$$\|P_A P_B\| = \|P_B P_A\| \le 1. \tag{2.3}$$

Indeed, the proof of the Lemma (put forward in the Appendix) shows a bit more: we actually have that $\|P_A P_B\|^2 = \|P_B P_A\|^2 = \rho(P_A P_B P_A) = \rho(P_B P_A P_B)$, the spectral radius.

Lemma 2. *Under the hypotheses of Lemma 1, there holds moreover*

$$\|P_A P_B\| = \sup_{f\in B} \frac{\|P_A f\|_{\mathcal{H}}}{\|f\|_{\mathcal{H}}} = \sup_{f\in A} \frac{\|P_B f\|_{\mathcal{H}}}{\|f\|_{\mathcal{H}}} = \|P_B P_A\|. \tag{2.4}$$

Proof. By definition, the operator norm of $P_A P_B : \mathcal{H} \to \mathcal{H}$ reads

$$\|P_A P_B\| = \sup_{f\in\mathcal{H}} \frac{\|P_A P_B f\|_{\mathcal{H}}}{\|f\|_{\mathcal{H}}}.$$

Since P_B is an orthogonal projection, one can split $\mathcal{H} = B \oplus B^\perp$ such that $f = P_B f + (\mathrm{Id} - P_B)f$ and $\|f\|^2_{\mathcal{H}} = \|P_B f\|^2_{\mathcal{H}} + \|(\mathrm{Id} - P_B)f\|^2_{\mathcal{H}}$. This yields

$$\|P_A P_B\|^2 = \sup_{f\in\mathcal{H}} \frac{\|P_A P_B(P_B f)\|^2_{\mathcal{H}}}{\|P_B f\|^2_{\mathcal{H}} + \|(\mathrm{Id} - P_B)f\|^2_{\mathcal{H}}},$$

and this expression is clearly maximized for $f \in B$. The same reasoning can be made with P_A and Lemma 1 allows us to conclude. □

At this level, it is of critical importance to be able to estimate as accurately as possible the quantity $\|\mathcal{Q}\mathcal{P}\|$ which controls both the invertibility of $\mathrm{Id} - \mathcal{Q}\mathcal{P}$ but also the error estimates (1.3) and (1.5). In both cases, the condition $\|\mathcal{Q}\mathcal{P}\| < 1$ expresses the fact that there exists no function belonging to V whose L^2 norm is not affected when being truncated to $\mathbb{R} \setminus T$ because

$$\|\mathcal{Q}\mathcal{P}\| = \sup_{f\in L^2(\mathbb{R})} \frac{\|\mathcal{Q}\mathcal{P}f\|_{L^2(\mathbb{R})}}{\|f\|_{L^2(\mathbb{R})}} = \sup_{f\in L^2(\mathbb{R})} \frac{\|\mathcal{Q}\mathcal{P}f\|_{L^2(\mathbb{R})}}{\|\mathcal{P}f\|_{L^2(\mathbb{R})}} = \sup_{g\in V} \frac{\|\mathcal{Q}g\|_{L^2(\mathbb{R})}}{\|g\|_{L^2(\mathbb{R})}}.$$

For instance, if one considers scale-limited extrapolation with the so-called discontinuous Haar basis ($\phi(t) = \chi_{[0,1]}(t)$), then $\|\mathcal{Q}\mathcal{P}\| = 1$ for $T = [2^{-J}k, 2^{-J}(k+1)]$, $V = V_J$ and any $k \in \mathbb{Z}$, hence stable recovery cannot be performed; see, however, the computations with this scaling function in [22]. In sharp contrast, if the scaling function ϕ is chosen to be a band-limited function (see [13], or the "prolate spheroidal wavelets" in [32]), then $\|\mathcal{Q}\mathcal{P}\| < 1$ because ϕ belongs to a Paley–Wiener space (see Theorem 4 in [13]). In [25], the authors proved the following result (see Theorem 2, page 340), which is a consequence of [4,5].

Theorem 2 (*See [25]*). *Let $V = PW_\omega$ for a given $\omega > 0$ and T an arbitrary measurable and bounded set of $\mathbb{R}$; then Eq. (1.2) is always invertible, that is, $\|\mathcal{Q}\mathcal{P}_\omega\| < 1$.*

We cannot expect such a strong result in the case $V = V_j$, a more general subspace belonging to a MRA of $L^2(\mathbb{R})$; especially, as soon as the scaling function ϕ has compact support and $|T|$ is large enough, it is possible to find non-trivial functions $f \in V_j$ such that $\|\mathcal{Q}f\| = \|f\|$.

2.3. Geometric interpretation of the composition of projections

Definition 2. Let A, B be two linear subspaces in a Hilbert space $\mathcal{H}$; the number $0 \le \theta(A, B) \le \frac{\pi}{2}$ is called the *minimal canonical angle* between A and B and satisfies

$$\cos\theta(A, B) = \sup_{a\in A, b\in B} \frac{|(a, b)|}{\|a\|\,\|b\|} \doteq \cos(A, B). \tag{2.5}$$

In particular, $\cos\theta(A, B) = 1$ when $A \subset B$ which is precisely the situation one wants to absolutely avoid in the context of an extrapolation problem because it means that, as they stand, the lacunary and possibly noised observations perfectly fit into the space of functions containing the original signal. In this case, there is no hope for recovery by means of alternating projections because $\operatorname{ran} P_A \subset \operatorname{ran} P_B$ implies $P_AP_B = P_BP_A = P_A$ and $\|P_A\| = 1$. Now, we can give a small result concerning an interpretation of the quantities involved in Lemma 1 as the cosine of linear subspaces in a general setting.

Lemma 3. *Under the hypotheses of Lemma 1, there holds $\|P_AP_B\|^2 = \cos^2(A, B)$.*

Proof. Thanks to (3) in the proof of Lemma 1, we have that $\|P_AP_B\|^2 = \rho(P_BP_AP_B)$ and since $P_BP_AP_B$ is self-adjoint this implies:

$$\begin{aligned}
\rho(P_BP_AP_B) &= \|P_BP_AP_B\| \\
&= \sup_{u\in\mathcal{H},\ \|u\|\le 1} |(P_BP_AP_Bu, u)| \\
&= \sup_{u\in\mathcal{H}} \frac{(P_AP_Bu, P_Bu)}{\|u\|^2_{\mathcal{H}}} \\
&= \sup_{u\in\mathcal{H}} \frac{(P_A^2P_Bu, P_Bu)}{\|u\|^2_{\mathcal{H}}} \\
&= \sup_{b\in B} \frac{(P_Ab, P_Ab)}{\|b\|^2_{\mathcal{H}}} \\
&= \sup_{b\in B}\sup_{a\in A} \frac{(a, b)^2}{\|a\|^2_{\mathcal{H}}\|b\|^2_{\mathcal{H}}} \\
&= \cos^2(A, B).
\end{aligned}$$

We used that P_A, P_B are self-adjoint, idempotent and have unit operator norm. □

The next statement already appears elsewhere, see for instance Theorem 2.1 in [33] or [34,5]. Part of it is proved in the standard textbook [35], pages 21–22.

Theorem 3 (*See [34,33]*). *Let A, B be two closed linear subspaces of a Hilbert space $\mathcal{H}$; the following statements are equivalent:*

(1) $\cos(A, B) < 1$;
(2) *$A + B$ is closed in $\mathcal{H}$, i.e. $\overline{A + B} = A + B$ and $A \cap B = \{0\}$;*
(3) *there exists $C > 0$ such that for all $a, b \in A \times B$, $\|a\| + \|b\| \le C\|a + b\|$.*

Clearly, statement (1) implies that $A \cap B = \{0\}$: otherwise it would suffice to pick $v \in A \cap B$, thus $P_AP_Bv = v$ and 1 is an eigenvalue of P_AP_B. In the context of band-limited extrapolation, the condition $A \cap B = \{0\}$ has a rather clear meaning: since A stands for the subspace of functions supported on T and B for the one of band-limited functions, by the Paley–Wiener theorem, it is equivalent to the statement that no non-zero analytic function can vanish on a positive measure interval of $\mathbb{R}$.

3. Reproducing the kernel Hilbert space approach to estimate $\|\mathcal{Q}\mathcal{P}\|$

We start with a classical definition (see [36,37] for more details):

Definition 3. A (separable) Hilbert space $\mathcal{H}$ is called a *Reproducing kernel Hilbert space* (RKHS) of functions $\mathbb{R} \to \mathbb{R}$ if for any $t \in \mathbb{R}$, there exists a continuous function $K(.,t) \in \mathcal{H}$, called the reproducing kernel, which satisfies

$$\forall t \in \mathbb{R},\ K(.,t) \in \mathcal{H}; \quad \forall (t,f) \in \mathbb{R} \times \mathcal{H},\ f(t) = \langle f, K(.,t)\rangle = \int K(s,t)f(s).\mathrm{d}s.$$

In other words, the point evaluation $f \mapsto f(t)$ is continuous as an application $\mathcal{H} \to \mathbb{R}$. The Riesz representation theorem guarantees, for every $t \in \mathbb{R}$, that the function $K(.,t)$ is unique.

The main point here is that, under mild assumptions and for $\mathcal{H} = L^2(\mathbb{R})$, the spaces V of interest for band-limited and scale-limited extrapolation are RKHS.

