

# Compressive Algorithms. Adaptive Solutions of PDEs and Variational Problems

Massimo Fornasier

Johann Radon Institute for Computational and Applied Mathematics,  
Alternbergerstrasse 69  
A-4040 Linz, Austria  
email: massimo.fornasier@oeaw.ac.at

**Abstract.** This paper is concerned with an overview of the main concepts and a few significant applications of a class of adaptive iterative algorithms which allow for dimensionality reductions when used to solve large scale problems. We call *Compressive Algorithms* this class of numerical methods. The introduction of this paper presents an historical excursus on the developments of the main ideas behind compressive algorithms and stresses the common features of diverse applications. The first part of the paper is addressed to the optimal performances of such algorithms when compared with known benchmarks in the numerical solution of elliptic partial differential equations. In the second part we address the solution of inverse problems both with sparsity and compressibility constraints. We stress how compressive algorithms can stem from variational principles. We illustrate the main results and applications by a few significant numerical examples. We conclude by pointing out future developments.

## 1 Introduction

*Compressive Algorithms* are a novel approach to efficient adaptive computing and take advantage of the property of solutions of certain PDE's and variational problems to be characterized by few *degrees of freedom*, which are recovered by adaptive nonlinear iterations. The approach to efficient computing via compressive algorithms responds to the need of addressing very large scale problems by means of a dimensionality reduction which requires the solution of a combinatorial optimization. Compressive algorithms can perform optimal complexity solutions since they tend to use the minimal number of degrees of freedom, and are simple to implement. They are already successfully applied in several problems. Their numerical analysis is yet very challenging.

Roughly speaking, compressive algorithms are generically formulated as *thresholded gradient iterations*

$$\left\{ \begin{array}{l} u^{(n+1)} = H_{\gamma_n} \left( \underbrace{u^{(n)} + L_n^*(f - L_n u^{(n)})}_{\text{gradient step}} \right), \quad n \in \mathbb{N} \\ L_n \approx L, \text{ for } n \rightarrow \infty \\ (\gamma_n)_n \text{ is a sequence of } \textit{shape} \text{ parameters ,} \end{array} \right. \quad (1)$$

to solve exactly, for well-posed problems, or approximatively, for regularized ill-posed problems, the equation

$$Lu = f.$$

Here  $L$  and  $L_n$  are suitable linear operators acting on the underlying solution space (usually a suitable Hilbert space), they depend on the particular problem at hand,  $f$  is the datum of the problem. In several instances  $L_n$  is a suitable approximation to a preconditioned version of  $L$ . The nonlinear function  $H_\gamma$  acts as a thresholding operator, i.e., *penalizes small/promotes relevant features* of the solution, depends on the problem, and on the “shape parameter”  $\gamma$ , which controls the thresholding type and level. Let us stress from now that the “relevant features” might not be merely, e.g., large *wavelet coefficients* (see, e.g. [17]), but they can be expressed in terms of more sophisticated representations of the solution. For example, we will consider solutions of degenerate PDE’s with discontinuities along curves, and these will be the interesting features to be recovered during the solution process.

The main task in compressive algorithms is to predetermine or to adapt the shape parameters  $(\gamma_n)_n$  and the approximations  $L_n$  in order to realize the best trade-off between rate of convergence to the solution and complexity.

Compressive algorithms have a long history of important successes. Their first formulations, starting with the concept of *thresholding*, can be traced back to the mid-nineties with the pioneering work of Chambolle, DeVore, Lee, and Lucier [17], Donoho and Johnstone [54, 55], Rudin, Osher, and Fatemi [86]. Indeed, on the one side, the design of bases (e.g., curvelets, local Fourier bases, wavelets) for sparse representations of digital signals has led to extremely efficient compression methods, such as JPEG 2000 [79]. Applications in signal de-noising appear in [49, 50]. On the other side, degenerate elliptic PDE’s for anisotropic diffusion, as appearing in total variation minimization, were also successfully applied for noise removal in digital images, since they essentially perform a suitable thresholding of derivatives and promote few edges. Degenerate PDE’s have revolutionized image processing with an enormous impact and consequences [3].

In the late-nineties, the attention moved from the compressibility of signals to the compressibility of functions that are only *implicitly* given as solutions of equations. A new generation of compressive algorithms was proposed by Cohen, Dahmen and DeVore in a sequence of fundamental papers [23–25] for the computation of compressed solutions of *elliptic* differential and integral equations, exploiting adaptive and greedy strategies. One of the innovations of their work is the *a priori* analysis of the optimality of such an adaptive scheme. Since solutions can be compressed, hence only few relevant degrees of freedom are sufficient to well-approximate it, one would like to have algorithms that approximate the solution performing a number of algebraic operations which are asymptotically proportional to the number of degrees of freedom of the best approximation. In other words, the complexity of the algorithm is  $\mathcal{O}(N)$ , where  $N$  is the minimal number of degrees of freedom for approximating the solution up to a given accuracy.

Although they were mostly interested in approximations of solutions by means of multiscale bases (e.g., wavelets), their work has recently influenced significantly also the approach to the analysis of adaptive finite element methods and the understanding of their optimal performances [5, 91, 92]. The latter optimality was previously evaluated only by *a posteriori* numerical tests [80, pag. 634].

The use of the compressibility in more general variational problems for signal recovery and solution of nonlinear equations is the most recent step of this concept's long career of "simplifying and understanding complexity", with an enormous potential in applications [25, 26, 45, 51, 78, 88, 89]. In particular, the observation that it is possible to reconstruct compressible signals from vastly incomplete information just seeking for the total variation or  $\ell_1$ -minimal solutions [10–12, 52] has led to a new line of research called *sparse recovery* or *compressed sensing*, with very fruitful mathematical and applied results. Again compressive algorithms play a fundamental role in this context [42, 46, 16, 18, 97, 98].

