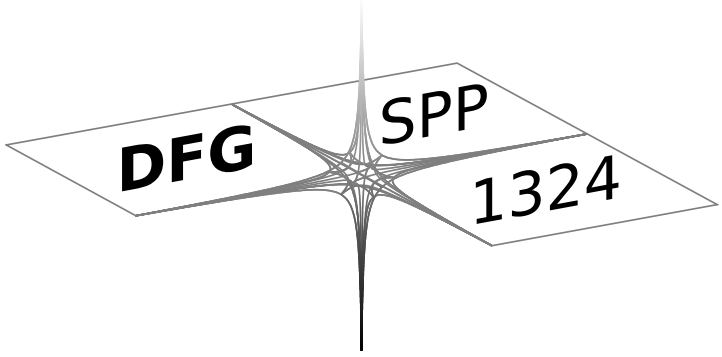
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Multilevel Preconditioning for Adaptive Sparse Optimization

S. Dahlke, M. Fornasier, T. Raasch

Preprint 25



Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

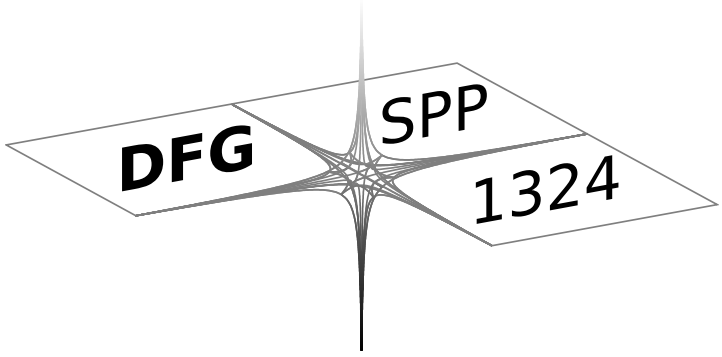
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Multilevel Preconditioning for Adaptive Sparse Optimization

S. Dahlke, M. Fornasier, T. Raasch

Preprint 25



The consecutive numbering of the publications is determined by their chronological order.

The aim of this preprint series is to make new research rapidly available for scientific discussion. Therefore, the responsibility for the contents is solely due to the authors. The publications will be distributed by the authors.

Multilevel Preconditioning for Adaptive Sparse Optimization

Stephan Dahlke*, Massimo Fornasier† and Thorsten Raasch‡

August 13, 2009

Abstract

We are concerned with the efficient numerical solution of minimization problems in Hilbert spaces involving sparsity constraints. These optimizations arise, e.g., in the context of inverse problems. In this work we analyze a very efficient variant of the well-known iterative soft-shrinkage algorithm for large or even infinite dimensional problems. This algorithm is modified in the following way. Instead of prescribing a fixed thresholding parameter, we use a decreasing thresholding strategy. Moreover, we use suitable variants of the adaptive schemes derived by Cohen, Dahmen and DeVore [11, 12] for the approximation of the infinite matrix-vector products. We derive a block multiscale preconditioning technique which allows for local well-conditioning of the underlying matrices and for extending the concept of *restricted isometry property* to infinitely labelled matrices. The careful combination of these ingredients gives rise to a numerical scheme that is guaranteed to converge with exponential rate, and which allows for a controlled inflation of the support size of the iterations. We also present numerical experiments that confirm the applicability of our approach which extends concepts from *compressed sensing* to large scale simulation.

AMS subject classification: 41A25, 65F35, 65F50, 65N12, 65T60.

Key Words: Operator equations, multiscale methods, adaptive algorithms, sparse optimization and compressed sensing, multilevel preconditioning, wavelets, iterative thresholding, restricted isometry property.

Contents

*The work of this author has been supported by Deutsche Forschungsgemeinschaft (DFG), Grant DA 360/12–1.

†The work of this author has been supported by the FWF project Y 432-N15 START-Preis "Sparse Approximation and Optimization in High Dimensions" and Deutsche Forschungsgemeinschaft (DFG), Grant DA 360/12–1.

‡The work of this author has been supported by Deutsche Forschungsgemeinschaft (DFG), Grants DA 360/7–1 and DA 360/12–1.

1 Introduction

In this paper we are concerned with the efficient minimization of functionals of the type:

$$J(u) := \|Ku - y\|_Y^2 + 2\|(\langle u, \tilde{\psi}_\lambda \rangle)_{\lambda \in \mathcal{I}}\|_{\ell_{1,\alpha}(\mathcal{I})}, \tag{1}$$

where $K : X \to Y$ is a bounded linear operator acting between two separable Hilbert spaces X and Y, $y \in Y$ is a given datum, and $\Psi := \{\psi_\lambda\}_{\lambda \in \mathcal{I}}$ is a prescribed countable basis for X with associated dual $\tilde{\Psi} := \{\tilde{\psi}_\lambda\}_{\lambda \in \mathcal{I}}$. For $1 \leq p < \infty$, the sequence norm $\|\mathbf{u}\|_{\ell_{p,\alpha}(\mathcal{I})} := \left(\sum_{\lambda \in \mathcal{I}} |u_\lambda|^p \alpha_\lambda\right)^{1/p}$ is the usual norm for weighted p-summable sequences, with weight $\alpha = (\alpha_\lambda)_{\lambda \in \mathcal{I}} \in \mathbb{R}_+^{\mathcal{I}}$. Associated to the basis, we are given the synthesis map $F : \ell_2(\mathcal{I}) \to X$ defined by

$$F\mathbf{u} := \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda, \quad \mathbf{u} \in \ell_2(\mathcal{I}). \tag{2}$$

We can re-formulate equivalently the functional in terms of sequences in $\ell_2(\mathcal{I})$ as follows:

$$J(\mathbf{u}) := J_\alpha(\mathbf{u}) = \|(K \circ F)\mathbf{u} - y\|_Y^2 + 2\|\mathbf{u}\|_{\ell_{1,\alpha}(\mathcal{I})}. \tag{3}$$

For ease of notation let us write $A := K \circ F$. Such a functional turns out to be very useful in many practical problems, where one cannot observe directly the quantities of most interest; instead their values have to be inferred from their effect on observable quantities. When this relationship between the observable y and the interesting quantity u is (approximately) linear the situation can be modeled mathematically by the equation

$$y = Ku \ , \tag{4}$$

If K is a "nice" (e.g., well-conditioned), easily invertible operator, and if the data y are free of noise, then this is a well-known task which can be addressed with standard numerical analysis methods. Often, however, the mapping K is not invertible or ill-conditioned. Moreover, typically (4) is only an idealized version in which noise has been neglected; a more accurate model is

$$y = Ku + e \ , \tag{5}$$

in which the data are corrupted by an (unknown) noise e. In order to deal with this type of reconstruction problem a *regularization* mechanism is required [27]. Regularization techniques try, as much as possible, to take advantage of (often vague) prior knowledge one may have about the nature of u. The approach modelled by the functional J in (1) is indeed tailored to the case when u can be represented by a *sparse* expansion, i.e., when u can be represented by a series expansion (2) with respect to a basis or a frame [9, 21] that has only a small number of large coefficients. In this paper, as in [22], we model the sparsity constraint by a regularizing ℓ_1-term in the functional to be minimized; it was shown in [22] that this assumption does indeed correspond to a regularization scheme,

and it promotes sparse minimal solutions.

It has been recognized for long [17–19] that for a large class of operators, including integral operators, multiscale bases, such as wavelets, give rise to matrix representations that are *compressible* (roughly speaking an infinitely indexed matrix is *compressible* when it can be well-approximated by small finitely indexed submatrices) and admit a simple diagonal preconditioner in the case of elliptic operators. In this paper we discuss the role of preconditioning in iterative schemes for the numerical minimization of the functional J associated to ill-posed problems (4). To do that we need first to address the compressibility of matrices A^*A. Then, similarly to the well-posed case, we propose diagonal and block-diagonal preconditioning strategies of matrix representations of operators K by means of multiscale bases, such as wavelets. Moreover we investigate how to take advantage of the compressibility in order to derive adaptive iterations which tend to work with the minimal number of degrees of freedom. In the following of this introduction we describe in detail how our paper introduces innovations with respect to the current state of the art, and, in particular, we clarify how we can develop efficient preconditioning mechanisms for ill-posed problems, by taking advantage of the expected sparsity of the minimal solution.

1.1 Iterative thresholding and rate of convergence

Several authors have proposed independently an iterative soft-thresholding algorithm to approximate minimizers $\mathbf{u}^* := \mathbf{u}^*_\alpha$ of the functional in (2), see [26, 30, 44, 45]. More precisely, $\mathbf{u}^*$ is the limit of sequences $\mathbf{u}^{(n)}$ defined recursively by

$$\mathbf{u}^{(n+1)} = \mathbb{S}_\alpha \left[\mathbf{u}^{(n)} + A^* y - A^* A \mathbf{u}^{(n)} \right] \quad , \tag{6}$$

starting from an arbitrary $\mathbf{u}^{(0)}$, where $\mathbb{S}_\alpha$ is the soft-thresholding operation defined by $\mathbb{S}_\alpha(\mathbf{u})_\lambda = S_{\alpha_\lambda}(u_\lambda)$ with

$$S_\tau(x) = \begin{cases} x - \tau & x > \tau \\ 0 & |x| \leq \tau \\ x + \tau & x < -\tau \end{cases} \quad . \tag{7}$$

This is our starting point and the reference iteration on which we want to work out several innovations. Strong convergence of this algorithm was proved in [22], under the assumption that $\|A\| < 1$ (actually, convergence can be shown also for $\|A\| < \sqrt{2}$ [14]; nevertheless, the condition $\|A\| < 1$ is by no means a restriction, since it can always be met by a suitable rescaling of the functional J, in particular of K, y, and α). Soft-thresholding plays a role in this problem because it leads to the unique minimizer of a functional combining ℓ_2 and ℓ_1-norms, i.e., (see [8, 22])

$$\mathbb{S}_\alpha(\mathbf{a}) = \arg\min_{\mathbf{u} \in \ell_2(\mathcal{I})} \left(\|\mathbf{u} - \mathbf{a}\|^2 + 2\|\mathbf{u}\|_{1,\alpha} \right). \tag{8}$$

We will call the iteration (6) the *iterative soft-thresholding algorithm* or the *thresholded Landweber iteration* (ISTA).

Recently, also the qualitative convergence properties of iterative soft-thresholding have been investigated. Note first that the aforementioned condition $\|A\| < \sqrt{2}$ does not guarantee contractivity of the iteration operator $I - A^*A$, since A^*A may not be boundedly invertible. The insertion of $\mathbb{S}_\alpha$ does not improve the situation since $\mathbb{S}_\alpha$ is nonexpansive (see (16)), but also noncontractive. Hence, for any minimizer $\mathbf{u}^*$ (which is also a fixed point of (6)), the estimate

$$\|\mathbf{u}^* - \mathbf{u}^{(n+1)}\|_{\ell_2(\mathcal{I})} \leq \big\|(I - A^*A)(\mathbf{u}^* - \mathbf{u}^{(n)})\big\|_{\ell_2(\mathcal{I})} \leq \|I - A^*A\|\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2(\mathcal{I})} \quad (9)$$

does not give rise to a linear error reduction. However, under additional assumptions on the operator A or on minimizers $\mathbf{u}^*$, linear convergence of (6) can be easily ensured. In particular, if A fulfills the so-called *finite basis injectivity* (FBI) condition (see [3] where this terminology is introduced), i.e., for any finite set $\Lambda \subset \mathcal{I}$, the restriction A_Λ is injective, then (6) converges linearly to a minimizer $\mathbf{u}^*$ of J. Here we used the notation A_Λ to indicate a submatrix extracted from A by retaining only the columns indexed in Λ. The following simple argument shows indeed that the FBI condition implies linear error reduction as soon as $\|A\| < 1$. In that case, we have strong convergence of the $\mathbf{u}^{(n)}$ to a finitely supported limit sequence $\mathbf{u}^*$. We can therefore find a finite index set $\Lambda \subset \mathcal{I}$ such that all iterates $\mathbf{u}^{(n)}$ and $\mathbf{u}^*$ are supported in Λ. By the FBI condition, A_Λ is injective and hence $A^*A|_{\Lambda\times\Lambda}$ is boundedly invertible, so that $I - A_\Lambda^* A_\Lambda$ is a contraction on $\ell_2(\Lambda)$. Using

$$\mathbf{u}_\Lambda^{(n+1)} = \mathbb{S}_\alpha\big(\mathbf{u}_\Lambda^{(n)} + A_\Lambda^*(v - A_\Lambda \mathbf{u}_\Lambda^{(n)})\big)$$

and an analogous argument as in (9), it follows that $\|\mathbf{u}^* - \mathbf{u}^{(n+1)}\|_{\ell_2(\mathcal{I})} \leq \gamma \|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2(\mathcal{I})}$, where $\gamma = \max\{|1 - \|(A^*A|_{\Lambda\times\Lambda})^{-1}\|^{-1}|, |\|A^*A|_{\Lambda\times\Lambda}\| - 1|\} \in (0,1)$. Typical examples where $A = K \circ F$ fulfills the FBI condition arise when K is injective and Ψ is either a Riesz basis for X or a so-called FBI frame, i.e., each finite subsystem of Ψ is linearly independent. However, depending on Λ, the matrix $A^*A|_{\Lambda\times\Lambda}$ can be arbitrarily badly conditioned, resulting in a constant error reduction γ, arbitrarily close to 1.

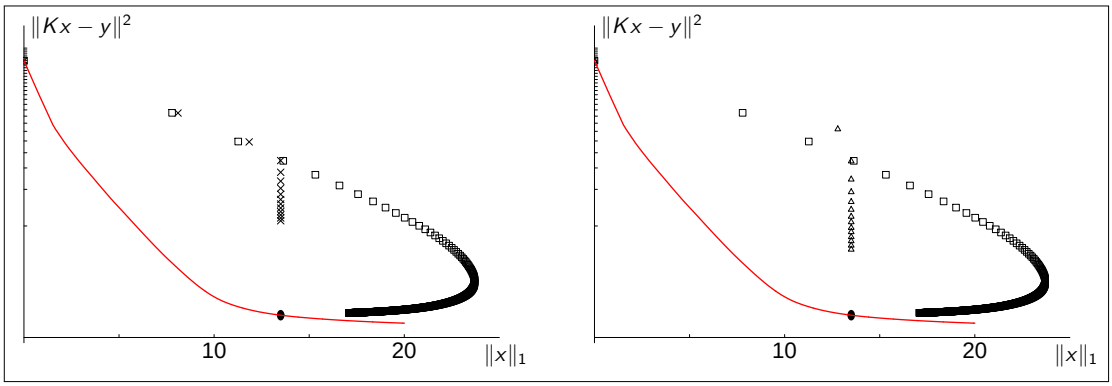
Our first innovation and contribution of this paper is to show that for several FBI operators K and for certain choices of Ψ, the matrix A^*A can be preconditioned by a matrix $D^{-1/2}$, resulting in the matrix $D^{-1/2}A^*AD^{-1/2}$, in such a way that any restriction $(D^{-1/2}A^*AD^{-1/2})_{\Lambda\times\Lambda}$ turns out to be well-conditioned as soon as $\Lambda \subset \mathcal{I}$ is a small set, but independently of its "location" within $\mathcal{I}$. Let us remark that, in particular, we do not claim to be able to have full well-conditioned matrices (as it happens in well-posed problems [17,18] by simple diagonal preconditioning), but that only small arbitrary finite dimensional submatrices are indeed well-conditioned. Let us say that we try to promote a "local" well-conditioning.

Typically we consider injective (non local) compact operators K with Schwartz kernel having certain polynomial decay properties of the derivatives, i.e.,

$$Ku(x) = \int_\Omega \Phi(x,\xi)u(\xi)d\xi, \quad x \in \tilde{\Omega},$$

for $\tilde{\Omega}, \Omega \subset \mathbb{R}^d$, $u \in X := H^t(\Omega)$, and

$$|\partial_x^\alpha \partial_\xi^\beta \Phi(x,\xi)| \leq c_{\alpha,\beta}|x-\xi|^{-(d+2t+|\alpha|+|\beta|)}, \quad t \in \mathbb{R}, \text{ and multi-indexes } \alpha, \beta \in \mathbb{N}^d.$$



(a) (b)

Figure 1: The path, in the $\|x\|_1$ vs. $\|Kx-y\|^2$ plane, followed by the iterates $\mathbf{u}^{(n)}$ of three different iterative algorithms. The operator K and the data y are taken from a seismic tomography problem [37]. The boxes (in both (a) and (b)) correspond to the thresholded Landweber algorithm. In this example, iterative thresholded Landweber (6) first overshoots the ℓ_1 norm of the limit (represented by the fat dot), and then requires a large number of iterations to reduce $\|\mathbf{u}^{(n)}\|_1$ again (500 are shown in this figure). In (a) the crosses correspond to the path followed by the iterates of the projected Landweber iteration (which is given as in (10) for $\beta^{(n)} = 1$); in (b) the triangles correspond to the projected steepest descent iteration (10); in both cases, only 15 iterates are shown. The discrepancy decreases more quickly for projected steepest descent than for the projected Landweber algorithm. The solid line corresponds to the limit *trade-off curve*, generated by $\mathbf{u}^*(\bar{\alpha})$ for decreasing values of $\bar{\alpha} > 0$. The vertical axes uses a logarithmic scale for clarity.

Moreover, for the proper definition of the discrete matrix $A^*A := F^*K^*KF$, we will show that multiscale bases Ψ, such as wavelets, do make a good job for us.

1.2 From the projected gradient method to the iterative soft-thresholding with decreasing thresholding parameter

With such a "local" well-conditioning, it should also be clear that iterating on small sets Λ will also improve the convergence rate. Unfortunately, the thresholded Landweber iteration does not iterate in general on small sets (see also Figure 3 and Figure 4), but it rather starts iterating on relatively large sets, slowly shrinking to the size of the support of the limit $\mathbf{u}^*$.

Let us take a closer look at the characteristic dynamics of the thresholded Landweber iteration in Figure 1. Let us assume for simplicity here that $\alpha_\lambda = \bar{\alpha} > 0$ for all $\lambda \in \mathcal{I}$. As this plot of the discrepancy $\mathcal{D}(\mathbf{u}^{(n)}) = \|Ku^{(n)} - y\|_Y^2 = \|A\mathbf{u}^{(n)} - y\|_Y^2$ versus $\|\mathbf{u}^{(n)}\|_1$ shows, the algorithm converges initially relatively fast, then it overshoots the value $\|\mathbf{u}^*\|_1$ and it takes very long to re-correct back. In other words, starting from $\mathbf{u}^{(0)} = 0$, the

algorithm generates a path $\{\mathbf{u}^{(n)}; n \in \mathbb{N}\}$ that is initially fully contained in the ℓ_1-ball $B_R := \{\mathbf{u} \in \ell_2(\Lambda); \|\mathbf{u}\|_1 \leq R\}$, with $R := \|\mathbf{u}^*\|_1$. Then it gets out of the ball to slowly inch back to it in the limit.

The way to avoid this long "external" detour was proposed in [23] by forcing the successive iterates to remain within the ball B_R. One method to achieve this is to substitute for the thresholding operations the projection $\mathbb{P}_{B_R}$, where, for any closed convex set C, and any $\mathbf{u}$, we define $\mathbb{P}_C(\mathbf{u})$ to be the unique point in C for which the ℓ_2-distance to $\mathbf{u}$ is minimal. With a slight abuse of notation, we shall denote $\mathbb{P}_{B_R}$ by $\mathbb{P}_R$; this will not cause confusion, because it will be clear from the context whether the subscript of $\mathbb{P}$ is a set or a positive number.

Furthermore, modifying the iterations by introducing an adaptive "descent parameter" $\beta^{(n)} > 0$ in each iteration, defining $\mathbf{u}^{(n+1)}$ by

$$\mathbf{u}^{(n+1)} = \mathbb{P}_R \left[\mathbf{u}^{(n)} + \beta^{(n)} A^*(y - A\mathbf{u}^{(n)})\right], \tag{10}$$

does lead, in numerical simulations, to much faster convergence. The typical dynamics of this modified algorithm are illustrated in Fig. 1(b), which clearly shows the larger steps and faster convergence (when compared with the projected Landweber iteration in Fig. 1(a) which is for $\beta^{(n)} = 1$). We shall refer to this modified algorithm as the *projected gradient iteration* or the *projected steepest descent* (PSD). The motivation of

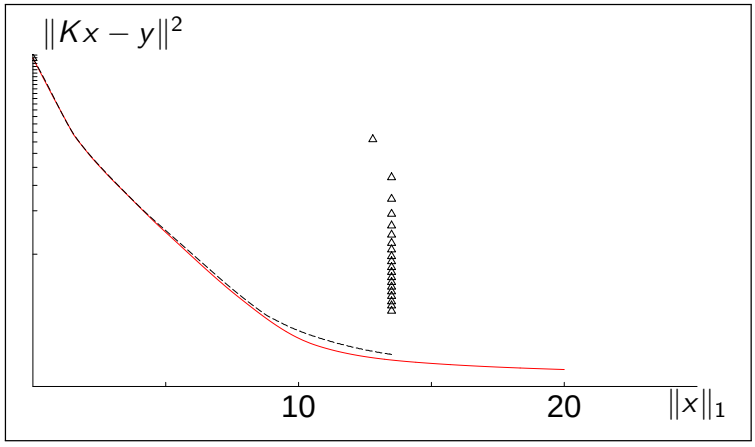


Figure 2: Trade-off curve and its approximation with algorithm (11) in 200 steps.

the faster convergence behavior is the fact that we never leave the target ℓ_1-ball, and we tend not to iterate on large index sets. On the basis of this intuition we find even more promising results for an 'interior' algorithm in which we still project on ℓ_1-balls, but now with a slowly increasing radius, i.e.

$$\mathbf{u}^{(n+1)} = \mathbb{P}_{R^{(n)}} \left(\mathbf{u}^{(n)} + \beta^{(n)} A^*(y - A\mathbf{u}^{(n)})\right) \qquad \text{and} \qquad R^{(n+1)} = (n+1)R/N \tag{11}$$

where N is the prescribed maximum number of iterations (the origin is chosen as the starting point of this iteration). The better performance of this algorithm can be explained by the fact that the projection $\mathbb{P}_R(\mathbf{u})$ onto an ℓ_1-ball of radius R do coincide with a thresholding $\mathbb{P}_R(\mathbf{u}) = \mathbb{S}_{\alpha(\mathbf{u};R)}(\mathbf{u})$ for a suitable thresholding parameter $\alpha = \alpha(\mathbf{u}; R)$ depending on $\mathbf{u}$ and R, which is larger for smaller R [23, Lemma 2]. This in particular implies that the algorithm (11) iterates initially on very small sets which inflate by growing during the process and approach the size of the support of the target minimizer $\mathbf{u}^*$. Unlike the thresholded Landweber iteration and the projected steepest descent [22, 23], unfortunately there is no proof yet of convergence of this 'interior' algorithm, being a very interesting open problem.