Theorem 4 (*See [38,37]*)**.** 1. *For any $\omega \in \mathbb{R}^+$, the Paley–Wiener subspace $PW_\omega(\mathbb{R})$ of $L^2(\mathbb{R})$ is an RKHS with the Shannon kernel $K_\omega(s,t) = \int_{-\omega}^{\omega} \exp(2i\pi\xi(t-s)).\mathrm{d}\xi = \frac{\sin 2\pi\omega(s-t)}{\pi(s-t)}$.*
2. *If $|\phi(t)| \le C(1+|t|)^{-\frac{1}{2}-\varepsilon}$ for $\varepsilon > 0$, any multi-resolution subspace V_j is an RKHS with kernel $K_j(s,t) = \sum_{n\in\mathbb{Z}} \phi_{j,n}(s)\phi_{j,n}(t) = 2^j \sum_{n\in\mathbb{Z}} \phi(2^j s - n)\phi(2^j t - n)$.*

3.1. Calculation of the Hilbert–Schmidt norm of $\mathcal{Q}\mathcal{P}_{V_j}$: Donoho–Stark criterion

At this point, one observes that for $K \in L^2(\mathbb{R}^2)$ continuous and any $t \in \mathbb{R}, f(t) = \int_{\mathbb{R}} K(s,t)f(s)\mathrm{d}s$ is a Hilbert–Schmidt operator; hence for band-limited extrapolation,

$$\mathcal{P}_\omega \mathcal{Q} f(t) = (\mathcal{Q}f, K_\omega(.,t)) = \int_{\mathbb{R}} K_\omega(s,t)\chi_T(s)f(s).\mathrm{d}s,$$

and a similar expression holds for scale-limited extrapolation. Thus, on the one hand it is well-known that in this case $\|\mathcal{P}_\omega\mathcal{Q}\| \le \|\mathcal{P}_\omega\mathcal{Q}\|_{HS} = \|K_\omega\chi_T\|_{L^2(\mathbb{R}^2)}$, and on the other hand, Lemma 1 ensures that $\|\mathcal{P}_\omega\mathcal{Q}\| = \|\mathcal{Q}\mathcal{P}_\omega\|$. So we obtain a convenient bound for the operator norm which controls the error estimate (1.5). We switch now to MRA subspaces.

Theorem 5. *For V_j being some MRA subspace of $L^2(\mathbb{R})$ associated with a continuous scaling function ϕ satisfying $\|\phi\|_{L^2(\mathbb{R})} = 1$ and $|\phi(t)| \le C(1+|t|)^{-\frac{1}{2}-\varepsilon}$, there holds*

$$\|\mathcal{P}_{V_j}\mathcal{Q}\|_{HS}^2 = \int_{2^j T} \mathcal{Z}(\widehat{|\phi|^2})(0,s).\mathrm{d}s, \tag{3.1}$$

with $\mathcal{Z}(f)(t,\xi)$ standing for the Zak transform of the function f (cf. Appendix B).

Proof. We want to compute $\|\mathcal{P}_{V_j}\mathcal{Q}\|_{HS}^2 = I_j(T) = \int_T \int_{\mathbb{R}} \left|\sum_{n\in\mathbb{Z}} \phi_{j,n}(s)\phi_{j,n}(t)\right|^2 .\mathrm{d}s.\mathrm{d}t$ for any $j \in \mathbb{Z}$. First, by a simple rescaling argument, we get that $I_j(T) = I_0(2^j T)$: hence we concentrate on the task of computing $I_0(T)$ which is split into several steps.

(1) First, for any $s \in \mathbb{R}$, we define the function $k_s : t \mapsto \sum_{n\in\mathbb{Z}} \phi(s-n)\phi(t-n)$ and we do the Fourier transform in the t variable:

$$\hat{k}_s(\xi) = \sum_{n\in\mathbb{Z}} \phi(s-n)\exp(-2i\pi n\xi)\hat{\phi}(\xi) = \hat{\phi}(\xi)\mathcal{Z}\phi(s,-\xi).$$

The Plancherel equality allows us to rewrite $I_0(T) = \int_T \int_{\mathbb{R}} |\phi(\xi)|^2 |\mathcal{Z}\phi(s,-\xi)|^2 \mathrm{d}\xi\,\mathrm{d}s$.

(2) We know that $\mathcal{Z}\phi(s,-\xi) = \exp(-2i\pi\xi s)\mathcal{Z}\hat{\phi}(\xi,s)$ and that $\mathcal{Z}\hat{\phi}(\xi,s+1) = \mathcal{Z}\hat{\phi}(\xi,s)$ for any s (cf. [39], p.161–163), so we get

$$I_0(T) = \int_T \int_{\mathbb{R}} |\phi(\xi)|^2 |\mathcal{Z}\hat{\phi}(\xi,s)|^2 \mathrm{d}\xi\,\mathrm{d}s = \int_T \int_0^1 |\mathcal{Z}\hat{\phi}(\xi,s)|^2 \underbrace{\sum_{k\in\mathbb{Z}} |\hat{\phi}(\xi+k)|^2}_{=1} \mathrm{d}\xi\,\mathrm{d}s,$$

thanks to the properties of the scaling function ϕ generating an MRA (see [40], p.173). In the case $T = [0,1]$, this is already enough to conclude that $I_0([0,1]) = \int_{\mathbb{R}} |\hat{\phi}(\xi)|^2 \mathrm{d}\xi = \|\phi\|_{L^2(\mathbb{R})}^2$. More generally, if $T = [a,b]$ with $a, b \in \mathbb{Z}^2$, then $I_j([a,b]) = 2^j |T| \|\phi\|_{L^2(\mathbb{R})}^2$.

(3) To estimate $\int_0^1 |\mathcal{Z}\hat{\phi}(\xi,s)|^2 \mathrm{d}\xi$, we must use the following fact (cf. [39], p.165):

$$\int_0^1 \mathcal{Z}\hat{\phi}(\xi,s)\overline{\mathcal{Z}\hat{\phi}(\xi,s)}\mathrm{d}\xi = \sum_{k\in\mathbb{Z}} (\hat{\phi}, \hat{\phi}(.-n))\exp(-2i\pi ks) = \mathcal{Z}(\hat{\phi} * \eta)(0,s),$$

where $\eta(\tau) = \overline{\hat{\phi}(-\tau)} = \hat{\phi}(\tau)$ because ϕ is real-valued. Hence the expression reduces to

$$\int_0^1 \mathcal{Z}\hat{\phi}(\xi, s)\overline{\mathcal{Z}\hat{\phi}(\xi, s)}\mathrm{d}\xi = \mathcal{Z}(\hat{\phi} * \hat{\phi})(0, s) = \mathcal{Z}(\widehat{|\phi|^2})(0, s).$$

It remains to integrate on T to obtain (3.1). □

As an consequence of (3.1), one can recover part of the result established by Donoho and Stark (Lemma 2 in [4]) for instance in the case $T = [0, n]$, $n \in \mathbb{N}$, and $\phi(t) = \frac{\sin \pi\omega t}{\pi t}$, $\omega = 2^j$: since $s \mapsto \mathcal{Z}(\widehat{|\phi|^2})(0, s)$ is 1-periodic and thanks to the inversion formula (cf. (B.3) or [39], p.163), one finds immediately that

$$\|\mathcal{P}_\omega \mathcal{Q}\|_{HS}^2 = n\widehat{|\phi|^2}(0) = |T|\|\phi\|_{L^2(\mathbb{R})}^2 = |T|\|\chi_{[-\frac{\omega}{2}, \frac{\omega}{2}]}\|_{L^2(\mathbb{R})}^2 = \omega|T| = 2^j|T|.$$

The third equality comes from the Plancherel identity.

3.2. Equivalent form of the Hilbert–Schmidt norm $\|\mathcal{Q}\mathcal{P}_{V_j}\|_{HS}$

The estimate (3.1) is difficult to use when T has many connected components, or even if T is an interval with non-integer extremities; the following result fixes this issue.

Corollary 2. *Under the hypotheses of Theorem 5, there holds:*

$$\|\mathcal{Q}\mathcal{P}_{V_j}\|^2 = \|\mathcal{P}_{V_j}\mathcal{Q}\|^2 \le \|\mathcal{P}_{V_j}\mathcal{Q}\|_{HS}^2 = \sum_{k\in\mathbb{Z}} (|\phi|^2 * \chi_{2^jT})(k). \tag{3.2}$$

Proof. This is a direct consequence of the Poisson summation formula:

$$\begin{aligned}
\int_T \mathcal{Z}(\widehat{|\phi|^2})(0, s)\mathrm{d}s &= \int_T \sum_{k\in\mathbb{Z}} \widehat{|\phi|^2}(k)\exp(-2\mathrm{i}\pi ks)\mathrm{d}s \\
&= \int_T \sum_{k\in\mathbb{Z}} |\phi|^2(k - s)\mathrm{d}s \\
&= \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} |\phi|^2(k - s)\chi_T(s)\mathrm{d}s \\
&= \sum_{k\in\mathbb{Z}} (|\phi|^2 * \chi_T)(k). \quad \square
\end{aligned}$$

In Fig. 1, we display the squares of several standard scaling functions to be used in (3.2).

Similarly to Theorem 10 in [4], one can question the sharpness of the sufficient criterion (3.2) and wonder whether it is possible to find sets T such that $\|\mathcal{Q}\mathcal{P}_{V_j}\| < 1$ and (3.2) is not satisfied. Actually, this is possible for band-limited extrapolation; however, a crucial ingredient in the proof of Theorem 10 in [4] lies in the fact that the reproducing kernel for the Paley–Wiener space PW_ω is the function "$\mathrm{sinc}_\omega(t - s)$" which decays when $|t - s|$ grows. This is not the case when V_j is an MRA subspace in the sense of Definition 1 as one has only the following simple estimate:

$$|K_0(s, t)| \le \sum_n |\phi(s - n)||\phi(t - n)| \le \frac{C}{\sqrt{1 + |s - t|}},$$

which is a consequence of the decay assumption on ϕ and the inequality for any $x, y \in \mathbb{R}$: $(1 + |x|)(1 + |y|) \ge 1 + |x - y|$. Indeed, let us consider $T = T_1 \cup T_2$: the core of the proof in [4] is to establish that, with straightforward notation, $\langle \mathcal{Q}_1 f, \mathcal{P}\mathcal{Q}_2 f\rangle$ is small when T_1 is far from T_2 for any $f \in L^2(\mathbb{R})$:

$$\begin{aligned}
\langle \mathcal{Q}_1 f, \mathcal{P}_{V_0}\mathcal{Q}_2 f\rangle &= \int_{T_1} f(t)\left(\sum_{n\in\mathbb{Z}} \int_{T_2} \phi(s - n)f(s).\mathrm{d}s\phi(t - n)\right).\mathrm{d}t \\
&= \sum_{n\in\mathbb{Z}} \left(\int_{T_1} \phi(t - n)f(t).\mathrm{d}t\right)\left(\int_{T_2} \phi(s - n)f(s).\mathrm{d}s\right) \\
&= \sum_{n\in\mathbb{Z}} [f\chi_{T_1} * \phi].[f\chi_{T_2} * \phi](n) \\
&= \langle \mathcal{P}_{V_0}\mathcal{Q}_1 f, \mathcal{P}_{V_0}\mathcal{Q}_2 f\rangle.
\end{aligned}$$

This quantity is the scalar product in $L^2(\mathbb{R})$ of $\mathcal{P}_{V_0}(f\chi_{T_1})$ and $\mathcal{P}_{V_0}(f\chi_{T_2})$: it does not decrease if T_1 and T_2 are far from each other. Hence the condition (3.2) is probably sharper than its analog for band-limited extrapolation studied in [4]. Moreover, it does not seem that analogs of the "large sieve" estimates studied in [18] allow us to improve (3.2) in the case where T is the union of many disjoint intervals.

