HAL
open science

Compressed sensing with preconditioning for sparse recovery with subsampled matrices of Slepian prolate functions

Laurent Gosse

To cite this version:

Laurent Gosse. Compressed sensing with preconditioning for sparse recovery with subsampled matrices of Slepian prolate functions. 2010. hal-00560962v1

HAL Id: hal-00560962

https://hal.science/hal-00560962v1

Preprint submitted on 31 Jan 2011 (v1), last revised 30 May 2011 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Compressed sensing with preconditioning for sparse recovery with subsampled matrices of Slepian prolate functions

Laurent Gosse *

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

Abstract

Efficient recovery of smooth functions which are s-sparse with respect to the base of so–called Prolate Spheroidal Wave Functions from a small number of random sampling points is considered. The main ingredient in the design of both the algorithms we propose here consists in establishing a uniform L^∞ bound on the measurement ensembles which constitute the columns of the sensing matrix. Such a bound provides us with the Restricted Isometry Property for this rectangular random matrix, which leads to either the exact recovery property or the "best s-term approximation" of the original signal by means of the ℓ^1 minimization program. The first algorithm considers only a restricted number of columns for which the L^∞ holds as a consequence of the fact that eigenvalues of the Bergman's restriction operator are close to 1 whereas the second one allows for a wider system of PSWF by taking advantage of a preconditioning technique. Numerical examples are spread throughout the text to illustrate the results.

Key words: Doubly orthogonal sequences, Slepian functions, Compressed sensing, restricted isometry property, Preconditioning, bandlimited extrapolation.
1991 MSC: 94a11; 94a12; 94a20; 60b20

1 Introduction

Compressed sensing (CS), or Compressive Sampling, is a recent field of research which hinges on a simple yet fascinating idea which led to a major rethinking of data acquisition protocols. It is well known that signals of interest are generally endowed with a specific structure making them concisely representable (that is, by means of quite a small number

* Corresponding Author

Preprint submitted to Elsevier Science

of generalized Fourier coefficients) in specific orthogonal bases. Hence, with the development of *e.g.* new wavelet systems, transmission of these signals has been progressively reduced to the one of a limited quantity of significant numbers. However, the acquisition protocols for these signals were still working in a massive and expensive manner: the totality of the amount of digital data was first acquired, and then later encoded in the appropriate orthogonal basis. It was at this level that filtering was applied in order to reduce the number of coefficients to be actually considered; in general, a big proportion of them was discarded which resulted in a wasteful processing. CS changed the whole framework by showing that, under the same hypothesis on the original scene (the possibility of being expressed by very few meaningful coefficients in a convenient basis, which is usually referred to as *sparsity*), the computational cost of the data acquisition protocol can be drastically reduced. Indeed, instead of acquiring the totality of the data to be later encoded, it limits itself to acquire directly all the important information by measuring the projections of this sparse signal onto a fixed set of well defined but randomly sampled basis vectors (which put together constitute the random *sensing matrix*). The signal can therefore be under-sampled according to the classical Shannon-Nyquist theory because the quantity of measurements is proportional to the number of non-zero coefficients, and not to the width of the support of the spectrum.

This collection of basis vectors, sometimes referred to as a *measurement ensemble*, was originally taken as a realization of Gaussian white noise or a sequence of Bernoulli random variables as this choice asks for a near-minimal amount of sampling points [5]. However, in practical applications, it is desirable to derive a CS strategy for signals which are sparse in structured measurement ensembles, like *e.g.* the discrete Fourier basis, a Multi-Resolution framework, or very recently the Legendre polynomial base [34]. In [7], the case of a sensing matrix which originates from an orthonormal matrix is studied and sufficient conditions ensuring the Restricted Isometry Property (RIP) [1] which leads to the exact recovery by ℓ^1 minimization are given. These conditions have been later refined in *e.g.* [10,18,19,33].

In this paper, we follow this research program and show that another class of matrices can be used in order to set up a CS algorithm: considering the so–called "Prolate Spheroidal Wave functions" (PSWF), one gets for any value of their Slepian parameter $c > 0$ an orthonormal base of $L^2(-1,1)$ [40,48]. These functions are restrictions to a compact interval of bandlimited functions, hence entire functions of exponential type when extended to the whole complex plane [42,41,22,23,29]. Entire functions, as they can be expressed as convergent series for any value of their argument, may be viewed as "infinite degree generalization" of polynomials. Moreover, it is a well-known fact that PSWF admit an expansion in the

Legendre polynomial basis, a feature which is useful for their practical computation [21].

In §2, the derivation of PSWF is recalled, starting from the inversion of Bergman's restriction operator with Seip's theorem on doubly-orthogonal sequences and composition of two orthogonal projections in Hilbert space [38,36,31]. We made this choice for 2 reasons: first, this emphasizes the very particular properties of this type of orthogonal bases, and second, it shows that the doubly-orthogonal Slepian's functions are by no means a unique object [50]. Besides that, these bases allow to perform extrapolation of signals even if this ill-conditioned problem has to be stabilized (see for instance [13,17]): hence in this perspective, the *exact recovery property* for sparse signals of CS algorithms may become very valuable as a limited amount of measurements can permit to reconstruct the signal's very disconnected spectrum with supposedly machine's accuracy and then allow to extrapolate observations made in, say, $[-1, 1]$ to a bigger interval (see Remark 6). Usual interpolation properties for PSWF are recalled in §2.2 together with error estimates for spectral approximation [4,11]; in particular, the recent estimates by Wang [46] are included.

§3 is devoted to proving a L^∞ bound on a subset of the PSWF base on $[-1, 1]$; like Legendre polynomials, PSWF can display sharp "tails" close to the edges of this interval. However, the situation here is better compared to polynomials because there exists a collection of indexes for which both a L^∞ bound and spectral accuracy hold as stated in Lemma 1; roughly speaking, it corresponds to the PSWF endowed with eigenvalues not too far from 1 (this statement can be made precise by means of the classical Landau-Widom estimate, see [26]). With this L^∞ bound at hand, it is possible to follow the canvas of [7] and estimate the *concentration measure* parameter μ which leads to the RIP under technical assumptions. We followed the approach of [33,34] which permits to reach similar conclusions under slightly milder restrictions. Once the RIP is established, the results by Foucart [18,19] ensure that either the exact recovery occurs, either accurate error estimates hold between the original signal and its best k-term approximation [12]. Numerical test-cases are performed in §3.4 and illustrate previous theoretical results.

In §4, we follow the original idea of Rauhut and Ward [34] of including preconditioning techniques inside a CS framework. Indeed, since the L^∞ bound of Lemma 1 blow up as more and more PSWF are added in the sensing matrix, it became desirable to derive another methodology to handle more complex problems. It turned out that a simple diagonal preconditioner (the same as used in [34]) allows to control the concentration measure of the sensing matrix columns in a more robust way, see Lemma 8. This comes from the fact

that Slepian's functions can be written as a series of Legendre polynomials for which fine estimates have been proven on the coefficients [48]. The exact recovery property is shown the same way as in §3, see Theorem 9, and the Corollary 2 is given concerning the recovery of functions belonging to the Hilbert spaces $\tilde{H}_c^r(-1,1)$ studied in [46] for which spectral accuracy always holds. Again, numerical tests are displayed in §4.4, involving more complex and possibly noisy signals.

To the best of the author's knowledge, the idea of performing CS with a sensing matrix built from PSWF seems to be new; classical sampling theorems involving PSWF already exist, see *e.g.* [25,43,44] and also [37] with references therein. However, no proof of the restricted isometry property or the exact recovery of sparse signals appears.

2 Bergman's restriction operator and Prolate spheroidal wave functions

2.1 Composition of orthogonal projections and doubly-orthogonal sequences

Our starting point is the Bergman-Shapiro problem for the inversion of a truncation operator in a separable Hilbert space with scalar product $< .,. >_{\mathcal{H}}$. Let us denote $\mathcal{H} = L^2(\mathbb{R})$, V a closed linear subspace of $\mathcal{H}$ with $\mathcal{P}$ the orthogonal projection $\mathcal{H} \to V$ and $\mathcal{T} : f \in V \mapsto f\chi_A$ with χ_A the characteristic function of a measurable set $A \subset \mathbb{R}$. In this framework, a special case of a result of Seip [36] can be stated:

Theorem 1 *Assume V is a reproducing kernel Hilbert space (RKHS) and let $(f_k)_{k\in\mathbb{N}}$ be an orthonormal basis of V. Then $(f_k)_{k\in\mathbb{N}}$ is furthermore orthogonal for the induced scalar product $< .,. >_A := < .,.\chi_A >_{\mathcal{H}}$ if and only if f_k are singular functions of $\mathcal{PT}$; in such a case, $(f_k)_{k\in\mathbb{N}}$ is said to be a* <u>*doubly orthogonal sequence*</u> *(DOS).*

Let us recall that the general structure of the composition product of 2 orthogonal projections in a Hilbert space is studied by Nees in [31]. Such DOS seem to have been first studied by Bergman in the context of analytic functions [38]; they are useful when it comes to recover a function from its values on a subset of its domain of definition [39]. However, in most cases, $\mathcal{TP}$ is a compact Hilbert-Schmidt operator (except if the Lebesgue measure $|\mathbb{R} \setminus A|$ is small enough to ensure that the operator norm $\|(Id - \mathcal{T})\mathcal{P}\| < 1$ which allows for **stable** inversion by means of Neumann series, see *e.g.* [49,16]) therefore performing inversion leads to a (possibly severely) ill-posed problem [15]; see however [17] for an original

regularization method and [13] for the recent approach called "geometric harmonics" [13].

A RKHS $V \subset L^2(\mathbb{R})$ with kernel K is a linear subspace which corresponding orthogonal projection $\mathcal{P}$ rewrites as an integral operator:

$$f \in V \Leftrightarrow \forall t,\ f(t) = \mathcal{P}f(t) =< f(.), K(.,t) >_{\mathcal{H}} = \int_{\mathbb{R}} K(s,t)f(s)ds. \tag{1}$$

Hence, given a measurable set $A \subset \mathbb{R}$, one can define the so–called concentration operator on A, $\mathcal{TP}$, for the functions of V; for V a subset of smooth functions and $|A|$ bounded, the Ascoli's theorem ensures that $\mathcal{TP}$ is compact and there holds:

$$\|\mathcal{TP}\|^2 = \|\mathcal{PT}\|^2 = \|\mathcal{PTP}\| = \lambda_0 \text{ with } \mathcal{PTP}\psi_0 = \lambda_0\psi_0 \text{ and } \mathcal{PT}f(t) =< f(.), K(.,t) >_A .$$

More generally, a sufficient condition for compactness is as follows [36]:

$$\int_A \int_A |K(s,t)|^2 ds\ dt = \int_A K(s,s)ds < \infty.$$

The first eigenfunction φ_0 is the function of V which realizes the maximum of concentration on the set A; its corresponding eigenvalue λ_0 can be understood as the concentration ratio:

$$\|\mathcal{TP}\| = \sup_{f \in L^2(\mathbb{R})} \frac{\|\mathcal{TP}f\|_{L^2(\mathbb{R})}}{\|f\|_{L^2(\mathbb{R})}} = \sup_{g \in V} \frac{\|\mathcal{T}g\|_{L^2(\mathbb{R})}}{\|g\|_{L^2(\mathbb{R})}} = \frac{\|\mathcal{T}\psi_0\|_{L^2(\mathbb{R})}}{\|\psi_0\|_{L^2(\mathbb{R})}} = \sqrt{\lambda_0} \leq 1.$$

This equation is the starting point of the presentation by Slepian in [41] in the particular case where V is chosen to be the Paley-Wiener space of ω-bandlimited functions:

$$PW_\omega = \left\{ f \in L^2(\mathbb{R}) \text{ such that } \hat{f}(\xi) := \int_{\mathbb{R}} f(t)\exp(2i\pi t\xi)dt \text{ is supported in } [-\omega,\omega] \right\}.$$

We stress that it is not the unique example to which these ideas can be applied: for instance, one can choose $V = V_j$, one of the nested linear subspaces of $L^2(\mathbb{R})$ constituting a Multi-Resolution Analysis (MRA). In this case, the reproducing kernel in (1) reads $K_j(s,t) = 2^j \sum_{n\in\mathbb{Z}} \phi(2^j s - n)\phi(2^j t - n)$ [28] where $j \in \mathbb{Z}$ is the scale index and ϕ is the so–called "scaling function" whose integer shifts generate V_0; see *e.g.* [27,20].