The second innovation and contribution of this paper is indeed the proposal and the proof of convergence of an algorithm which mimics the behavior of (11), i.e., it starts with large thresholding parameters $\alpha^{(n)}$ and geometrically reduces them during the iterations to a target limit $\alpha > 0$

$$\mathbf{u}^{(n+1)} = \mathbb{S}_{\alpha^{(n)}}\left[\mathbf{u}^{(n)} + A^*y - A^*A\mathbf{u}^{(n)}\right]. \tag{12}$$

For matrices A for which the restrictions $A^*A|_{\Lambda\times\Lambda}$ are uniformly well-conditioned with respect to Λ of small size, our analysis provides also a prescribed linear rate of convergence of the iteration (12). The strategy of choosing slowly decreasing thresholding parameters is not at all new. Several other contributions, for example [29], proposed similar approaches. However, none of them quantifies rigorously the convergence improvement by relating the rate of convergence with the local conditioning of the matrix A^*A.

1.3 Adaptivity

In the case the action of A cannot be exactly operated, e.g., when A is the forward solution operator from a partial differential equation, the iteration (12) can not be exactly implementable. Instead, one will have to replace the exact applications of A and A^* by suitable numerical approximations. In particular, we assume that for any finitely supported vector $\mathbf{w} \in \ell_2(\mathcal{I})$, we are able to compute an approximation of $A^*A\mathbf{w}$. For this, we will rely on the existence of a numerical routine **APPLY** which, for given input data $\mathbf{w}$ and $\epsilon > 0$, outputs a finitely supported vector $\mathbf{w}_\epsilon$ such that $\|A^*A\mathbf{w} - \mathbf{w}_\epsilon\|_{\ell_2(\mathcal{I})} \leq \epsilon$. We refer, e.g., to [11, 46], for concrete examples of such a procedure, when the matrix A^*A can be assumed compressible [17, §9.5], see §4. Moreover, we assume that there exists a routine **RHS** which, for given input data ε, outputs a finitely supported vector $\mathbf{r}_\epsilon$ such that $\|A^*y - \mathbf{r}_\epsilon\|_{\ell_2(\mathcal{I})} \leq \epsilon$.

Instead of (12), we will consider then the numerically implementable algorithm

$$\tilde{\mathbf{u}}^{(0)} = \mathbf{0}, \quad \tilde{\mathbf{u}}^{(n+1)} = \mathbb{S}_{\alpha^{(n)}}\big(\tilde{\mathbf{u}}^{(n)} + \mathbf{RHS}[A^*y, \delta_n] - \mathbf{APPLY}[A^*A, \tilde{\mathbf{u}}^{(n)}, \gamma_n]\big), \quad n = 0, 1, \ldots, \tag{13}$$

with certain vanishing tolerances δ_n and γ_n. Such an *adaptive* scheme has been first proposed in the context of sparse optimization and inverse problems in the paper [31],

"borrowing a leaf" from the analysis of adaptive schemes for well-posed problems [11,12, 15,16,46]; in the present work we would like to greatly extend those preliminary results and make them more rigorous. More recent contributions, e.g., [2,41], catched also this idea and proposed similar approaches.

Our final contribution of this paper is the proof of linear convergence of (13) when the matrix A^*A has restrictions $A^*A|_{\Lambda\times\Lambda}$ which are uniformly well-conditioned with respect to Λ of small size.

1.4 Related work and innovations

There exist by now several iterative methods that can be used for the minimization problem (3) in *finite dimensions*. We shall account a few of the most recently analyzed and discussed:

(a) the *GPSR-algorithm* (gradient projection for sparse reconstruction), another iterative projection method, in the auxiliary variables $\mathbf{x}, \mathbf{y} \geq 0$ with $\mathbf{u} = \mathbf{x} - \mathbf{y}$ [28].

(b) the $\ell_1 - \ell_s$ *algorithm*, an interior point method using preconditioned conjugate gradient substeps (this method solves a linear system in each outer iteration step) [35].

(c) *FISTA* (fast iterative soft-thresholding algorithm) is a variation of the iterative soft-thresholding [1]. Define the operator $\Gamma(\mathbf{u}) = \mathbb{S}_\alpha(\mathbf{u} + A^*(y - A\mathbf{u}))$. The FISTA is defined as the iteration, starting for $\mathbf{u}^{(0)} = 0$,

$$\mathbf{u}^{(n+1)} = \Gamma\left(\mathbf{u}^{(n)} + \frac{t^{(n)} - 1}{t^{(n+1)}}\left(\mathbf{u}^{(n)} - \mathbf{u}^{(n-1)}\right)\right),$$

where $t^{(n+1)} = \frac{1+\sqrt{1+4(t^{(n)})^2}}{2}$ and $t^{(0)} = 1$.

Besides iterative solvers, one disposes also of direct methods where the solution can be in principle found in a finite number of steps, for instance by using the *homotopy-LARS* method [26]. As is addressed in the recent paper [38] which accounts a very detailed comparison of these different algorithms, they do perform quite well when the regularization parameter α is sufficiently large, with a small advantage for GPSR. When α gets quite small all the algorithms, except for FISTA, deteriorate significantly their performances. Moreover, local conditioning properties of the linear operator A seem particularly affecting the performances of iterative algorithms.

While these methods are particularly suited for finite dimensional problems, we reiterate that our goal in the present paper is to produce an effective strategy, for any range of the parameter α, for a large class of infinite dimensional problems, where we stress two fundamental ingredients:

- *multiscale preconditioning* allows for a local well-conditioning of the matrix A and therefore reproduces at infinite dimension the conditions of best performances for iterative algorithms;

- *adaptivity* combined with a *decreasing thresholding strategy* allow for a *controlled* inflation of the support size of the iterations, promoting the minimal computational cost in terms of number of algebraic equations, as well as the exploitation from the very beginning of the iteration of the local well-conditioning of the matrix A.

In [41] the authors propose an adaptive method similar to (13) where, instead of the soft-thresholding, a *coarsening function*, i.e., a compressed hard-thresholding procedure, is implemented. The emphasis in the latter contribution is on the regularization properties of such an adaptive method which does not dispose of a reference energy functional (1) as it is for (13). We stress that our analysis is mainly addressed to a careful presentation of the relationship among multiscale preconditioning, adaptivity, and rate of convergence to minimizers of (1). Moreover, our analysis cannot be straightforwardly generalized to the algorithm proposed in (13) because we make heavily use of the non-expansiveness of the soft-thresholding, whereas hard-thresholding operations are expansive.

1.5 Outline of the paper

In Section 2 we collect a few technical results on sparse and compressible vectors as well as on properties of the soft-thresholding operator. Section 3 is dedicated to the analysis of the decreasing iterative soft-thresholding algorithm when both exact and adaptive matrix-vector multiplications are involved and local well-conditioning of the matrix A^*A is assumed. In Section 3 we address the issue of multiscale preconditioning and we show that for a large class of integral operators it is possible to choose a multiscale basis for which the resulting discretization matrices can be locally preconditioned by simple block-diagonal preconditioners. In Section 5 we clarify how preconditioning can be used in ill-posed problems and how can we take advantage of the sparsity of minimal solutions and of adaptive matrix-vector multiplications in order to avoid topological issues. Since our results on the efficiency of the proposed strategy depends on the sparsity of the minimal solution, in Section 6 we quantify how the support size of minimal solutions may increase as the thresholding parameter α is made smaller and smaller. We conclude the paper with Section 7 where we show numerical results which demonstrate and confirm the theoretical achievements of the previous sections.

2 Technical Lemmas

We are particularly interested in computing approximations with the smallest possible number of nonzero entries. As a benchmark, we recall that the most economical approximations of a given vector $\mathbf{v} \in \ell_2(\mathcal{I})$ are provided by the *best N-term approximations* $\mathbf{v}_N$, defined by discarding in $\mathbf{v}$ all but the $N \in \mathbb{N}_0$ largest coefficients in absolute value. The error of best N-term approximation is defined as

$$\sigma_N(\mathbf{v}) := \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\mathcal{I})}. \tag{14}$$

The subspace of all ℓ_2 vectors with best N-term approximation rate $s > 0$, i.e., $\sigma_N(\mathbf{v}) \lesssim N^{-s}$ for some decay rate $s > 0$, is commonly referred to as the *weak ℓ_τ space*

$\ell_\tau^w(\mathcal{I})$, for $\tau = (s+\frac{1}{2})^{-1}$, which, endowed with

$$|\mathbf{v}|_{\ell_\tau^w(\mathcal{I})} := \sup_{N\in\mathbb{N}_0} (N+1)^s \sigma_N(\mathbf{v}), \tag{15}$$

becomes the quasi-Banach space $(\ell_\tau^w(\mathcal{I}), |\cdot|_{\ell_\tau^w(\mathcal{I})})$. Moreover, for any $0 < \epsilon \leq 2-\tau$, we have the continuous embedding $\ell_\tau(\mathcal{I}) \hookrightarrow \ell_\tau^w(\mathcal{I}) \hookrightarrow \ell_{\tau+\epsilon}(\mathcal{I})$, justifying why $\ell_\tau^w(\mathcal{I})$ is called weak $\ell_\tau(\mathcal{I})$.

When it comes to the concrete computations of good approximations with a small number of active coefficients, one frequently utilizes certain thresholding procedures. Here small entries of a given vector are simply discarded, whereas the large entries may be slightly modified. In this paper, we shall make use of *soft-thresholding* that we already introduced in (7). It is well-known, see [22], that $\mathbb{S}_\alpha$ is non-expansive for any $\alpha \in \mathbb{R}_+^{\mathcal{I}}$,

$$\|\mathbb{S}_\alpha(\mathbf{v}) - \mathbb{S}_\alpha(\mathbf{w})\|_{\ell_2(\mathcal{I})} \leq \|\mathbf{v}-\mathbf{w}\|_{\ell_2(\mathcal{I})}, \quad \text{for all } \mathbf{v},\mathbf{w} \in \ell_2(\mathcal{I}). \tag{16}$$

Moreover, for any fixed $x \in \mathbb{R}$, the mapping $\tau \mapsto S_\tau(x)$ is Lipschitz continuous with

$$|S_\tau(x) - S_{\tau'}(x)| \leq |\tau - \tau'|, \quad \text{for all } \tau,\tau' \geq 0. \tag{17}$$

We readily infer the following technical estimate.

Lemma 2.1. *Assume $\mathbf{v} \in \ell_2(\mathcal{I})$, $\alpha,\beta \in \mathbb{R}_+^{\mathcal{I}}$ such that $\bar{\alpha} = \inf_\lambda \alpha_\lambda = \inf_\lambda \beta_\lambda = \bar{\beta} > 0$, and define $\Lambda_{\bar{\alpha}}(\mathbf{v}) := \{\lambda \in \mathcal{I} : |v_\lambda| > \bar{\alpha}\}$. Then*

$$\|\mathbb{S}_\alpha(\mathbf{v}) - \mathbb{S}_\beta(\mathbf{v})\|_{\ell_2(\mathcal{I})} \leq \Big(\#\Lambda_{\bar{\alpha}}(\mathbf{v})\Big)^{1/2} \max_{\lambda\in\Lambda_{\bar{\alpha}}(\mathbf{v})} |\alpha_\lambda - \beta_\lambda|. \tag{18}$$

Proof. By (17) we have the estimate

$$\begin{aligned}
\|\mathbb{S}_\alpha(\mathbf{v}) - \mathbb{S}_\beta(\mathbf{v})\|_{\ell_2(\mathcal{I})} &= \left(\sum_{\lambda\in\mathcal{I}} |S_{\alpha_\lambda}(v_\lambda) - S_{\beta_\lambda}(v_\lambda)|^2\right)^{1/2} \\
&= \left(\sum_{\lambda\in\{\mu\in\mathcal{I}:|v_\mu|>\min\{\alpha_\mu,\beta_\mu\}\}} |S_{\alpha_\lambda}(v_\lambda) - S_{\beta_\lambda}(v_\lambda)|^2\right)^{1/2} \\
&= \left(\sum_{\lambda\in\Lambda_{\bar{\alpha}}(\mathbf{v})} |S_{\alpha_\lambda}(v_\lambda) - S_{\beta_\lambda}(v_\lambda)|^2\right)^{1/2} \\
&\leq \Big(\#\Lambda_{\bar{\alpha}}(\mathbf{v})\Big)^{1/2} \max_{\lambda\in\Lambda_{\bar{\alpha}}(\mathbf{v})} |\alpha_\lambda - \beta_\lambda|
\end{aligned}$$

□

Let $\mathbf{v} \in \ell_\tau^w(\mathcal{I})$, it is well-known [24, §7]

$$\#\Lambda_{\bar{\alpha}}(\mathbf{v}) \leq C|\mathbf{v}|^\tau_{\ell_\tau^w(\mathcal{I})}\bar{\alpha}^{-\tau}, \tag{19}$$

and, for $\alpha_\lambda = \bar{\alpha}$ for all $\lambda \in \mathcal{I}$, we have

$$\|\mathbf{v} - \mathbb{S}_\alpha(\mathbf{v})\|_{\ell_2(\mathcal{I})} \leq C|\mathbf{v}|^{\tau/2}_{\ell^w_\tau(\mathcal{I})}\bar{\alpha}^{1-\tau/2}, \tag{20}$$

where the constants are given by $C = C(\tau) > 0$. Let $\mathbf{v} \in \ell_0(\mathcal{I}) := \cap_{\tau>0}\ell^w_\tau(\mathcal{I})$, and $|\mathbf{v}|_{\ell_0} := \#\operatorname{supp}(\mathbf{v}) < \infty$. Then we have the straightforward estimate

$$\#\Lambda_{\bar{\alpha}}(\mathbf{v}) \leq |\mathbf{v}|_{\ell_0} \tag{21}$$

and, for $\alpha_\lambda = \bar{\alpha}$ for all $\lambda \in \mathcal{I}$, we have

$$\|\mathbf{v} - \mathbb{S}_\alpha(\mathbf{v})\|_{\ell_2(\mathcal{I})} \leq |\mathbf{v}|^{1/2}_{\ell_0}\bar{\alpha}, \tag{22}$$

which is easily shown by a direct computation. In the sequel, we shall also use the following support size estimate.

Lemma 2.2 ([11, Lemma 5.1, formula (5.5)]). *Let $\mathbf{v} \in \ell^w_\tau(\mathcal{I})$ and $\mathbf{w} \in \ell_2(\mathcal{I})$ with $\|\mathbf{v} - \mathbf{w}\|_{\ell_2(\mathcal{I})} \leq \epsilon$. Assume $\alpha = (\alpha_\lambda)_{\lambda\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}_+$ and $\inf_\lambda \alpha_\lambda = \bar{\alpha} > 0$. Then it holds*

$$\#\operatorname{supp}\mathbb{S}_\alpha(\mathbf{w}) \leq \#\Lambda_{\bar{\alpha}}(\mathbf{w}) \leq \frac{4\epsilon^2}{\bar{\alpha}^2} + 4C|\mathbf{v}|^\tau_{\ell^w_\tau(\mathcal{I})}\bar{\alpha}^{-\tau}, \tag{23}$$

where $C = C(\tau) > 0$. In particular if $\mathbf{v} \in \ell_0(\mathcal{I})$ then the estimate is refined

$$\#\operatorname{supp}\mathbb{S}_\alpha(\mathbf{w}) \leq \#\Lambda_{\bar{\alpha}}(\mathbf{w}) \leq \frac{4\epsilon^2}{\bar{\alpha}^2} + |\mathbf{v}|_{\ell_0(\mathcal{I})}. \tag{24}$$

Proof. For the reader convenience we include the proof of this simple lemma. We consider two sets $\mathcal{I}_1 = \{\lambda \in \mathcal{I} : |\mathbf{w}_\lambda| \geq \bar{\alpha}, \text{ and } |\mathbf{v}_\lambda| > \bar{\alpha}/2\}$, and $\mathcal{I}_2 = \{\lambda \in \mathcal{I} : |\mathbf{w}_\lambda| \geq \bar{\alpha}, \text{ and } |\mathbf{v}_\lambda| \leq \bar{\alpha}/2\}$. Then from (19)

$$\#\mathcal{I}_1 \leq \#\{\lambda \in \mathcal{I} : |\mathbf{v}_\lambda| > \bar{\alpha}/2\} \leq 2^\tau C|\mathbf{v}|^\tau_{\ell^w_\tau(\mathcal{I})}\bar{\alpha}^{-\tau} \leq 4C|\mathbf{v}|^\tau_{\ell^w_\tau(\mathcal{I})}\bar{\alpha}^{-\tau},$$

and

$$(\bar{\alpha}/2)^2(\#\mathcal{I}_2) \leq \sum_{\lambda\in\mathcal{I}_2} |\mathbf{v}_\lambda - \mathbf{w}_\lambda|^2 \leq \varepsilon^2.$$

These estimates imply (23), and similarly one gets (24). □

3 Adaptive Iterative Soft-Thresholding

In this section we address first the analysis of the iteration (12) where exact matrix-vector multiplications involving A^*A are assumed to be realizable. In particular, we will introduce spectral conditions on such a matrix in order to ensure a linear rate of convergence of the algorithm to a minimizer $\mathbf{u}^*$. Then we address a particular example of the use of the modified iteration (12) for a *compressed sensing* problem, where A is a random matrix with Gaussian i.i.d. entries. Local well-conditioning of such matrices is

established now for long [6, 42]. They do provide a perfect benchmark for our algorithm which shows a dramatic improvement, in terms of speed of convergence, with respect to the thresholded Landweber iteration with fixed thresholding parameter α. In the second part of the section, we address the problem of the use of such an iteration when inexact applications of the matrix A^*A are necessary and we provide again a fine analysis of the rate of convergence.

3.1 A variant of iterative soft-thresholding

For the case of exact operator evaluations and threshold parameters $\alpha, \alpha^{(n)} \in \mathbb{R}^{\mathcal{I}}_+$, where $\alpha^{(n)} \geq \alpha$, i.e., $\alpha^{(n)}_\lambda \geq \alpha_\lambda$ for all $\lambda \in \Lambda$, and $\bar{\alpha} = \inf_{\lambda\in\mathcal{I}} \alpha_\lambda > 0$, we consider the iteration

$$\mathbf{u}^{(0)} = \mathbf{0}, \quad \mathbf{u}^{(n+1)} = \mathbb{S}_{\alpha^{(n)}}\big(\mathbf{u}^{(n)} + A^*(y - A\mathbf{u}^{(n)})\big), \quad n = 0, 1, \dots \tag{25}$$

which we call the *decreasing iterative soft-thresholding algorithm* (D-ISTA).