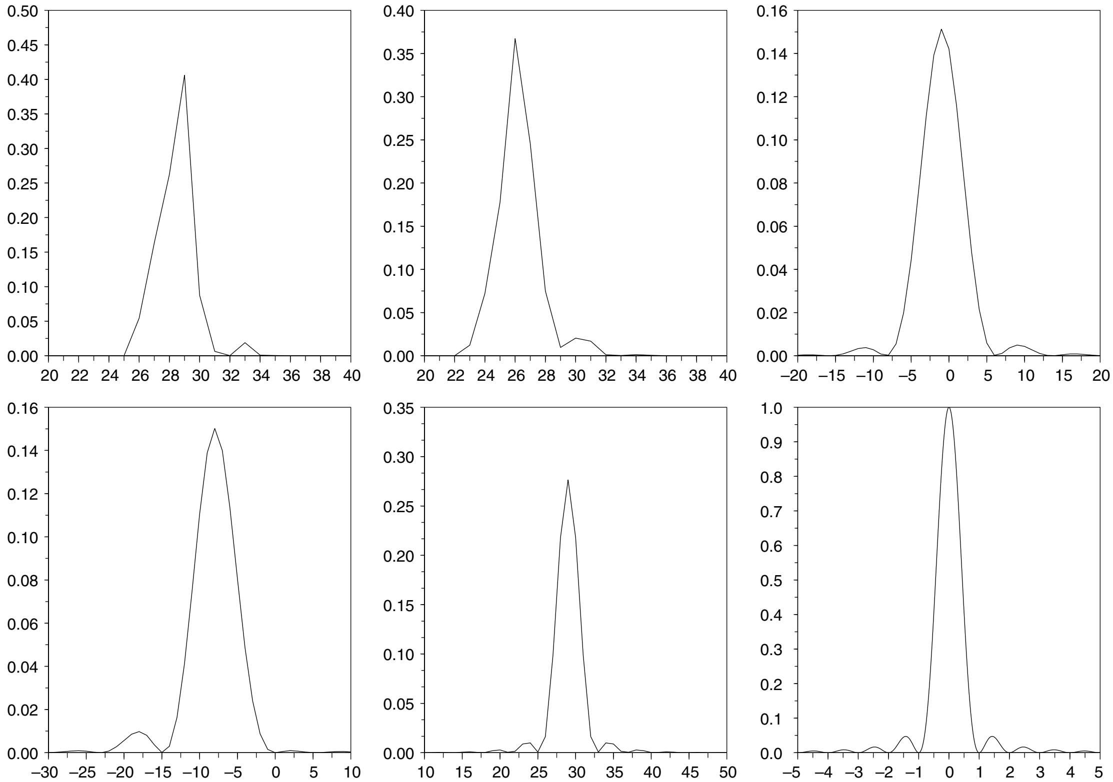

Fig. 1. Scaling functions $|\phi|^2$: Daubechies 4 (top left), Daubechies 6 (top middle), Coiflet 5 (top right), Symmlet 10 (bottom left), Meyer 3 (bottom middle) and sinc (bottom right).

Lemma 4. *Let $T \subset \mathbb{R}$ and V be a closed linear subspace of $L^2(\mathbb{R})$ such that the orthogonal projections $\mathcal{P}$ and $\mathcal{Q}$ satisfy $\|\mathcal{Q}\mathcal{P}\| < 1$. Then, for any $x \in V = \operatorname{ran}(\mathcal{P})$, $\langle \mathcal{T}x, x\rangle = 0$ if and only if $x = 0$.*

Proof. Any $x \in \operatorname{ran}(\mathcal{P})$ rewrites $x = \mathcal{P}f$ for some $f \in L^2(\mathbb{R})$, so

$$\langle \mathcal{T}x, x\rangle = \langle \mathcal{T}x, \mathcal{P}x\rangle = \langle \mathcal{P}\mathcal{T}\mathcal{P}f, \mathcal{P}f\rangle = \langle \mathcal{T}^2\mathcal{P}f, \mathcal{P}f\rangle = \|\mathcal{T}\mathcal{P}f\|^2.$$

Hence, assuming that $\|\mathcal{T}\mathcal{P}f\|^2 = 0$ yields that $\mathcal{P}f = \mathcal{Q}\mathcal{P}f$. But, from the contents of Lemma 2, this implies that for such an x, one has

$$1 = \frac{\|\mathcal{Q}\mathcal{P}f\|}{\|\mathcal{P}f\|} = \|\mathcal{Q}\mathcal{P}\|,$$

which contradicts the hypothesis. □

Corollary 3. *Under the general hypotheses of Theorem 5, as soon as the sufficient condition $\sum_{k\in\mathbb{Z}}(|\phi|^2 * \chi_{2^jT})(k) < 1$ is met, the following hold:*

(1) $\|\mathcal{Q}\mathcal{P}_{V_j}\| = \|\mathcal{P}_{V_j}\mathcal{Q}\| < 1$ *and* $\cos(\operatorname{ran}\mathcal{P}_{V_j}, \operatorname{ran}\mathcal{Q}) < 1$;
(2) $\operatorname{ran}(\mathcal{P}_{V_j}) \cap \operatorname{ran}(\mathcal{Q}) = \{0\}$ *and* $\operatorname{ran}(\mathcal{P}_{V_j}) \oplus \operatorname{ran}(\mathcal{Q})$ *is closed in* $L^2(\mathbb{R})$;
(3) $V_j = \operatorname{ran}(\mathcal{P}_{V_j}\mathcal{T})$; *especially* $\operatorname{ran}(\mathcal{P}_{V_j}\mathcal{T})$ *and* $\operatorname{ran}(\mathcal{T}\mathcal{P}_{V_j})$ *are closed and the operator* $\mathcal{T}\mathcal{P}_{V_j}$ *is not compact;*
(4) $\operatorname{ran}(\mathcal{P}_{V_j} + \mathcal{Q}) = \operatorname{ran}(\mathcal{P}_{V_j}) \oplus \operatorname{ran}(\mathcal{Q})$; *in particular, the orthogonal projection onto* $\operatorname{ran}(\mathcal{P}_{V_j}) \oplus \operatorname{ran}(\mathcal{Q})$ *reads* $(\mathrm{Id} - \mathcal{Q})(\mathrm{Id} - \mathcal{P}_{V_j}\mathcal{Q})^{-1}\mathcal{P}_{V_j} + (\mathrm{Id} - \mathcal{P}_{V_j})(\mathrm{Id} - \mathcal{Q}\mathcal{P}_{V_j})^{-1}\mathcal{Q}$;
(5) $\operatorname{ran}((\mathrm{Id} - \mathcal{P}_{V_j})\mathcal{Q})$ *and* $\operatorname{ran}(\mathcal{Q}(\mathrm{Id} - \mathcal{P}_{V_j}))$ *are closed.*

Proof. Points (1) and (2) follow from Theorem 3. For (3), the property "$\operatorname{ran}(\mathcal{P}_{V_j}(\mathrm{Id} - \mathcal{Q}))$ closed" is a consequence of Lemma 2.4 in [19] as soon as $\operatorname{ran}(\mathcal{P}_{V_j}) + \operatorname{ran}(\mathcal{Q})$ is closed in $L^2(\mathbb{R})$; clearly, $\operatorname{ran}(\mathcal{P}_{V_j}\mathcal{T}) \subset V_j$. In order to prove the converse, it suffices to observe that $\ker(\mathcal{T}\mathcal{P}_{V_j}) = \{0\}$ from the proof of Lemma 4; moreover, $\operatorname{ran}(\mathcal{T}\mathcal{P}_{V_j})$ is closed because $\operatorname{ran}(\mathcal{P}_{V_j}\mathcal{T})$ is closed. Theorem II.19 in [35] allows us to conclude that $\mathcal{P}_{V_j}\mathcal{T}$ is onto. Points (4) and (5) also come from Lemma 3.4 of [19]. □

According to [11] (see also [13]), $\mathrm{ran}(\mathcal{P}_{V_j}\mathcal{T})$ is precisely the space $\mathcal{U}_j$ written in Theorem 1 in the context of scale-limited extrapolation. Taking into account for the non-zero angle hypothesis allows us to refine their result by showing that $\mathcal{U}_j = V_j$ as long as $\|\mathcal{Q}\mathcal{P}_{V_j}\| < 1$ for general scaling functions inside an orthogonal wavelet framework. In these former works, the property $\mathcal{U}_j = V_j$ was proved only for band-limited scaling functions.

Remark 1. Here, we let $\mathcal{P}$ be any orthogonal projection $L^2(\mathbb{R}) \to \mathrm{ran}(\mathcal{P})$: from Corollary 3.2 in [19], it comes that both conditions $\|\mathcal{P}\mathcal{Q}\| < 1$ and $\|(\mathrm{Id} - \mathcal{P})\mathcal{T}\| < 1$ imply that $L^2(\mathbb{R}) = \mathrm{ran}(\mathcal{P}) \oplus \mathrm{ran}(\mathcal{Q})$ because the second one ensures that $(\mathrm{Id} - (\mathrm{Id} - \mathcal{P})(\mathrm{Id} - \mathcal{Q}))$ is invertible. Unfortunately we are not able to present a situation for which $\|\mathcal{P}_{V_j}\mathcal{Q}\| < 1$ and $\|(\mathrm{Id} - \mathcal{P}_{V_j})\mathcal{T}\| < 1$ hold for V_j an MRA subspace and a measurable set T. However, it is rather easy to visualize their meaning: the first condition expresses the fact that, apart from zero, no function supported on T belongs to V_j, and the second, that no function supported on $\mathbb{R} \setminus T$ (that is, the measurements in (1.1)) belongs to the direct sum of wavelet subspaces $\oplus_{\ell > j} W_\ell$; functions belonging to $\oplus_{\ell > j} W_\ell$ generally have a certain number of vanishing moments, [40].