The reproducing kernel for PW_ω is the standard "sinc" function,

$$K_\omega(s,t) = \frac{\sin(2\pi\omega(t-s))}{\pi(t-s)},$$

and one is led to seek the singular value decomposition (SVD) of the compact operator,

$$\mathcal{PT}f(t) = \int_A f(s)\frac{\sin(2\pi\omega(t-s))}{\pi(t-s)}ds, \qquad A = [-T,T], \tag{2}$$

which has been thoroughly studied by Slepian, Landau and Pollak: see *e.g.* [22,23,29,42]. In particular, the integral operator (2) commutes with a second order differential operator: Daubechies gives a geometric explanation of this "lucky accident" in [14]. Such a property eases considerably the task of computing numerically the DOS associated to $V = PW_\omega$:

$$\mathcal{TPT}\varphi_k = \lambda_k\varphi_k,\ \mathcal{PTP}\psi_k = \lambda_k\psi_k, \qquad k \in \mathbb{N}.$$

Many efforts have been dedicated to the derivation of efficient algorithms to compute the peculiar PSWF system associated to PW_ω: see [21,24,25,45,48]. As PW_ω contains entire functions of exponential type, $0 < \lambda_k < 1$ and an easy computation shows that:

$$\psi_k = \frac{\mathcal{P}\varphi_k}{\sqrt{\lambda_k}}, \qquad \mathcal{T}\psi_k = \sqrt{\lambda_k}\varphi_k.$$

Trying to solve directly the integral equations (2) is a very difficult task because the non-negative eigenvalues λ_k display a sharp behaviour depending on $c = 2\pi\omega T$, the so–called Slepian parameter representing the area in the time-frequency plane, which makes them decay supergeometrically [3] (see also [9,23]) to zero:

$$\lambda_k \simeq 1 \text{ for } k \leq \frac{2c}{\pi}. \tag{3}$$

2.2 Spectral approximation with Prolate Spheroidal wave functions

Prolate Spheroidal wave functions (PSWF) constitute an orthogonal base of $L^2(-T,T)$: see [48,41,23,22,4]. Thus they can serve as an interpolator on any compact interval of $\mathbb{R}$ as an alternative choice which can enjoy spectral accuracy instead of classical polynomial

systems like Legendre, see [46] for very precise error estimates in this direction. The following theorem (taken from [48]) summarizes the main properties of $(\varphi_k)_{k\in\mathbb{N}}$ as an interpolator:

Theorem 2 *For any positive real value c, the eigenfunctions φ_k, $k \in \mathbb{N}$ are purely real and complete in $L^2(-T,T)$. The even-numbered eigen functions are even, and the odd-numbered are odd according to the order of decreasing eigenvalues which are non-zero and simple. The eigenfunctions constitute a Chebyshev system on $[-T,T]$, in particular, φ_k has exactly k zeros in this interval.*

PSWF satisfy also another eigenvalue problem which reads [24,44,48]:

$$\mu_k \varphi_k(t) = \int_{-T}^{T} \varphi_k(\xi) \exp(icts) ds, \qquad \lambda_k = \frac{c}{2\pi} |\mu_k|^2. \tag{4}$$

This formulation has the drawback of involving imaginary eigenvalues μ_k when k is odd; however, it involves a kernel $\tilde{K}$ which depends only on the product st (compare with (1)). At this point, it is **important** in the context of PSWF to make a clear distinction between

- approximation of bandlimited functions on $\mathbb{R}$, that is, approximation in PW_ω by means of ψ_k, $k \in \mathbb{N}$ which are normalized so as to have $\|\psi_k\|_{L^2(\mathbb{R})} = 1$ (which implies that $\|\varphi_k\|_{L^2(-T,T)} = \sqrt{\lambda_k} \to 0$ when k grows) as studied for instance in [40],
- and approximation in the space $L^2(-T,T)$, as presented in Theorem 2 or in [4,48,46], which is made with functions φ_k normalized so as $\|\varphi_k\|_{L^2(-T,T)} = 1$ (which implies that $\|\psi_k\|_{L^2(\mathbb{R})} = \frac{1}{\sqrt{\lambda_k}} \to +\infty$, especially for very delocalized functions such that $k \gg 2c/\pi$).

Theorem 3 *Let $f \in H^r(-T,T)$ the Sobolev space of functions of $L^2(-T,T)$ having all derivatives up to order $r \geq 0$ in $L^2(-T,T)$ have the expansion: $f(t) = \sum_{k\in\mathbb{N}} f_k \varphi_k(t)$. Then, if the parameter c <u>associated to φ_k</u> is such that $\varrho_N(c) := \frac{2c}{\pi(N+\frac{1}{2})} < 1$, there holds:*

$$|f_N| \leq C \left\{ N^{-\frac{2}{3}r} \|f\|_{H^r(-T,T)} + \varrho_N(c)^{\delta N} \|f\|_{L^2(-T,T)} \right\}. \tag{5}$$

The constants C, δ are independent of f, N, and the Slepian parameter c.

This error estimate appears in [11,4,46] in slightly different forms. Following Wang [46], we recall the Sturm-Liouville operator associated with the system of PSWF φ_k, $k \in \mathbb{N}$, for a

fixed value $c > 0$ of the Slepian parameter and u a smooth function,

$$\forall t \in [-1,1], \qquad \mathcal{D}_c u(t) = -\frac{d}{dt}\left(w(t)\frac{du(t)}{dt}\right) + (ct)^2 u(t) \text{ with } w(t) = 1 - t^2,$$

which is clearly self-adjoint and positive and for any integer $r \geq 0$, the Hilbert space:

$$\tilde{H}_c^r(-1,1) = \left\{ u \in L^2(-1,1) \text{ such that } \|u\|_{\tilde{H}_c^r} = \sqrt{< (\mathcal{D}_c)^r u, u >_{[-1,1]}} < +\infty \right\}. \qquad (6)$$

Thanks to the "lucky accident", PSWF also satisfy $\mathcal{D}_c \varphi_k = \chi_k \varphi_k$ and it turns out that:

$$\|u\|^2_{\tilde{H}_c^r(-1,1)} = \sum_{k \geq 0} (\chi_k)^r \left| < u, \varphi_k >_{[-1,1]} \right|^2, \qquad c > 0.$$

Theorem 4 *Let $f \in \tilde{H}_c^r(-1,1)$ with $r \geq 0$ as defined in (6); it holds that,*

$$\left\| f - \sum_{k=0}^{N} < f, \varphi_k >_{[-1,1]} \varphi_k \right\|_{L^2(-1,1)} \leq N^{-r} \|f\|_{\tilde{H}_c^r(-1,1)}. \qquad (7)$$

Hence spectral accuracy holds without restriction for functions belonging to the Hilbert spaces (6). The first PSWF ψ_0 has been used as a scaling function to construct a Multi-resolution analysis (MRA) of $L^2(\mathbb{R})$ enjoying specific properties; see [44]. On Fig. 1, we display the first 10 PSWF φ_k with $c = 130$, $T = 1$ and 256 grid points in the t variable:

3 Restricted Isometry Property (RIP) and sparse ℓ^1-recovery

In the sequel, we shall only be interested in the "single" orthogonal system of PSWF φ_k, $k \in \mathbb{N}$, complete in $L^2(-T,T)$ and normalized so as to have $\|\varphi_k\|_{L^2(-T,T)} = 1$. Without loss of generality, the value of T can be fixed to $T = 1$ as suggested in [44], Remark 1.

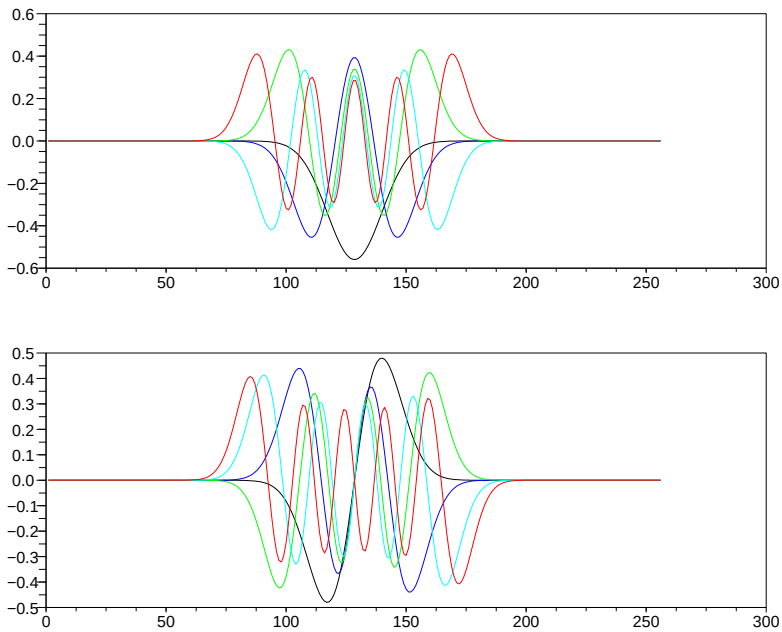

Fig. 1. First 10 PSWF (top: even, bottom: odd) with $c = 130$ and $T = 1$.

3.1 Preliminaries on Compressed Sensing (CS)

Let us begin by recalling the ℓ^p norm in $\mathbb{R}^N$ for $N \in \mathbb{N}$:

$$\|x\|_{\ell^p} = \left(\sum_{k=1}^{N} |x_k|^p \right)^{\frac{1}{p}}, \qquad 0 < p < \infty, \ x := (x_k)_{k=1,\dots,N}.$$

The 2 extreme cases $p = 0$ and $p = \infty$ are defined as follows:

$$\|x\|_{\ell^\infty} = \max_{k=1,\dots,N} |x_k|, \qquad \|x\|_{\ell^0} = \#\{x_k, \ x_k \neq 0\} \in \mathbb{N},$$

with the symbol # denoting the cardinal number of a set. Any given vector x in $\mathbb{R}^N$ is said to be sparse as soon as $\|x\|_{\ell^0}$ is smaller than a certain threshold value, generally denoted by $s \in \mathbb{N}$. An important quantity is the **best s-term approximation** [12] of a vector $x \in \mathbb{R}^N$ in ℓ^p which reads:

$$\sigma_s(x)_{\ell^p} := \inf_{y: \ \|y\|_{\ell^0} \leq s} \|y - x\|_{\ell^p}.$$

A direct consequence of this definition is that $\sigma_s(x)_{\ell^p} = 0$ for all p if x is s-sparse; similarly, x is called **compressible** if $\sigma_s(x)_{\ell^1}$ decays at a fast rate with increasing values of s. Actually, it has been proved rigorously that, for any $q < p$, there holds $\sigma_s(x)_{\ell^p} \leq s^{\frac{1}{p}-\frac{1}{q}} \|x\|_{\ell^q}$. Care

must be taken because sparse signals don't generate a linear space as the sum of two s-sparse vectors may only be $2s$-sparse. The error estimate (5) implies that functions in $H^r(-1,1)$ yield compressible vectors in convenient PSWF bases for which $\varrho_N(c) < 1$.