Theorem 3.1. *Let $\|A\| < \sqrt{2}$ and let $\bar{\mathbf{u}} := (I - A^*A)\mathbf{u}^* + A^*y \in \ell^w_\tau(\mathcal{I})$ for some $0 < \tau < 2$. Moreover, let $L = L(\alpha) := \frac{4\|\mathbf{u}^*\|^2_{\ell_2(\mathcal{I})}}{\bar{\alpha}^2} + 4C\|\bar{\mathbf{u}}\|^\tau_{\ell^w_\tau(\mathcal{I})}\bar{\alpha}^{-\tau}$, and assume that for $S^* := \operatorname{supp}\mathbf{u}^*$ and all finite subsets $\Lambda \subset \mathcal{I}$ with at most $\#\Lambda \leq 2L$ elements, the operator $(I - A^*A)|_{(S^*\cup\Lambda)\times(S^*\cup\Lambda)}$ is contractive on $\ell_2(S^*\cup\Lambda)$, i.e., $\|(I - A^*A)|_{S^*\cup\Lambda\times S^*\cup\Lambda}\mathbf{w}\|_{\ell_2(S^*\cup\Lambda)} \leq \gamma_0\|\mathbf{w}\|_{\ell_2(S^*\cup\Lambda)}$, for all $\mathbf{w} \in \ell_2(S^* \cup \Lambda)$, or*

$$\|(I - A^*A)|_{S^*\cup\Lambda\times S^*\cup\Lambda}\| \leq \gamma_0, \tag{26}$$

where $0 < \gamma_0 < 1$. Then, for any $\gamma_0 < \gamma < 1$, the iterates $\mathbf{u}^{(n)}$ from (25) fulfill $\#\operatorname{supp}\mathbf{u}^{(n)} \leq L$ and they converge to $\mathbf{u}^*$ at a linear rate

$$\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2(\mathcal{I})} \leq \gamma^n\|\mathbf{u}^*\|_{\ell_2(\mathcal{I})} =: \epsilon_n \tag{27}$$

whenever the $\alpha^{(n)}$ are chosen according to

$$\alpha_\lambda \leq \alpha^{(n)}_\lambda \leq \alpha_\lambda + (\gamma - \gamma_0)L^{-1/2}\epsilon_n, \text{ for all } \lambda \in \Lambda. \tag{28}$$

Proof. We develop the proof by induction. For the initial iterate, we have $\mathbf{u}^{(0)} = \mathbf{0}$, so that $\#\operatorname{supp}\mathbf{u}^{(0)} \leq L$ and (27) is trivially true. Assume as an induction hypothesis that $S^{(n)} := \operatorname{supp}(\mathbf{u}^{(n)})$ is such that $\#S^{(n)} \leq L$, and $\|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2(\mathcal{I})} \leq \epsilon_n$. Abbreviating $\mathbf{w}^{(n)} := \mathbf{u}^{(n)} + A^*(y - A\mathbf{u}^{(n)})$, by $\|A^*A\| \leq 2$ and the induction hypothesis, it follows that

$$\|\bar{\mathbf{u}} - \mathbf{w}^{(n)}\|_{\ell_2(\mathcal{I})} = \big\|(I - A^*A)(\mathbf{u}^* - \mathbf{u}^{(n)})\big\|_{\ell_2(\mathcal{I})} \leq \|\mathbf{u}^* - \mathbf{u}^{(n)}\|_{\ell_2(\mathcal{I})} \leq \epsilon_n. \tag{29}$$

Hence, using (23), we obtain the estimate

$$\#S^{(n+1)} = \#\operatorname{supp}\mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)}) \leq \Lambda_{\bar{\alpha}}(\mathbf{w}^{(n)}) \leq \frac{4\epsilon_n^2}{\bar{\alpha}^2} + 4C|\bar{\mathbf{u}}|^\tau_{\ell^w_\tau(\mathcal{I})}\bar{\alpha}^{-\tau} \leq L. \tag{30}$$

Since also $\#S^{(n)} \leq L$ by induction hypothesis, the set $\Lambda^{(n)} := S^{(n)} \cup S^{(n+1)}$ has at most $2L$ elements, so that, by assumption, $(I - A^*A)|_{S\cup\Lambda^{(n)}\times S\cup\Lambda^{(n)}}$ is contractive with contraction constant γ_0. Using the identities

$$\begin{aligned}\mathbf{u}^*_{S\cup\Lambda^{(n)}} &= \mathbb{S}_\alpha(\bar{\mathbf{u}}_{S\cup\Lambda^{(n)}}) \\ &= \mathbb{S}_\alpha(\mathbf{u}^*_{S\cup\Lambda^{(n)}} + A^*_{S\cup\Lambda^{(n)}}(y - A_{S\cup\Lambda^{(n)}}\mathbf{u}^*_{S\cup\Lambda^{(n)}})),\end{aligned}$$

and

$$\begin{aligned}\mathbf{u}^{(n+1)}_{S\cup\Lambda^{(n)}} &= \mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)}_{S\cup\Lambda^{(n)}}) \\ &= \mathbb{S}_{\alpha^{(n)}}(\mathbf{u}^{(n)}_{S\cup\Lambda^{(n)}} + A^*_{S\cup\Lambda^{(n)}}(y - A_{S\cup\Lambda^{(n)}}\mathbf{u}^{(n)}_{S\cup\Lambda^{(n)}})),\end{aligned}$$

it follows from (16), (18), (9), and $\alpha^{(n)} \geq \alpha$ that

$$\begin{aligned}&\|\mathbf{u}^* - \mathbf{u}^{(n+1)}\|_{\ell_2(\mathcal{I})} \\ &= \|(\mathbf{u}^* - \mathbf{u}^{(n+1)})_{S\cup\Lambda^{(n)}}\|_{\ell_2(S\cup\Lambda^{(n)})} \\ &= \big\|\mathbb{S}_\alpha(\bar{\mathbf{u}}_{S\cup\Lambda^{(n)}}) - \mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)}_{S\cup\Lambda^{(n)}})\big\|_{\ell_2(S\cup\Lambda^{(n)})} \\ &\leq \big\|\mathbb{S}_\alpha(\bar{\mathbf{u}}_{S\cup\Lambda^{(n)}}) - \mathbb{S}_\alpha(\mathbf{w}^{(n)}_{S\cup\Lambda^{(n)}})\big\|_{\ell_2(S\cup\Lambda^{(n)})} + \big\|\mathbb{S}_\alpha(\mathbf{w}^{(n)}_{S\cup\Lambda^{(n)}}) - \mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)}_{S\cup\Lambda^{(n)}})\big\|_{\ell_2(S\cup\Lambda^{(n)})} \\ &\leq \big\|(I - A^*A|_{S\cup\Lambda^{(n)}\times S\cup\Lambda^{(n)}})(\mathbf{u}^* - \mathbf{u}^{(n)})_{S\cup\Lambda^{(n)}}\big\|_{\ell_2(S\cup\Lambda^{(n)})} \\ &\qquad + \Big(\#\Lambda_{\bar\alpha}(\mathbf{w}^{(n)})\Big)^{1/2}\Big(\max_{\lambda\in\Lambda_{\bar\alpha}(\mathbf{w}^{(n)})}|\alpha_\lambda - \alpha^{(n)}_\lambda|\Big) \\ &\leq \gamma_0\epsilon_n + \Big(\#\Lambda_{\bar\alpha}(\mathbf{w}^{(n)})\Big)^{1/2}\Big(\max_{\lambda\in\Lambda_{\bar\alpha}(\mathbf{w}^{(n)})}|\alpha_\lambda - \alpha^{(n)}_\lambda|\Big).\end{aligned}$$

Using (30) we obtain $\|\mathbf{u} - \mathbf{u}^{(n+1)}\|_{\ell_2(\mathcal{I})} \leq \gamma_0\epsilon_n + \sqrt{L}\Big(\max_{\lambda\in\Lambda_{\bar\alpha}(\mathbf{w}^{(n)})}|\alpha^{(n)}_\lambda - \alpha_\lambda|\Big)$, and, since the $\alpha^{(n)}$ are chosen according to (28), the claim follows. □

3.2 Random matrices, Restricted Isometry Property, and compressed sensing

In this section we would like to show a comparison of the performances of the algorithm proposed in (25) and the ones of the iterative soft-thresholding algorithm in a problem of sparse vector recovery from finite linear measurement data. Let us describe the problem we would like to solve. Given a rectangular matrix $A = \{a_i \in \mathbb{R}^n : i = 1, ..., N\}$ for $n \ll N$ and a vector $y \in \mathbb{R}^n$ we would like to compute $\mathbf{u}^* \in \mathbb{R}^N$ with $\|\mathbf{u}^*\|_{\ell_0} := \#\operatorname{supp}(\mathbf{u}) \ll N$ such that

$$A\mathbf{u}^* = y,$$

where

$$\mathbf{u}^* = \arg\min_{A\mathbf{z}=y}\|\mathbf{z}\|_{\ell_0}. \tag{31}$$

It is known that this problem is in general NP-hard. However, stable recovery procedures with polynomial complexity have been recently proposed [4, 5, 7, 25] under special conditions on the matrix A. One of the crucial properties for the polynomial complexity solution of (31) is the so-called *Restricted Isometry Property* (RIP). We say that A has the RIP of order k if there is a $0 < \gamma_k < 1$ such that

$$(1-\gamma_k)\|\mathbf{u}_\Lambda\|_{\ell_2^N}^2 \leq \|A_\Lambda \mathbf{u}_\Lambda\|_{\ell_2^n}^2 \leq (1+\gamma_k)\|\mathbf{u}_\Lambda\|_{\ell_2^N}^2, \tag{32}$$

for all $\mathbf{u} \in \mathbb{R}^N$ and for all $\Lambda \subset \{1, \ldots, N\}$ with $\#\Lambda \leq k$. This property essentially requires that every set of columns with cardinality less than k is approximatively an orthonormal system. The relationship between RIP property and the stable recovery of solutions to (31) is provided by the following result.

Theorem 3.2 ([13])**.** *Let A be a matrix with the RIP of order $2k$ for $\gamma_{2k} \leq \sqrt{2}-1$. Then, for any $\mathbf{u} \in \mathbb{R}^N$, the decoder*

$$\Delta(y) := \arg \min_{A\mathbf{z}=y=A\mathbf{u}} \|\mathbf{z}\|_{\ell_1},$$

is such that

$$\|\mathbf{u} - \Delta(A\mathbf{u})\|_{\ell_1} \leq C\sigma_k(\mathbf{u})_{\ell_1},$$

where $\sigma_k(\mathbf{u})_X = \|\mathbf{u} - \mathbf{u}_k\|_X$, and $\mathbf{u}_k$ is the best k-term approximation of $\mathbf{u}$ in norm X. In particular, if $\#\operatorname{supp}\mathbf{u} \leq k$ then $\sigma_k(\mathbf{u})_{\ell_1} = 0$.

We shall now have a simple reformulation of the RIP in order to show how it does fit with the proposal of algorithm (25), see [6, Definition 1.2].

Proposition 3.3. *The following conditions are equivalent:*

(i) A has the RIP property of order k and constant γ_k;

(ii) the symmetric matrix $A^*A|_{\Lambda\times\Lambda}$ is positive definite with eigenvalues in $[1-\gamma_l, 1+\gamma_k]$, for all $\Lambda \subset \{1, \ldots, N\}$ with $\#\Lambda \leq k$;

(iii) the matrix $(I - A^*A)|_{\Lambda\times\Lambda}$ has spectral norm $\|(I - A^*A)|_{\Lambda\times\Lambda}\|$ bounded above by γ_k, for all $\Lambda \subset \{1, \ldots, N\}$ with $\#\Lambda \leq k$.

Note that the equivalent condition (iii) above is precisely the one requested in (26). The key question is, given a fixed n, N, how large can we take k and still have matrices with the RIP property of order k. It was shown in [6] as well as in [25] that certain families of random matrices will, with high probability, satisfy the RIP of order k with constant $\gamma_k \leq \gamma_0 < 1$, for a prescribed γ_0 independent of N, provided the upper bound $k \leq c_0 \frac{n}{\log(N/k)}$. In particular, the entries of A can be generated as realizations of i.i.d. Gaussian and Bernoulli variables with a proper normalization. See [42] for further insights.

In practice we will never be in the idealized situation of the problem (31) and noise will affect the data y. In this case, one may want to solve instead the following problem

$$\operatorname*{argmin}_{\mathbf{u}\in\mathbb{R}^N} \|\mathbf{u}\|_{\ell_1^N} \text{ subject to } \|A\mathbf{u}-y\|_{\ell_2^n} \le \epsilon, \text{ or}$$
$$\operatorname*{argmin}_{\mathbf{u}\in\mathbb{R}^N} \|A\mathbf{u}-y\|_{\ell_2^n}^2 + 2\alpha\|\mathbf{u}\|_{\ell_1^N}, \tag{33}$$

for a suitable $\alpha = \alpha(\epsilon)$ depending on the noise level ϵ. Note that problem (33) coincides precisely with the minimization of a functional of the type (3). Moreover, we have the following regularization result.

Theorem 3.4 (Theorem 1 [4])**.** *Let k be such that $\gamma_{3k} + 3\gamma_{4k} < 2$. Then for any vector $\mathbf{u}$ supported on Λ with $\#\Lambda \le k$, and any perturbation $e = A\mathbf{u} - y$ with $\|e\|_{\ell_2^n} \le \epsilon$, the solution $\mathbf{u}^*$ of (33) obeys*

$$\|\mathbf{u} - \mathbf{u}^*\|_{\ell_2^N} \le c_k\epsilon,$$

where the constant c_k may only depend on γ_{4k}.

With these results at hand, we are in the situation of estimating the relevant parameters in order to apply Theorem 3.1 and proceed to a numerical comparison of the algorithm D-ISTA in (25) and the iterative soft-thresholding ISTA. In Figure 3 we show the behavior of the algorithms in the computation of a sparse minimizer $\mathbf{u}^*$ for A being a 500×2500 matrix with i.i.d. Gaussian entries, $\alpha = 10^{-3}$, $\gamma_0 = 0.1$ and $\gamma = 0.95$. In Figure 4 we show the behavior of the algorithms in the same situation but for parameters $\alpha = 10^{-4}$, $\gamma_0 = 0.01$ and $\gamma = 0.998$. In both the cases, related to small values of α (we reiterate that a small range of α is the most crucial situation for the efficiency of iterative algorithms, see §1.4), ISTA tends to iterate initially on vectors with a large number of nonzero entries, while D-ISTA inflates slowly the support size of the iterations to eventually converge to the right support of $\mathbf{u}^*$. The iteration on an inflating support allows D-ISTA to take advantage of the local well-conditioning of the matrix A from the very beginning of the iterations. This effect results in a *controlled* linear rate of convergence which is much steeper than the one of ISTA. In particular in Figure 4 after 1500 D-ISTA has correctly detected the support of the minimizer $\mathbf{u}^*$ and reached already an accuracy of $10^{-0.5}$, whereas it is clear that the convergence of ISTA is simply dramatically slow. The following sections are devoted to extend these results to infinite dimensions, by employing proper preconditioning and adaptivity strategies.

3.3 Iterative Thresholding with Inexact Operator Evaluations

In the case of A being the forward solution operator, e.g., from a partial differential equation, iteration (25) will not be implementable as it is. Instead, one will have to replace the applications of A and A^* by suitable approximations. In order to control the resulting additional approximation errors, we assume that for any finitely supported vector $\mathbf{w} \in \ell_2(\mathcal{I})$, we are able to compute an approximation of $A^*A\mathbf{w}$. In particular, we will rely on the existence of a numerical routine **APPLY** which, for a given input data

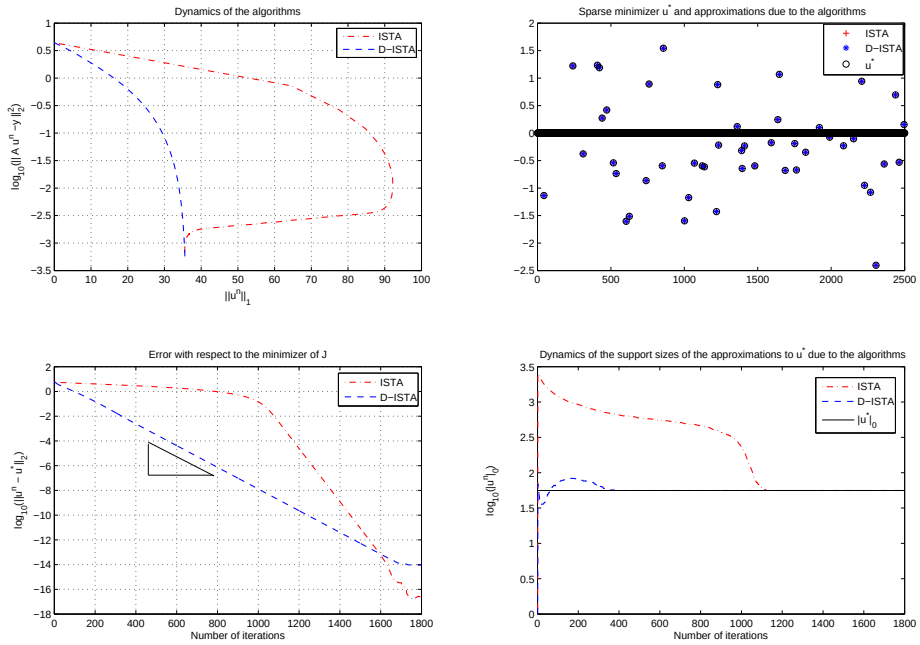


Figure 3: We show the behavior of the algorithms ISTA and D-ISTA in the computation of a sparse minimizer $\mathbf{u}^*$ for A being a 500×2500 matrix with i.i.d. Gaussian entries, $\alpha = 10^{-3}$, $\gamma_0 = 0.1$ and $\gamma = 0.95$. In the top left figure we presents the dynamics of the algorithms in the plane $\|\mathbf{u}\|_{\ell_1} - \log(\|A\mathbf{u} - y\|_2^2)$. On the bottom left, we show the absolute error to the precomputed minimizer $\mathbf{u}^*$ with respect to the number of iterations. On the bottom right we show how the size of the supports of the iterations grow with the number of iterations. The figure on the top right shows the vector $\mathbf{u}^*$, and the approximations due to the algorithms. In this case both the algorithms approximate with very high accuracy the minimizer $\mathbf{u}^*$ after 1800 iterations.

$\mathbf{w}$ and $\epsilon > 0$, outputs a finitely supported vector $\mathbf{w}_\epsilon$ such that $\|A^*A\mathbf{w} - \mathbf{w}_\epsilon\|_{\ell_2(\mathcal{I})} \leq \epsilon$. Moreover, we assume that there exists a routine **RHS** which, for given input data ϵ, outputs a finitely supported vector $\mathbf{r}_\epsilon$ such that $\|A^*y - \mathbf{r}_\epsilon\|_{\ell_2(\mathcal{I})} \leq \epsilon$.

Instead of (25), we consider the implementable algorithm

$$\tilde{\mathbf{u}}^{(0)} = \mathbf{0}, \quad \tilde{\mathbf{u}}^{(n+1)} = \mathbb{S}_{\alpha^{(n)}}\big(\tilde{\mathbf{u}}^{(n)} + \mathbf{RHS}[A^*y, \delta_n] - \mathbf{APPLY}[A^*A, \tilde{\mathbf{u}}^{(n)}, \gamma_n]\big), \quad n = 0, 1, \ldots, \tag{34}$$

with certain tolerances δ_n and γ_n. We call this iteration the *adaptive iterative soft-thresholding algorithm* (A-ISTA). As we shall see, the proof of Theorem 3.1 still works also in the case of inexact operator evaluations.

Proposition 3.5. *Let* $\|A\| < \sqrt{2}$ *and let* $\bar{\mathbf{u}} := (I - A^*A)\mathbf{u}^* + A^*y \in \ell^w_\tau(\mathcal{I})$ *for some* $0 < \tau < 2$. *Assume* $0 < \gamma_0 < \gamma < \tilde{\gamma} < 1$ *and* $\rho \geq \frac{1}{1-\tilde{\gamma}}$, *i.e.,* $\frac{\gamma}{\tilde{\gamma}} + \frac{1}{\rho} \leq 1$. *Moreover, let* $\tilde{L} =$

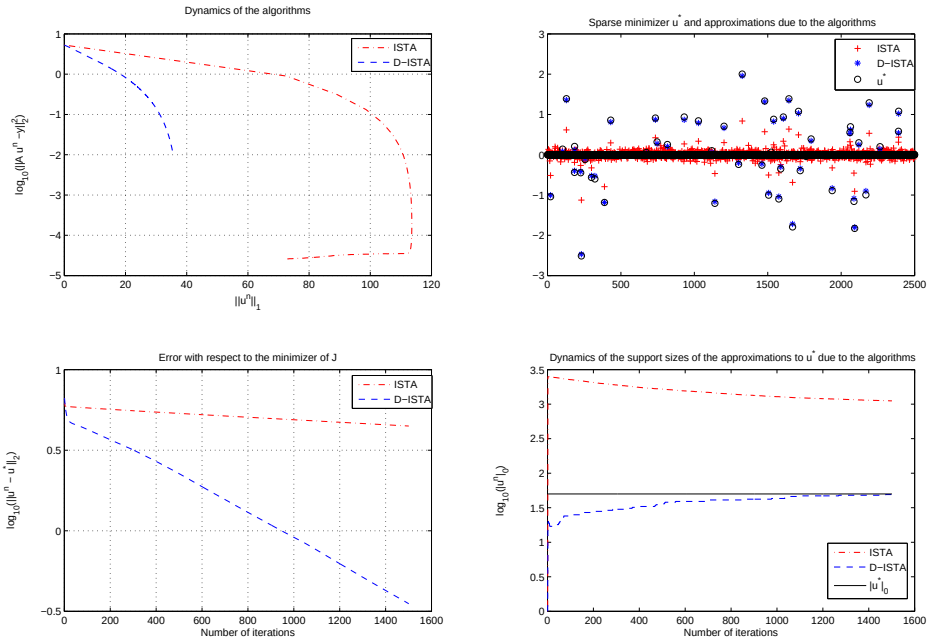


Figure 4: We show the behavior of the algorithms ISTA and D-ISTA in the computation of a sparse minimizer $\mathbf{u}^*$ for A being a 500×2500 matrix with i.i.d. Gaussian entries, $\alpha = 10^{-4}$, $\gamma_0 = 0.01$ and $\gamma = 0.998$. In the top left figure we presents the dynamics of the algorithms in the plane $\|\mathbf{u}\|_{\ell_1} - \log(\|A\mathbf{u} - y\|_2^2)$. On the bottom left, we show the absolute error to the precomputed minimizer $\mathbf{u}^*$ with respect to the number of iterations. On the bottom right we show how the size of the supports of the iterations grow with the number of iterations. The figure on the top right shows the vector $\mathbf{u}^*$, and the approximations due to the algorithms. In this case D-ISTA detects the right support of $\mathbf{u}^*$ after 1500 iterations, whereas ISTA keeps dramatically far behind.