3.3. Relation with the Minimum-Norm (MN) solution

Point (2) in Corollary 3 has an interesting consequence; namely, considering the so-called *Minimum-Norm (MN) solution* as proposed in Theorem 1 in [12], it is shown that the iteration limit $\bar{s}$ of (1.4) admits the following minimization formulation:

$$\|\bar{s}\|_{L^2(\mathbb{R})} = \inf_{f \in V}\{\|f\|_{L^2(\mathbb{R})} \text{ such that } \mathcal{T}f = \mathcal{T}s \text{ for } s \in V\}. \tag{3.3}$$

This is one of the "best approximation problems" considered in [41], Sections 5 and 6. First, as soon as the invertibility condition $\|\mathcal{Q}\mathcal{P}\| < 1$ is met (and in particular, for any band-limited extrapolation problem, see [25]), this formulation is not relevant. In the special case where one deals with a "zero-angle" problem for which $\cos(\mathrm{ran}(\mathcal{P}), \mathrm{ran}(\mathcal{Q})) = 1$, the dimension of $\ker(\mathrm{Id} - \mathcal{Q}\mathcal{P})$ is strictly positive and one must restrict Eq. (1.2) to $r \in \mathrm{ran}(\mathrm{Id} - \mathcal{Q}\mathcal{P}) = \ker(\mathrm{Id} - \mathcal{P}\mathcal{Q})^\perp$ thus satisfying a finite number of orthogonality conditions; see especially the Comment and Corollary 2 in [5] (page 698).

Let us begin by recalling a result from *e.g.* [42]:

Lemma 5. *Let $\mathcal{H}$ be a Hilbert space and $M_1, M_2, \ldots, M_K$ be a family of closed linear subspaces of $\mathcal{H}$; if $M \doteq \cap_{i=1}^K M_i$ denotes the (closed) intersection of the M_i's and P_{M_i} is the orthogonal projection on M_i, then for all $x \in \mathcal{H}$, there holds:*

$$\lim_{n \to +\infty} \|(P_{M_K} \circ \cdots \circ P_{M_2} \circ P_{M_1})^n x - P_M x\| = 0. \tag{3.4}$$

Since $\mathcal{Q}$ is an orthogonal projection and thanks to the assumption $r \in \mathrm{ran}(\mathrm{Id} - \mathcal{Q}\mathcal{P})$, we have that $r = s - \mathcal{Q}\mathcal{P}s$ and the corresponding $s \in V$ decomposes into $s = (\mathrm{Id} - \mathcal{Q})s + \mathcal{Q}s = r + \mathcal{Q}\mathcal{P}s$, inside which one can plug again the decomposition $s = r + \mathcal{Q}\mathcal{P}s$ in order to obtain $s = r + \mathcal{Q}\mathcal{P}(r + \mathcal{Q}\mathcal{P}s) = r + \mathcal{Q}\mathcal{P}r + (\mathcal{Q}\mathcal{P})^2 s$. Denoting $s^{(k)}$ the kth iterate of (1.4), one gets

$$s = \sum_{i=0}^{k} (\mathcal{Q}\mathcal{P})^i r + (\mathcal{Q}\mathcal{P})^{k+1} s \Leftrightarrow s^{(k)} = s - (\mathcal{Q}\mathcal{P})^{k+1} s.$$

The orthogonal projections $\mathcal{Q}$ and $\mathcal{P}$ satisfy the hypotheses of Lemma 5, hence we can deduce that $\bar{s} = \lim_{k \to +\infty} s^{(k)} = s - P_{\mathrm{ran}(\mathcal{P}) \cap \mathrm{ran}(\mathcal{Q})}(s)$ where $P_{\mathrm{ran}(\mathcal{P}) \cap \mathrm{ran}(\mathcal{Q})}$ stands for the orthogonal projection onto the intersection of V and the subspace of functions supported in T. At this level, one observes that the condition $\|\mathcal{Q}\mathcal{P}\| < 1$ implies that $\mathrm{ran}(\mathcal{P}) \cap \mathrm{ran}(\mathcal{Q}) = \{0\}$, so $\lim_{k \to +\infty} s^{(k)} = s$. For a zero-angle extrapolation problem, this property does not hold and the limit $\bar{s}$ can be characterized by the minimal property of any orthogonal projection:

$$\|\bar{s}\|_{L^2(\mathbb{R})} = \|s - P_{\mathrm{ran}(\mathcal{P}) \cap \mathrm{ran}(\mathcal{Q})}(s)\|_{L^2(\mathbb{R})} = \inf_{f \in V \cap \mathrm{ran}(\mathcal{Q})} \|s - f\|_{L^2(\mathbb{R})}.$$

As a consequence of both the preceding equality and the orthogonal decomposition $L^2(\mathbb{R}) = \mathrm{ran}(\mathcal{Q}) \oplus \mathrm{ran}(\mathcal{Q})^\perp$, picking any $g \in V$ such that $(\mathrm{Id} - \mathcal{Q})g = (\mathrm{Id} - \mathcal{Q})s$ yields $s - g = \mathcal{Q}(s - g) \in V \cap \mathrm{ran}(\mathcal{Q})$ which leads to $\|\bar{s}\|_{L^2(\mathbb{R})} = \inf_{g \in V} \|s - (s - g)\|_{L^2(\mathbb{R})} = \inf_{g \in V} \|g\|_{L^2(\mathbb{R})}$.

This MN solution emerging from (1.4) is unstable in the very general situation considered in [5]. Here we limit ourselves to a somewhat simpler case for which $\mathcal{Q}\mathcal{P}$ is compact which yields that $\mathrm{Id} - \mathcal{Q}\mathcal{P}$ is a Fredholm operator with closed range. Hence a slight perturbation $r + \nu$ of $r \in \mathrm{ran}(\mathrm{Id} - \mathcal{Q}\mathcal{P})$ will still belong to the range of $\mathrm{Id} - \mathcal{Q}\mathcal{P}$ if ν is small enough. However, this notion of solution does not allow us to treat problems like (1.2) for which $\|\mathcal{Q}\mathcal{P}\| = 1$ and r admits an orthogonal projection onto $\ker(\mathrm{Id} - \mathcal{P}\mathcal{Q})$; they are in the next section.

3.4. Singular Value Decomposition (SVD) and eigenfunctions of $\mathcal{Q}\mathcal{P}_{V_j}\mathcal{Q}$

With the previous notations, let us now look at $\mathcal{A} = \mathcal{Q}\mathcal{P}$ as a bounded operator defined on the Hilbert space $\mathcal{H} = L^2(\mathbb{R})$ which is assumed to be compact and non-self-adjoint. It is easy to see that $\mathrm{ran}(\mathcal{P})^\perp = \mathrm{ran}(\mathrm{Id} - \mathcal{P}) \subset \ker(\mathcal{Q}\mathcal{P})$. Thus, if we define

$$\mathcal{A} = \mathcal{Q}\mathcal{P} : V = \mathrm{ran}(\mathcal{P}) \to \mathrm{ran}(\mathcal{Q}), \qquad \mathcal{A}^* = \mathcal{P}\mathcal{Q} : \mathrm{ran}(\mathcal{Q}) \to V = \mathrm{ran}(\mathcal{P}), \tag{3.5}$$

the standard SVD theory for compact operators in Hilbert spaces gives

$$\mathcal{A}^*\mathcal{A}\psi_k = \mathcal{P}\mathcal{Q}\mathcal{P}\psi_k = \lambda_k\psi_k, \qquad \mathcal{A}\mathcal{A}^*\varphi_k = \mathcal{Q}\mathcal{P}\mathcal{Q}\varphi_k = \lambda_k\varphi_k, \quad \lambda_k \geq 0,\ k \in \mathbb{N}, \tag{3.6}$$

where $\mathcal{A}^* = \mathcal{P}\mathcal{Q}$ is the adjoint of $\mathcal{A}$. Moreover,

$$\mathcal{A}\psi_k = \mathcal{Q}\mathcal{P}\psi_k = \sqrt{\lambda_k}\varphi_k, \qquad \mathcal{A}^*\varphi_k = \mathcal{P}\mathcal{Q}\varphi_k = \sqrt{\lambda_k}\psi_k.$$

Clearly, definition (3.5) implies that $\mathcal{A}^*\mathcal{A} = \mathcal{P}\mathcal{Q}\mathcal{P} : V \to V$, $\mathcal{A}\mathcal{A}^* : \mathrm{ran}(\mathcal{Q}) \to \mathrm{ran}(\mathcal{Q})$ are self-adjoint and therefore:

- $(\psi_k)_{k\in\mathbb{Z}}$ is an orthonormal base of $\overline{\mathrm{ran}(\mathcal{P}\mathcal{Q}\mathcal{P})} = \overline{\mathrm{ran}(\mathcal{P}\mathcal{Q})}$,
- $(\varphi_k)_{k\in\mathbb{Z}}$ is an orthonormal base of $\overline{\mathrm{ran}(\mathcal{Q}\mathcal{P}\mathcal{Q})} = \overline{\mathrm{ran}(\mathcal{Q}\mathcal{P})}$.

The singular values λ_k are smaller than 1 since $\mathcal{A}$ is a composition of 2 orthogonal projections; there also holds $\|\mathcal{A}\| = \sqrt{\lambda_0}$. In the context of standard band-limited extrapolation, the collection of functions (φ_k, ψ_k), normalized[2] as $\|\psi_k\|_{L^2(\mathbb{R})} = 1$ is called the Prolate Spheroidal Wave functions (PSWF). They have been studied in detail by Slepian, Landau and Pollak; see [44,22], the book [45] and the surveys [29,46] for more on this topic. The function ψ_0 is an extremal function of the type studied in Section 4 of [41].