We need a bit of terminology as we aim at computing efficiently a sparse solution of a linear system of the following form: $y = \Phi x$ where $y \in \mathbb{R}^m$ is the **information vector**, m is therefore the number of samples. The matrix Φ is usually not square but $m \times N$ instead; it is referred to as to the **encoder** or the **measurement matrix**. In the present context, it is a portion of the matrix A written in Theorem 7.1 in [48]. The integer N stands for the maximum number of components which is necessary to represent a signal of $L^2(-1,1)$ in the PSWF base: according to [4,11,46], it depends linearly on the Slepian parameter c:

$$N = \frac{2c}{\pi} \Rightarrow \varrho_N(c) < 1.$$

In order to recover the sparse (or at least, compressible) vector x, we need a **decoder** Δ; we stress that Δ is not required to be linear, but is must have the **exact recovery property** for s-sparse signals ($s \ll N$):

$$\|x\|_{\ell^0} \leq s \Rightarrow \Delta(\Phi x) = x.$$

At last, the number of samples m must be smaller than the total number of grid points used to represent a signal in the time variable: we call $M \gg m$ this number of grid points.

A fundamental stepping stone for establishing the exact recovery property for sparse signals is the so–called **restricted isometry property** (RIP) (see [5,1]):

Definition 1 *For any integer $s \leq N$, the restricted isometry constant δ_s of a (possibly rectangular) measurement matrix Φ is the smallest nonnegative number such that it holds for every s-sparse vector:*

$$(1-\delta_s)\|x\|_{\ell^2}^2 \leq \|\Phi x\|_{\ell^2}^2 \leq (1+\delta_s)\|x\|_{\ell^2}^2.$$

The following theorem is fundamental as it expresses the fact that if an encoder Φ is endowed with a restricted isometry constant which is small enough, then the corresponding decoding can be done simply by solving the ℓ^1 minimization problem:

Theorem 5 *(see [5,6,18]) Let δ_s stand for the restricted isometry constant (an increasing function of $s \ll N$) of the $m \times N$ measurement matrix Φ. Assume further that there holds:*

$$\delta_{2s} < \frac{2}{3+\sqrt{\frac{7}{4}}} \simeq 0.4627$$

Then, if x^* denotes the solution of the ℓ^1 minimization problem,

$$\inf \|z\|_{\ell^1} \text{ such that } \Phi z = \Phi x, \tag{8}$$

then the following error estimate holds:

$$\|x - x^*\|_{\ell^2} \leq C \frac{\sigma_s(x)_{\ell^1}}{\sqrt{s}}. \tag{9}$$

The constant C depends only on δ_{2s} and the recovery is exact in case the vector is s-sparse.

3.2 Uniform bound for a subset of N first PSWF and exact recovery property

Another step can be made when the encoder Φ is a matrix whose columns contains sampled values of an orthonormal system of functions, as it is the case for the PSWF φ_k normalized such that $\|\varphi_k\|_{L^2(A)} = 1$:

$$\int_A \varphi_i(t)\varphi_j(t).d\nu(t) = \delta_{i,j}, \tag{10}$$

with ν is a probability measure on the measurable space A (in most of the cases $A = [-1, 1]$ and ν is simply the Lebesgue measure) and $\delta_{i,j}$ stands for the Kronecker symbol.

Theorem 6 *(see [33,34]) Assume that the orthonormal system φ_k is bounded in $L^\infty(A)$:*

$$\sup_{k<N} \sup_{t \in A} |\varphi_k(t)| \leq K_N, \qquad K_N \geq 1. \tag{11}$$

If the $m \times N$ measurement matrix Φ is built from this bounded basis functions, then for

$$\frac{m}{\log m} \geq C \frac{K_N^2}{\delta^2} s (\log s)^2 \log N, \tag{12}$$

it holds that, with probability at least $1 - N^{-\gamma(\log s)^2 \log m}$, the restricted isometry constant

δ_s of $\frac{1}{\sqrt{m}}\Phi$ satisfies $\delta_s \leq \delta$. The constants $C, \gamma > 0$ are universal.

Remark 1 *The bound (12) which gives only that $\delta_s \leq \delta$. However, in order to apply the results of [18] which ensure the exact recovery property, one needs to have a bound on δ_{2s}. Hence it turns out that (12) rigorously implies that the "ℓ^1 minimization decoder" Δ recovers exactly $\frac{s}{2}$-sparse signals. The stronger restriction $m \geq C\mu^2 s(\log N)^4$ where μ is the* **concentration measure** *parameter (as studied in e.g. [7]) yields directly the exact recovery property with higher probability $1 - N^{-\gamma(\log N)^3}$ (which is independent of s).*

All in all, the preceding results suggest that a crucial ingredient toward establishing the exact recovery property for the encoder $\Phi \in \mathbb{R}^{m\times N}$ being built out of the PSWF base φ_k is a uniform bound like (11). This is the purpose of the following result:

Lemma 1 *Let $0 < \alpha < 1$ and N be the higher integer such that $\lambda_{N-1} \geq \alpha$. There hold:*

(1) $\sup_{k<N} \sup_{t\in[-T,T]} |\varphi_k(t)| \leq \sqrt{\frac{cT}{\pi\alpha}} \|\varphi_k\|_{L^2(-T,T)} := K_N(c)$,

(2) let U stand for the orthonormal $N \times N$ matrix whose entry at $(j,k) \in [0, N-1]^2$ reads $\varphi_k(-T + jh)$, $h = \frac{2T}{N}$, then $\mu(U) := \sqrt{N} \max_{j,k} |U_{j,k}| \leq T\sqrt{\frac{2c}{\pi\alpha}}$ where $c = 2\pi T\omega$.

The bound (1) of Lemma 1 is quite logical: it grows together with ω because increasing the cut-off frequency allows for more concentrated functions in $[-T, T]$. Increasing T means that a larger interval is considered and with the same cut-off frequency, a major concentration ratio can also be achieved.

Proof: From (4) and for all $t \in [-T, T]$, we get by Hölder's inequality:

$$|\varphi_k(t)| \leq \frac{1}{|\mu_k|}\left(\int_{-T}^{T} |\exp(ict\xi)|^2 d\xi\right)^{1/2} \|\varphi_k\|_{L^2(-T,T)} \leq \frac{\sqrt{2T}}{|\mu_k|}\|\varphi_k\|_{L^2(-T,T)}.$$

The idea is now to consider $\|\varphi_k\|_{L^2(-T,T)}$ as a degree of freedom. Based on a result by Landau and Widom [26] (recalled in *e.g.* [23,48]), given $0 < \alpha < 1$, the <u>number</u> of eigenvalues greater that α reads:

$$N = \frac{2c}{\pi} + \frac{\log c}{\pi^2} \log\left(\frac{1-\alpha}{\alpha}\right) + O(\log c).$$

Thus, since $\frac{1}{|\mu_k|} = \sqrt{\frac{c}{2\pi\lambda_k}}$, we have for $0 \leq k < N$,

$$|\varphi_k(t)| \leq \sqrt{\frac{cT}{\pi\alpha}}\|\varphi_k\|_{L^2(-T,T)},$$

and the first point is proved. At this point, we use the degree of freedom: having N discretization points uniformly griding $[-T,T]$, we have $\|\varphi_k\|^2_{L^2(-T,T)} = \sum_{j=0}^{N-1} h|\varphi_k(-T+jh)|^2 + o(h)$ with $h = 2T/N$. In order to ensure that U is an orthonormal matrix, a sufficient condition is to ask for $\|\varphi_k\|^2_{L^2(-T,T)} = h$. We can now define the parameter μ as in [5,7] and it comes that:

$$\mu \leq \sqrt{N}\sqrt{\frac{cT}{\pi\alpha}}\sqrt{\frac{2T}{N}} = T\sqrt{\frac{2c}{\pi\alpha}}.$$

□

Remark 2 • *The bound on μ can be compared with the one where U is the discrete Fourier transform matrix like in §3.4 of [5]. We insist on the fact that according to Theorem 3, the region where $\lambda_k \simeq 1$ allows for high accuracy approximation for smooth functions belonging to $H^r(-T,T)$. For the choice $T = 1$, one gets easily that $\mu^2 \leq 4\omega/\alpha$ which implies a first bound on the minimum amount of samples for the exact recovery of a s-sparse signal in the N-component PSWF base with cut-off frequency $\omega > 0$:*

$$m \geq C\left(\frac{4\omega}{\alpha}\right) s(\log N)^4.$$

- *A direct argument hints that one shouldn't expect any uniform bound in the infinity norm to exist for $k \in \mathbb{N}$. Indeed, according to Proposition 1 in [2], the following estimate holds:*

$$\|\varphi_k\|_{H^1(-1,1)} = \|\varphi_k\|_{L^2(-1,1)} + \|\varphi_k'\|_{L^2(-1,1)} \leq 1 + \frac{c}{\sqrt{\lambda_k}}, \tag{13}$$

thanks to Bernstein's inequality for bandlimited functions. It is well-known that in dimension 1, the H^1 norm controls the L^∞ one. The authors state ([2], page 266) that the existence of a uniform bound K_c for all $k = 0, 1, \ldots$ can be proved by observing that PSWFs approach Legendre polynomials for $j \gg c$ (see Theorem 5.2 in [4]); however, a sharp bound on these Legendre polynomials (recalled in [34]) yields that their infinity norm grows like $\sqrt{2k+1}$ hence this cannot lead to proving the existence of such a quantity K_c. The growth of the amplitude of Legendre polynomials is fully compatible with the uniform bound on PSWFs proposed in [40] (see formula (11)) with identical normalization:

$$\forall k \in \mathbb{N}, \qquad \sup_{t\in[-1,1]} |\varphi_k(t)| \leq 2\sqrt{k}.$$

- *The proof of Lemma 1 uses two main ingredients: the integral equation satisfied by φ_k (and we know that the case of the Paley-Wiener subspace of L^2 is not unique) and the estimate*

on the number of eigenvalues greater than a certain threshold. Thus it may hopefully be extended in various directions: for instance, the case where A is a finite union of disjoint intervals allows for the same Landau-Widom estimate: see Theorem 3 in [23] and [25] for the construction of corresponding PSWF. Moreover, assuming a similar estimate can be established, it may also be applied to the case where the reproducing kernel is $K_j(s,t)$, the one corresponding to a j-scale-limited subspace of L^2, see [28,27,20].

Clearly, Lemma 1 shows a big difference with the case of Legendre polynomials studied in [34] as they don't allow for any useful bound on the quantity μ which controls the minimum amount of samples m allowing for the exact recovery property. This drawback comes from the fact that sharp "tails" appear very quickly with k in the vicinity of $t = \pm 1$; on the contrary, these "tails" appear only for PSWF with an index k bigger than $2c/\pi$, and these are precisely the ones that one doesn't need to have high accuracy. Hence the bigger amount of computational work which is required in order to generate the PSWF basis is somehow paid back through their nice properties.

Theorem 7 *Let $T = 1$ and the normalization be $\|\varphi_k\|_{L^2(-1,1)} = 1$ for all $k \in \mathbb{N}$. Suppose Φ is the $m \times N$ measurement random matrix built from φ_k, $k = 0, 1, ..., N-1$ and,*

$$\frac{m}{\log m} \geq 9.35C\left(\frac{\omega}{\alpha}\right) s(\log s)^2 \log N, \qquad \lambda_{k=0,1,\ldots,N-1} \geq \alpha \in]0,1[, \tag{14}$$

then, with probability at least $1 - N^{-\gamma(\log s)^2 \log m}$, the ℓ^1-minimization program (8) recovers exactly $\frac{s}{2}$-sparse vectors $x \in \mathbb{R}^N$; otherwise the error estimate (9) holds accordingly.

We recall that $\omega > 0$ is the cut-off frequency defining the Paley-Wiener space PW_ω whose orthonormal base reads (with the normalization of Theorem 7) $\mathcal{P}\varphi_k$, $k \in \mathbb{N}$. Moreover, the "good number" of PSWF N depends on c and α and we have that $\frac{2c}{\pi} \leq N < c$; a convenient value for α is $\frac{1}{2}$ because in this case, the Landau-Widom estimate gives $N = \frac{2c}{\pi} + O(\log(c))$ and 18.7 replaces 9.35 in (14).