$\tilde{L}(\alpha) := \frac{4(1+\tilde{\gamma}/\rho)\|\mathbf{u}^*\|^2_{\ell_2(\mathcal{I})}}{\bar{\alpha}^2} + 4C\|\bar{\mathbf{u}}\|^{\tau}_{\ell^w_\tau(\mathcal{I})}\bar{\alpha}^{-\tau}$, *and assume that for* $S^* := \operatorname{supp}\mathbf{u}^*$ *and all finite subsets* $\Lambda \subset \mathcal{I}$ *with at most* $\#\Lambda \leq 2\tilde{L}$ *elements, the operator* $(I - A^*A)|_{S^*\cup\Lambda\times S^*\cup\Lambda}$ *is contractive on* $\ell_2(S^*\cup\Lambda)$, *i.e.,* $\|(I - A^*A|)_{S^*\cup\Lambda\times S^*\cup\Lambda}\mathbf{w}\|_{\ell_2(S^*\cup\Lambda)} \leq \gamma_0\|\mathbf{w}\|_{\ell_2(S^*\cup\Lambda)}$, *for all* $\mathbf{w} \in \ell_2(S^*\cup\Lambda)$, *or*

$$\|(I - A^*A)|_{S^*\cup\Lambda\times S^*\cup\Lambda}\| \leq \gamma_0. \tag{35}$$

If $\delta_n = \gamma_n = \frac{\tilde{\epsilon}_{n+1}}{2\rho}$ *and* $\tilde{\epsilon}_n = \tilde{\gamma}^n\|\mathbf{u}^*\|_{\ell_2(\mathcal{I})}$, *then the iterates* (34) *fulfill*

$$\#\operatorname{supp}\tilde{\mathbf{u}}^{(n)} \leq \tilde{L} \text{ and } \|\mathbf{u}^* - \tilde{\mathbf{u}}^{(n)}\|_{\ell_2(\mathcal{I})} \leq \tilde{\epsilon}_n. \tag{36}$$

Proof. Since $\tilde{\mathbf{u}}^{(0)} = \mathbf{0}$ and $\tilde{\epsilon}_0 = \|\mathbf{u}^*\|_{\ell_2(\mathcal{I})}$, it suffices to discuss the n-th iteration step.

We abbreviate $\mathbf{w}^{(n)} := (I - A^*A)\tilde{\mathbf{u}}^{(n)} + A^*y$ and $\tilde{\mathbf{w}}^{(n)} := \tilde{\mathbf{u}}^{(n)} + \mathbf{RHS}[\frac{\tilde{\epsilon}_{n+1}}{2\rho}] - \mathbf{APPLY}[A^*A, \tilde{\mathbf{u}}^{(n)}, \frac{\tilde{\epsilon}_{n+1}}{2\rho}]$. Using $\|A^*A\| \leq 2$ and the induction hypothesis, it follows

that

$$\|\bar{\mathbf{u}} - \tilde{\mathbf{w}}^{(n)}\|_{\ell_2(\mathcal{I})} \leq \left\|(I - A^*A)(\mathbf{u}^* - \tilde{\mathbf{u}}^{(n)})\right\|_{\ell_2(\mathcal{I})} + \tfrac{\tilde{\epsilon}_{n+1}}{\rho} \leq (1 + \tfrac{\tilde{\gamma}}{\rho})\tilde{\epsilon}_n, \tag{37}$$

and

$$\|\bar{\mathbf{u}} - \mathbf{w}^{(n)}\|_{\ell_2(\mathcal{I})} \leq \left\|(I - A^*A)(\mathbf{u}^* - \tilde{\mathbf{u}}^{(n)})\right\|_{\ell_2(\mathcal{I})} \leq \tilde{\epsilon}_n. \tag{38}$$

By an application of (23) we obtain

$$\#\operatorname{supp}(\tilde{\mathbf{u}}^{(n)}) = \#\operatorname{supp}\mathbb{S}_{\alpha^{(n)}}(\tilde{\mathbf{w}}^{(n-1)}) \leq \frac{4\left[(1 + \tilde{\gamma}/\rho)\tilde{\epsilon}_{n-1}\right]^2}{\bar{\alpha}^2} + 4C|\bar{\mathbf{u}}|^{\tau}_{\ell^w_\tau}\bar{\alpha}^{-\tau}.$$

Note that

$$\begin{aligned}(1 + \tilde{\gamma}/\rho)\tilde{\epsilon}_{n-1} &= (1 + \tilde{\gamma}/\rho)\tilde{\gamma}^{n-1}\|\mathbf{u}^*\|_{\ell_2(\mathcal{I})} \\ &\leq (1 + \tilde{\gamma}/\rho)\|\mathbf{u}^*\|_{\ell_2(\mathcal{I})},\end{aligned}$$

Hence $\#\operatorname{supp}(\tilde{\mathbf{u}}^{(n)}) \leq \tilde{L}$, and similarly we obtain also $\#\operatorname{supp}\mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)}) \leq L \leq \tilde{L}$. We can restrict the iterations onto small sets of entries and use (35) to obtain

$$\left\|\mathbf{u}^* - \mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)})\right\|_{\ell_2(\mathcal{I})} \leq \gamma\tilde{\epsilon}_n.$$

So the claim follows from

$$\begin{aligned}\|\mathbf{u}^* - \tilde{\mathbf{u}}^{(n+1)}\|_{\ell_2(\mathcal{I})} &\leq \|\mathbf{u}^* - \mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)})\|_{\ell_2(\mathcal{I})} + \|\mathbb{S}_{\alpha^{(n)}}(\mathbf{w}^{(n)}) - \mathbb{S}_{\alpha^{(n)}}(\tilde{\mathbf{w}}^{(n)})\|_{\ell_2(\mathcal{I})} \\ &\leq \gamma\tilde{\epsilon}_n + \tfrac{\tilde{\epsilon}_{n+1}}{\rho} \\ &= (\tfrac{\gamma}{\tilde{\gamma}} + \tfrac{1}{\rho})\tilde{\epsilon}_{n+1} \\ &\leq \tilde{\epsilon}_{n+1}.\end{aligned}$$

□

4 A Multilevel Preconditioning

We discussed in §3.2 that finite dimensional matrices with an optimal *restricted isometry property* exist. Unfortunately, for compact operators in Hilbert spaces, the uniform contractivity assumption fails to hold in general. Indeed anytime we represent such operators with respect to a quasi-diagonalizing basis (which is the one we would prefer in order to maximally sparsify the matrix), then the diagonal entries of the resulting matrix will decay according to the behavior of the spectrum of the operator. As a remedy, we suggest a rescaling strategy which essentially consists of a block-diagonal preconditioning of certain classes of such operators, which include several concrete and interesting cases (see §4.2). Note that in general a block-diagonal preconditioning is indeed usually quite effective when the matrix representation of an operator with respect to a given basis is diagonal dominant, with diagonal entries decreasing to zero.

4.1 Block-diagonal preconditioning implies a Restricted Isometry Property for infinite matrices

Throughout this section, we have to be more specific concerning the operator K and the generating system $\Psi = \{\psi_\lambda\}_{\lambda\in\mathcal{I}}$. Typically, we shall be concerned with the following situation. Let us assume that the Hilbert space X and its dual X', together with $L_2(\Omega)$, $\Omega \subset \mathbb{R}^d$, forms a Gelfand triple

$$X \subset L_2(\Omega) \subset X' \tag{39}$$

and that the operator K is a bounded linear operator from X' to $L_2(\Omega)$. Then, the operator K^*K is a well-defined, bounded operator from X' to X. Moreover, we assume that the generating system is given by a *compactly supported* basis or a frame of *wavelet type* [10, §2.12] for $L_2(\Omega)$. In this case, the index $\lambda = (j, k, e)$ typically encodes several types of information, namely the *scale* $j \in \mathbb{Z}$, often denoted by $|\lambda|$, the *spatial location* $k \in \mathbb{Z}^d$, and the *type* of the wavelet indexed by e. We refer to [10,11] for more details on the notation. Within this setting, we shall make the following technical assumptions:

- The entries in the stiffness matrix of K^*K satisfy the following decay estimate

$$|\langle K^*K\psi_\lambda, \psi_\mu\rangle| \leq c_1 \frac{2^{-s||\lambda|-|\mu||}2^{-\sigma(|\lambda|+|\mu|)}}{\left(1 + 2^{\min(|\lambda|,|\mu|)}\mathrm{dist}(\Omega_\mu, \Omega_\lambda)\right)^r} \tag{40}$$

  where $c_1, s, \sigma, r \in \mathbb{R}_+, r > d$ and Ω_μ denotes the support of ψ_μ.

- For the diagonal entries, i.e., for $\mu = \lambda$, we require an additional estimate from below,

$$|\langle K^*K\psi_\lambda, \psi_\lambda\rangle| \geq c_2 2^{-2\sigma|\lambda|}. \tag{41}$$

- For the same scale, i.e., for $|\mu| = |\lambda|, \lambda = (|\lambda|, k, e), \mu = (|\lambda|, k', e')$, we assume that

$$|\langle K^*K\psi_\lambda, \psi_\mu\rangle| \leq \frac{c_3\, 2^{-2\sigma|\lambda|}}{(1 + |k - k'|)^r}. \tag{42}$$

Remark 4.1.

i.) At first sight, the conditions (40), (41) and (42) seem to be rather restrictive. However, in §4.2 we present some nontrivial examples for which condition (40) is indeed satisfied, whereas the others may be checked case by case.

ii.) For the special case $\sigma = 0$, the estimate (40) usually holds for zero order operators with Schwartz kernels. The parameters s and r depend on the smoothness of the wavelet basis, the mapping properties of the underlying operator and on the number of vanishing moments of the wavelet basis. Typically, increasing the smoothness and the number of vanishing moments produces larger values of r and s. We refer to [17] for details.

With the assumptions above, we can now prove the following theorem

Theorem 4.2. *Let $A^*A = F^*K^*KF = (\langle K^*K\psi_\lambda, \psi_\mu\rangle)_{\lambda,\mu\in\mathcal{I}}$ denote the stiffness matrix of K^*K. Let $D_j^b = (\langle K^*K\psi_\lambda, \psi_\mu\rangle)_{|\lambda|=|\mu|=j}$ denote the diagonal block of A^*A corresponding to the refinement level j, and let D^b denote the block diagonal matrix $D^b = (D_0^b, D_1^b, \ldots)$. Suppose that (40), (41), and (42) are satisfied with $c_2 > c_3/(r-d)$. Then there exists a constant $C = C(c_1, c_2, c_3, r, d)$ such that for each finite set $\Lambda \subset \mathcal{I}$ with $|\Lambda| < 2^s C^{-1}$ the sub-matrix $((D^b)^{-1/2}\, A^*A\, (D^b)^{-1/2})|_{\Lambda\times\Lambda}$ satisfies*

$$\|(I - (D^b)^{-1/2} A^*A (D^b)^{-1/2})|_{\Lambda\times\Lambda}\| < C 2^{-s}|\Lambda| \tag{43}$$

and

$$\mathcal{K}((D^b)^{-1/2}\, A^*A\, (D^b)^{-1/2}|_{\Lambda\times\Lambda}) \le \frac{1 + C\, 2^{-s}|\Lambda|}{1 - C\, 2^{-s}|\Lambda|}. \tag{44}$$

Remark 4.3. *i) Obviously, Theorem 4.2 implies that for increasing values of s (which can, e.g., be achieved by increasing the smoothness of the wavelet basis), larger index sets Λ can be used.*

ii) In Theorem 4.2, we have been purposely not very precise concerning the concrete value of C for ease of presentation. It depends on all the various parameters, and these relationships will become clear in the proof of Theorem 4.2.

iii) In the proof of Theorem 4.2, we will ignore the dependence on the type of the wavelets, i.e., on the parameter e, since this dependence only produces an additional constant.

iv) The use of a diagonal preconditioner works as well in practice, see § 7.1 and Figure 9. Unfortunately, Gerschgorin's theorem is too pessimistic in this case to allow us for an estimate of the type (43).

Proof of Theorem 4.2. The proof is essentially based on Gerschgorin's theorem, see, e.g., [32, Theorem 7.2.1]. The first step is to estimate the decay of the entries of $(D_j^b)^{-1/2}$. To this end, we will use the following theorem which has been proved by Jaffard [34], see [34, formula (7)] in particular.

Theorem 4.4. *Let $A = (a_{k,\ell})_{k,k'\in\mathbb{Z}^d}$ be a symmetric matrix satisfying $\|A\| < 1$ and*

$$|a_{k,k'}| \le \frac{c_4}{(1 + |k - k'|)^r}. \tag{45}$$

Then the entries of $B = (b_{k,k'})_{k,k'\in\mathbb{Z}^d}, B = A^{-1}$, satisfy

$$|b_{k,k'}| \le \frac{c_5\, \lambda_{\min}(A)^{-1}}{(1 + |k - k'|)^r}. \tag{46}$$

Moreover, the entries of $B^{1/2} = (b_{k,\ell}^{1/2})_{k,\ell\in\mathbb{Z}^d}$ satisfy

$$|b_{k,k'}^{1/2}| \le \frac{c_6\, \lambda_{\min}(A)^{-1/2}}{(1 + |k - k'|)^r}. \tag{47}$$

To apply Theorem 4.4, we need to estimate the minimal eigenvalue of $D_j^b = ((D_j^b)_{\lambda,\lambda'})_{|\lambda|=|\lambda'|=j} := (d^j_{k,k'})_{k,k'\in\mathbb{Z}^d}, \lambda = (j,k), \lambda' = (j,k')$. By Gerschgorin's theorem, we may conclude that

$$\lambda_{\min} \geq \min\Big\{ d^j_{k,k} - \sum_{k'\neq k} |d^j_{k,k'}| \Big\}. \tag{48}$$

By using (41) and (42) we obtain

$$\begin{aligned} d^j_{kk} - \sum_{k'\neq k} |d^j_{k,k'}| &\geq c_2\, 2^{-2j\sigma} - \sum_{k'\neq k} |d^j_{k,k'}| \\ &\geq c_2\, 2^{-2j\sigma} - \sum_{k'\neq k} \frac{c_3\, 2^{-2\sigma j}}{(1+|k-k'|)^r} \\ &\geq \Big(c_2 - \frac{c_3}{r-d}\Big) 2^{-2j\sigma} =: c_7\, 2^{-2j\sigma}. \end{aligned}$$

Consequently by (47), we get for $(D_j^b)^{-1/2} = (\tilde{d}^j_{k,k'})$ the estimate

$$|\tilde{d}^j_{k,k'}| \leq \frac{c_8\, 2^{j\sigma}}{(1+|k-k'|)^r}. \tag{49}$$

The next step is to estimate one entry of $(D^b)^{-1/2} A^* A (D^b)^{-1/2}$. For simplicity, let us assume that $|\mu| > |\lambda|$. We get

$$\begin{aligned} \big|((D^b)^{-1/2} A^* A (D^b)^{-1/2})_{\lambda,\mu}\big| &= \Big| \sum_{|\lambda'|=|\lambda|} (D^b)^{-1/2}_{\lambda,\lambda'} (A^* A\, (D^b)^{-1/2})_{\lambda',\mu} \Big| \\ &= \Big| \sum_{|\lambda'|=|\lambda|} (D^b)^{-1/2}_{\lambda,\lambda'} \sum_{|\mu'|=|\mu|} (A^*A)_{\lambda',\mu'} (D^b)^{-1/2}_{\mu',\mu} \Big| \\ &\leq \sum_{|\mu'|=|\mu|} \sum_{|\lambda'|=|\lambda|} |(D^b)^{-1/2}_{\lambda,\lambda'}||(A^*A)_{\lambda',\mu'}||(D^b)^{-1/2}_{\mu',\mu}|. \end{aligned}$$

Recall that for compactly supported wavelets $\mathrm{supp}(\psi_\lambda) \subset 2^{-|\lambda|}k + 2^{-|\lambda|}Q$, where Q is a suitable cube centered at the origin [10, §2.12]. Combining this observation with (49), (40) and the fact that $|\mu| > |\lambda|$, we obtain by setting $\lambda' = (|\lambda|, k'), \mu' = (|\mu|, l')$

$$\begin{aligned} \big|((D^b)^{-1/2}\, A^* A\, (D^b)^{-1/2})_{\lambda,\mu}\big| &\leq \sum_{|\mu'|=|\mu|} \sum_{k'} \frac{c_9\, 2^{|\lambda|\sigma}}{(1+|k-k'|)^r} \frac{2^{-s||\lambda|-|\mu||}\, 2^{-\sigma(|\lambda|+|\mu|)}}{(1+2^{|\lambda|}|2^{-|\lambda|}k' - 2^{-|\mu|}l'|)^r} (D^b)^{-1/2}_{\mu',\mu} \\ &\leq \sum_{|\mu'|=|\mu|} \sum_{k'} \frac{c_9\, 2^{-s||\lambda|-|\mu||}\, 2^{-\sigma|\mu|}}{(1+|k-k'|)^r (1+|k' - 2^{-|\mu|+|\lambda|}l'|)^r} \cdot (D^{b^{-1/2}}_{\mu',\mu}). \end{aligned}$$

Invoking the fact that the matrices satisfying (45) form an algebra, see again [34], and using (49) for another time yields

$$\begin{aligned}
\left|((D^b)^{-1/2}A^*A(D^b)^{-1/2})_{\lambda,\mu}\right| &\leq \sum_{l'} \frac{c_{10}\, 2^{-s||\lambda|-|\mu||}\, 2^{-\sigma|\mu|}}{(1+|k-2^{-|\mu|+|\lambda|}l'|)^r} \cdot \frac{c_8\, 2^{\sigma|\mu|}}{(1+|l'-l|)^r} \\
&\leq c_{10}c_8 2^{-s||\lambda|-|\mu||} \sum_{l'} \frac{1}{(1+|k-2^{-|\mu|+|\lambda|}l'|)^r(1+|l'-l|)^r} \qquad (50)\\
&\leq c_{10}c_8 2^{-s||\lambda|-|\mu||} \sum_{l'} \frac{1}{(1+|l'-l|)^r} \\
&\leq C2^{-s||\lambda|-|\mu||}
\end{aligned}$$

The case $|\lambda| < |\mu|$ can be treated similarly, hence

$$|((D^b)^{-1/2}A^*A\,(D^b)^{-1/2})_{\lambda,\mu}| \leq C\,2^{-s}, \quad |\lambda| \neq |\mu|. \qquad (51)$$

Now another application of Gerschgorin's theorem yields

$$\begin{aligned}
&\lambda_{\min}((D^b)^{-1/2}A^*A\,(D^b)^{-1/2}|_{\Lambda\times\Lambda}) \geq 1 - C\,2^{-s}|\Lambda| \\
&\text{and} \qquad (52)\\
&\lambda_{\max}((D^b)^{-1/2}A^*A\,(D^b)^{-1/2}|_{\Lambda\times\Lambda}) \leq 1 + C\,2^{-s}|\Lambda|.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\|(I-(D^b)^{-1/2}A^*A(D^b)^{-1/2})|_{\Lambda\times\Lambda}\| &= \lambda_{\max}((I-(D^b)^{-1/2}A^*A(D^b)^{-1/2})|_{\Lambda\times\Lambda}) \\
&= 1-\lambda_{\min}((I-(D^b)^{-1/2}A^*A(D^b)^{-1/2})|_{\Lambda\times\Lambda}) \\
&\leq C2^{-s}|\Lambda|
\end{aligned}$$

and

$$\mathcal{K}((D^b)^{-1/2}A^*A\,(D^b)^{-1/2}|_{\Lambda\times\Lambda}) = \frac{\lambda_{\max}((D^b)^{-1/2}A^*A\,(D^b)^{-1/2}|_{\Lambda\times\Lambda})}{\lambda_{\min}((D^b)^{-1/2}A^*A\,(D^b)^{-1/2}|_{\Lambda\times\Lambda})} \leq \frac{1+C\,2^{-s}|\Lambda|}{1-C\,2^{-s}|\Lambda|}.$$

□

Remark 4.5. *Note that from* (50) *we have also (for* $|\lambda| < |\mu|$*),*

$$\begin{aligned}
\left|((D^b)^{-1/2}A^*A(D^b)^{-1/2})_{\lambda,\mu}\right| &\leq C2^{-s||\lambda|-|\mu||} \sum_{l'} \frac{1}{(1+|k-2^{-|\mu|+|\lambda|}l'|)^r(1+|l'-l|)^r} \\
&\leq C2^{-s||\lambda|-|\mu||} \int_{\mathbb{R}^d} \frac{1}{(1+|k-2^{-|\mu|+|\lambda|}x|)^r(1+|x-l|)^r}dx \\
&\leq C2^{-s||\lambda|-|\mu||} \int_{\mathbb{R}^d} \frac{1}{(1+|k-2^{-|\mu|+|\lambda|}x|)^r(1+2^{-|\mu|+|\lambda|}|x-l|)^r}dx \\
&= C2^{-s||\lambda|-|\mu||}2^{d||\lambda|-|\mu||} \int_{\mathbb{R}^d} \frac{1}{(1+|k-\xi|)^r(1+|\xi-2^{-|\mu|+|\lambda|}l|)^r}d\xi \\
&\leq C\frac{2^{(d-s)||\lambda|-|\mu||}}{(1+2^{\min\{|\lambda|,|\mu|\}}|2^{-|\lambda|}k-2^{-|\mu|}l|)^r}.
\end{aligned}$$

This means that for $s > d$ the preconditioned matrix $((D^b)^{-1/2}A^*A(D^b)^{-1/2})$ is a zero-order operator, see the details in [17, formula (9.4.10)]. In particular this matrix is compressible in the sense described in [17, §9.5]. For such matrices, an efficient **APPLY** *routine, as required in §3.3, is provided, see [11, 46].*

Theorem 4.2 is very well suited for classical integral operators with Schwartz kernels such as potential operators, see §4.2.1 for details. However, we would like to treat also slightly different situations like the ones described in §4.2.2 which includes, for instance, the Biot-Savart operator. To this end, we have to slightly modify our assumptions as follows:

- The entries in the stiffness matrix of K^*K satisfy the following decay estimate

$$|\langle K^*K\psi_\lambda, \psi_\mu\rangle| \le c_1 \frac{2^{-s||\lambda|-|\mu||}2^{-\sigma\min(|\lambda|,|\mu|)}}{\left(1+2^{\min(|\lambda|,|\mu|)}\mathrm{dist}(\Omega_\mu,\Omega_\lambda)\right)^r} \tag{53}$$

  where $c_1, s, \sigma, r \in \mathbb{R}_+, r > d$.