Lemma 6. *If $\mathcal{Q}\mathcal{P}$ is a compact operator, then its singular functions (3.6) satisfy*

$$\forall k \in \mathbb{N}, \quad \frac{\mathcal{P}\varphi_k}{\sqrt{\lambda_k}} = \psi_k. \tag{3.7}$$

Proof. From $\mathcal{Q}\mathcal{P}\mathcal{Q}\varphi_k = \lambda_k\varphi_k$, we get from (3.6) that

$$\lambda_k\mathcal{P}\varphi_k = \mathcal{P}\mathcal{Q}(\mathcal{P}\mathcal{Q}\varphi_k) = \mathcal{P}\mathcal{Q}\mathcal{P}(\sqrt{\lambda_k}\varphi_k) = \lambda_k\sqrt{\lambda_k}\varphi_k. \quad \square$$

Here, we try to compute numerically the analogs of the functions φ_k when $\mathcal{P}$ is assumed to be the orthogonal projection onto an MRA subspace with a given index $j \in \mathbb{Z}$ for various choices of the scaling function. In Figs. 2 and 3, we display the first 10 functions satisfying $\mathcal{Q}\mathcal{P}_{V_j}\mathcal{Q}\varphi_k = \lambda_k\varphi_k$ splitting between the even and odd ones. One can easily see that the shape of the scaling function appears very clearly in these eigenfunctions which are quite different from one another according to the choice of the scaling function: see in particular the ones emerging from the Daubechies 4 compared to the Symmlet 10. The eigenfunctions coming out of the Coiflet 5 scaling functions have also a particular shape. The behavior of the eigenvalues λ_k is presented for each choice of the scaling function; however, even if we displayed only the 10 first eigenfunctions (corresponding to eigenvalues very close to 1), we chose to show the whole set of numerical eigenfunctions. Possible inaccuracies may be present because the linear system is ill-conditioned and difficult to diagonalize efficiently. The matrices were 256×256 or 512×512 and the scale index $j = 4$ or $j = 5$; the discrete wavelet transform involves a periodization of the signal. These numerical results follow early computations displayed in [20]. Classical PSWF were computed using the algorithms proposed in [43] with 512 points and the Slepian parameter $c = 13$.

4. The case $\|\mathcal{Q}\mathcal{P}\| = 1$: Minimum-Norm Least Squares (MNLS) solution

In sharp contrast with band-limited extrapolation for which the property $\|\mathcal{Q}\mathcal{P}\| < 1$ generally holds (which is another form of the analytic extension principle for functions belonging to the Paley–Wiener space PW_ω), it is easy to see that for MRA subspaces, the "bad case" $\|\mathcal{Q}\mathcal{P}_{V_j}\| = 1$ can happen. Indeed, pick $T = [-a, a]$ and a scaling function $\phi_{j,n}$ of compact support: clearly $\|\mathcal{Q}\mathcal{P}_{V_j}\| = 1$ for $2a \gg |\mathrm{supp}(\phi_{j,0})|$. Throughout this section, we shall write $\mathcal{P}$ instead of $\mathcal{P}_{V_j}$ for simplicity.

4.1. Least squares solution and normal equations

In the context of band-limited extrapolation, MNLS solutions have been studied numerically in [3] who tackled directly Eqs. (1.1).

Definition 4. Let $F : X \to Y$ be a bounded linear operator with X, Y two Hilbert spaces.

(1) $\bar{x} \in X$ is called a *least squares solution* of $Fx = y$ if

$$\|F\bar{x} - y\|_Y = \inf_{z\in X}\{\|Fz - y\|_Y\}.$$

(2) $x^\dagger \in X$ is called the "best approximate solution" if it is a least squares solution with minimum norm, that is

$$\|x^\dagger\|_X = \inf_{z\in X}\{\|z\|_X \text{ with } z \text{ is a least squares solution of } Fx = y\}.$$

[2] Numerically, the normalization is generally taken as $\|\varphi_k\|_{L^2(\mathbb{R})} = 1$ (see *e.g.* [43]), which implies $\|\psi_k\|_{L^2(\mathbb{R})} \to \infty$ as $k \to \infty$ as a consequence of Lemma 6.

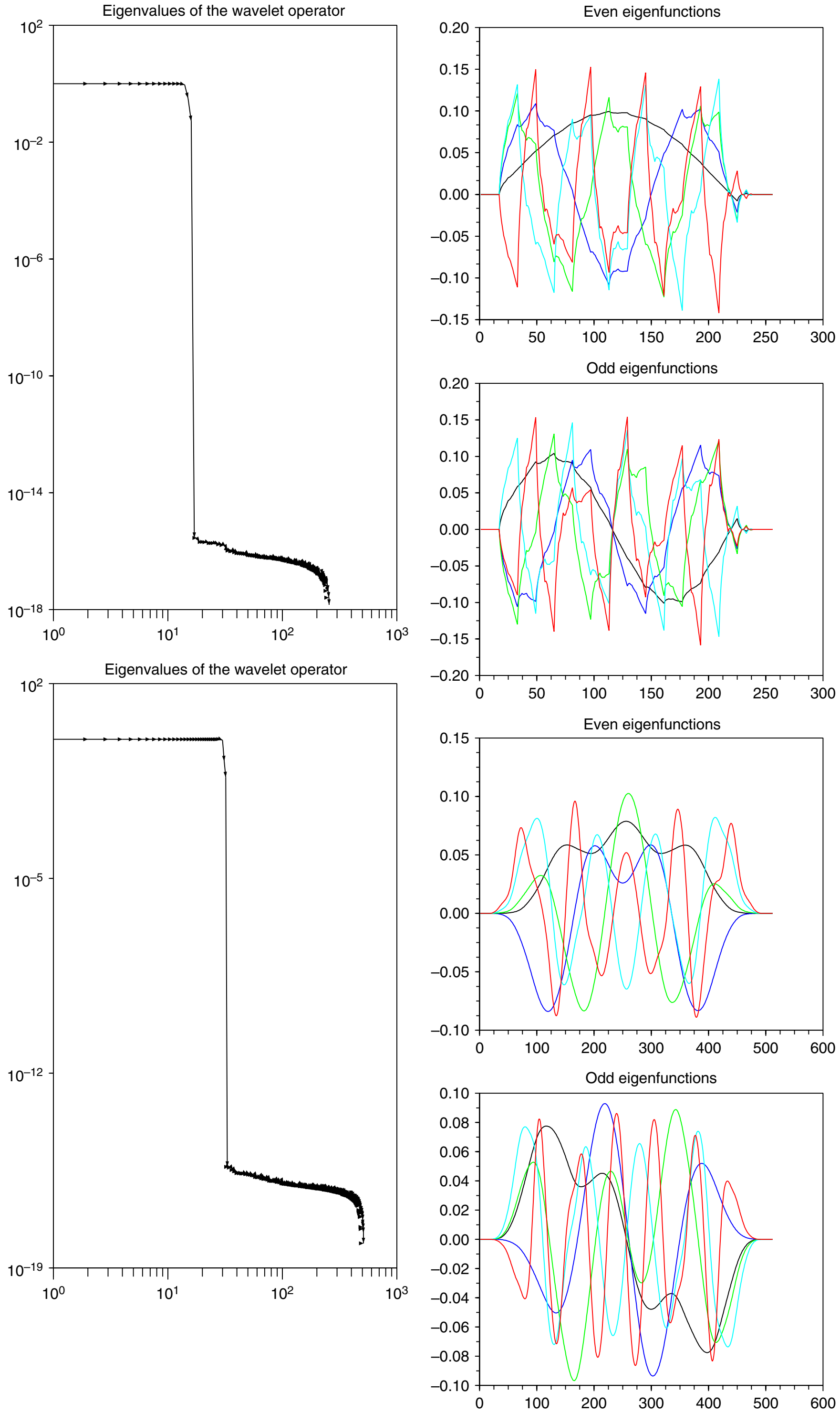

Fig. 2. Eigenvalues and eigenfunctions for $\mathcal{Q}\mathcal{P}_{V_j}\mathcal{Q}$ with Daubechies 4 (top) and Coiflet 5 (bottom) scaling functions.

It is well-known that x is a least squares solutions if and only if it satisfies the so-called "normal" equation $F^*Fx = F^*y$; in infinite dimension, this modified problem may have no solution. However, in the case where $\text{ran}(F)$ is closed, the set of all least squares solutions is a nonempty convex set which therefore admits a unique element of minimum norm. Hence in this context, it makes sense to speak about "the best approximate solution" $x^\dagger$ of $Fx = y$ which is also referred to as its *Minimum-Norm Least Squares* (MNLS) solution.