Proof: From point (1) in Lemma 1, it suffices to plug $T = 1$ and $\|\varphi_k\|_{L^2(-1,1)} = 1$ in order to derive that $K_N(c)^2 \leq \frac{c}{\pi\alpha} = \frac{2\omega}{\alpha}$. Now, inserting this value inside (12) gives that the restricted isometry constant of Φ satisfies $\delta_s \leq \delta$ for which we impose $\delta = 0.4627$ by taking advantage of Theorem 5. It comes therefore that $2/\delta^2 < 9.35$ and since we have a bound on δ_s with high probability, the results of Foucart [18] ensure that we exactly recover $\frac{s}{2}$-sparse vectors through ℓ^1 minimization. The constants C, γ are identical in Theorems 6 and 7. □

Theorem 7 deals with the exact recovery of a $\frac{s}{2}$-sparse solution out of a small number of m random measurements. This extends in a straightforward manner to functions which are either sparse or compressible in the PSWF base:

Corollary 1 *Let $T = 1$, $c = 2\pi\omega$ be the Slepian parameter, $f \in H^r(-1,1)$ such that,*

$$\forall t \in [-1,1], \qquad f(t) = \sum_{k=0}^{N-1} f_k \varphi_k(t), \tag{15}$$

and $y = [f(t_1), f(t_2), ..., f(t_m)]$ a vector of m measurement values taken at the random locations t_i drawn independently from the uniform probability measure on $[-1,1]$. If Φ is the $m \times N$ measurement matrix whose k^{th} column contains the values $\varphi_k(t_i)$ with $\|\varphi_k\|_{L^2(-1,1)} = 1$ and if m meets the requirement (14), then the "best $\frac{s}{2}$-term approximation" of the coefficient vector $\vec{f} := (f_k)_{k=0,\dots,N-1}$ is recoverable with probability exceeding $1 - N^{-\gamma(\log s)^2 \log m}$ by means of the ℓ^1 minimization program (8). More precisely, if $\|\vec{f}\|_{\ell^0} \leq \frac{s}{2}$, the recovery is exact; otherwise, the following error estimate holds:

$$\|\vec{f} - \vec{f}^*\|_{\ell^2} \leq C_2 \frac{\sigma_{\frac{s}{2}}(\vec{f})_{\ell^1}}{\sqrt{\frac{s}{2}}}, \qquad \vec{f}^* = \arg\min_{y=\Phi z} \|z\|_{\ell^1}. \tag{16}$$

Proof: Having f exactly writable as a N-term expansion in the PSWF base corresponding to the Slepian parameter c allows to move from a continuous problem to the discrete one consisting in recovering $\vec{f} \in \mathbb{R}^N$ from $y \in \mathbb{R}^m$ which can be handled by the methods recalled in §3.1. With the measurement matrix Φ built as indicated, the criterion (14) ensures that $\delta_s \leq 0.4626$ with high probability and this is a sufficient condition allowing to apply Theorem 5 with a sparsity level equal to $\frac{s}{2}$. The error estimate (16) follows. □

Remark 3 *Relying on Theorem 3, we know that any function in $H^r(-1,1)$ which writes $f(t) = \sum_{k\geq 0} f_k \varphi_k(t)$ is strongly compressible (see for instance Figure 3 in [3]) because its "best N-term approximation" is spectrally accurate as soon as $\varrho_N(c) < 1$ meaning that $N > \frac{2c}{\pi} - \frac{1}{2}$. A strategy for treating functions written as an infinite PSWF series for which <u>the parameter c is already fixed</u> by means of a coefficient vector $\vec{f} \in \mathbb{R}^N$ only consists in:*

(1) adjusting N in order to satisfy both the spectral accuracy criterion, in such a way that the best N-term approximation becomes simply the "first N terms approximation"

written in (15), and a reasonable smallness level for the corresponding α;

(2) recovering a "best $\frac{s}{2}$-term approximation" out of the m random measurements with m satisfying (14); both L^2 errors between f and its N-term approximation and between $\vec{f}$ and $\vec{f}^*$ can be controlled by (5) and (16) respectively.

(3) in case the strong criterion $\varrho_{\frac{s}{2}}(c) < 1$ is met, which means that $c < \frac{\pi(s+1)}{4}$, then the coefficients f_k will strongly decay for $k \geq s$ and either the recovery of $\vec{f}$ through (8) will be exact, either the error (16) will be very small.

Actually, point (1) is delicate as it asks for finding a compromise between two antagonist requirements. We shall study in the next section a preconditioning method which lightens this situation by allowing to handle much smaller values of α.

In order to fully exploit the potential of Corollary 1, the best choice is probably to select the functions belonging to the spaces $\tilde{H}_c^r(-1,1)$ as defined in (6) because of the very fast convergence of finite sums expressed in Theorem 4. In this case, the approximation (15) is exponentially accurate and even for moderate values of s, it can be considered as being exact in practice. The ℓ^1 minimization process recovers therefore a "best $\frac{s}{2}$-term approximation" of the coefficient vector in $\mathbb{R}^N$ out of a collection of m samples taken at random; in case N is too low, a modified procedure adapted to noisy measurements should be used [5,8].

3.4 First set of numerical experiments

We intend now to show examples to demonstrate both the feasibility and the efficiency of the numerical processing presented in Theorem 7; in particular, we carried out the ℓ^1 minimization program by means of the algorithm called ROMP [30] to lower the CPU cost. Other numerical strategies exist for this minimization task: see also [10].

3.4.1 Verification of the exact recovery property for sparse signals

First, we set up a numerical test on the exact recovery of a strongly under-sampled signal admitting a M-term approximation in the PSWF base on $[-1,1]$ with $c = 130$ and 256 grid points. The code generates at random $M = 10$ coefficients between -1 and 1 together with 10 random indexes corresponding to integer positions between 0 and $2c/\pi \simeq 82 := N$. Then the signal $f(t)$ is formed by linear combination of the form (15) but with only $M \ll N$ non-zero coefficients. Numerically, it seems that the restriction (14) is somehow too restrictive and we selected instead a number of samples equal to $m = \frac{3}{2}M\log(N)$ to obtain the results

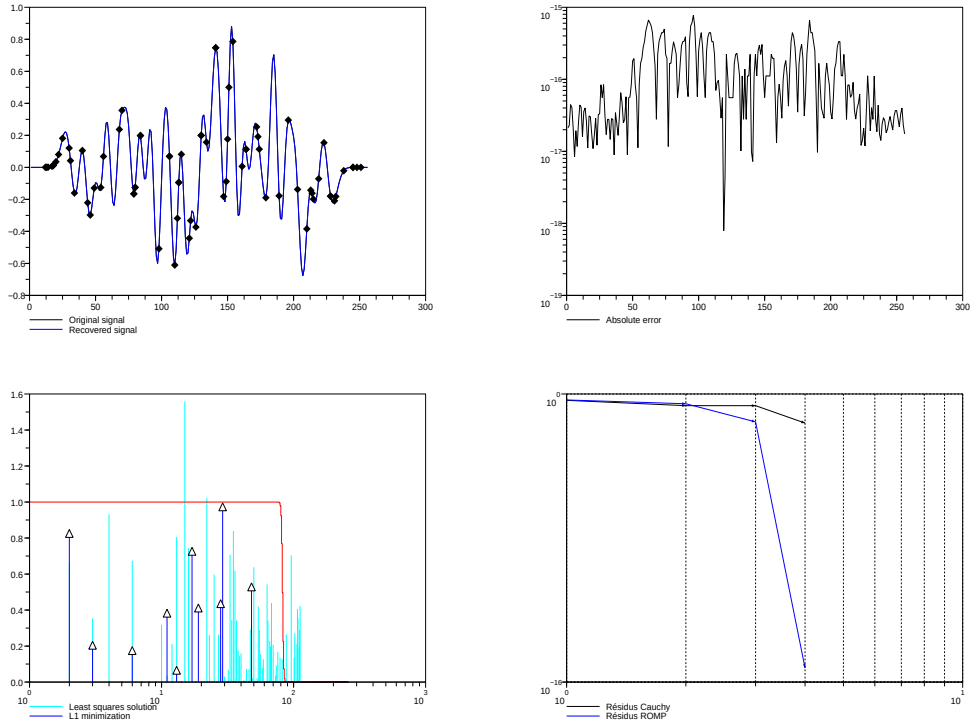

Fig. 2. Exact recovery of $\vec{f}$ with $\|\vec{f}\|_{\ell^0} = 10$.

displayed on Fig. 2. The ROMP algorithm had a very quick convergence and the pointwise absolute error is close to machine precision and it is impossible to distinguish between the original and the recovered signals. The location of the random m samples is indicated on the top left picture; on the bottom left one, one can see the exact value of the absolute value of the coefficients f_k (the small triangles) compared to the recovered ones (the blue lines) and also to the approximation obtained by a least squares approximation (in light blue) which is known to not promote sparsity. The red curve displays the behavior of the eigenvalues λ_k; beyond $N \simeq 2c/\pi$, the bounds of Lemma 1 are likely to blow up and the minimum number of samples m to maintain the exact recovery property should grow up quickly before becoming greater than N, which makes the whole approach useless.

3.4.2 Case of a signal involving too many components in the PSWF base

Corollary allows to treat cases for which the number m is too small according to the complexity of the original signal; in other words, there are too many non-zero coefficients $M' = \|\vec{f}\|_{\ell^0}$ and only a "best s-sparse approximation" can be recovered through ℓ^1 minimization from the limited number m of samples. For this numerical test, we set up a similar framework than in the preceding subsection, but this time we took $c = 85$ and $M' = 19 = M + 9$ with

still $m = \frac{3}{2} M \log(N)$. The coefficients are again chosen in random locations corresponding to indexes smaller that the transition value $2c/\pi \simeq 55 := N$. Two sets of numerical results