- For $\mu = \lambda$, we require

$$|\langle K^*K\psi_\lambda, \psi_\lambda\rangle| \ge c_2 2^{-\sigma|\lambda|}. \tag{54}$$

- For $|\mu| = |\lambda|$, we assume that

$$|\langle K^*K\psi_\lambda, \psi_\mu\rangle| \le \frac{c_3\, 2^{-\sigma|\lambda|}}{(1+|k-k'|)^r}. \tag{55}$$

With these assumptions, we obtain the following theorem, whose proof can be performed by following the lines of the proof of Theorem 4.2. Therefore, we omit the details.

Theorem 4.6. *Let $A^*A = F^*K^*KF = (\langle K^*K\psi_\lambda, \psi_\mu\rangle)_{\lambda,\mu\in\mathcal{I}}$ denote the stiffness matrix of K^*K. Suppose that (53), (54), and (55) are satisfied with $c_2 > c_3/(r-d)$ and $s > \sigma/2$. Then there exists a constant $C = C(c_1, c_2, c_3, r, d)$ such that for each finite set $\Lambda \subset \mathcal{I}$ with $|\Lambda| < 2^s C^{-1}$ the sub-matrix $((D^b)^{-1/2}\, A^*A\, (D^b)^{-1/2})|_{\Lambda\times\Lambda}$ satisfies*

$$\|(I - (D^b)^{-1/2}A^*A(D^b)^{-1/2})|_{\Lambda\times\Lambda}\| < C2^{-(s-\sigma/2)}|\Lambda| \tag{56}$$

and

$$\mathcal{K}((D^b)^{-1/2}\, A^*A\, (D^b)^{-1/2}|_{\Lambda\times\Lambda}) \le \frac{1 + C\, 2^{-(s-\sigma/2)}|\Lambda|}{1 - C\, 2^{-(s-\sigma/2)}|\Lambda|}. \tag{57}$$

4.2 Operators which allow for an effective block-diagonal preconditioning

We consider now classes of operators with property (40) and (53), respectively. We shall mainly discuss two important families of examples.

4.2.1 Integral operators with Schwartz kernels

In this section, we are concerned with a large class of operators of negative order which allows for suitable preconditioning techniques. As we shall see, the classical layer potential operators such as the single and double layer potentials fall into this category. As usual, we are concerned with a Gelfand triple

$$X \subset L_2(\Omega) \subset X', \tag{58}$$

where here in particular X is a Sobolev space on a domain or a closed manifold Ω, i.e., $X = H^t(\Omega),\ X' = H^{-t}(\Omega)$. We consider a bounded operator $S\ :\ H^{-t}(\Omega) \longrightarrow H^t(\Omega)$ of the form

$$(Su)(x) = \int_\Omega \Phi(x,\xi)u(\xi)d\xi, \quad x \in \Omega, \tag{59}$$

where we assume that the kernel Φ is of Schwartz type, i.e,

$$|\partial_x^\alpha \partial_\xi^\beta \Phi(x,\xi)| \le c_{\alpha,\beta}|x-\xi|^{-(d+2t+|\alpha|+|\beta|)}, \quad t \in \mathbb{R}, \text{ and multi-indexes } \alpha,\beta \in \mathbb{N}^d. \tag{60}$$

We are interested in solving operator equations of the form $Su = f$. In practice, due to noisy data, it might happen that the right–hand side is not contained in $H^t(\Omega)$, but only in $L_2(\Omega)$. Therefore we consider the problem

$$Ku = f \in L_2(\Omega), \quad Ku(x) = \int_\Omega \Phi(x,\xi)u(\xi)d\xi, \quad x \in \Omega. \tag{61}$$

We changed purposely the symbol $S \to K$ of the operator in order to point out its different action, in particular, its image space. We want to discretize (61) by means of a tight wavelet frame $\{\psi_\lambda\}_{\lambda\in\mathcal{I}}$ on Ω. We do not enter into the details of these bases; we will just assume a few useful properties, which we describe in a simplified form for ease of presentation (valid for $\Omega = [0,1]^d$):

(i) as already recalled before, the index $\lambda = (|\lambda|, k, e)$ encodes several different properties, respectively, the scale $|\lambda|$, the spatial location $k \in \mathbb{R}^d$, and the wavelet label e (without loss of cogency in the following we ignore this latter parameter);

(ii) $\Omega_\lambda := \mathrm{supp}(\psi_\lambda)$, $|\Omega| \sim 2^{-|\lambda|}$; we can also assume for simplicity that $\Omega_\lambda \subset 2^{-|\lambda|}k + 2^{-|\lambda|}Q$, where Q is a suitable cube centered at the origin [10, §2.12];

(iii) $\int_\Omega \xi^\alpha\, \psi_\lambda(\xi)\, d\xi = 0,\ \alpha = 0,\dots,d^* \in \mathbb{N}$;

(iv) $\|\psi_\lambda\|_\infty \le C\, 2^{d/2|\lambda|}$.

Actually the latter two conditions implies the more general *cancellation property* [17, formula 9.3.4] [43]:

$$\left|\int_\Omega \phi(\xi)\psi_\lambda(\xi)d\xi\right| \le c_{0,\beta}2^{-|\lambda|(d/2+d^*+1)}\|\phi\|_{W^{\infty,d^*+1}(\Omega_\lambda)},$$

which we may want to assume directly for more general domains Ω. It can be shown that if (60) holds, then the decay of the entries $\langle K\psi_{\lambda'}, \psi_\lambda\rangle$ is typically governed by the following basic estimate

$$2^{(|\lambda'|+|\lambda|)t}|\langle K\psi_{\lambda'}, \psi_\lambda\rangle| \le C\frac{2^{-\eta||\lambda|-|\lambda'||)}}{\big(1+2^{\min(|\lambda|,|\mu|)}\mathrm{dist}(\Omega_{\lambda'}, \Omega_\lambda)\big)^{d+2m-2t}}, \tag{62}$$

where η depends on the mapping properties of S and the smoothness of the wavelet frame whereas m is related with the number of vanishing moments d^*, see, e.g. [19, 20, 43] for details. Equation (62) means that the matrix $(2^{(|\lambda'|+|\lambda|)t)}|\langle K\psi_{\lambda'}, \psi_\lambda\rangle|))_{\lambda,\lambda'\in\mathcal{I}}$ is contained in the *Lemarié algebra* $\mathcal{M}$ which is the class of all matrices $\mathbf{N} = (n_{\lambda,\lambda'})_{\lambda,\lambda'\in\mathcal{I}}$ such that

$$|n_{\lambda,\lambda'}| \le C\frac{2^{-\eta||\lambda|-|\lambda'||}}{\big(1+2^{\min(|\lambda|,|\mu|)}\mathrm{dist}(\Omega_{\lambda'}, \Omega_\lambda)\big)^{r}}, \quad \text{for all } \lambda, \lambda' \in \mathcal{I}, \tag{63}$$

for some constant C and suitable parameters $r > d$ and $\eta > d/2$. We refer to [36] for further information concerning this class. Due to the usual norm equivalences of wavelet bases or frames, the discretization of (61) leads to the following biinfinite matrix equation on ℓ_2:

$$(F \circ K \circ F^* \circ \mathbf{D}^{-1})\mathbf{u} = F(f). \tag{64}$$

where $\mathbf{D} = \mathrm{diag}(2^{-t|\lambda|}, \lambda \in \mathcal{I})$ and F denotes the analysis operator associated with the tight frame, i.e.,

$$F(v) := (\langle v, \psi_\lambda\rangle)_{\lambda\in\mathcal{I}}.$$

We refer to [15] for details. Now the fundamental estimate (62) implies that

$$\mathbf{D}^{-1}F \circ K \circ F^* \circ \mathbf{D}^{-1} = \mathbf{M} \in \mathcal{M}$$

and hence

$$F \circ K \circ F^* \circ \mathbf{D}^{-1} = \mathbf{DM}.$$

The Lemarie algebra is stable under taking adjoints, therefore

$$\mathbf{D}^{-1}F \circ K^* \circ F^*\mathbf{D}^{-1} = \mathbf{M}^* \in \mathcal{M}.$$

Since we are working with a tight frame, this yields:

$$\begin{aligned}\mathbf{D}^{-1}(\langle K^*K\psi_{\lambda'}, \psi_\lambda\rangle)_{\lambda,\lambda'\in\mathcal{I}}\mathbf{D}^{-1} &= \mathbf{D}^{-1} \circ F \circ K^*K \circ F^* \circ \mathbf{D}^{-1} \\ &= (\mathbf{D}^{-1} \circ F \circ K^* \circ F^*) \circ (F \circ K \circ F^* \circ \mathbf{D}^{-1}) \\ &= \mathbf{M}^*\mathbf{DDM}. \end{aligned} \tag{65}$$

Since $\mathbf{DD} = \mathrm{diag}(2^{2t|\lambda|}, \lambda \in \mathcal{I})$, we may estimate $|(\mathbf{M}^*\mathbf{DDM})_{\lambda,\lambda'}| \le (|\mathbf{M}^*\mathbf{M})_{\lambda,\lambda'}|$, being $\mathbf{M}^*, \mathbf{M} \in \mathcal{M}$, also the product $\mathbf{M}^*\mathbf{M} \in \mathcal{M}$, maybe with a slightly smaller parameters $r' < r$ and $\eta' < \eta$. Altogether, we arrive at

$$|\langle K^*K\psi_{\lambda'}\psi_\lambda\rangle| \le C'\frac{2^{-\eta'||\lambda|-|\lambda'|)}2^{-2t(|\lambda'|+|\lambda|)}}{\big(1+2^{\min(|\lambda|,|\mu|)}\mathrm{dist}(\Omega_{\lambda'}, \Omega_\lambda)\big)^{r'}}, \quad \text{for all } \lambda, \lambda' \in \mathcal{I},$$

i.e., (40) holds.

4.2.2 Integral operators with respect to disjoint domains

Throughout this section let us assume that we are given an operator K

$$Ku(x) = \int_{\Omega} \Phi(x,\xi)u(\xi)d\xi, \quad x \in \tilde{\Omega},$$

for $\tilde{\Omega}$, $\Omega \subset \mathbb{R}^d$, two relatively compact domains such that $\mathrm{dist}(\Omega, \Omega') = \delta > 0$, $u \in X := H^t(\Omega)$, and $\Phi : \tilde{\Omega} \times \Omega \to \mathbb{R}$. We will again assume that the kernel Φ is of Schwartz type, i.e., (60) holds.

Lemma 4.7. *The operator K^*K has again a smooth kernel $\Psi : \Omega \times \Omega \to \mathbb{R}$ such that*

$$|\partial_\eta^\alpha \partial_\xi^\beta \Psi(\eta,\xi)| \le c_{\alpha,\beta}|\eta-\xi|^{-(d+2t+\min\{|\alpha|,|\beta|\})}, \quad t \in \mathbb{R}, \text{ and multi-indexes } \alpha,\beta \in \mathbb{N}^d. \tag{66}$$

Proof. Let us express K^*K in integral form:

$$\begin{aligned} K^*Ku(\eta) &= \int_{\tilde{\Omega}} \Phi(x,\eta)\left(\int_{\Omega}\Phi(x,\xi)u(\xi)d\xi\right)dx \\ &= \int_{\Omega} \underbrace{\left(\int_{\tilde{\Omega}}\Phi(x,\eta)\Phi(x,\xi)dx\right)}_{:=\Psi(\eta,\xi)} u(\xi)d\xi. \end{aligned}$$

Hence, we have the derivative estimates

$$\begin{aligned} |\partial_\eta^\alpha\partial_\xi^\beta\Psi(\eta,\xi)| &\le \int_{\tilde{\Omega}} \left|\partial_\eta^\alpha\Phi(x,\eta)\partial_\xi^\beta\Phi(x,\xi)\right| dx \\ &\le \int_{\tilde{\Omega}} c_{0,\alpha}c_{0,\beta}|x-\eta|^{-(d+2t+|\alpha|)}|x-\xi|^{-(d+2t+|\beta|)}dx. \end{aligned}$$

Thus, by using the fact that $\frac{1}{|\eta-x||\xi-x|} \le C_{x,\delta}\frac{1}{|\eta-\xi|}$, for $|\eta-x| \ge \delta, |\xi-x| \ge \delta$, we immediately obtain (66). □

For the proper definition of the discrete matrix $A^*A := F^*K^*KF$, we will use again multiscale bases $\Psi = \{\psi_\lambda\}$ for $L_2(\Omega)$ as introduced in §4.2.1.

We develop Φ into power series around the point $\xi_0 \in \Omega_\lambda$:

$$\Phi(x,\xi) = \sum_{|\beta|\le d^*+1} c_\beta(x,\xi_0)(\xi-\xi_0)^\beta + R_{d^*+1}(x,\xi_0,\xi),$$

where

$$R_{d^*+1}(x,\xi_0,\xi) = \sum_{|\beta|=d^*+1} \frac{d^*+1}{\beta!}(\xi-\xi_0)^\beta \int_0^1 (1-t)^{d^*}\partial_\xi^\beta\Phi(x,\xi_0+t(\xi-\xi_0))dt,$$

the remainder. For $\xi \in \Omega_\lambda$, we have the estimates

$$\begin{aligned} |R_{d^*+1}(x,\xi_0,\xi)| &\leq \left(\sum_{|\beta|=d^*+1} \frac{d^*+1}{\beta!} |\xi-\xi_0|^{d^*+1} \sup_{\xi\in\Omega_\lambda} |\partial_\xi^\beta \Phi(x,\xi)| \right) \\ &\leq c_{0,\beta} \mathrm{dist}(x,\Omega_\lambda)^{-(d+2t+d^*+2)} \left(\sum_{|\beta|=d^*+1} \frac{d^*+1}{\beta!} |\xi-\xi_0|^{d^*+1} \right). \end{aligned}$$

By the latter estimates, we have

$$\begin{aligned} |K\psi_\lambda(x)| &= \left| \int_{\Omega_\lambda} \Phi(x,\xi)\psi_\lambda(\xi) d\xi \right| \\ &\leq c_{0,\beta} \mathrm{dist}(x,\Omega_\lambda)^{-(d+2t+d^*+2)} \int_{\Omega_\lambda} \sum_{|\beta|=d^*+1} \frac{d^*+1}{\beta!} |\xi-\xi_0|^{d^*+1} |\psi_\lambda(\xi)| d\xi. \end{aligned}$$

By the basic properties of wavelets

$$\begin{aligned} &\int_{\Omega_\lambda} |\xi-\xi_0|^{d^*+1} |\psi_\lambda(\xi)| d\xi = \int_{\mathbb{R}^d} |\xi-\xi_0|^{d^*+1} |\psi_\lambda(\xi)| d\xi \\ \leq\ & C \int_Q |2^{-|\lambda|}\eta - \xi_0|^{d^*+1} 2^{d/2|\lambda|} 2^{-d|\lambda|} d\eta \leq C 2^{-|\lambda|(d/2+d^*+1)}. \end{aligned}$$

Finally this implies

$$|K\psi_\lambda(x)| = \left| \int_\Omega \Phi(x,\xi)\psi_\lambda(\xi) d\xi \right| \leq C_{0,\beta} 2^{-|\lambda|(d/2+d^*+1)} \mathrm{dist}(x,\Omega_\lambda)^{-(d+2t+d^*+2)}.$$

Thus, by using the fact that $\frac{1}{|\eta-x||\xi-x|} \leq C_{x,\delta} \frac{1}{|\eta-\xi|}$, for $|\eta-x| \geq \delta, |\xi-x| \geq \delta$, we show immediately, for $\lambda \neq \mu$,

$$\begin{aligned} |(A^*A)_{\lambda,\mu}| &= |\int_{\Omega'} (K\psi_\lambda(x))(K\psi_\mu(x)) dx| \\ &\leq C'_{0,\beta} 2^{-(|\lambda|+|\mu|)(d/2+d^*+1)} \int_{\Omega'} \mathrm{dist}(x,\Omega_\lambda)^{-(d+2t+d^*+2)} \mathrm{dist}(x,\Omega_\mu)^{-(d+2t+d^*+2)} \\ &= C'_{0,\beta} \frac{2^{-||\lambda|-|\mu||(d/2+d^*+1)} 2^{\min\{|\lambda|,|\mu|\}(2t-d^*)}}{\left(2^{\min\{|\lambda|,|\mu|\}} \mathrm{dist}(\Omega_\lambda,\Omega_\mu)\right)^{d+2t+d^*+2}}, \end{aligned}$$

so that (53) holds.

Example 4.8. *In this example we deal with operators which appears in magnetic tomography. In such applications, one is interested to reconstruct a current distribution density in volume, provided a few external measurements of the generated magnetic field. The procedure usually does not furnish a full measurement of an external magnetic field, but only of its normal component at an external surface, e.g., as in magnetoencephalography, on a helmet out of the head (see [33]). In order to obtain a model easily tractable,*

but still sufficiently realistic, we assume that a vector valued current density $\mathbf{j}$ *is confined in a spherical volume* Ω*. If* $\mathbf{e}_r(x)$ *is the radially oriented unit vector at point* $x \notin \Omega$*, then* $B_r(x) := \mathbf{B}(x) \cdot \mathbf{e}_r(x)$ *is given by*

$$(K\mathbf{j})(x) := B_r(x, \mathbf{j}) \quad = \quad \frac{\mu_0}{4\pi} \int_\Omega \frac{\mathbf{j}(\xi) \times (x-\xi) \cdot \mathbf{e}_r(x)}{|x-\xi|^3}\, d\xi. \tag{67}$$

This formula relates the normal component of the magnetic field directly to the current density $\mathbf{j}$*. Thanks to well-known results in potential theory, the normal component of the magnetic field on an external surface uniquely determines the magnetic field out of the volume as soon as the current flux on the volume surface is known. In this case the measurements of* B_r *on an external surface are a sufficient information to determine the whole magnetic field.*

Since the Biot-Savart operator can be expressed as

$$\mathcal{B}(x, \mathbf{j}) \quad = \quad \frac{\mu_0}{4\pi} \int_\Omega \frac{\mathbf{j}(\xi) \times (x-\xi)}{|x-\xi|^3} d\xi = \int_\Omega [\nabla_x \Phi(x,\xi)] \times \mathbf{j}(\xi) d\xi,$$

where Φ *is the single layer potential kernel* $\Phi(x,\xi) = \frac{\mu_0}{4\pi} \frac{1}{|x-\xi|}$, $x \neq \xi$*, then* B_r *can be shown to have a kernel which fulfills the assumptions* (60)*, and therefore it has an associated compressible matrix* A^*A*. We refer to [31] for more details.*

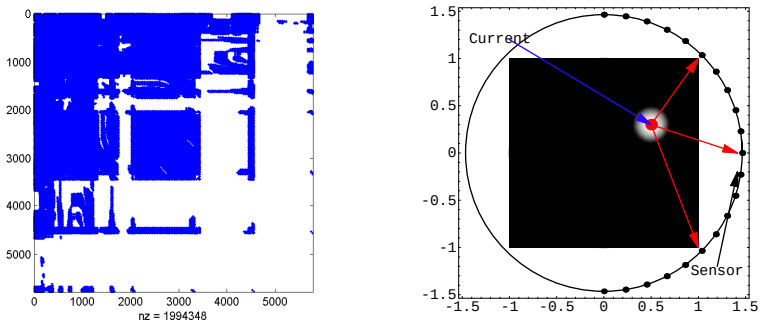


Figure 5: **Right**: Compressed matrix A^*A associated to the Biot-Savart operator in 2D, i.e., $\mathcal{B}(x,j) := \int_\Omega \frac{(x-\xi)^\perp}{|x-\xi|^2} j(\xi) d\xi$, $\Omega \subset \mathbb{R}^2$, with respect to Daubechies wavelets with $d^* = 4$ vanishing moments. We retain only the entries which exceed 10^{-6}. Only the 5.97% of the matrix is non-zero. **Left**: 2D model of a magnetic tomography. Sensors are distributed on a semicircle around the area where current densities are distributed.