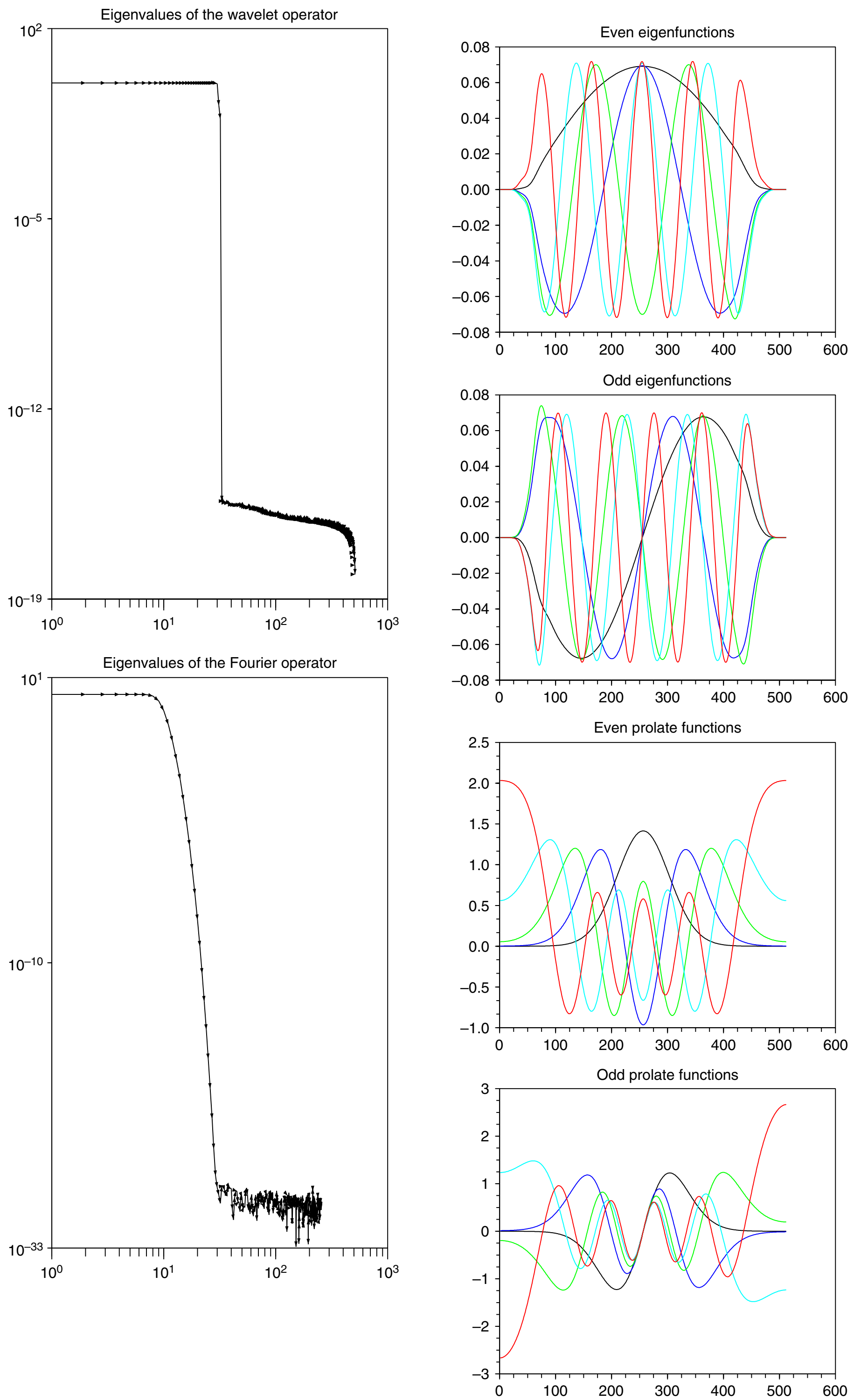

Fig. 3. Eigenvalues and eigenfunctions for $\mathcal{Q}\mathcal{P}_{V_j}\mathcal{Q}$ with Symmlet 10 (top) and sin c (bottom) scaling functions.

Dealing with operators with a closed range brings many advantages when it comes to solving equations like (1.1); however, except in the case where ran$(\mathcal{T}\mathcal{P})$ is finite dimensional (thus closed, which is an assumption in [3]), the operator $\mathcal{A} = \mathcal{T}\mathcal{P}$ is compact and its range is generally not closed (see [2] for more details). Thus it is forbidden to speak about the MNLS of (1.1) without supplementary assumptions. It is therefore interesting to once again switch to the formulation (1.2) involving a Fredholm operator, which may be singular in the sense that $\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})\neq\{0\}$, but for which $\mathrm{ran}(\mathrm{Id}-\mathcal{Q}\mathcal{P})$ is always closed.

The normal equations for (1.2) read

$$(\mathrm{Id}-\mathcal{P}\mathcal{Q})(\mathrm{Id}-\mathcal{Q}\mathcal{P})s=(\mathrm{Id}-[\underbrace{\mathcal{P}\mathcal{Q}+\mathcal{Q}\mathcal{P}-\mathcal{P}\mathcal{Q}\mathcal{P}}_{\mathcal{P}\mathcal{Q}\circ(\mathrm{Id}-\mathcal{P})+\mathcal{Q}\circ\mathcal{P}}])s=(\mathrm{Id}-\mathcal{P}\mathcal{Q})r. \tag{4.1}$$

Clearly, since $\ker(\mathrm{Id}-\mathcal{P}\mathcal{Q})^{\perp}=\mathrm{ran}(\mathrm{Id}-\mathcal{Q}\mathcal{P})$, the right-hand side satisfies

$$(\mathrm{Id}-\mathcal{P}\mathcal{Q})r=(\mathrm{Id}-\mathcal{P}\mathcal{Q})(r-\mathbf{P}_{\ker(\mathrm{Id}-\mathcal{P}\mathcal{Q})}r)=(\mathrm{Id}-\mathcal{P}\mathcal{Q})\mathbf{P}_{\mathrm{ran}(\mathrm{Id}-\mathcal{Q}\mathcal{P})}r.$$

Because of the hypothesis $\|\mathcal{Q}\mathcal{P}\|=\|\mathcal{P}\mathcal{Q}\|=1$, the formal Neumann series for inverting (4.1) may not converge since for any $x\in\mathrm{ran}(\mathcal{P})$, $[\mathcal{P}\mathcal{Q}+\mathcal{Q}\mathcal{P}-\mathcal{P}\mathcal{Q}\mathcal{P}]x=\mathcal{Q}\mathcal{P}x$ and for $x'\in\mathrm{ran}(\mathcal{P})^{\perp}$, $[\mathcal{P}\mathcal{Q}+\mathcal{Q}\mathcal{P}-\mathcal{P}\mathcal{Q}\mathcal{P}]x'=\mathcal{P}\mathcal{Q}x'$. However, this formal series is equivalent to the following iterative scheme:

$$s^0=(\mathrm{Id}-\mathcal{P}\mathcal{Q})r,\qquad s^{k+1}=(\mathrm{Id}-\mathcal{P}\mathcal{Q})r+[\mathcal{P}\mathcal{Q}+\mathcal{Q}\mathcal{P}-\mathcal{P}\mathcal{Q}\mathcal{P}]s^k,$$

which is a special case of the following "steepest descent" algorithm:

$$s^{k+1}=s^k-\alpha_k(\mathrm{Id}-\mathcal{P}\mathcal{Q})[(\mathrm{Id}-\mathcal{Q}\mathcal{P})s^k-r], \tag{4.2}$$

with the particular choice $\alpha_k\equiv 1$ and $s^0=(\mathrm{Id}-\mathcal{P}\mathcal{Q})r$. By its construction, all the iterates of the algorithm (4.2) belong to $\mathrm{ran}(\mathrm{Id}-\mathcal{Q}\mathcal{P})^*$ as soon as the initial value s^0 does; this fact is used in [3] in order to reduce the complexity for band-limited extrapolation when $|T|$ is large. Indeed, one sees that starting from $z^0=r$ and then defining an auxiliary sequence,

$$z^{k+1}=z^k-\alpha_k[(\mathrm{Id}-\mathcal{Q}\mathcal{P})(\mathrm{Id}-\mathcal{P}\mathcal{Q})z^k-r],$$

one recovers any of the s^k values in (4.2) for some $k\in\mathbb{N}$ by computing $s^k=(\mathrm{Id}-\mathcal{P}\mathcal{Q})z^k$; this allows us to iterate inside $\mathrm{ran}(\mathrm{Id}-\mathcal{Q}\mathcal{P})$ only. Related references are [9,17].

4.2. *Gradient algorithms for singular operators with closed range*

General convergence results of gradient algorithms for singular operators with closed range in Hilbert spaces have been proved in [21,47]: we are about to adapt them now to our particular extrapolation/recovery problem.

Theorem 6. *Consider the Fredholm operator* $\mathrm{Id}-\mathcal{Q}\mathcal{P}:\mathrm{ran}(\mathcal{P})\to\mathrm{ran}(\mathcal{T})$ *with* $\|\mathcal{Q}\mathcal{P}\|=1$*: the sequence* $(s^k)_{k\in\mathbb{N}}$ *generated by* (4.2) *with*

$$q_k=(\mathrm{Id}-\mathcal{P}\mathcal{Q})[(\mathrm{Id}-\mathcal{Q}\mathcal{P})s^k-r],\qquad \alpha_k=\frac{\|q_k\|^2}{\|(\mathrm{Id}-\mathcal{Q}\mathcal{P})q_k\|^2},$$

converges in $L^2(\mathbb{R})$ *toward a least squares solution* $\bar{s}$ *which depends on the initial value:*

$$\bar{s}=s^{\dagger}+\mathbf{P}_{\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})}s^0,$$

where $\mathbf{P}_{\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})}$ *stands for the orthogonal projection onto the kernel of* $\mathrm{Id}-\mathcal{Q}\mathcal{P}$*. In the case where the initial value satisfies*

$$s^0\in\mathrm{ran}(\mathrm{Id}-\mathcal{P}\mathcal{Q})=\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})^{\perp},$$

the sequence $(s^k)_{k\in\mathbb{N}}$ *converges toward the MNLS* $s^{\dagger}$ *of the equation* $(\mathrm{Id}-\mathcal{Q}\mathcal{P})s=r$.

The conclusion of Theorem 6 still holds for the simplified version of the algorithm obtained by fixing a constant value of α_k as long as it is smaller than $2/\|\mathrm{Id}-\mathcal{Q}\mathcal{P}\|$; clearly, $\|\mathrm{Id}-\mathcal{Q}\mathcal{P}\|=\|(\mathrm{Id}-\mathcal{Q})\mathcal{P}+(\mathrm{Id}-\mathcal{P})\|\leq\|(\mathrm{Id}-\mathcal{Q})\mathcal{P}\|+\|\mathrm{Id}-\mathcal{P}\|=2$. Hence the case $\alpha_k\equiv 1$ is admissible and the formal Neumann series coming from the normal equation is convergent:

$$s^{\dagger}=\sum_{k\geq 0}[\mathcal{P}\mathcal{Q}+\mathcal{Q}\mathcal{P}-\mathcal{P}\mathcal{Q}\mathcal{P}]^k\,(\mathrm{Id}-\mathcal{P}\mathcal{Q})r.$$

Remark 2. (1) The "closed range hypothesis" is crucial here: for some results valid in the case where it is bypassed, see *e.g.* [48]. They allow us to invert equations involving a compact operator but the obtained solution is unstable.