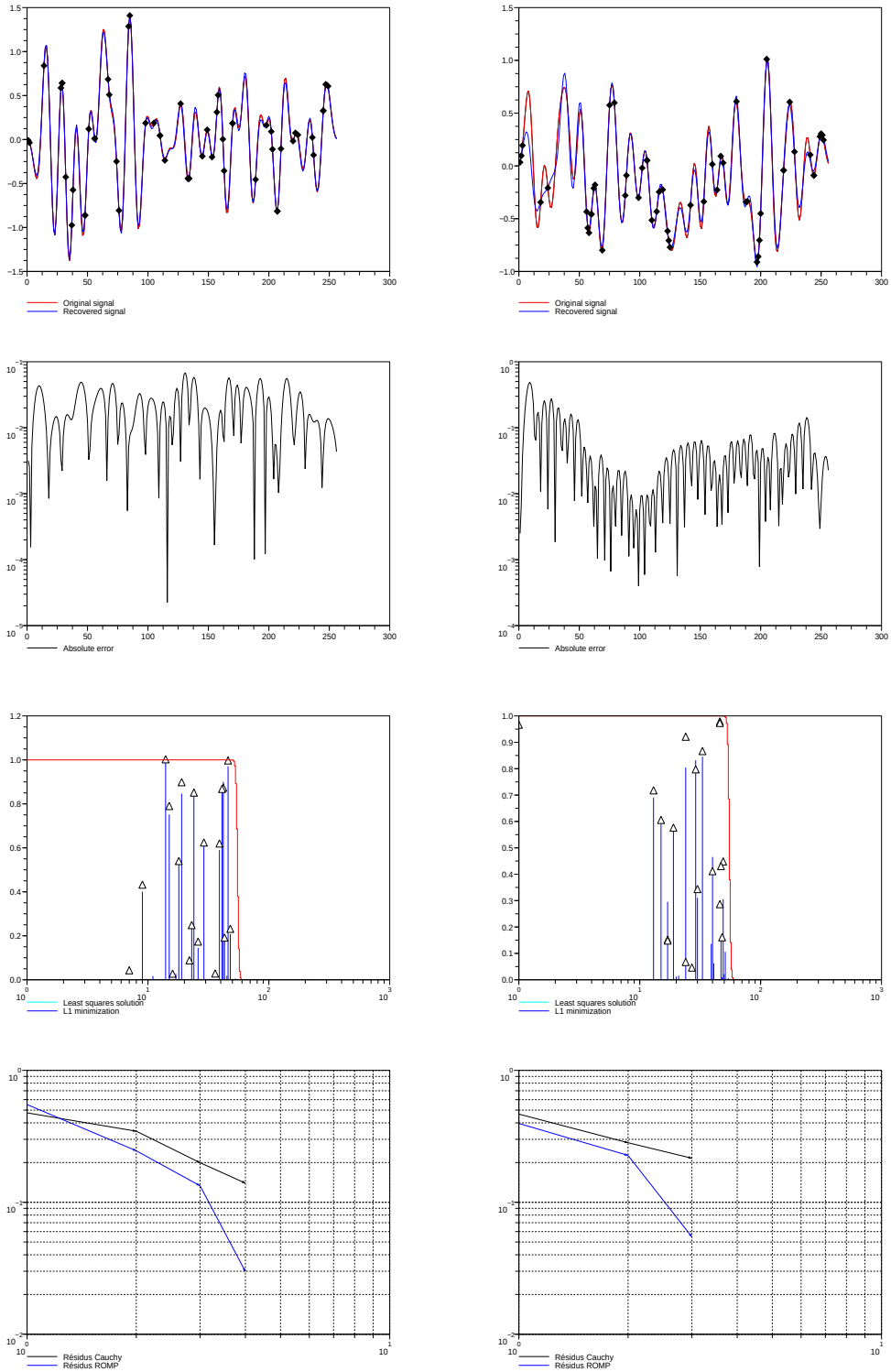

Fig. 3. Two recovery experiments for $\|\vec{f}\|_{\ell^0} = 19$ but $s = 9$ only with $c = 85$.

are displayed on Fig. 3: the left column shows a result where the recovery was rather good whereas the right one corresponds to a less satisfying one. Especially, the pointwise error on the right column grows substantially close to the left border and the recovery of coefficients suffers from noticeable errors. The number of samples used is roughly the same as in the preceding test, but it didn't allow for a good convergence of the ROMP algorithm, especially on the second test-case.

4 Preconditioning the "tails" and non-uniformly distributed random samples

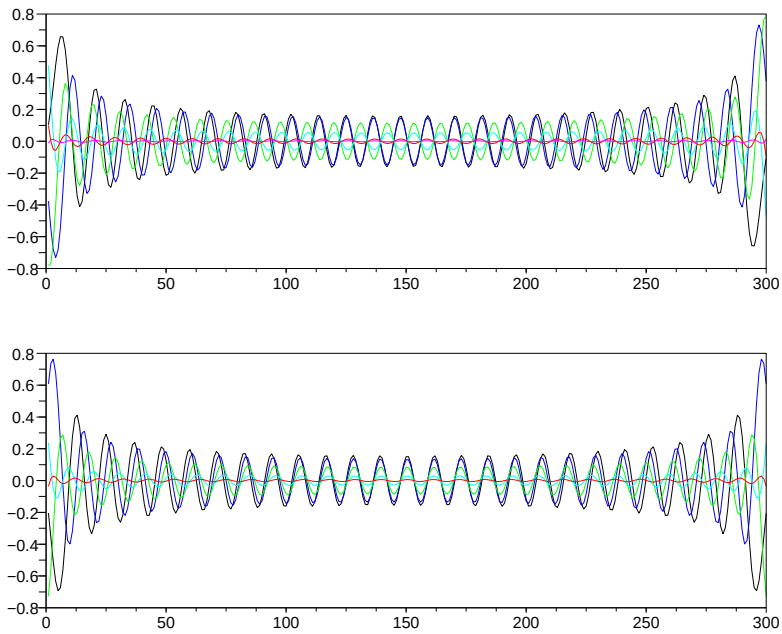

Fig. 4. Delocalized PSWF with $k \simeq 2c/\pi$ and $c = 85$ (top: odd, bottom: even).

4.1 Amplitude estimate on PSWF with series of Legendre polynomials

It is a classical fact [29,48] that any PSWF φ_k defined on $A = [-1, 1]$ admits an expansion based on normalized Legendre polynomials; we denote $\bar{P}_j(t)$ the orthonormal system of normalized Legendre polynomials which satisfy:

$$\forall i, j \in \mathbb{N}^2, \qquad < \bar{P}_i, \bar{P}_j >_A = \int_{-1}^{1} \bar{P}_i(t)\bar{P}_j(t)dt = \delta_{i,j}.$$

Accordingly, the Plancherel equality yields:

$$\varphi_k(t) = \sum_{j\geq 0} < \varphi_k, \bar{P}_j >_A \bar{P}_j(t), \qquad \|\varphi_k\|^2_{L^2(-1,1)} = \sum_{j\geq 0} | < \varphi_k, \bar{P}_j >_A |^2.$$

Usually, the coefficients rewrite $\beta_j^k :=< \varphi_k, \bar{P}_j >_A$ and any sequence β^k belongs to $\ell^2(\mathbb{N})$. From [34,35], we get the following pointwise estimates for all $t \in [-1,1]$, $j,k \in \mathbb{N}^2$:

$$|\bar{P}_j(\pm 1)| \leq \sqrt{j+\frac{1}{2}}, \qquad (1-t^2)^{\frac{1}{4}}|\bar{P}_j(t)| \leq \frac{1}{\sqrt{\pi}}, \qquad |\varphi_k(\pm 1)| \leq \sqrt{k+\frac{1}{2}}. \tag{17}$$

In order to improve the pointwise estimate of Lemma 1, it sounds appealing to compute:

$$\sqrt{\pi}(1-t^2)^{\frac{1}{4}}|\varphi_k(t)| \leq \sum_{j\geq 0} | < \varphi_k, \bar{P}_j >_A | = \sum_{j\geq 0} |\beta_j^k|.$$

Lemma 8 *Let $\lfloor a \rfloor$ denote the integer part of $a \in \mathbb{R}^+$, that is, $\lfloor a \rfloor = \max\{n \in \mathbb{N},\ n \leq a\}$. For $T = 1$ and with the normalization $\|\varphi_k\|_{L^2(-1,1)} = 1$, there holds for any $k \in \mathbb{N}$:*

$$\forall t \in [-1,1], \qquad \sqrt{\pi}(1-t^2)^{\frac{1}{4}}|\varphi_k(t)| \leq \sum_{j\geq 0} |\beta_j^k| \leq 2(\lfloor ec \rfloor + 1) + \frac{1}{2^{2\lfloor ec \rfloor}\mu_k}. \tag{18}$$

Proof: The basic estimate is the one given inside Theorem 3.4 of [48]: for $j \geq 2(\lfloor ec \rfloor + 1)$, one has $\beta_j^k < 2^{1-j}/\mu_k$. Thus, for any PSWF index $k \in \mathbb{N}$, we can split the infinite summation as follows:

$$\sum_{j\geq 0} |\beta_j^k| \leq \sum_{j=0}^{2\lfloor ec \rfloor + 1} |\beta_j^k| + \sum_{j \geq 2\lfloor ec \rfloor + 2} \frac{2^{1-j}}{\mu_k} := I + II.$$

The term I is but a finite summation: for $A = [-1,1]$, Cauchy-Schwarz inequality yields,

$$|\beta_j^k| = | < \varphi_k, \bar{P}_j >_A | \leq \|\varphi_k\|_{L^2(-1,1)} \|\bar{P}_j\|_{L^2(-1,1)} = 1,$$

which gives automatically:

$$I = \sum_{j=0}^{2\lfloor ec \rfloor + 1} |\beta_j^k| \leq \sum_{j=0}^{2\lfloor ec \rfloor + 1} 1 = 2(\lfloor ec \rfloor + 1).$$

Now, the second term rewrites as a geometric series:

$$II = \frac{2^{-2\lfloor ec \rfloor - 1}}{\mu_k} \sum_{j \geq 0} 2^{-j} = \frac{2^{-2\lfloor ec \rfloor - 1}}{\mu_k} \frac{1}{1 - 1/2}.$$

□

Remark 4 *The algorithms by Rauhut and Ward [34] efficiently recover signals which are sparse in the Legendre polynomial basis. But since PSWF are* **not** *sparse in this basis (as spectacularly shown in Theorem 11 of [35]), it turns out that a signal which is sparse in the PSWF basis may not be sparse in the Legendre basis and vice-versa.*

The estimate (18) is somewhat reminiscent of the expression of the H^1 norm given in (13). It isn't as nice as the uniform bound obtained on Legendre polynomials in [34] which is valid for any value of $j \in \mathbb{N}$, however, it compares interestingly with the one obtained in Lemma 1. Indeed, this former one explodes quickly with $\sqrt{\lambda_k} \simeq \mu_k$ as k grows beyond $2c/\pi$ and there is nothing in the denominator which may dampen this process whereas in (18), the μ_k in the denominator is multiplied by a factor $2^{2\lfloor ec \rfloor}$ which helps in keeping it from going to zero too fast (but it will do eventually as k grows with c being fixed). This can be stated quantitatively once again thanks to the estimate by Landau and Widom:

$$\tilde{N} = \frac{2c}{\pi} + \frac{\log c}{\pi^2} \log\left(\frac{1-\alpha}{\alpha}\right) + O(\log c), \qquad \alpha = 2^{-2\lfloor ec \rfloor}.$$

For applying Lemma 1, the easiest choice is to take $\alpha = \frac{1}{2}$; for Lemma 8, we get an extra number of usable indexes $k > N$, this number being approximately for $c \gg 1$,

$$\log\left(\frac{1 - 2^{-2\lfloor ec \rfloor}}{2^{-2\lfloor ec \rfloor}}\right) = \log(2^{2\lfloor ec \rfloor} - 1) \simeq 2\lfloor ec \rfloor \log(2).$$

4.2 Preconditioning $\tilde{N}$ first PSWF gives RIP and exact recovery property

Here we follow completely the ideas presented in §5 of [34], that is to say we consider the Chebyshev probability measure $\nu(dt) = \frac{dt}{\pi\sqrt{1-t^2}}$ for $t \in (-1, 1)$ and we observe that since PSWF are orthogonal with respect to the standard Lebesgue measure $< \varphi_i, \varphi_j >_A = \delta_{i,j}$ (the Kronecker symbol), multiplying by $\sqrt{\pi}(1-t^2)^{\frac{1}{4}}$ makes then orthogonal with respect

to the Chebyshev measure,

$$\int_{-1}^{1} \pi(1-t^2)^{\frac{1}{2}} \varphi_i(t)\varphi_j(t) \frac{dt}{\pi\sqrt{1-t^2}} = \int_{-1}^{1} \varphi_i(t)\varphi_j(t)dt = \delta_{i,j}, \tag{19}$$

and we recover a particular case of (10). We can therefore apply Theorem 6 to deduce:

Theorem 9 *Let $T = 1$ and the normalization be $\|\varphi_k\|_{L^2(-1,1)} = 1$ for all $k \in \mathbb{N}$. Suppose Ψ is the $m \times N$ measurement matrix whose columns read $\sqrt{\pi}(1-t^2)^{\frac{1}{4}}\varphi_k(t_i)$, t_i chosen independently according to Chebyshev probability for $i = 1, ..., m$, $k = 0, 1, ..., \tilde{N} - 1$ and*

$$\frac{m}{\log m} \geq C \left(\frac{2(\lfloor ec \rfloor + 1) + 1}{0.4626} \right)^2 s(\log s)^2 \log \tilde{N}, \qquad \mu_{k=0,1,...,\tilde{N}-1} \geq 2^{-2\lfloor ec \rfloor},$$

then, with probability at least $1 - \tilde{N}^{-\gamma(\log s)^2 \log m}$, the preconditioned ℓ^1-minimization,

$$\inf \|z\|_{\ell^1} \text{ subject to } \Psi z = \Psi x, \tag{20}$$

recovers exactly $\frac{s}{2}$-sparse vectors $x \in \mathbb{R}^{\tilde{N}}$; otherwise the error estimate (9) holds accordingly.