5 Reduction to a Finite Dimensional Problem

5.1 Reformulation of the problem after preconditioning

In this section we would like to show how preconditioned matrices can be indeed employed in the adaptive iteration (34). We start first by reformulating the functional J

as follows:

$$\begin{aligned} J(\mathbf{u}) &= \|A\mathbf{u} - y\|_Y^2 + 2\alpha\|\mathbf{u}\|_{\ell_1(\mathcal{I})} \\ &= \|AD^{-1/2}\underbrace{D^{1/2}\mathbf{u}}_{:=\mathbf{z}} - y\|_Y^2 + 2\alpha\|D^{-1/2}\underbrace{D^{1/2}\mathbf{u}}_{:=\mathbf{z}}\|_{\ell_1(\mathcal{I})} \\ &= \|AD^{-1/2}\mathbf{z} - y\|_Y^2 + 2\alpha\|D^{-1/2}\mathbf{z}\|_{\ell_1(\mathcal{I})} := J^D(\mathbf{z}). \end{aligned}$$

Hence,

$$\operatorname*{argmin}_{\mathbf{u}\in\ell_2(\mathcal{I})} J(\mathbf{u}) = D^{-1/2}\left(\operatorname*{argmin}_{\mathbf{z}\in D^{1/2}\ell_2(\mathcal{I})} J^D(\mathbf{z})\right).$$

Here, we assume that $D^{1/2} : \ell_2(\mathcal{I}) \to D^{1/2}\ell_2(\mathcal{I})$ is a suitable preconditioning matrix which has a well-defined formal inverse $D^{-1/2}$ on its image $D^{1/2}\ell_2(\mathcal{I})$. Moreover, we also assume that this matrix $D^{1/2}$ is symmetric, block-diagonal, diagonal dominant, in the sense that $\|D^{-1/2}\mathbf{z}\|_{\ell_1(\mathcal{I})} \sim \|\operatorname{diag}(D^{-1/2})\mathbf{z}\|_{\ell_1(\mathcal{I})}$ for all $\mathbf{z} \in \ell_0(\mathcal{I})$, and with positive diagonal elements decreasing to zero, so that the diagonal entries of its inverse $D^{-1/2}$ do increase to infinity. Moreover, since the treatment of the term $\|D^{-1/2}\mathbf{z}\|_{\ell_1(\mathcal{I})}$ can be difficult in practice, for the sake of simplicity we may want to exploit the approximation $\|D^{-1/2}\mathbf{z}\|_{\ell_1(\mathcal{I})} \sim \|\operatorname{diag}(D^{-1/2})\mathbf{z}\|_{\ell_1(\mathcal{I})}$, and, with a slight abuse, we re-define

$$J^D(\mathbf{z}) := \|AD^{-1/2}\mathbf{z} - y\|_Y^2 + 2\|\mathbf{z}\|_{\ell_{1,\alpha\operatorname{diag}(D^{-1/2})}(\mathcal{I})},$$

which is again in the form (3). The use of a diagonal preconditioner is also justified by the numerical evidences we reported n §7.1 and Figure 9.

We provided examples of such matrices $D^{1/2}$ in §4, and we have also observed in Remark 4.5 that, when used, they produce preconditioned matrices $D^{-1/2}A^*AD^{-1/2}$ which represents zero-order operators. In particular, in these cases, $D^{-1/2}A^*AD^{-1/2}$ is a bounded operator on $\ell_2(\mathcal{I})$, a so we will consider it throughout this section. However, in our discussion we will not be too concerned with the topology where the minimization $\operatorname{argmin}_{\mathbf{z}} J^D(\mathbf{z})$ should take place, because we will clarify below that, due to the thresholding action, this apparently *infinite* dimensional problem is actually a *finite* dimensional one. Hence no topological issue have to be taken into account.

5.2 Equivalence of the infinite dimensional problem to a finite dimensional one

Let us consider $\mathbf{z}^* = \operatorname{argmin}_{\mathbf{z}} J^D(\mathbf{z}) = D^{1/2}\mathbf{u}^*$. Since $\mathbf{u}^*$ is a finitely supported vector and $D^{1/2}$ is a block-diagonal matrix, then also $\mathbf{z}^*$ is a finitely supported vector. Let $\Lambda = \operatorname{supp}(\mathbf{z}^*)$ and $\Lambda^\circ \subset \mathcal{I}$ any finite subset of indexes sufficiently large, and $\Lambda \subset \Lambda^\circ$. We want to emphasize that we know the existence of the set Λ°, but we shall not define it precisely a priori.
It is clear that the finite vector $\mathbf{z}^*|_{\Lambda^\circ}$ is also a minimizer of the functional

$$J^D|_{\Lambda^\circ}(\mathbf{z}) := \|(AD^{-1/2})|_{\Lambda^\circ}\mathbf{z} - y\|_Y^2 + 2\alpha\|\operatorname{diag}(D^{-1/2})|_{\Lambda^\circ}\mathbf{z}\|_{\ell_1(\Lambda^\circ)}, \tag{68}$$

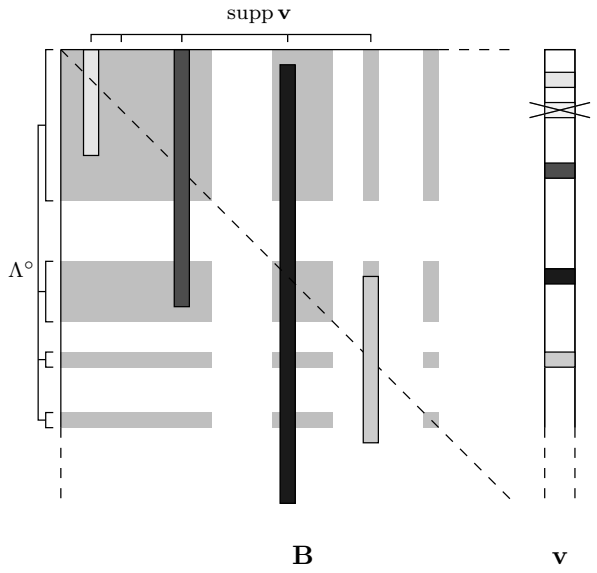


Figure 6: Scheme of approximate matrix-vector products $\mathbf{Bv}$, provided by the **APPLY** subroutine. The entries of $\mathbf{v}$ are multiplied with adaptively compressed columns of $\mathbf{B}$. Large entries of $\mathbf{v}$ meet only slightly compressed columns, whereas small entries are multiplied with strongly compressed columns or are even discarded. In case that $\mathbf{B} = (D^{-1})^* A^* A D^{-1}$ and $\mathbf{v} = \mathbf{z}^{(n)}$, the support of the output may exceed Λ°.

defined for $\mathbf{z} \in \mathbb{R}^{\Lambda^\circ}$, which is a finite dimensional space. Here, we used the notation $T|_{\Lambda^\circ}$ to indicate a matrix for which all the columns indexed in $\mathcal{I} \setminus \Lambda^\circ$ are deleted.

5.3 The adaptive numerical solution of the finite dimensional problem equals the one of the infinite dimensional problem

Note that now the infinite dimensional optimization problem has been reformulated in this way into a finite dimensional one, for which the norm-topology on the solution space $\mathbb{R}^{\Lambda^\circ}$ is not a relevant issue anymore. Let us assume, for instance, that the topology on $\mathbb{R}^{\Lambda^\circ}$ is simply the Euclidean. In such topology the adjoint operator of $AD^{-1/2}|_{\Lambda^\circ}$ is given simply by the transposed matrix. This implies that the minimizer $\mathbf{z}^*|_{\Lambda^\circ}$ may be computed via the following iterative thresholding algorithm:

$$\mathbf{z}^{(n+1)} = \mathbb{S}_{\alpha \operatorname{diag}(D^{-1/2})_{\Lambda^\circ}}(\mathbf{z}^{(n)} + (AD^{-1/2})^*_{\Lambda^\circ} y - (D^{-1/2} A^* A D^{-1/2})|_{\Lambda^\circ \times \Lambda^\circ} \mathbf{z}^{(n)})), \quad (69)$$

where the rate of convergence of the iterations is now governed by the conditioning of the iteration matrix $D^{-1/2} A^* A D^{-1/2}|_{\Lambda^\circ \times \Lambda^\circ}$, which is precisely the one we analyzed in Theorem 4.2.

Of course, the iteration (69) cannot be implemented, unless we know a priori precisely the set Λ°, which is, as previously noted, out of our capability. Nevertheless, in the

following we want to show that the implementation of the *adaptive* algorithm (34), where the iteration matrix A^*A is substituted by $D^{-1/2}A^*AD^{-1/2}$ and the thresholding operations are suitably adapted to $\mathbb{S}_{\alpha^{(n)}\,\mathrm{diag}(D^{-1/2})}$, will turn out to be *equivalent* to applying *the same algorithm* to the resolution of the problem of minimizing the functional $J^D|_{\Lambda^\circ}$ in (68), as we were perfectly in possession of Λ°. In this preconditioned context we will denote by $\tilde{\mathbf{z}}^{(n)}$ the iterations of the adaptive algorithm and by $\mathbf{z}^*$ the minimizer, or fixed point of (69) in finite dimensions.

In particular, we will show that the use of the **APPLY** routine to approximate the matrix-vector multiplication involving $D^{-1/2}A^*AD^{-1/2}$ or $D^{-1/2}A^*AD^{-1/2}|_{\Lambda^\circ\times\Lambda^\circ}$ is simply equivalent, despite the fact that Λ° is not given to us. Actually, it turns out that we can chose $\Lambda^\circ \supset \bigcup_{n=0}^{\infty}\Lambda^{(n)}$, where $\Lambda^{(n)} = \mathrm{supp}(\tilde{\mathbf{z}}^{(n)})$ is the support of the n^{th}-iteration. Therefore, the adaptive algorithm will construct for us the set Λ°.

5.4 The procedure APPLY and its role in the reduction to finite dimensions

In order to clarify our argument we need to have a more precise understanding of the functioning of the **APPLY** routine. In particular, for helping the explanation we refer us to Figure 6. For the precise definition of **APPLY** we refer the reader to [11, 46].

First of all, the main assumption for the efficient use of **APPLY** is the *compressibility* of the matrix $D^{-1/2}A^*AD^{-1/2}$. As shown in Remark 4.5 we can assume for several interesting cases that $D^{-1/2}A^*AD^{-1/2}$ is indeed a compressible matrix. Hence we are indeed allowed to consider this procedure, and not only we can assume $D^{-1/2}A^*AD^{-1/2}$ bounded on ℓ_2, but also on ℓ_τ^w for $0 < \tau < 2$ (compressible matrices have this boundedness property [11, 46]).

The procedure **APPLY** receives as an input a finite vector $\tilde{\mathbf{z}}^{(n)}$ and then it reorders its entries by their magnitude in absolute value. Of the columns corresponding to the largest entries of $\tilde{\mathbf{z}}^{(n)}$ **APPLY** selects and activates around the diagonal a larger number of row-entries, simply by setting to zero the other ones, whereas for the smallest entries of $\tilde{\mathbf{z}}^{(n)}$ only a fewer number of row-entries will be used. The height of the partial columns depends on the accuracy we want eventually to achieve. Note that such accuracy increases during the iteration of the algorithm (34). The higher the accuracy, the larger will be the height of the partial columns. This partial columns so constructed are then multiplied by the vector $\tilde{\mathbf{z}}^{(n)}$. The resulting vector $\mathbf{w}^{(n)}$ of these partial columns-vector multiplications is an approximation of the matrix-vector multiplication $D^{-1/2}A^*AD^{-1/2}\tilde{\mathbf{z}}^{(n)}$. Moreover, $\mathbf{w}^{(n)}$ is again a finite vector. However, it is clear that when the requested accuracy will be very fine, the height of the columns will also be very large and therefore the support of the resulting vector will necessarily exceed the set Λ°. Let us assume now that Λ° is sufficiently large so that, at the very beginning, none of the first, say, N^{th} iterations did exceed, after **APPLY**, the set Λ°. Hence, without loss of generality we may assume that $\cup_{n=0}^{N}\Lambda^{(n)} \cup S^* \subset \Lambda^\circ$. We would like to show now that, thanks to the action of the thresholding operator, $\Lambda^{(n)} \subset \Lambda^\circ$ for all $n \geq N$ as soon as Λ° was chosen sufficiently large. Hence, regardless of the size of the

support of the resulting vectors after an application of the **APPLY** routine on iterations $n \geq N$, the thresholding will produce the restriction of the output of **APPLY** to Λ°. Hence, it would be as **APPLY** was outputting directly on Λ°.

Theorem 5.1. *For convenience of notations let us fix $M = D^{-1/2}A^*AD^{-1/2}$ and $P = D^{-1/2}A^*$. We can also assume $\|M\| < 2$ without loss of generality (remind that M is a bounded operator on $\ell_2(\mathcal{I})$ as well as on $\ell_\tau^w(\mathcal{I})$). Moreover, as a technical assumption we require that $Py \in \ell_2(\mathcal{I}) \cap \ell_\tau^w(\mathcal{I})$, for $0 < \tau < 2$. Let us assume that we are also in the conditions of applicability of Proposition 3.5 for the convergence of the algorithm* (34) *applied for the minimization of the finite dimensional functional* (68)*, but actually without restricting yet neither the matrices nor the vectors to any set Λ° a priori fixed.*

We define $\bar{\mathbf{z}} := \mathbf{z}^* + (Py - M\mathbf{z}^*) \in \ell_2(\mathcal{I})$ and

$$\Lambda^\circ = \cup_{n=0}^N \Lambda^{(n)} \cup S^* \cup \Lambda_\delta(\bar{\mathbf{z}}),$$

where $N \in \mathbb{N}$ is such that

$$\|\mathbf{z}^* - \tilde{\mathbf{z}}^{(N)}\| \leq \tilde{\epsilon}_N \leq \varepsilon, \tag{70}$$

for

$$\left(1 + \frac{1}{\rho}\right)\varepsilon + \delta \leq \alpha \left(\inf_{\lambda \in \mathcal{I}} \operatorname{diag}(D^{-1/2})_\lambda\right), \quad \delta > 0. \tag{71}$$

(We recall the notations $\Lambda^{(n)} = \operatorname{supp}(\tilde{\mathbf{z}}^{(n)})$, $S^* = \operatorname{supp}(\mathbf{z}^*)$, and $\Lambda_\delta(\bar{\mathbf{z}}) = \{\lambda \in \mathcal{I} : \delta < \mathbf{z}^*_\lambda + (Py - M\mathbf{z}^*)_\lambda\}$. The constant ρ is as in Proposition 3.5.)
Then the supports of the iterations never exceed Λ°, i.e.,

$$\Lambda^{(n)} \subset \Lambda^\circ, \text{ for all } n \geq 0.$$

Proof. Note that $\Lambda^{(n)} \subset \Lambda^\circ$, for all $n \leq N$ by definition of Λ°. Let us assume that $n > N$. Let $\lambda \notin \Lambda^\circ$, hence $\lambda \notin S^* = \operatorname{supp}(\mathbf{z}^*)$, and $|\tilde{\mathbf{z}}^{(n)}_\lambda| \leq \varepsilon$. Moreover, we have

$$\begin{aligned}
& \|\mathbf{z}^* + Py - M\mathbf{z}^* - \big(\tilde{\mathbf{z}}^{(n)} + \mathbf{RHS}[\delta_n] - \mathbf{APPLY}[M, \tilde{\mathbf{z}}^{(n)}, \gamma_n]\big)\| \\
\leq\ & \|\mathbf{z}^* + Py - M\mathbf{z}^* - \big(\tilde{\mathbf{z}}^{(n)} + Py - M\tilde{\mathbf{z}}^{(n)}\big)\| \\
+\ & \|\big(\tilde{\mathbf{z}}^{(n)} + Py - M\tilde{\mathbf{z}}^{(n)}\big) - \big(\tilde{\mathbf{z}}^{(n)} + \mathbf{RHS}[\delta_n] - \mathbf{APPLY}[M, \tilde{\mathbf{z}}^{(n)}, \gamma_n]\big)\| \\
\leq\ & \varepsilon + \frac{\varepsilon}{\rho} = \left(1 + \frac{1}{\rho}\right)\varepsilon
\end{aligned} \tag{72}$$

Since $\mathbf{z}^*$ is a fixed point of (69) (in the finite dimensional environment) we have

$$|\mathbf{z}^*_\lambda + (Py - M\mathbf{z}^*)_\lambda| \leq \alpha \operatorname{diag}(D^{-1/2})_\lambda.$$

If $|\mathbf{z}^*_\lambda + (Py - M\mathbf{z}^*)_\lambda| \leq \delta$, then from (72) and (71) we obtain also

$$\begin{aligned}
& |\big(\tilde{\mathbf{z}}^{(n)} + \mathbf{RHS}[\delta_n] - \mathbf{APPLY}[M, \tilde{\mathbf{z}}^{(n)}, \gamma_n]\big)_\lambda| \leq \left(1 + \frac{1}{\rho}\right)\varepsilon + \delta \\
\leq\ & \alpha \left(\inf_{\lambda \in \mathcal{I}} \operatorname{diag}(D^{-1/2})_\lambda\right) \leq \alpha^{(n)} \operatorname{diag}(D^{-1/2})_\lambda.
\end{aligned}$$

Hence $\tilde{\mathbf{z}}_\lambda^{(n+1)} = 0$. Moreover, since $\mathbf{z}^*$ is a finite dimensional vector, M is a bounded operator on $\ell_2(\mathcal{I})$, and $Py \in \ell_2(\mathcal{I})$, then $\bar{\mathbf{z}} := \mathbf{z}^* + (Py - M\mathbf{z}^*) \in \ell_2(\mathcal{I})$ and therefore

$$\Lambda_\delta(\bar{\mathbf{z}}) = \{\lambda \in \mathcal{I} : \delta < \mathbf{z}^*_\lambda + (Py - M\mathbf{z}^*)_\lambda\},$$

has finite cardinality. Hence, if eventually we define now $\Lambda^\circ = \cup_{n=0}^N \Lambda^{(n)} \cup S^* \cup \Lambda_\delta(\bar{\mathbf{z}})$ then we have that $\Lambda^{(n)} \subset \Lambda^\circ$ for all $n \geq 0$. □

Remark 5.2. *1. We remark that in the previous theorem the boundedness properties of the matrix $M = D^{-1/2}A^*AD^{-1/2}$ on $\ell_2(\mathcal{I})$ and $\ell^w_\tau(\mathcal{I})$ play a crucial role. These properties are ensured by the compressibility of M as mentioned in Remark 4.5. Therefore, neither for every operator K nor for any preconditioning matrices $D^{1/2}$ we can expect Theorem 5.1 to hold.*

2. Preconditioning might create some topological troubles and complicate significantly the setting where the algorithm (34) *can work. In simple words Theorem 5.1 establishes that for operators and block-diagonal preconditioning as in §4 the use of the adaptive algorithm* (34) *to compute a sparse solution $\mathbf{u}^*$ is allowed and it will converge as expected, because in practice it will behave as being used for a finite dimensional problem, although the finite dimensional reference space is built on the fly.*

6 Convergence to Compressible and Sparse Solutions

In this section, for the sake of simplicity, we assume $\alpha_\lambda = \bar{\alpha}$ for all $\lambda \in \mathcal{I}$. Hence, we may assume that $\alpha = \bar{\alpha}$ is a scalar. The analysis we developed so far is valid when the parameter α is not too small. Indeed, we may expect that for α becoming smaller, the support of the minimizer $\mathbf{u}^*_\alpha$ of J_α (as in (3)) is becoming also larger and larger. Therefore, we have to expect that the property (26) deteriorates in the sense that $\gamma_0 = \gamma_0(\alpha) \to 1$ for $\alpha \to 0$. In this section we would like to explore the main features of such limit behavior.