(2) The expression of $\mathbf{P}_{\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})}$ can be made more explicit by observing that $\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})=\mathrm{ran}(\mathcal{P})\cap\mathrm{ran}(\mathcal{Q})\neq\{0\}$ (see Lemma 2.2 in [19]). Since these ranges are closed,

$$\begin{aligned}\mathrm{ran}(\mathrm{Id}-\mathcal{P}\mathcal{Q})&=\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})^{\perp}\\&=(\mathrm{ran}(\mathcal{P})\cap\mathrm{ran}(\mathcal{Q}))^{\perp}\supset\overline{\mathrm{ran}(\mathcal{P})^{\perp}+\mathrm{ran}(\mathcal{Q})^{\perp}}\\&=(\mathrm{ran}(\mathrm{Id}-\mathcal{P})^{\perp}\cap\mathrm{ran}(\mathrm{Id}-\mathcal{Q})^{\perp})^{\perp}\\&=\overline{\mathrm{ran}(\mathrm{Id}-\mathcal{P})+\mathrm{ran}(\mathrm{Id}-\mathcal{Q})}\\&=\overline{\ker(\mathcal{P})+\ker(\mathcal{Q})}.\end{aligned}$$

According to Corollary 3, the condition $\|\mathcal{Q}\mathcal{P}\|<1$ ensures that $\mathrm{ran}(\mathcal{P})\cap\mathrm{ran}(\mathcal{Q})=\{0\}=(\ker(\mathcal{P})+\ker(\mathcal{Q}))^{\perp}$, but here it does not hold, thus $\mathrm{ran}(\mathrm{Id}-\mathcal{P}\mathcal{Q})\neq L^2(\mathbb{R})$.

(3) The MNLS $s^\dagger$ also belongs to $\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})^\perp = \overline{\ker(\mathcal{P})+\ker(\mathcal{Q})}$, which contains all the iterates s^k as soon as $s^0 \in \ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})^\perp$.
(4) Stability of the MNLS $s^\dagger$ in the presence of additive noise $\nu(t)$ in the observations is ensured by the boundedness of the so-called Moore–Penrose generalized inverse of (4.1); see *e.g.* [3,48,47,19,2] for details.
(5) The issue of the continuous dependence of $s^\dagger$ with respect to the measure of T or the scale index j is delicate, as explained in Section 4 of [49].

The convergence of the steepest descent algorithm can be slow despite it already being faster than the standard Gerchberg–Papoulis iterates. Hence it makes sense to speed it up by setting up a Conjugate Gradient (CG) routine as follows: let $s^0 \in \mathrm{ran}(\mathcal{P})$ and compute $v^0 = p^0 = (\mathrm{Id}-\mathcal{P}\mathcal{Q})((\mathrm{Id}-\mathcal{Q}\mathcal{P})s^0 - r)$. If $p^0 \neq 0$, then $s^1 = s^0 - \alpha_0 p^0$ with $\alpha_0 = \|v^0\|^2/\|(\mathrm{Id}-\mathcal{Q}\mathcal{P})v^0\|^2$ like in the former algorithm. Now, for $k \in \mathbb{N}$,

$$\alpha_{k-1} = \frac{\langle v^{k-1}, p^{k-1}\rangle}{\|(\mathrm{Id}-\mathcal{Q}\mathcal{P})p^{k-1}\|^2}, \quad v^k = v^{k-1} - \alpha_{k-1}(\mathrm{Id}-\mathcal{P}\mathcal{Q})(\mathrm{Id}-\mathcal{Q}\mathcal{P})p^{k-1}, \tag{4.3}$$

and as long as $v^k \neq 0$ of $\|v^k\| \geq \varepsilon$ with ε a small positive number, compute

$$\beta_{k-1} = \frac{\langle v^k, (\mathrm{Id}-\mathcal{P}\mathcal{Q})(\mathrm{Id}-\mathcal{Q}\mathcal{P})p^{k-1}\rangle}{\|(\mathrm{Id}-\mathcal{Q}\mathcal{P})p^{k-1}\|^2}, \quad p^k = v^k - \beta_{k-1}p^{k-1}. \tag{4.4}$$

When $\|v^k\| < \varepsilon$, it remains to set $s^k = s^0 - \sum_{\ell=0}^k \alpha_\ell p^\ell$. Along with the computation of the iterates, it is interesting to compute the following function:

$$g(s^k) = \langle v^k, s^k - s^\dagger\rangle, \quad s^\dagger \text{the MNLS of (1.2)},$$

as it satisfies the following relation (see [47]):

$$g(s^k) = g(s^{k-1}) - \alpha_{k-1}\|v^{k-1}\|^2.$$

At last, we define the two positive numbers $0 < m \leq M$ as the spectral bounds of the operator R which is defined as the restriction of $(\mathrm{Id}-\mathcal{P}\mathcal{Q})(\mathrm{Id}-\mathcal{Q}\mathcal{P})$ to $\mathrm{ran}(\mathrm{Id}-\mathcal{P}\mathcal{Q})$:

$$\forall x \in \mathrm{ran}(\mathrm{Id}-\mathcal{P}\mathcal{Q}), \quad m\|x\| \leq \|Rx\| \leq M\|x\|.$$

We can adapt the convergence result of [47] to our context as follows.

Theorem 7. *Let $s^0 \in \mathrm{ran}(\mathcal{P})$, the iterates (4.3)–(4.4) generate a sequence $(s^k)_{k\in\mathbb{N}}$ which converges monotonically toward a least squares solution of (1.2) $\bar{s} = s^\dagger + \mathbf{P}_{\ker(\mathrm{Id}-\mathcal{Q}\mathcal{P})}s^0$. In the case where $s^0 \in \mathrm{ran}(\mathrm{Id}-\mathcal{P}\mathcal{Q}) = \overline{\ker(\mathcal{P})+\ker(\mathcal{Q})}$, one has the decay estimate:*

$$\|s^k - s^\dagger\| \leq \sqrt{\frac{g(s^0)}{m}}\left(\frac{M-m}{M+m}\right)^k.$$

In general, one gives an initial value $s^0 \in \mathrm{ran}(\mathcal{T})$; for scale-limited extrapolation, $\mathcal{P} = \mathcal{P}_{V_j}$ and $\ker(\mathcal{P}_{V_j}) = \mathrm{ran}(\mathrm{Id}-\mathcal{P}_{V_j}) = \oplus_{\ell>j} W_\ell$, the wavelet subspaces satisfying $V_{j+1} = V_j \oplus W_j$ and containing functions with vanishing moments. The simplest choice is of course $s^0 = 0$.

Corollary 4. *Let V_j be an MRA subspace and T such that $\|\mathcal{Q}\mathcal{P}_{V_j}\| = 1$: the MNLS solution $s^\dagger$ of (1.2) satisfies*

$$s^\dagger \in \mathrm{ran}(\mathrm{Id}-\mathcal{P}_{V_j}\mathcal{Q}) = \overline{\oplus_{\ell>j} W_\ell + \mathrm{ran}(\mathcal{T})}.$$

In particular, it can happen that $s^\dagger \notin V_j$ and $\mathcal{T}\mathcal{P}_{V_j}s^\dagger = (\mathrm{Id}-\mathcal{Q})\mathcal{P}_{V_j}s^\dagger \neq (\mathrm{Id}-\mathcal{Q}\mathcal{P}_{V_j})s^\dagger$.

Proof. This is mainly a consequence of Point (2) in Remark 2. □

In [47], the convergence of the algorithm (4.3)–(4.4) in the case where one deals with a bounded singular operator with non-closed range is also established under some supplementary hypotheses on r and s^0. This is the situation arising when trying to compute the least squares solution of the Eqs. (1.1) in infinite dimension: this has been done numerically (in finite dimension) for band-limited extrapolation in [3] and for scale-limited extrapolation in [12].

5. Conclusion

We presented in this paper a rather simple and explicit criterion allowing us to estimate the operator norm $\|\mathcal{Q}\mathcal{P}\|$ which controls the stability of the extrapolation process in the particular case where $\mathcal{P}$ is the orthogonal projector onto one of the nested subspaces of an MRA. The choice of the scaling function does not appear in the computation, and it is not required that the "hole" T should be an interval of $\mathbb{R}$. Geometric harmonics for several choices of the scaling functions are also displayed together with their corresponding eigenvalues which show a sharp decay from nearly 1 to zero beyond a certain level. These results allow us to give a precise answer to a question raised in [11] and also [50] in the context of a peculiar application. Concerning the extension of these Donoho–Stark type criteria, the cases of finite-dimensional problems and sparse Compressed Sensing situations have been treated in [29,28]. More elaborate integral transforms are studied in [26] and PSWF for the fractional Fourier transform are computed in [23,24] to which a similar approach might be applied.

Appendix A. Proof of Lemma 1

Clearly, $\|P_A P_B\| \leq \|P_A\| \, \|P_B\| \leq 1$; we split the proof into several steps:

(1) For any bounded operator T on $\mathcal{H}$, let $\rho(T)$ stand for the spectral radius of T, *i.e.* $\rho(T) = \sup |\lambda|$ for the λ such that $T - \lambda I$ is not invertible. Then $\rho(T) = \limsup \|T^n\|^{1/n}$ when $n \to +\infty$. This yields in particular $\rho(T) \leq \|T\|$.
(2) For T self-adjoint on $\mathcal{H}$, one has $|Tf|^2 = (Tf, Tf) = (f, T^2 f) \leq |f||T^2 f| \leq |f|^2 \|T^2\|$, hence $\|T\| \leq \|T^2\|^{1/2}$, and $\|T\| \leq \|T^n\|^{1/n}$ for $n = 2^p$, which implies $\|T\| \leq \rho(T)$ and by the preceding step, $\|T\| = \rho(T)$.
(3) If T, U are self-adjoint and invertible, $\|TU\| = \rho(UTTU)^{1/2}$, $\|UT\| = \rho(TUUT)^{1/2}$. But $TUUT = TU(UTTU)(TU)^{-1}$, so $TUUT$ and $UTTU$ are similar and have the same spectrum (thus the same spectral radius). This implies that $\|TU\| = \|UT\|$.
(4) The preceding step is still valid when T, U are self-adjoint and limits of self-adjoint and invertible operators; this is the case for any two orthogonal projections P_A and P_B which can be approximated by themselves plus εId with $\varepsilon > 0$ a small real number and Id the identity mapping. □

Appendix B. The Zak transform: definition and properties

In this appendix section, we limit ourselves to recalling some basic facts about the Zak transform originally introduced in the context of solid-state physics. Following [39], we have:

Definition 5. Let f be a continuous function decaying at least like $C(1+|t|)^{-1-\varepsilon}$ with $\varepsilon > 0$ as $|t| \to +\infty$. The *Zak transform* of f is defined as

$$\forall t, \xi \in \mathbb{R}^2, \quad \mathcal{Z}f(t,\xi) = \sum_{k\in\mathbb{Z}} f(t+k)\exp(-2i\pi k\xi). \tag{B.1}$$