Proof: From the expression of Chebyshev probability, the $m \times m$ diagonal preconditioning matrix P built from $\sqrt{\pi}(1-t_i^2)^{\frac{1}{4}}$ with $i = 1, ..., m$ is invertible. Formula (19) implies that,

$$\left\{ \sqrt{\pi}(1-t^2)^{\frac{1}{4}} \varphi_k(t) \right\}_{k \in \mathbb{N}},$$

is an orthonormal complete system of $L^2(-1, 1; d\nu)$. Moreover, (18) gives a L^∞ bound on a subset of this system for all indexes $k < \tilde{N}$ such that $\mu_{\tilde{N}-1} \geq 2^{-2\lfloor ec \rfloor}$. Hence we are in position to apply Theorem 6 with K_N given by (18), $\delta = 0.4626$ and $\Psi = P\Phi$: this yields with high probability the restricted isometry constant of $\Psi/\sqrt{m}$ being $\delta_s < 0.4627$ and thus allows to conclude the proof by invoking Theorem 5 with a sparsity index $\frac{s}{2}$. □

It is interesting to compare the contents of Theorems 7 and 9 with both give sufficient conditions for exact recovery of $\frac{s}{2}$-sparse vectors: the first one involves only the N first PSWF which are endowed with eigenvalues greater than $\frac{1}{2}$ which, after a random sampling according to the uniform probability on $[-1, 1]$, constitute the columns of the measurement matrix Φ. The minimum number of samples is a consequence of the pointwise bound proved in Lemma 1. The "miracle" comes from the Theorem 3 which states that spectral accuracy

holds in particular for this (small) subset of N first PSWF hence such a strategy can be considered reliable for recovering smooth functions belonging to $H^r(-1,1)$. The second one, instead, takes advantage from both the preconditioning technique proposed in [34] and the classical trick consisting in rewriting PSWF as a Legendre polynomials expansion. Then, preconditioning the Legendre system allows to derive another (possibly larger) bound for a larger subset of $\tilde{N} > N$ PSWF and this leads to similar recovery properties valid for a supposedly wider class of signals. On Fig.5 a comparison of the location of "big values" inside

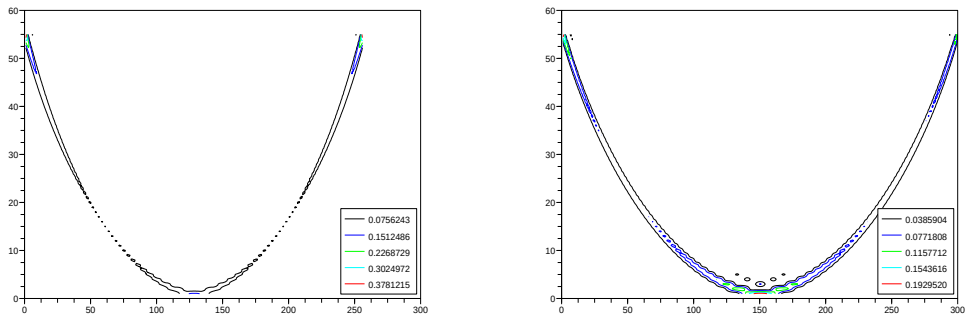

Fig. 5. Biggest values appearing in the PSWF sensing 256×55 matrix (left, used in §3) and in the preconditioned 300×55 one (right).

the sensing matrices is shown: on the left, this is the Φ sensing matrix built in Theorem 7 for 256 points griding the interval $[-1,1]$, $c = 85$ and $N = 2c/\pi$ which corresponds to $\alpha \simeq \frac{1}{2}$. On the right, this is the Ψ preconditioned sensing matrix used in Theorem 9 for 300 points griding the interval $[-1,1]$. The biggest value in modulus is visualized with the red points: on the left, it is located on the "tails" of the PSWF which correspond to the eigenvalues in the transition zone, dropping sharply from one to zero whereas on the right, it is on the mean value of the most concentrated eigenfunction $\varphi_0(0) = \frac{1}{\mu_0}\int_{-1}^{1}\varphi_0(t)dt$.

As recalled in [34], computing a set of locations t_i drawn from Chebyshev probability on $[-1,1]$ is easy: it suffices to get first a set of values τ_i according to the uniform probability on $[0,\pi]$, then $t_i = \cos(\tau_i)$ meets the requirement.

4.3 Preconditioning and efficient recovery of functions in $\tilde{H}_c^r(-1,1)$

First, we state a variant of a result shown by Foucart and Lai [19] (see also [6,8]):

Theorem 10 *Let Φ and P be matrices of sizes $m \times N$ and $m \times m$, respectively, and denote*

$\Psi = P\Phi$ *for* $|P| \neq 0$. *Suppose that the restricted isometry constant of* Ψ *satisfies:*

$$\delta_{2s} < 0.4627.$$

Then, if $x \in \mathbb{R}^N$ *and* $y := \Phi x + \eta$, $\eta \in \mathbb{R}^m$ *being a noise vector, with* $\|\eta\|_{\ell^2} \leq \varepsilon$, *the solution of the* ℓ^1 *minimization program,*

$$\min_{z \in \mathbb{R}^N} \|z\|_{\ell^1} \text{ subject to } \|\Psi z - Py\|_{\ell^2} \leq \|P\|\varepsilon, \tag{21}$$

satisfies the following error estimates:

$$\|x - x^*\|_{\ell^1} \leq C_1 \sigma_s(x)_{\ell^1} + D_1 \|P\| \varepsilon \sqrt{s}, \qquad \|x - x^*\|_{\ell^2} \leq C_2 \frac{\sigma_s(x)_{\ell^1}}{\sqrt{s}} + D_2 \|P\| \varepsilon. \tag{22}$$

We recall that since $\|P\|$ is a subordinated matrix norm, it can be defined as:

$$\|P\| := \inf_{\alpha \in \mathbb{R}^+} \left\{ \|Pv\|_{\ell^2(\mathbb{R}^m)} \leq \alpha \|v\|_{\ell^2(\mathbb{R}^m)} \text{ for } v \in \mathbb{R}^m \right\}.$$

And it is very natural to fix the noise level of the preconditioned program (21) as $\|P\|\|\eta\|_{\ell^2}$. With this result at hand, it is now possible to prove the following corollary of Theorem 9:

Corollary 2 *Let* $c > 0$ *be the Slepian parameter,* $T = 1$, *the normalization* $\|\varphi_k\|_{L^2(-1,1)} = 1$ *for* $k \in \mathbb{N}$, $f \in \tilde{H}^r_c(-1,1)$, *and fix* $s, N \in \mathbb{N}^2$ *with* $s \ll N$. *Assume further that they satisfy,*

$$\frac{m}{\log m} \geq C \left(\frac{2(\lfloor ec \rfloor + 1) + 1}{0.4626} \right)^2 s (\log s)^2 \log N, \qquad \mu_{k=0,1,\ldots,N-1} \geq 2^{-2\lfloor ec \rfloor}, \tag{23}$$

and $(t_1, \ldots, t_m) \in [-1,1]^m$ *are* m *sampling points drawn at random from the Chebyshev probability measure. Then, with probability exceeding* $1 - N^{-\gamma(\log s)^2 \log m}$, *the "best* $\frac{s}{2}$*-term approximation" of the vector* $\vec{f} := (< f, \varphi_k >_{[-1,1]})_{k=0,\ldots,N-1} \in \mathbb{R}^N$ *can be recovered out of the set of values* $y := (f(t_i))_{i=1,\ldots,m} \in \mathbb{R}^m$ *by solving the* ℓ^1 *minimization program:*

$$\vec{f}^* = \arg\min \|z\|_{\ell^1(\mathbb{R}^N)} \text{ subject to } \|\Psi z - Py\|_{\ell^2(\mathbb{R}^m)} \leq \sqrt{\pi}\, \|f\|_{\tilde{H}^r_c(-1,1)}\, N^{-r}.$$

Precisely, the following error estimates hold:

$$\|\vec{f} - \vec{f}^*\|_{\ell^1} \leq C_1 \sigma_{\frac{s}{2}}(\vec{f})_{\ell^1} + D_1 N^{-r} \|f\|_{\tilde{H}^r_c(-1,1)} \sqrt{\frac{\pi s}{2}},$$

$$\|\vec{f} - \vec{f}^*\|_{\ell^2} \leq C_2 \frac{\sigma_{\frac{s}{2}}(\vec{f})_{\ell^1}}{\sqrt{\frac{s}{2}}} + D_2 \sqrt{\pi}\ N^{-r} \|f\|_{\tilde{H}^r_c(-1,1)}.$$

The constants $C, \gamma, C_1, D_1, C_2, D_2$ are universal.

Proof: The system φ_k, $k \in \mathbb{N}$ is complete in $L^2(-1,1)$ for any $c > 0$ thus $f \in \tilde{H}^r_c \subset L^2(-1,1)$ admits an expansion of the form: (we dropped the $[-1,1]$ for clarity)

$$f = \sum_{k \geq 0} < f, \varphi_k > \varphi_k = \left(\sum_{k=0}^{N-1} + \sum_{k \geq N} \right) < f, \varphi_k > \varphi_k := \sum_{k=0}^{N-1} < f, \varphi_k > \varphi_k + \eta.$$

Thanks to the results of [46] recalled in Theorem 4, we get that $\|\eta\|_{L^2(-1,1)} \leq N^{-r} \|f\|_{\tilde{H}^r_c}$. We know from Theorem 6 that the condition (23) is enough to ensure that the restricted isometry constant for $\Psi/\sqrt{m}$ satisfies $\delta_s < 0.4627$ with high probability. With probability 1, the diagonal preconditioning $m \times m$ matrix P is invertible and its operator norm is $\|P\| = \sqrt{\pi}$. So, relying on Theorem 10, the ℓ^1 minimization program furnishes a "best $\frac{s}{2}$-term solution" $\vec{f}^*$ satisfying the error estimates (22) for $\varepsilon := N^{-r} \|f\|_{\tilde{H}^r_c} \|P\|$. The constants C, γ are still the same as in the previous results. □

Remark 5 *The approach in [34] is different: Rauhut and Ward establish the existence of a restricted isometry constant for their preconditioned matrix $\Psi/\sqrt{m}$ with high probability, but they solve a non-preconditioned ℓ^1 program involving a supposedly higher noise level (multiplied by $\sqrt{m}$). Here, we prove the same type of restricted isometry property, and then, for improving numerical efficiency, the preconditioned ℓ^1 program involving Ψ is treated.*

4.4 Second set of numerical experiments (with preconditioning)

In this section too, ROMP is still used in order to lower as much as possible the CPU cost.