6.1 Convergence to compressible solutions

Theorem 6.1. *Assume $\|A\| < \sqrt{2}$ and that there exists a compressible solution $\mathbf{u}^\circ \in \ell^w_\tau(\mathcal{I})$, for $0 < \tau < 2$, such that $A\mathbf{u}^\circ = y$. Let us define the function*

$$\Gamma(\alpha) := \frac{\|(I - A^*A)(\mathbf{u}^*_\alpha - \mathbf{u}^\circ)\|_{\ell_2(\mathrm{supp}(\mathbf{u}^*_\alpha) \cup \mathrm{supp}(\mathbf{u}^\circ))}}{\|\mathbf{u}^*_\alpha - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})}}. \tag{73}$$

The function Γ is certainly bounded above by 1; here we assume that $\Gamma(\alpha) < 1$. Then we have the following estimates:

(i) there exist a constant $C = C(\tau)$ such that

$$\|\mathbf{u}^*_\alpha - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} \leq \frac{C}{1 - \Gamma(\alpha)} |\mathbf{u}^\circ|^{\tau/2}_{\ell^w_\tau} \alpha^{1-\tau/2}; \tag{74}$$

(ii) for another constant $C' = C'(\tau) > 0$ we have

$$\#\operatorname{supp}(\mathbf{u}_\alpha^*) \le \left(\frac{4C^2}{(1-\Gamma(\alpha))^2} + 4C'\right) |\mathbf{u}^\circ|_{\ell_\tau^w}^\tau \alpha^{-\tau}. \tag{75}$$

Proof. Let us write

$$\mathbf{u}_\alpha^* - \mathbf{u}^\circ = \mathbb{S}_\alpha(\mathbf{u}_\alpha^* - A^*A\mathbf{u}_\alpha^* + A^*A\mathbf{u}^\circ) - \mathbb{S}_\alpha(\mathbf{u}^\circ) + \mathbb{S}_\alpha(\mathbf{u}^\circ) - \mathbf{u}^\circ.$$

Then, from (16) and (20), we have the estimates

$$\begin{aligned}
\|\mathbf{u}_\alpha^* - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} &\le \|\mathbb{S}_\alpha(\mathbf{u}_\alpha^* - A^*A\mathbf{u}_\alpha^* + A^*A\mathbf{u}^\circ) - \mathbb{S}_\alpha(\mathbf{u}^\circ)\|_{\operatorname{supp}(\mathbf{u}_\alpha^*)\cup\operatorname{supp}(\mathbf{u}^\circ)} + \|\mathbb{S}_\alpha(\mathbf{u}^\circ) - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} \\
&\le \|(I - A^*A)(\mathbf{u}_\alpha^* - \mathbf{u}^\circ)\|_{\ell_2(\operatorname{supp}(\mathbf{u}_\alpha^*)\cup\operatorname{supp}(\mathbf{u}^\circ))} + C|\mathbf{u}^\circ|_{\ell_\tau^w}^{\tau/2}\alpha^{1-\tau/2} \\
&\le \Gamma(\alpha)\|\mathbf{u}_\alpha^* - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} + C|\mathbf{u}^\circ|_{\ell_\tau^w}^{\tau/2}\alpha^{1-\tau/2}.
\end{aligned}$$

The latter estimate immediately shows (i). Note now that

$$\|\,(\mathbf{u}_\alpha^* - A^*A\mathbf{u}_\alpha^* + A^*y) - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} = \|(I - A^*A)(\mathbf{u}_\alpha^* - \mathbf{u}^\circ)\|_{\ell_2(\mathcal{I})} \le \frac{C}{1-\Gamma(\alpha)}|\mathbf{u}^\circ|_{\ell_\tau^w}^{\tau/2}\alpha^{1-\tau/2}.$$

A straightforward application of (23) yields

$$\begin{aligned}
\#\operatorname{supp}(\mathbf{u}_\alpha^*) &\le \frac{4C^2|\mathbf{u}^\circ|_{\ell_\tau^w}^\tau \alpha^{2-\tau}}{(1-\Gamma(\alpha))^2\alpha^2} + 4C'|\mathbf{u}^\circ|_{\ell_\tau^w}^\tau \alpha^{-\tau} \\
&\le \left(\frac{4C^2}{(1-\Gamma(\alpha))^2} + 4C'\right)|\mathbf{u}^\circ|_{\ell_\tau^w}^\tau \alpha^{-\tau}.
\end{aligned}$$

This concludes the proof of (ii). □

6.2 Convergence to sparse solutions

Corollary 6.2. *Under the hypothesis and notation of Theorem 6.1, assume now that $\mathbf{u}^\circ$ is a sparse solution, i.e., $\mathbf{u}^\circ \in \ell_0(\mathcal{I}) := \cap_{\tau>0}\ell_\tau^w(\mathcal{I})$, and $|\mathbf{u}^\circ|_{\ell_0} = \#\operatorname{supp}(\mathbf{u}^\circ) < \infty$. Let us denote $S^\circ = \operatorname{supp}(\mathbf{u}^\circ)$ and $|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})} := \#S^\circ$. We define the function $\mathscr{S} : \mathbb{N} \to \mathbb{R}_+$ by*

$$\mathscr{S}(k) := \sup_{\Lambda\subset\mathcal{I},\#\Lambda\le k} \|(I - A^*A)|_{(\Lambda\cup S^\circ)\times(\Lambda\cup S^\circ)}\|. \tag{76}$$

Then we have the following estimates

(i)

$$\|\mathbf{u}_\alpha^* - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} \le \frac{1}{1-\Gamma(\alpha)}|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})}^{1/2}\alpha; \tag{77}$$

(ii)

$$\#\operatorname{supp}(\mathbf{u}_\alpha^*) \le \left(\frac{4}{(1-\Gamma(\alpha))^2} + 1\right)|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})}; \tag{78}$$

(iii) the function Γ *fulfills the following functional inequality*

$$\Gamma(\alpha) \leq \mathscr{S}\left(\left\lfloor\left(\frac{4}{(1-\Gamma(\alpha))^2}+1\right)|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})}\right\rfloor\right). \tag{79}$$

Proof. The estimates (i) and (ii) follows immediately from Theorem 6.1 by taking the limit $\tau \to 0$ (see formulas (21) and (22); note the modification of the constants for $\tau = 0$ in Lemma 2.2). The estimate (iii) follows from (ii) by observing that, by definition, $\Gamma(\alpha) \leq \mathscr{S}(\#\operatorname{supp}(\mathbf{u}^*_\alpha))$. □

Remark 6.3. *1. An estimation of the values that the function* Γ *can realize allows for a more precise understanding of the rate of convergence* (77) *and of the support bound* (78)*. This can be achieved by estimating the spectrum of the operator* K *by assuming additional smoothing properties, as it is done in [22, Proposition 4.7]. For general operators* K *little can be said: the functional inequality* (79) *establishes a bound on the range of the values that the function* Γ *can realize, as soon as* $\mathscr{S}$ *is given. In fact, the function* $\mathscr{S}$ *can be heuristically estimated by means of Montecarlo experiments. Therefore the values* $x \in [0,1]$ *for which*

$$x \leq \mathscr{S}\left(\left\lfloor\left(\frac{4}{(1-x)^2}+1\right)\mathscr{A}\right\rfloor\right)$$

are admissible values of Γ *when a sparse solution* $\mathbf{u}^\circ$ *with* $|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})} = \mathscr{A}$ *exists. In Figure*

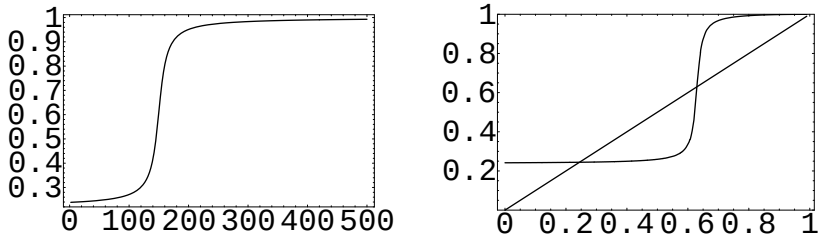


Figure 7: On the left we show a plausible function $\mathscr{S}$ which grows moderately for small values, whereas it saturates to 1 rather quickly for large values. On the right we show the corresponding S-shaped function $\mathscr{S}\left(\left\lfloor\left(\frac{4}{(1-x)^2}+1\right)\mathscr{A}\right\rfloor\right)$ for $\mathscr{A} = 5$, and we highlight the locations where this function is larger than the identity function. These locations correspond to admissible values of Γ.

7 we show an example where a plausible function $\mathscr{S}$ *is given, and the values* $x \in [0,1]$ *where* $x \leq \mathscr{S}\left(\left\lfloor\left(\frac{4}{(1-x)^2}+1\right)\mathscr{A}\right\rfloor\right)$ *holds. Note further that inequality* (79) *to estimate* $\Gamma(\alpha)$ *can be re-used in the previous estimate of the support* (78)*. Hence, we can iterate formula* (79) *with this additional estimation of the support. This iteration can produce a gain of information in certain cases. Unfortunately, it seems that not much more can be said a priori about* $\Gamma(\alpha)$ *in such an abstract setting.*

6.3 Equivalent conditions for the uniform linear convergence of the algorithms as $\alpha \to 0$.

We conclude this section with a result which establishes conditions for a uniform behavior of algorithms (25) and (34) as $\alpha \to 0$.

Corollary 6.4. *Assume that there exists* $\mathbf{u}^\circ \in \ell_0(\mathcal{I})$ *such that* $A\mathbf{u}^\circ = y$, *and*

$$\|I - A^*A|_{\Lambda\times\Lambda}\| \le \gamma_0 < 1 \tag{80}$$

for all $\Lambda \subset \mathcal{I}$ *such that* $\#\Lambda \le \left(\frac{4}{(1-\gamma_0)^2} + 2\right)|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})}$. *Then the following conditions are equivalent:*

(a) $\sup_{\alpha>0}\Gamma(\alpha) \le \gamma_0 < 1$;

(b) $\mathbf{u}_\alpha^*$ *converges linearly to* $\mathbf{u}^\circ$, *i.e.,* $\|\mathbf{u}_\alpha^* - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} \le \frac{1}{1-\gamma_0}|\mathbf{u}^\circ|^{1/2}_{\ell_0(\mathcal{I})}\alpha$;

(c) the support sizes of $\mathbf{u}_\alpha^*$ *are uniformly bounded, i.e.,* $\sup_{\alpha>0}\#\operatorname{supp}(\mathbf{u}_\alpha^*) \le \left(\frac{4}{(1-\gamma_0)^2} + 1\right)|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})}$.

Proof. Assume that (a) holds. Then by (77) we have

$$\|\mathbf{u}_\alpha^* - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})} \le \frac{1}{1-\gamma_0}|\mathbf{u}^\circ|^{1/2}_{\ell_0(\mathcal{I})}\alpha,$$

which implies (b). Let us assume (b) now. Then, again an application of Lemma 2.2 (for the constants adjusted to $\tau = 0$) yields (c):

$$\#\operatorname{supp}(\mathbf{u}_\alpha^*) \le \frac{4|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})}\alpha^2}{(1-\gamma_0)^2\alpha^2} + |\mathbf{u}^\circ|_{\ell_0(\mathcal{I})} = \left(\frac{4}{(1-\gamma_0)^2} + 1\right)|\mathbf{u}^\circ|_{\ell_0(\mathcal{I})}.$$

If (c) holds, then from (80) we have

$$\Gamma(\alpha) = \frac{\|(I - A^*A)(\mathbf{u}_\alpha^* - \mathbf{u}^\circ)\|_{\ell_2(\operatorname{supp}(\mathbf{u}_\alpha^*)\cup\operatorname{supp}(\mathbf{u}^\circ))}}{\|\mathbf{u}_\alpha^* - \mathbf{u}^\circ\|_{\ell_2(\mathcal{I})}} \le \gamma_0.$$

□

7 Numerical Experiments

For instructive and simple numerical experiments in an infinite-dimensional setting, we will consider the Volterra integral operator $K : L_2(0,1) \to L_2(0,1)$,

$$Ku(t) = \int_0^t u(s)\,\mathrm{d}s, \quad K^*Ku(t) = \int_0^1 \big(1 - \max(s,t)\big)u(s)\,\mathrm{d}s. \tag{81}$$

The integration operator can be regarded as a model case for more general Fredholm-type integral operators. K is injective and bounded with norm $\|K\| = 2/\pi \approx 0.64$. The

nonzero eigenvalues of K^*K are explicitly available as $\lambda_n = 1/(\pi(n+\frac{1}{2}))^2$, see [39]. In the following, the discretization of K will be performed using a biorthogonal, piecewise linear spline wavelet basis for $L_2(0,1)$ with 2 vanishing moments from [40]. Due to the piecewise linear kernel of K^*K and the compression properties of the wavelet basis, the system matrix A^*A is quasi-sparse, we refer to Figure 8 for its nonzero pattern. In the experiments, A^*A is scaled to have $\mathcal{L}(\ell_2)$-norm 1.

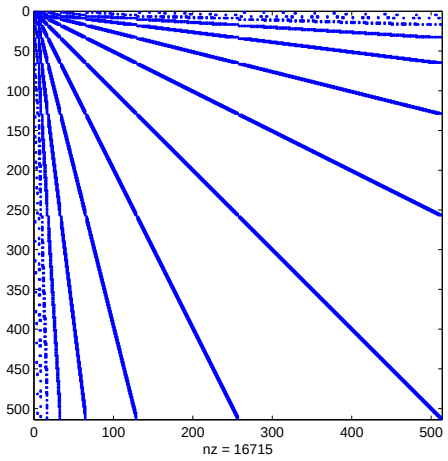


Figure 8: Nonzero pattern of the system matrix A^*A, using piecewise linear spline wavelets up to level 8

7.1 Local well-conditioning

As a first numerical test, we investigate the effect of preconditioning strategies onto the spectral properties of small submatrices of A^*A. For each $N \in \mathbb{N}$, we randomly select 10 times a support set $T \subset \mathcal{I}$ of size $\#T \leq N$. The arithmetic mean over the spectral condition numbers of these submatrices is plotted against the support size N in Figure 9.

As we clearly see, both preconditioning strategies drastically improve the spectral properties of small submatrices. In particular, for square submatrices with less than 100 columns, their condition numbers stay below 10 for diagonal preconditioning, whereas they exceed 10^6 in case of the original matrix. Note that the spectral improvement via diagonal preconditioning is already significant, whereas switching to block-diagonal preconditioning does not further improve the conditioning considerably. Therefore, in practice, the additional computational work of block-diagonal preconditioning can be avoided at no noticeable loss. Of course, due to the clustering of the singular values at 0, larger submatrices of A^*A will fail to have good spectral properties in both preconditioned cases as well.

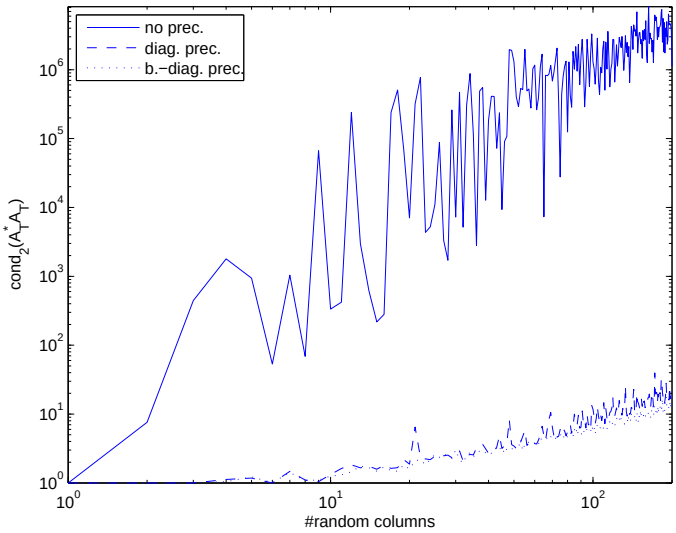


Figure 9: Average spectral condition numbers $\mathrm{cond}_2(A^*A|_T)$ of small $N \times N$-submatrices of A^*A, without preconditioning (*solid line*), with diagonal preconditioning (*dashed line*), and with block-diagonal preconditioning (*dotted line*)

7.2 Recovery of sparse solutions

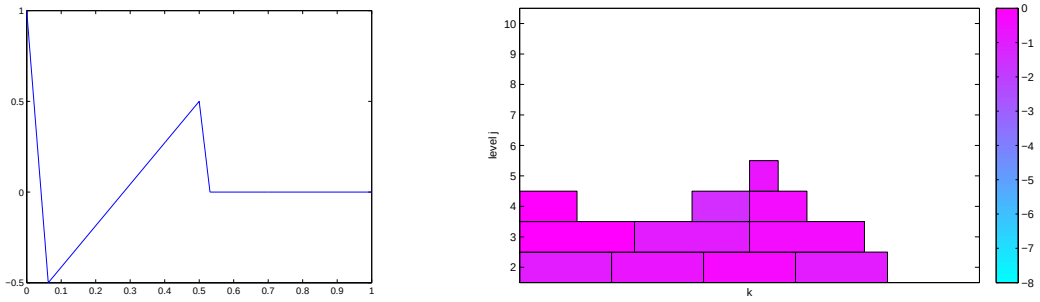
In order to validate the numerical performance of the proposed recovery algorithms, we will fix the following function $u \in L_2(0,1)$ as the solution of $Ku = y$:

$$u(x) = \begin{cases} -24x + 1 & , \ 0 \le x < \frac{1}{16} \\ \frac{16}{7}x - \frac{9}{14} & , \ \frac{1}{16} \le x < \frac{1}{2} \\ -16x + \frac{17}{2} & , \ \frac{1}{2} \le x < \frac{17}{32} \\ 0 & , \ \text{otherwise} \end{cases} . \tag{82}$$

By construction, u is a linear spline function with nodes at dyadic points, therefore its spline wavelet expansion has only finitely many nonzero coefficients, see also Figure 10. In particular, we have $\|\mathbf{u}\|_{\ell_0} = 11$, $\|\mathbf{u}\|_{\ell_1} \approx 1.23628$ and $\|\mathbf{u}\|_{\ell_2} \approx 0.49473$.

7.2.1 Choice of the regularization parameter

The exact minimizer $\mathbf{u}^* = \mathbf{u}^*(\alpha)$ depends on the regularization parameter α in a non-linear way. Due to the presence of roundoff errors in the computation of the system matrix and of the right-hand side, and due to the truncation of the thresholding process after finitely many steps, we may not expect that the numerical reconstructions $\tilde{\mathbf{u}} \approx \mathbf{u}^*$ perfectly match the unknown solution $\mathbf{u}$ as α gets arbitrarily small. In Table 1, we report the attained residual errors and the values of the penalty term after 10000



(a) graph (b) wavelet coefficient distribution

Figure 10: Unknown function u to be recovered

iterations of iterative thresholding (ISTA), with truncation to all wavelets up to level 12. The residual errors apparently begin to stagnate for values of α significantly smaller than 10^{-6}. Therefore, in the numerical experiments, we choose parameters $\alpha \geq 10^{-8}$.

α	$\|A\tilde{\mathbf{u}} - Ku\|_{L_2(0,1)}$	$\|\tilde{\mathbf{u}}\|_{\ell_1}$
1e-04	5.94068e-05	0.37960
1e-05	1.27318e-05	0.96679
1e-06	8.38322e-06	1.58118
1e-07	7.17095e-06	2.56212
1e-08	6.99264e-06	3.02167

Table 1: Residual errors and penalty terms for certain values of the regularization parameter, after 10000 iterations of ISTA

7.2.2 Linear convergence to the minimizer

By the injectivity of the forward operator K and by the finite support of the minimizer $\mathbf{u}^*$ for each given $\alpha > 0$, any of the considered thresholding algorithms should converge linearly. However, the particular error reduction constants will most likely depend on the magnitude of the index set which the particular iterates are supported in. Due to the spectral properties of A^*A, the choice of a large support set within a given iterative scheme will inhibit a significant error reduction per iteration step, while smaller support sets will be beneficial. Therefore, we hope that an iteration with decreasing threshold parameters converges faster than an algorithm with fixed parameters. Moreover, based on the observation from Figure 9, preconditioning strategies should further reduce the error reduction constants considerably.

In order to quantify the linear convergence to the minimizer, we first compute a very good approximation to the exact minimizer $\mathbf{u}^*$ with sufficiently many preconditioned iterations. Then we restart from $\mathbf{u}^{(0)} = \mathbf{0}$, and we measure the ℓ_2 error of the iterands

with respect to that limit. In the experiments, the threshold parameters are chosen according to the rule $\alpha^{(n)} = \alpha + \eta\gamma^n\epsilon_0$, for certain parameters $0 < \gamma, \eta < 1$ and the choice $\epsilon_0 := 5\|\mathbf{u}\|_{\ell_1}$. The iteration is allowed to use all wavelet indices λ with level $|\lambda| \leq 12$. Note that the neglected entries $(A^*A)_{\lambda,\mu}$ are already smaller than 10^{-8} in absolute value, so this is not really a limitation.

In Figures 11 and 12, we plot the ℓ_2 error histories for different iterative thresholding algorithms. We clearly observe linear convergence of all iterative schemes. The

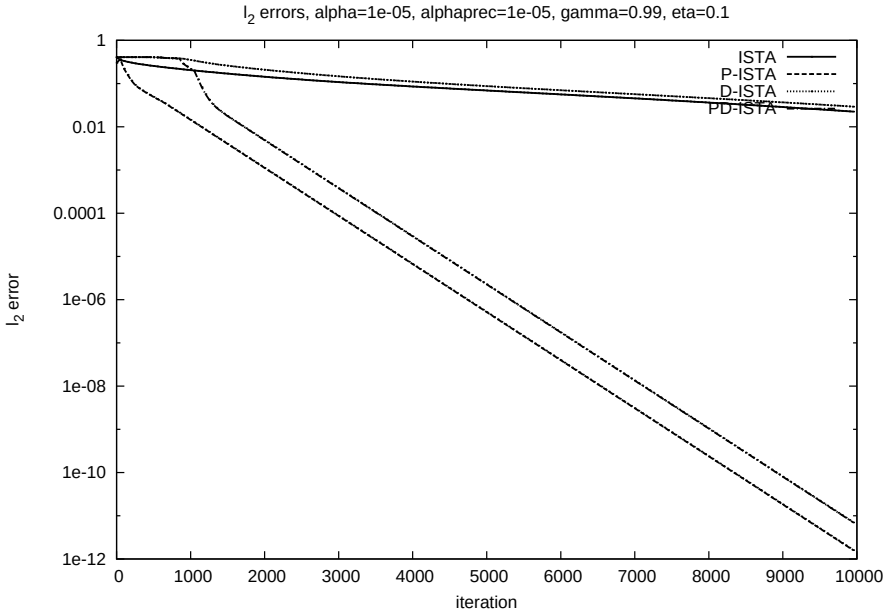


Figure 11: ℓ_2 error histories for different iterative thresholding algorithms, $\alpha = 10^{-5}$, $\gamma = 0.99$ and $\eta = 0.1$: ISTA (*solid line*), D-ISTA (*dashed line*), P-ISTA (*dotted line*), and PD-ISTA (*dash-dotted line*)

original iterative thresholding method (ISTA) and its variant with decreasing threshold parameters (D-ISTA) show a similar asymptotic behavior. Both preconditioned variants P-ISTA and PD-ISTA have a significantly better error reduction per iteration for any choice of the regularization parameter α.