It is clear that $\mathcal{Z}f$ is periodic:

$$\mathcal{Z}f(t,\xi+1) = \mathcal{Z}f(t,\xi), \qquad \mathcal{Z}f(t+1,\xi) = \exp(2\pi i\xi)\mathcal{Z}f(t,\xi).$$

Hence $\mathcal{Z}$ maps a function defined on $\mathbb{R}$ to another which is fully determined by its restriction to the torus $\mathbb{T} = [0,1]^2$ in the time/frequency plane. Let $\hat{f}(\xi) = \int_{\mathbb{R}} f(t)\exp(-2i\pi t\xi)\mathrm{d}t$ be the Fourier transform of $f \in L^2(\mathbb{R})$, there holds: $\forall t, \xi$,

$$\mathcal{Z}f(t,\xi) = \exp(2i\pi t\xi)\mathcal{Z}\hat{f}(-\xi,t), \qquad \mathcal{Z}\hat{f}(\xi,t) = \exp(2i\pi t\xi)\mathcal{Z}\hat{f}(t,-\xi). \tag{B.2}$$

Moreover, there are inversion formulas

$$f(t) = \int_0^1 \mathcal{Z}f(t,\xi)\mathrm{d}\xi, \qquad \hat{f}(\xi) = \int_0^1 \exp(-2i\pi\xi t)\mathcal{Z}f(t,\xi)\mathrm{d}t. \tag{B.3}$$

The Zak transform is an isometry from $L^2(\mathbb{R})$ onto $L^2(\mathbb{T})$ because for any f, g,

$$\int_{\mathbb{T}} \mathcal{Z}f(t,\xi)\overline{\mathcal{Z}g(t,\xi)}\mathrm{d}t.\mathrm{d}\xi = \int_{\mathbb{R}} f(t)\bar{g}(t).\mathrm{d}t.$$

References

[1] C.H. Chang, C.W. Ha, Sharp inequalities of singular values of smooth kernels, Integral Equations Operator Theory 35 (1999) 20–27.
[2] S.H. Kulkarni, M.T. Nair, A characterization of closed range operators, Indian J. Pure Appl. Math. 31 (4) (2000) 353–361.
[3] A.K. Jain, S. Ranganath, Extrapolation algorithm for discrete signal with application in spectral estimation, IEEE Trans. Acoust. Speech Signal Process. 29 (1981) 830–845.
[4] D.L. Donoho, P.B. Stark, Uncertainty principles and signal recovery, SIAM J. Appl. Math. 49 (1989) 906–931.
[5] D.C. Youla, Generalized image restoration by the method of alternating orthogonal projections, IEEE Trans. Circuits Syst. 25 (1978) 694–702.
[6] R.W. Gerchberg, Super-resolution through error energy reduction, Opt. Acta 21 (1974) 709–720.
[7] A. Papoulis, A new algorithm in spectral analysis and band-limited extrapolation, IEEE Trans. Circuits Syst. 22 (1975) 735–742.
[8] Ch. Byrne, A unified treatment of some iterative algorithms in signal processing and image reconstruction, Inverse Problems 20 (2004) 103–120.
[9] T.J. Connolly, K.A. Landman, L.R. White, About Gerchberg's method for the inverse Fourier problem, J. Aust. Math. Soc. Ser. B 37 (1995) 26–44.
[10] A.K. Katsaggelos, S.N. Efstratiadis, A class of iterative signal restoration algorithms, IEEE Trans. Acoust. Speech Signal Process. 38 (1990) 778–786.
[11] L.-C. Lin, C.-C. Jay Kuo, On theory and regularization of scale-limited extrapolation, Signal Process. 54 (1996) 225–237.
[12] L.-C. Lin, X.-G. Xia, C.-C. Jay Kuo, On the convergence of wavelet-based iterative signal extrapolation algorithms, Signal Process. 48 (1996) 51–65.
[13] X.-G. Xia, C.-C. Jay Kuo, Z. Zhang, Signal extrapolation in wavelet subspaces, SIAM J. Sci. Comput. 16 (1995) 50–73.
[14] X.-G. Xia, L.-C. Lin, C.-C. Jay Kuo, On the limit of sampled signal extrapolation with a wavelet signal model, Numer. Funct. Anal. Optim. 16 (1995) 1255–1272.
[15] H.G. Feichtinger, K. Gröchenig, Theory and Practice of Irregular Sampling, in: J. Benedetto, M. Frazier (Eds.), Wavelets: Mathematics and Applications, CRC Press, 1993, pp. 305–363.
[16] K. Gröchenig, Irregular sampling, Toeplitz matrices and the approximation of entire functions of exponential type, Math. Comp. 68 (1999) 749–765.
[17] Th. Strohmer, On discrete band-limited signal extrapolation, Contemp. Math. 190 (1995) 323–337.
[18] D.L. Donoho, B.F. Logan, Signal recovery and the large sieve, SIAM J. Appl. Math. 52 (1992) 577–591.
[19] J.J. Koliha, V. Rakočević, Fredholm properties of the difference of orthogonal projections in a Hilbert space, Integral Equations Operator Theory 52 (2005) 125–134.

[20] M.J. Croft, J.A. Hogan, Wavelet-based signal extrapolation, in: Fourth International Symposium on Signal Processing and its Applications, ISSPA 96, vol. 2, 25–30 August 1996, pp. 752–755.
[21] M.Z. Nashed, Steepest descent for singular linear operator equations, SIAM J. Numer. Anal. 7 (1970) 358–362.
[22] R.R. Coifman, S. Lafon, Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions, Appl. Comput. Harmon. Anal. 21 (2006) 31–52.
[23] S.-C. Pei, J.-J. Ding, Generalized prolate spheroidal wave functions for optical finite fractional Fourier and linear canonical transforms, J. Opt. Soc. Amer. A 22 (2005) 460–474.
[24] L.-L. Wang, J. Zhang, A new generalization of the PSWF with applications to spectral approximations on quasi-uniform grids, Appl. Comput. Harmon. Anal. 29 (3) (2010) 523–545.
[25] T. Moumni, A. Karoui, Fourier and Hankel bandlimited signal recovery, Integral Transforms Spec. Funct. 21 (2010) 337–349.
[26] Vu Kim Tuand, Semyon Yakubovich, Donoho–Stark and Paley–Wiener theorems for the *G*-transform, Adv. in Appl. Math. 45 (2010) 108–124.
[27] K. Drouiche, D. Kateb, C. Noiret, Regularization of the ill-posed problem of extrapolation with the Malvar–Wilson wavelets, Inverse Problems 17 (2001) 1513–1533.
[28] J. Tropp, On the independence of spikes and sines, J. Fourier Anal. Appl. 14 (2008) 838–858.
[29] Scott Izu, J.J. Lakey, Time-frequency localization and sampling of multiband signals, Acta Appl. Math. 107 (2009) 399–435 (Issue dedicated to: Applied Harmonic Analysis: Wavelets and Sampling).
[30] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, 1998.
[31] Manuela Nees, Products of orthogonal projections as Carleman operators, Integral Equations Operator Theory 35 (1999) 85–92.
[32] Gilbert G. Walter, Xiaoping Shen, Wavelet based on prolate spheroidal wave functions, J. Fourier Anal. Appl. 10 (2004) 1–26.
[33] Wai-Shing Tang, Oblique projections, biorthogonal Riesz bases and multi-wavelets in Hilbert spaces, Proc. Amer. Math. Soc. 128 (1999) 463–473.
[34] F. Deutsch, The angle between subspaces of a Hilbert space, in: S.P. Singh (Ed.), Approximation Theory: Wavelets and Applications, Kluwer, 1995, pp. 107–130.
[35] H. Brezis, Analyse fonctionnelle: théorie et applications, Éditions Masson, Paris, 1983 (ISBN 9782225771989).
[36] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–404.
[37] K. Yao, Applications of reproducing kernel Hilbert spaces–Band-limited signals models, Inf. Control 11 (1967) 429–444.
[38] Y.M. Liu, G. Walter, Irregular sampling in wavelet subspaces, J. Fourier Anal. Appl. 2 (1995) 181–189.
[39] A.I. Zayed, P. Mikusinski, On the extension of the Zak transform, Methods Appl. Anal. 2 (1995) 160–172.
[40] I.C. Daubechies, A.C. Gilbert, Harmonic analysis, wavelets, and applications, in: Luis Cafarelli, E. Weinan (Eds.), Hyperbolic Equations and Frequency Interactions, in: IAS/Park City Mathematics Series, vol. 5, 1998.
[41] Tsuyoshi Ando, Saburou Saitoh, Restrictions of reproducing kernel Hilbert spaces to subsets, Preliminary Reports, Suri Kaiseki Kenkyu Jo, Koukyu Roku, vol. 743, 1991, pp. 164–187. Available at: http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/0743-18.pdf.
[42] H.H. Bauschke, F. Deutsch, H. Hundal, S.H. Park, Accelerating the method of alternating projections, Trans. Amer. Math. Soc. 355 (2003) 3433–3461.
[43] H. Xiao, V. Rokhlin, N. Yarvin, Prolate spheroidal wave functions, quadrature and interpolation, Inverse Problems 17 (2001) 805–838.
[44] D. Slepian, Some comments on Fourier analysis, uncertainty and modeling, SIAM Rev. 25 (1983) 379–393.
[45] J.A. Hogan, J.D. Lakey, Time-Frequency and Time-Scale Methods, Birkhäuser, 2005.
[46] I.C. Moore, M. Cada, Prolate spheroidal wave functions, an introduction to the Slepian series and its properties, Appl. Comput. Harmon. Anal. 16 (2004) 208–230.
[47] W.J. Kammerer, M.Z. Nashed, On the convergence of the conjugate gradient method for singular linear operator equations, SIAM J. Numer. Anal. 9 (1972) 165–181.
[48] W.J. Kammerer, M.Z. Nashed, Steepest descent for singular linear operators with non-closed range, Appl. Anal. 1 (1971) 143–159.
[49] V. Rakočević, On continuity of the Moore–Penrose and Drazin inverses, Mat. Vesnik 49 (1997) 197–206.
[50] L. Gosse, Analysis and short-time extrapolation of stock market indexes through projection onto discrete wavelet subspaces, Nonlinear Anal. RWA 11 (2010) 3139–3154.