4.4.1 Exact recovery property for sparse signals with $N < 2c/\pi$

First, we verify that the exact recovery property works fine in practice: the same framework is set up than in §3.4.1 with $c = 85$, $T = 1$ and 300 grid points. We generate the same

type of random signals which are sparse in the PSWF system of φ_k, $k \leq 2c/\pi \simeq 55$. The random sampling points are taken according to the Chebyshev probability measure which tends to give more importance to the areas close to the edges of the computational domain. The presentation of the numerical results is identical to the one of §3.4.1 except for the

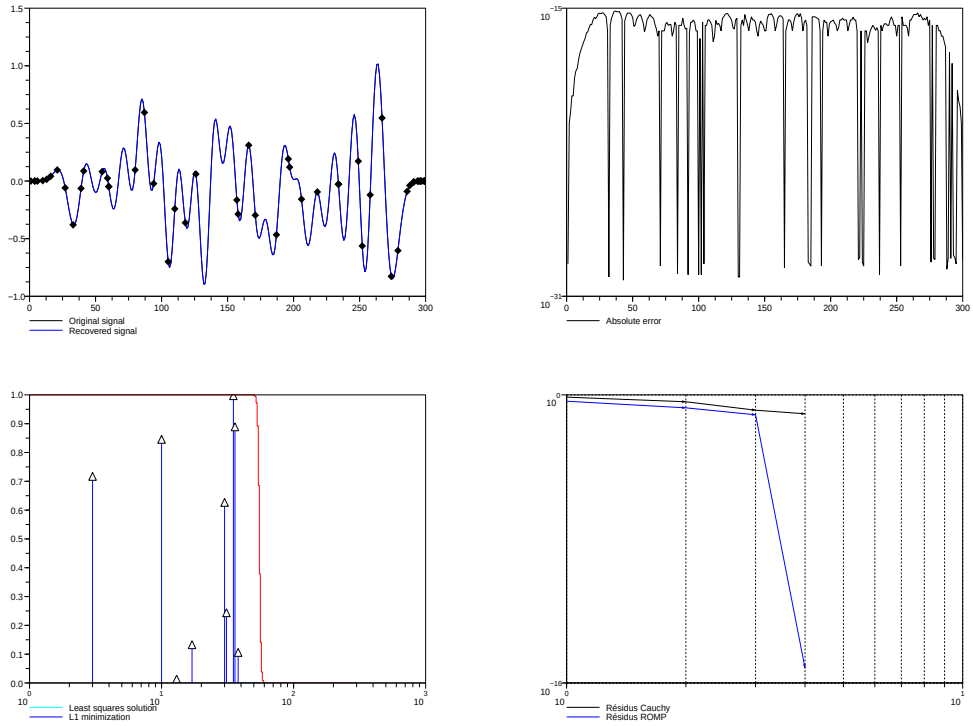

Fig. 6. Exact recovery of $\vec{f}$ with $\|\vec{f}\|_{\ell^0} = 9$ from 54 samples taken as random.

comparison with the least-squares selection procedure as we know that it doesn't deliver the expected sparse representation. Instead, the processing based on ROMP converges very quickly and the absolute pointwise error is below 10^{-15}, which can be considered as fully satisfying. The values of the 9 PSWF coefficients is good too; $m = \frac{3}{2} 9 \log(55) \simeq 54$ samples have been used.

4.4.2 Taking advantage of preconditioning with $N > 2c/\pi$

We stress that for the present example, the bound on the non-preconditioned sensing matrix as computed in Lemma 1 would cease to be useful for practical computations because of the quick decay of $\sqrt{\lambda_k}$. Nevertheless, we set up the same framework than in the preceding subsection, but his time, we allow the index k to grow up to $2c/\pi + 35 \simeq 90$, which results in the value $\sqrt{\lambda_{90}} \simeq 10^{-15}$. The compensating factor $2^{2\lfloor ec \rfloor} = 2^{462} \simeq 1.19.10^{139}$ would probably permit to go much beyond this limit. We built up a random signal such that

$\vec{f} \in \mathbb{R}^{90}$, $\|\vec{f}\|_{\ell^0} = 9$ and $m = \frac{3}{2}9\log(90) \simeq 61$ samples have been taken independently according to Chebyshev probability. On Fig. 7, we see that the ROMP algorithm converged

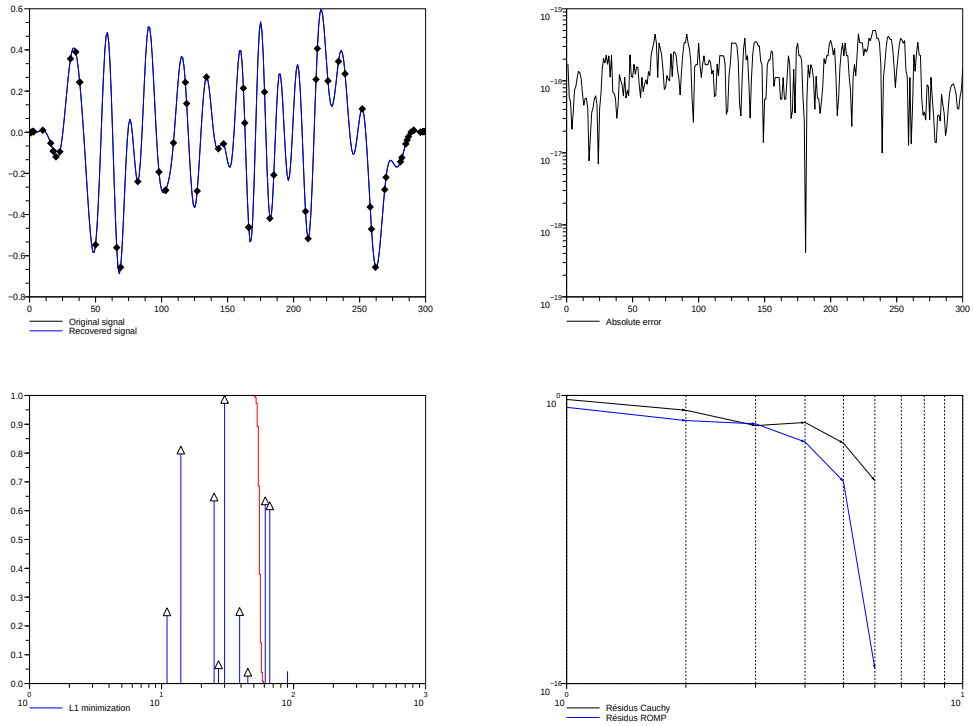

Fig. 7. Exact recovery of $\vec{f}$ with $\|\vec{f}\|_{\ell^0} = 9$ from 61 samples taken as random.

nicely, but more iterations are needed when compared to former test-cases. Out of 9, 2 PSWF were located beyond the critical index $2c/\pi$ and they have been well recovered. The absolute errors are still of the order of 10^{-15} thus the original and the recovered signals are indistinguishable from one another.

Remark 6 *The PSWF functions inside the range $\frac{2c}{\pi} \leq k \leq c$ are usually the most useful in order to carry out an extrapolation process based on the PSWF, seen this time as a DOS in the sense of Theorem 1 (see [41] p.388): let $f \in PW_\omega$ with $\omega = \frac{c}{2\pi}$, and suppose only its restriction $\mathcal{T}f$ to $A = [-1, 1]$ is available:*

$$\mathcal{T}f = f\chi_{[-1,1]} = \sum_{k\geq 0}\left(\int_{-1}^{1} f(t)\varphi_k(t)dt\right)\varphi_k \Rightarrow f = \sum_{k\geq 0}\left(\int_{-1}^{1} f(t)\varphi_k(t)dt\right)\frac{\psi_k}{\sqrt{\lambda_k}},$$

is the most direct way to recover values of such a bandlimited function outside the observations interval. Now, it is in general hopeless to try to recover coefficients corresponding to very small eigenvalues at indexes $k > c$, and even if it were feasible, numerical truncations

errors would be highly amplified by the very small divisors $\sqrt{\lambda_k}$. *However, if the function* f *under consideration is sparse in the base* ψ_k, *or equivalently, if* $\mathcal{T}f$ *is sparse in* φ_k, *then hopefully the preconditioned* ℓ^1 *minimization process will be able to recover its coefficients exactly for* $2c/\pi \leq k \leq c$ *and its bandlimited extrapolation might give good results.*

4.4.3 Case of a noisy signal with $N < 2c/\pi$

We now aim at recovering a signal which has been slightly corrupted by noise. Our noise term $\eta(t)$ is generated by a random number generator working with the uniform probability on $[-1, 1]$. Hence, we still generate a random s-sparse signal with a coefficient vector $\vec{f}_N$ belonging to $[-1, 1]^N$, but we perturb the resulting observations $\Phi\vec{f}_N$ with Φ being $m \times N$ by the vector of noisy terms $0.1\eta \in [-0.1, 0.1]^m$: these noisy observations are shown by means of the red curve on Fig. 8 (top, left). It is on this red curve that the m samples are

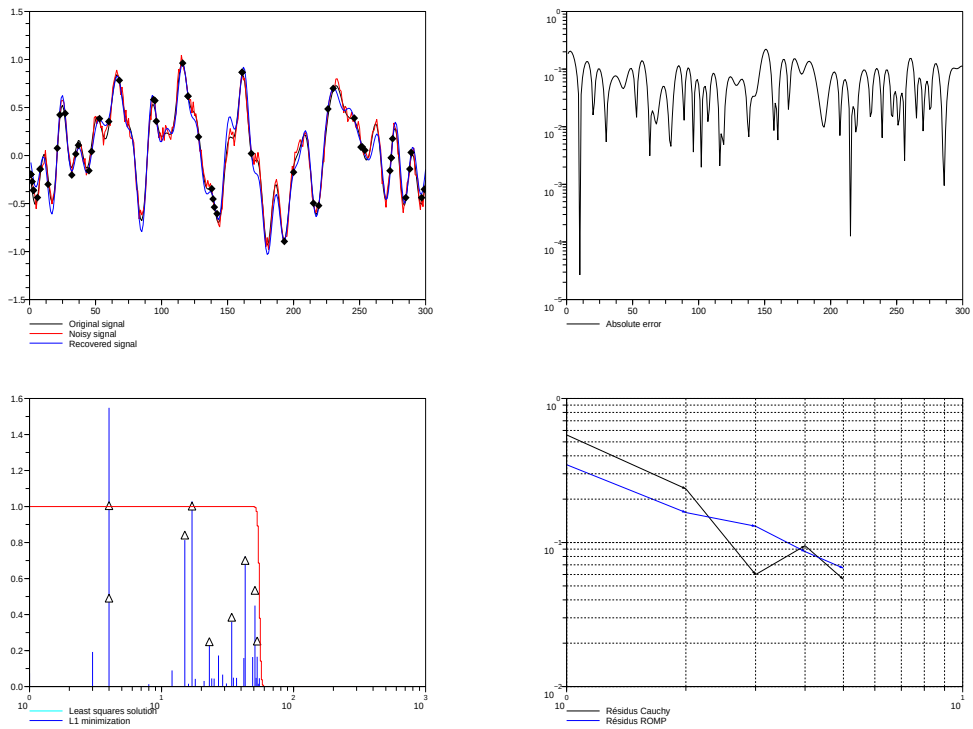

Fig. 8. Recovery involving 10% of noise with $s = 9$ and $c = 85$.

taken, and the recovery algorithm recovers an approximate signal which is depicted in blue. The absolute difference between the original signal and its recovered version is displayed on the top right graphic of Fig. 8: clearly, the machine accuracy error level has been lost, but the general error level around 10^{-1} matches $\|\eta\|_{\ell^\infty}$. The recovery of coefficients can also be considered satisfying.