A decreasing choice of the threshold parameters only pays off for medium values of $\alpha \approx 10^{-6}$. An explanation of the varying slopes in Figure 11 can be found when considering the support sizes of the iterands during the iteration, see Figure 14 from the next subsection. For $\alpha = 10^{-7}$, e.g., PD-ISTA asymptotically works on a slightly larger index set than P-ISTA, resulting in a marginally worse error reduction per iteration.

7.2.3 Support dynamics

One of the core ideas towards the acceleration of iterative thresholding methods is to keep the number of active degrees of freedom as small as possible. The rationale is

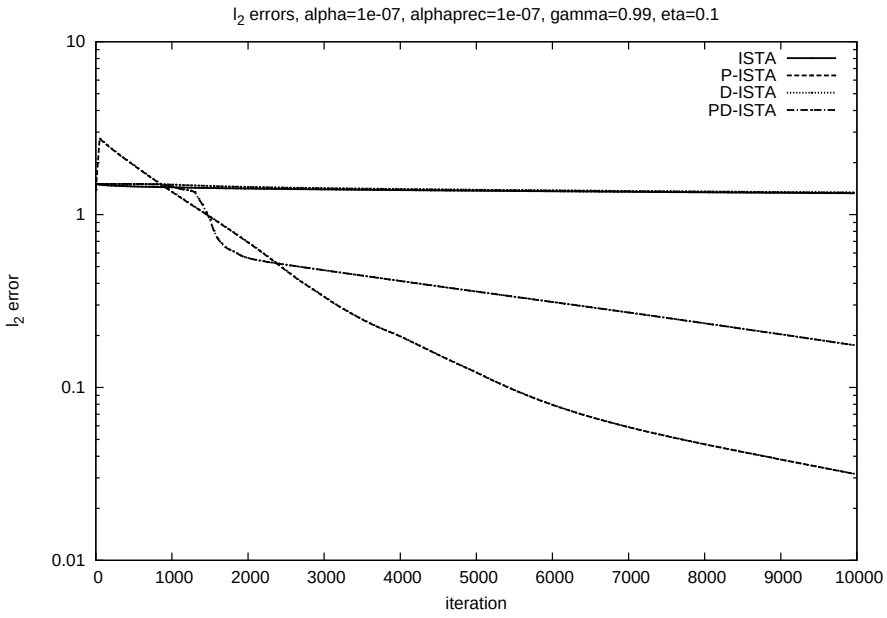


Figure 12: ℓ_2 error histories for different iterative thresholding algorithms, $\alpha = 10^{-7}$, $\gamma = 0.99$ and $\eta = 0.1$: ISTA (*solid line*), D-ISTA (*dashed line*), P-ISTA (*dotted line*), and PD-ISTA (*dash-dotted line*)

to take advantage of the spectral bounds of the associated submatrices of A^*A. In Figures 13 and 14, we track the support size histories for different iterative thresholding algorithms and different regularization parameters α. Apparently, in the course of the iteration, all algorithms gradually reduce the support size of the respective iterands. A decreasing choice of the threshold parameters α_n can efficiently reduce the active degrees of freedom in the transient phase of the first iterations. What is more important, the considered thresholding algorithms behave in a quantitatively different way as concerns the number of active coefficients. As soon as the regularization parameter α is sufficiently small, the un-preconditioned algorithms iterate on significantly larger index sets than their preconditioned counterparts. In Figure 14 we show the first 10000 iterations at which we reached at best c.a. 100 active coefficients, instead of 11 as in the solution (82). As a matter of fact, further iterations will shrink more the support, but, due to accumulation of round-off errors, the algorithm seems unable to compensate them and it seems impossible to match perfectly the right support size. This is because the thresholding parameter $\alpha = 10^{-7}$ is too small to be able to make the round-off errors vanish. However, this effect should not be too surprising because, for an ill-posed problem the use of an inexact iterative algorithm which produces round-offs does make the error to the exact solution blow up when too many iterations are performed. In this paper we do not take into account this problem, and in particular we do not investigate the choice of a stopping rule. We may refer to the paper [41] where such matter was considered.

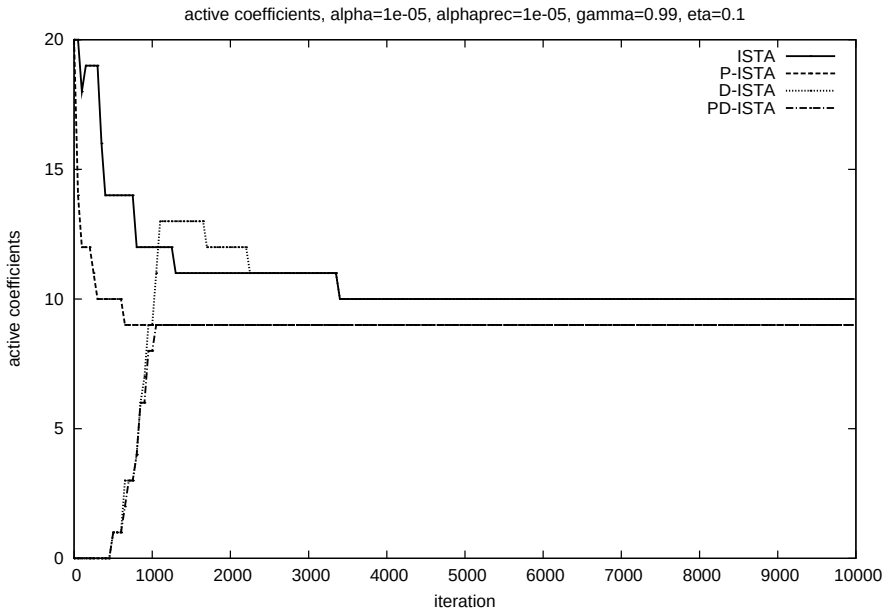


Figure 13: Support size histories for different iterative thresholding algorithms, $\alpha = 10^{-5}$, $\gamma = 0.99$ and $\eta = 0.1$: ISTA (*solid line*), D-ISTA (*dashed line*), P-ISTA (*dotted line*), and PD-ISTA (*dash-dotted line*)

7.2.4 Dynamics

In order to further analyze the response of the considered thresholding algorithms, we investigate the orbit of the iterates towards the minimizer $\mathbf{u}^*$ in the so-called *dynamics plane*, i.e., the relationship between the residuals $\|A\mathbf{u}^{(n)} - y\|_{L_2(0,1)}$ and the respective penalty terms $\|\mathbf{u}^{(n)}\|_{\ell_1}$. We refer to Figure 15 for a dynamics plot. All algorithms do have the same limit sequence $\mathbf{u}^*$, however the associated paths in the dynamics plane look quite different. Apart from the preconditioned iterative thresholding algorithm with constant parameters (P-ISTA), which approaches the minimizer from the right, all other algorithms yield iterates with increasing penalty norms. However, since the plot only shows the first 10000 iterations, it becomes clear again that both preconditioned iterations converge much faster to the limit coefficient array than ISTA.

7.2.5 Adaptivity

As shown in Theorem 5.1, we have theoretical guarantees of convergence of the preconditioned algorithm only if the adaptive applications of the forward operator A^*A are implemented by the procedure **APPLY**. Moreover, the efficiency in terms of complexity of the considered iterative thresholding method can be further improved by these inexact operations. In particular we can exploit that the biinfinite matrix A^*A is compressible in the aforementioned sense. More precisely, due to the piecewise smooth kernel of K^*K,

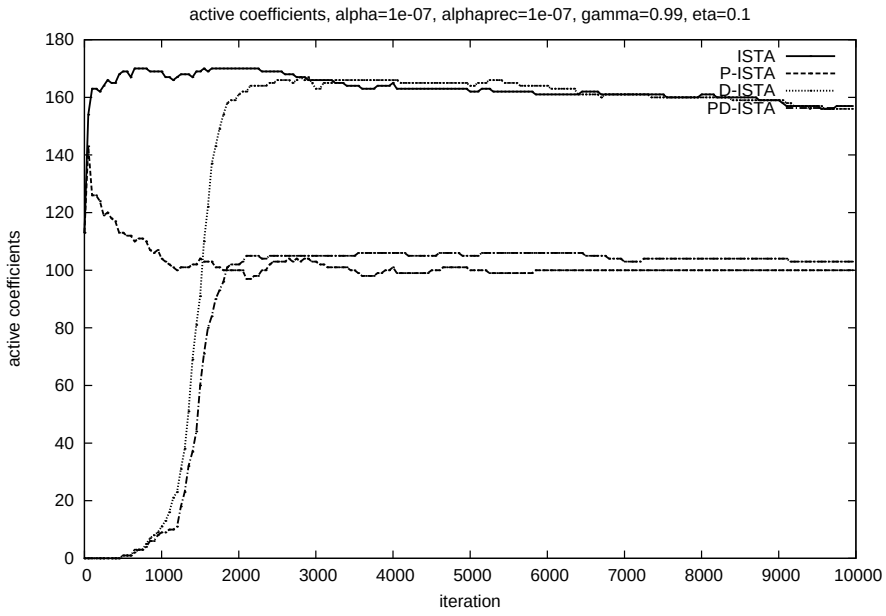


Figure 14: Support size histories for different iterative thresholding algorithms, $\alpha = 10^{-7}$, $\gamma = 0.99$ and $\eta = 0.1$: ISTA (*solid line*), D-ISTA (*dashed line*), P-ISTA (*dotted line*), and PD-ISTA (*dash-dotted line*)

the *compressibility exponent* s in formula (40) does not exceed $s^* = 1.5$, even for smooth wavelet bases. We refer to Table 2 for a numerical validation.

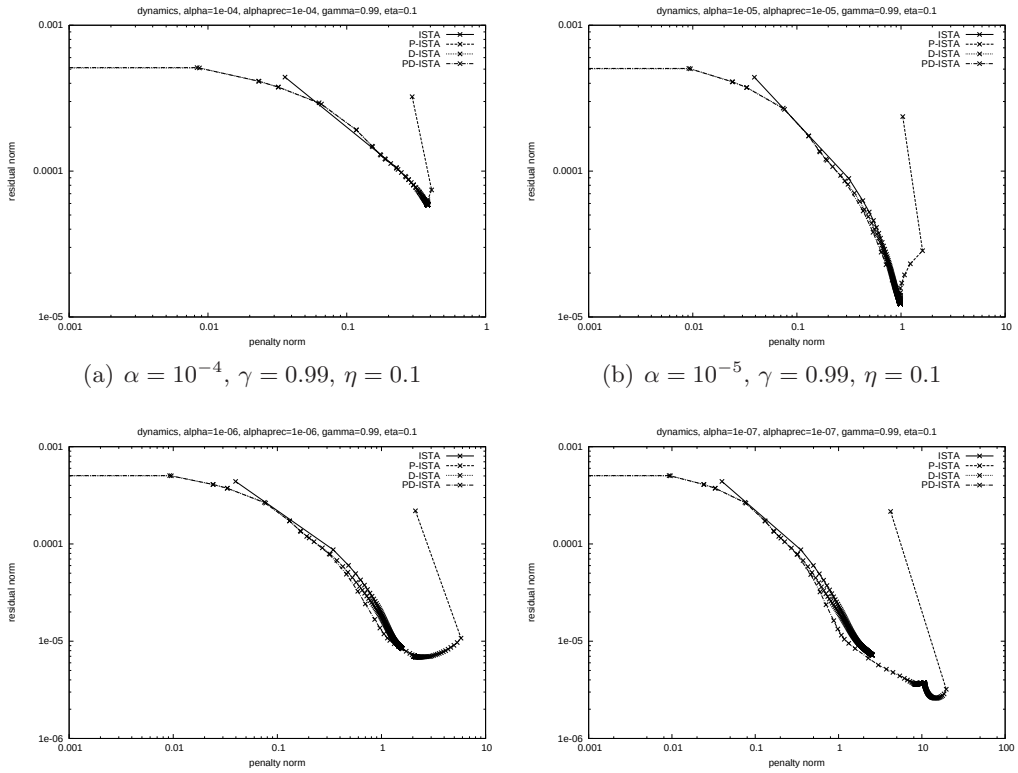
J_k	$\|\mathbf{B} - \mathbf{B}_{J_k}\|$	num. slope
0	5.22783e-03	
2	3.75366e-04	1.900
4	3.94610e-05	1.625
6	5.00009e-06	1.490
8	5.93117e-07	1.538

Table 2: Convergence rate for the approximation of $\mathbf{B} = \mathbf{A}^*\mathbf{A}$ by submatrices $\mathbf{B}_J$ with at most 2^J entries per row and column

We compare iterative thresholding algorithms with adaptive and nonadaptive operator evaluations in Figure 16.

Acknowledgement

The authors would like to thank Christine De Mol for a useful discussion on issues concerning the convergence to compressible solutions §6, and Manuel Werner for providing us with Figure 6.



(a) $\alpha = 10^{-4}$, $\gamma = 0.99$, $\eta = 0.1$

(b) $\alpha = 10^{-5}$, $\gamma = 0.99$, $\eta = 0.1$

(c) $\alpha = 10^{-6}$, $\gamma = 0.99$, $\eta = 0.1$

(d) $\alpha = 10^{-7}$, $\gamma = 0.99$, $\eta = 0.1$

Figure 15: Dynamics of the first 10000 iterations of ISTA (*solid line*), D-ISTA (*dashed line*), P-ISTA (*dotted line*), and PD-ISTA (*dash-dotted line*)

References

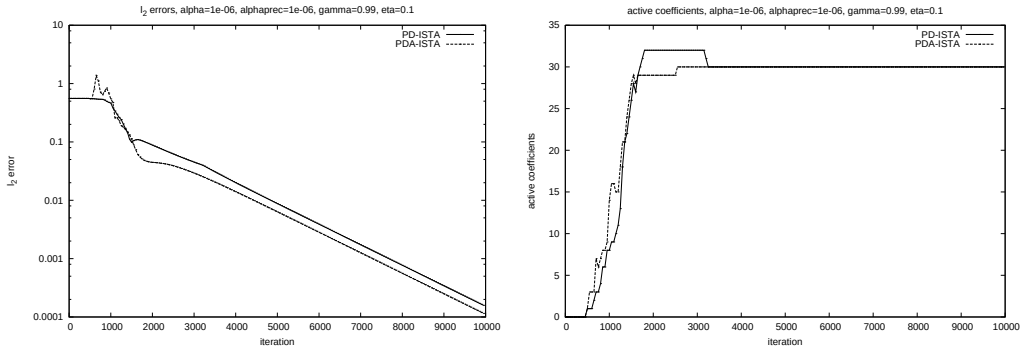
[1] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. **2** (2009), no. 1, 183–202.

[2] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch, *Adaptive wavelet methods and sparsity reconstruction for inverse heat conduction problems*, Preprint series DFG-SPP 1324, Philipps University of Marburg (2009).

[3] K. Bredies and D. Lorenz, *Linear convergence of iterative soft-thresholding*, J, Fourier Anal. Appl. **14** (2008), no. 5-6, 813–837.

[4] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59** (2006), no. 8, 1207–1223.

[5] E. J. Candès, J. Romberg, and T. Tao, *Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inf. Theory **52** (2006), no. 2, 489–509.



(a) ℓ_2 error history (b) degrees of freedom

Figure 16: Convergence and complexity of iterative thresholding algorithms with exact and inexact operator applications for summable tolerances: PD-ISTA (*solid line*) and PDA-ISTA (*dashed line*)

[6] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4203–4215.

[7] E. J. Candès and T. Tao, *Near optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425.

[8] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process. **7** (1998), no. 3, 319–335.

[9] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, 2003.

[10] A. Cohen, *Numerical Analysis of Wavelet Methods.*, Studies in Mathematics and its Applications 32. Amsterdam: North-Holland., 2003.

[11] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations — Convergence rates*, Math. Comp. **70** (2001), 27–75.

[12] ———, *Adaptive wavelet methods II: Beyond the elliptic case*, Found. Comput. Math. **2** (2002), no. 3, 203–245.

[13] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k-term approximation*, J. Amer. Math. Soc. **22** (2009), no. 1, 211–231.

[14] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul. **4** (2005), no. 4, 1168–1200.

[15] S. Dahlke, M. Fornasier, and T. Raasch, *Adaptive frame methods for elliptic operator equations*, Adv. Comput. Math. **27** (2007), no. 1, 27–63.

[16] S. Dahlke, M. Fornasier, T. Raasch, R. Stevenson, and M. Werner, *Adaptive frame methods for elliptic operator equations: The steepest descent approach*, IMA J. Numer. Anal. **27** (2007), no. 4, 717–740.

[17] W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica **6** (1997), 55–228.

[18] W. Dahmen and A. Kunoth, *Multilevel preconditioning*, Numer. Math **63** (1992), 315–344.

[19] W. Dahmen, S. Prössdorf, and R. Schneider, *Wavelet approximation methods for pseudodifferential operators II: Matrix compression*, Adv. Comput. Math. (1993), 259–335.

[20] _____, *Multiscale methods for pseudo-differential equations on smooth manifolds*, Proceedings of the International Conference on Wavelets: Theory, Applications, and Algorithms, Academic Press, 1994, pp. 385–424.

[21] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.

[22] I. Daubechies, M. Defrise, and C. DeMol, *An iterative thresholding algorithm for linear inverse problems*, Comm. Pure Appl. Math. **57** (2004), no. 11, 1413–1457.

[23] I. Daubechies, M. Fornasier, and I. Loris, *Accelerated projected gradient methods for linear inverse problems with sparsity constraints*, J. Fourier Anal. Appl. **14** (2008), no. 5-6, 764–792.

[24] R. A. DeVore, *Nonlinear approximation*, Acta Numerica **7** (1998), 51–150.

[25] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory **52** (2006), no. 4, 1289–1306.

[26] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Statist. **32** (2004), no. 2, 407–499.

[27] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems.*, Mathematics and its Applications (Dordrecht). 375. Dordrecht: Kluwer Academic Publishers., 1996.

[28] M. Figueiredo, R. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE J. Selected Topics in Signal Process. **4** (2007), no. 1, 586–597.

[29] _____, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Proc. (2009).

[30] M. A. T. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration.*, IEEE Trans. Image Proc. **12** (2003), no. 8, 906–916.

[31] M. Fornasier and F. Pitolli, *Adaptive iterative thresholding algorithms for magnetoenceophalography (MEG)*, J. Comput. Appl. Math. **221** (2008), no. 2, 386–395.

[32] G. H. Golub and C. F. van Loan, *Matrix Computations*, John Hopkins University Press, 1989.

[33] C. Del Gratta, V. Pizzella, F. Tecchio, and G. L. Romani, *Magnetoencephalography - a noninvasive brain imaging method with 1ms time resolution*, Rep. Prog. Phys. **64** (2001), 1759–1814.

[34] S. Jaffard, *Propriétés des matrices "bien localisées" près de leur diagonale et quelques applications*, Ann. Inst. H. Poincaré Anal. Non Linéaire **7** (1990), no. 5, 461–476.

[35] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, *A method for large-scale ℓ_1-regularized least squares problems with applications in signal processing and statistics*, IEEE Journal on Selected Topics in Signal Process. (2007).

[36] P.-G. Lemarié, *Bases de'ondelettes sur les groupes de Lie stratifié*, Bulletin de la Société Mathématique de France **117** (1989), no. 2, 213–232.

[37] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen, *Tomographic inversion using ℓ_1-norm regularization of wavelet coefficients*, Geophysical Journal International **170** (2007), no. 1, 359–370.

[38] Ignace Loris, *On the performance of algorithms for the minimization of* 1*-penalized functionals*, Inverse Problems **25** (2009), 035008.

[39] A. K. Louis, *Inverse und schlechtgestellte Probleme*, Teubner, Stuttgart, 1989.

[40] M. Primbs, *Stabile biorthogonale Spline-Waveletbasen auf dem Intervall*, Dissertation, Universität Duisburg-Essen, 2006.

[41] R. Ramlau, G. Teschke, and M. Zhariy, *A compressive Landweber iteration for solving ill-posed inverse problems*, Inverse Problems **24** (2008), no. 6, 065013.

[42] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Communications on Pure and Applied Mathematics (to appear).

[43] R. Schneider, *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur effizienten Lösung grosser vollbesetzter Gleichungssysteme*, Advances in Numerical Mathematics, Teubner Stuttgart, 1998.

[44] J.-L. Starck, E. J. Candès, and D. L. Donoho, *Astronomical image representation by curvelet transform*, Astronomy and Astrophysics **298** (2003), 785–800.

[45] J.-L. Starck, M. K. Nguyen, and F. Murtagh, *Wavelets and curvelets for image deconvolution: a combined approach*, Signal Proc. **83** (2003), 2279–2283.

[46] R. Stevenson, *Adaptive solution of operator equations using wavelet frames*, SIAM J. Numer. Anal **41** (2003), no. 3, 1074–1100.

Stephan Dahlke, Thorsten Raasch
Philipps-Universität Marburg
FB 12 Mathematik und Informatik
Hans-Meerwein Straße
Lahnberge
35032 Marburg
Germany
email: {dahlke,raasch}@mathematik.uni-marburg.de
WWW: http://www.mathematik.uni-marburg.de/∼{dahlke,raasch}

Massimo Fornasier
Johann Radon Institute for Computational and Applied Mathematics (RICAM)
Austrian Academy of Sciences
Altenbergerstrasse 69
A-4040 Linz
Austria
email: massimo.fornasier@oeaw.ac.at
WWW: http://www.ricam.oeaw.ac.at/people/page/fornasier/

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.