4.4.4 *Case of a noisy signal with $N > 2c/\pi$*

This is a very unstable case because the recovering process "tries" to express the random noise term η by means of a linear combination of the supplementary PSWF we allow itself to consider inside the fast decay area (as shown on Fig.3 of [3]). Hence it produces very big coefficients associated with indexes k which are associated with very small eigenvalues λ_k, the columns of the sensing matrix remaining bounded nonetheless thanks to the preconditioning. Since these PSWF with high index $k > \frac{2c}{\pi}$ don't have a big influence on the

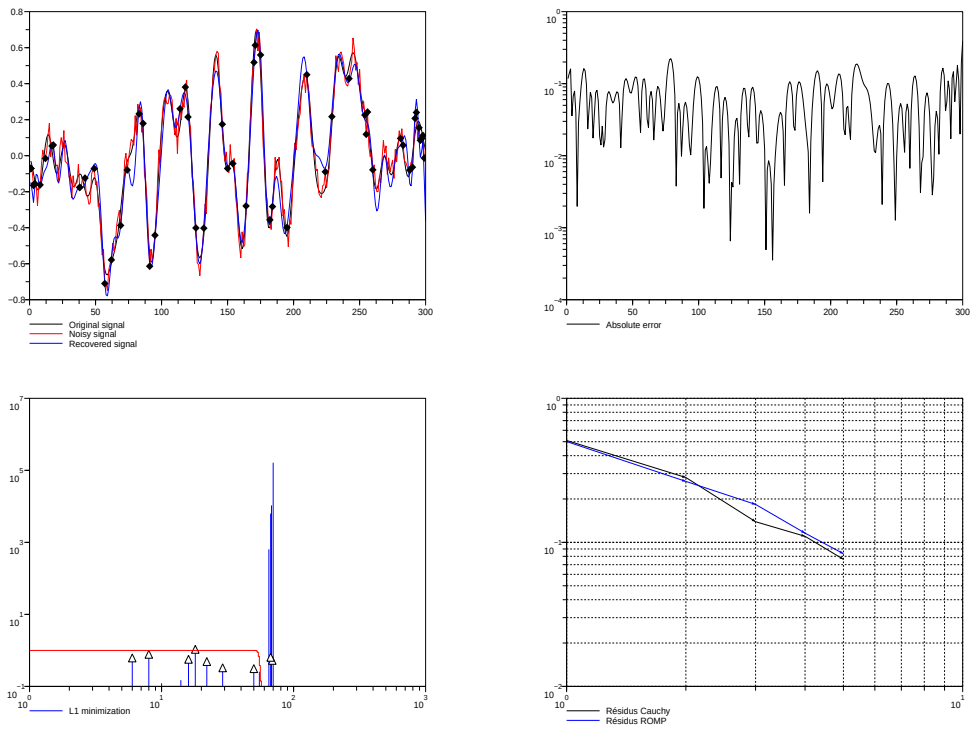

Fig. 9. Recovery with 10% of noise, $s = 9$, $c = 85$ and $N = 15 + \frac{2c}{\pi}$.

general shape of the signal inside the observations interval $[-1, 1]$, the absolute error in the recovery is roughly the same compared to the preceding test-case. However, the situation when looking at the recovery of the coefficient vector on Fig. 9 shows a much worse picture: in particular, setting up the extrapolation algorithm suggested in Remark 6 is completely doomed to failure because the division by $\sqrt{\lambda_k}$ will amplify even more the spurious coefficients which result from the noise term (which perturbs Picard's conditions [15]).

5 Conclusion and outlook

We presented in this paper a general way to apply Compressed Sensing methodology in order to perform an efficient recovery of signals having the peculiar feature of admitting a sparse representation in the PSWF base. Two distinct ways to proceed have been studied, either establishing a direct L^∞ bound on the sensing matrix (see Lemma 1) which is a technique which relies only on manipulating the eigenvalue problem satisfied by the measurement ensemble, either taking advantage of the particular expansion of Slepian's functions in the Legendre polynomials in order to take advantage of the preconditioning technique proposed in [34] and proving an extended L^∞ bound (see Lemma 8) in this special case. The advantage in proceeding this way lies in the possibility to recover PSWF coefficients in the area corresponding to small eigenvalues hence to try to extrapolate the observed signal outside the observations interval (as suggested in Remark 6). Clearly, this program cannot be set up in case the observations are corrupted by a small noise term.

This approach can be generalized to the newly introduced variants of classical Slepian functions: for instance, the PSWF associated to the fractional Fourier transform [32] and the ones in [47] which are "bandlimited in a weighted sense" and therefore generalize Chebyshev polynomials. This second construction could be the most straightforward one for extending the preconditioned framework proposed in §4 of the present paper, even at the price of heavier computations. Finally, the "wavelet prolate functions" studied in [27,20] may also be convenient, at least for the techniques developed in our §3.

Acknowledgments

The author is grateful to Prof. E.J. Candès and H. Rauhut for advices and discussions.

References

[1] R.G. Baraniuk, M. Davenport, R. DeVore, M. Wakin, *A simple proof of the restricted isometry property for random matrices*, Construct. Approx. **28** (2008) 253–263.

[2] G. Beylkin, K. Sandberg, *Wave propagation using bases for bandlimited functions*, Wave Motion **41** (2005) 263–291.

[3] John P. Boyd, *Approximation of an analytic function on a finite real interval by a band-limited function and conjectures on properties of prolate spheroidal*

functions, Appl. Comp. Harm. Anal. **15** (2003) 168–176.

[4] John P. Boyd, *Prolate spheroidal wave functions as an alternative to Chebyshev and Legendre polynomials for spectral element and pseudo-spectral algorithms*, J. Comp. Phys. **199** (2004) 688–716

[5] E.J. Candès, *Compressive sampling*, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.

[6] E.J. Candès, *The restricted isometry property and its implications for compressed sensing* C.R. Acad. Sci. Paris, Serie I, **346** 589–592.

[7] E.J. Candès and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems **23** (2007) 969–985.

[8] E.J. Candès, J. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** 1207–1223.

[9] C.H. Chang, C.W. Ha, *Sharp inequalities of singular values of smooth kernels*, Integral Equations and Operator Theory 35 (1999) 20–27.

[10] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Processing Letters, **14** (2007) 707–710.

[11] Q.Y. Chen, D. Gottlieb, J.S. Hesthaven, *Spectral methods based on Prolate spheroidal wave functions for hyperbolic PDEs*, SIAM J. Numer. Anal. **43** (2005) 1912–1933.

[12] A. Cohen, W. Dahmen, R. DeVore, *Compressed sensing and best k-term approximation*, J. Amer. Math. Soc. **22** (2009) 211-231.

[13] R.R. Coifman, S. Lafon, *Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions*, Applied and Computational Harmonic Analysis **21** (2006) 31–52.

[14] I. Daubechies, *Time-frequency localization operators: a geometric phase-space approach*, IEEE Trans. Inform. Theory **34** (1988) 605–612.

[15] J.B. Diaz, F.T. Metcalf, *On iteration procedures for equations of the first kind, $Ax = y$ and Picard's criterion for the existence of a solution*, Math. Comp. **24** (1970) 923–935.

[16] D.L. Donoho, P.B. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. **49** (1989) 906–931.

[17] K. Drouiche, D. Kateb and C. Noiret, *Regularization of the ill-posed problem of extrapolation with the Malvar-Wilson wavelets*, Inverse Problems **17** (2001) 1513-1533.

[18] S. Foucart, *A note on guaranteed sparse recovery via ℓ_1-minimization*, preprint (2010).

[19] S. Foucart, M. Lai, *Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for* $0 < q \leq 1$, Appl. Comput. Harm. Anal. **26** (2009) 395-407.

[20] L. Gosse, *A Donoho-Stark criterion for stable signal recovery in discrete wavelet subspaces*, preprint (2010).

[21] John A. Gubner, *A Simple Method for Computing Projections onto Subspaces of Prolate Spheroidal Wave Functions*, sumbitted to IEEE Trans. Comm., see `http://eceserv0.ece.wisc.edu/ gubner/pswfProjIpSubmitted2col.pdf`

[22] J.A. Hogan, J.D. Lakey, **Time-frequency and time-scale methods**, Birkhauser, 2005.

[23] Scott Izu, J.J. Lakey, *Time-Frequency Localization and Sampling of Multiband Signals*, Acta applicandae mathematicae **107** (2009) 399-435 [Issue dedicated to: Applied Harmonic Analysis: Wavelets and Sampling].

[24] A. Karoui, T. Moumni, *New efficient methods of computing the prolate spheroidal wave functions and their corresponding eigenvalues*, Appl. Comp. Harm. Anal. **24** (2008) 269–289.

[25] K. Khare, *Bandpass sampling and bandpass analogues of prolate spheroidal wave functions*, Signal Proc. **86** (2006), 1550–1558.

[26] H.J. Landau, H. Widom, *Eigenvalue distribution of time and frequency limiting*, J. Math. Anal. 77 (1980) 469–481.

[27] L.-C. Lin, C.-C. Jay Kuo, *On theory and regularization of scale-limited extrapolation*, Signal Processing **54** (1996) 225–237.

[28] Y.M. Liu, G. Walter, *Irregular sampling in wavelet subspaces*, J. Fourier Anal. Applic. **2** (1995), 181–189.

[29] I.C. Moore, M. Cada, *Prolate Spheroidal wave functions, an introduction to the Slepian series and its properties*, Appl. Comput. Harmonic Anal. **16** (2004) 208–230.

[30] D. Needell and R. Vershynin, *Uniform Uncertainty Principle and signal recovery via Regularized Orthogonal Matching Pursuit*, Foundations of Computational Mathematics, 9 pp. 317-334, 2009.

[31] Manuela Nees, *Products of orthogonal projections as Carleman operators*, Integral Equations and Operator Theory 35 (1999) 85–92.

[32] 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.

[33] H. Rauhut, *Compressive sensing and structured random matrices*, in "Theoretical foundations and numerical methods for sparse recovery" (M. Fornassier Ed.) **9** (2010) Radon series in computational and applied mathematics, deGruyter.

[34] H. Rauhut, R. Ward, *Sparse Legendre expansions via ℓ^1-minimization*, Preprint.

[35] Vladimir Rokhlin, Hong Xiao, *Approximate formulae for certain prolate spheroidal wave functions valid for large values of both order and band-limit*, Appl. Comp. Harm. Anal. **22** (2007) 105–123.

[36] Kristian Seip, *Reproducing formulas and double orthogonality in Bargmann and Bergman spaces*, SIAM J. Math. Anal. **22** (1991) 856–876.

[37] S. Senay, L.F. Chaparro, L. Durak, *Reconstruction of nonuniformly sampled time-limited signals using prolate spheroidal wave functions*, Signal Proc. **89** (2009) 2585-2595.

[38] H. Shapiro, *Stefan Bergman's theory of doubly-orthogonal functions: an operator-theoretic approach*, Proc. R.I.A. **89** Sect. A (1989) 49-??.

[39] H. Shapiro, *Reconstructing a function from its values on a subset of its domain– A Hilbert space approach*, J. Approx. Theory **46** (1986) 385-402.

[40] Yoel Shkolniskya, Mark Tygertb, Vladimir Rokhlin, *Approximation of bandlimited functions*, Appl. Comp. Harm. Anal. **21** (2006) 413–420.

[41] D. Slepian, *Some comments on Fourier analysis, uncertainty and modeling*, SIAM Rev. **25** (1983) 379–393.

[42] Slepian D., Pollak H.O., *Prolate spheroidal wave functions, Fourier analysis and uncertainty. I.*, Bell System Technical Journal **40** (1961) 43–63.

[43] G.G. Walter, X. Shen, *Sampling with Prolate Spheroidal Functions*, J. Sampling Theory Signal Image Proc. **2** (2003) 25-52.

[44] Gilbert G Walter, Xiaoping Shen, *Wavelet based on prolate spheroidal wave functions*, J. Fourier Anal. Appl. **10** (2004), 1–26

[45] G. Walter and T. Soleski, *A new friendly method of computing prolate spheroidal wave functions and wavelets*, Appl. Comput. Harmon. Anal. **19** (2005)

[46] L.L. Wang, *Analysis of spectral approximations using prolate spheroidal wave functions*, Math. Comp. **79** (2010) 807–827

[47] L.-L. Wang, J. Zhang, *A new generalization of the PSWF with applications to spectral approximations on quasi-uniform grids*, Appl. Comp. Harm. Anal. (2010), to appear.

[48] H. Xiao, V. Rokhlin, N. Yarvin, *Prolate spheroidal wave functions, quadrature and interpolation*, Inverse Problems **17** (2001) 805–838.

[49] D.C. Youla, *Generalized image restoration by the method of alternating orthogonal projections*, IEEE Trans. on Circ. & Syst. **25** (1978) 694- 702.

[50] A. Zayed, *A generalization of the prolate spheroidal wave functions*, Proc. Amer. Math. Soc. **135** (2007) 2193–2203.