This historical *excursus* motivates our understanding that these instances of compressing algorithms appearing in different contexts indeed belong to the same *family of numerical methods* with very similar underlying concepts and technical approaches.

We present a non-exhaustive overview of these methods and significant examples of relevant applications. We particularly stress the variational nature of these algorithms, which often can be derived as related to nonlinear *subgradient iterations* in a nonsmooth minimization process. Often such compressive algorithms stem also from multiple-minimization strategies: the initial variational problem is solved by the alternating minimization of an augmented functional with multiple variables. Sometimes this strategy is employed to perform the original minimization in an easier way by introducing useful auxiliary variables, whereas at other times we want to realize clever subspace corrections in order to accelerate the convergence and to reduce the dimensionality of the problem.

We would like to illustrate two kinds of results, without stressing too much the rigor of the presentation. We will include exhaustive references for the reader in search of further details. Not all the references are recalled in the text. We would like instead to convey the main principles, also by illustrating them with numerical examples.

**Optimality.** The first kind of results is addressed in Section 2 to the optimal performances of compressive algorithms. In particular, we consider the difficult task of realizing optimal approximations to solutions of elliptic operator equations discretized by redundant decompositions (e.g., wavelet frames).

**Variational formulation of compressing algorithms.** The second kind of results is addressed in Section 3 to the formulation of new compressive algorithms for the solution of variational problems involving sparsity constraints, in particular when the dimension of the problem becomes exaggeratedly large.

In order to support the understanding of the impact for applications, we provide

a collection of a few numerical examples for the adaptive solution of PDE's and for image processing.

## 2 Optimal Adaptive Frame Solvers for Operator Equations

### 2.1 The state of the art

As already mentioned in the short introductory overview, compressive algorithms were proposed in order to compute with optimal complexity approximations to solutions of well-posed elliptic operator equations. They are generally realized by iterative adaptive gradient steps, where the matrix-vector multiplications are performed taking advantage of the compressibility of the *stiffness matrices*  $L$  resulting by discretizations via suitable multiscale bases (e.g., wavelets), followed by thresholding operations, called *coarsening* in this context,

$$u^{(n+1)} = H_{\gamma_n} \left( u^{(n)} + \beta^{(n)}(f - L_n u^{(n)}) \right), \quad \beta^{(n)} > 0, \quad n \in \mathbb{N}. \quad (2)$$

The role of the latter operations is to enforce the elimination of negligible quantities and to select the most relevant coordinates, eventually ensuring the right balance between number of degrees of freedom/complexity and rate of convergence. The coarsening corresponds to a *hard-thresholding*  $H_{\gamma_n}$  with threshold  $\gamma_n \rightarrow 0$  for  $n \rightarrow \infty$ , see Figure 1 for an illustrative explanation of soft- and hard-thresholding. This means that progressively more and more details of the solution are retained. In practice, the thresholding determines a procedure  $\mathbf{COARSE}[v, \varepsilon] = H_{\gamma_\varepsilon}(v)$ , such that

$$\|\mathbf{COARSE}[v, \varepsilon] - v\| \leq \varepsilon, \quad \varepsilon > 0$$

for  $\varepsilon > 0$ , and  $\mathbf{COARSE}[v, \varepsilon]$  has asymptotically minimal support for  $\varepsilon \rightarrow 0$ . The compressed matrix-vector multiplications  $L_n u^{(n)}$  are also realized with the help of a suitable adaptive procedure  $\mathbf{APPLY}[v, \varepsilon]$ , which determines from a finitely supported  $v$ , an asymptotically minimal support vector (for  $\varepsilon \rightarrow 0$ )  $z_\varepsilon$  such that

$$\|Lv - z_\varepsilon\| \leq \varepsilon, \quad \varepsilon > 0.$$

In order to promote the minimal complexity in computing the coarsening, also the datum  $f$  is compressed by means of a suitable procedure  $\mathbf{RHS}[f, \varepsilon]$  which returns an asymptotically minimal support vector (for  $\varepsilon \rightarrow 0$ ) such that

$$\|\mathbf{RHS}[f, \varepsilon] - f\| \leq \varepsilon, \quad \varepsilon > 0.$$

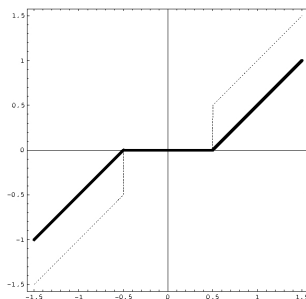
The optimal complexity of such procedures is discussed, e.g., in [23, 90, 93], and they allow for an implementable version of (2):

$\mathbf{SOLVE}[\eta, \varepsilon] \rightarrow u$ :

% Input should satisfy  $\eta > 0$  large enough.

% Define the parameters  $\alpha_{opt} := \frac{2}{\|L\| + \|L^T\| - 1}$  and  $\rho := \frac{\kappa(L) - 1}{\kappa(L) + 1}$ .

% Let  $\theta$  and  $K$  be constants with  $2\rho^K < \theta < 1/2$ .



**Fig. 1.** We depict soft- and hard-thresholding curves, with solid and dashed lines respectively. Small absolute values are mapped to zero, while values exciding in modulus the threshold  $\gamma$  are slightly diminished and just kept respectively.

```

u := 0;
while  $\eta > \varepsilon$  do
  for j := 1 to K do
    u := u +  $\alpha_{opt}$  ( RHS[f,  $\frac{\rho^j \eta}{2\alpha K}$ ] - APPLY[u,  $\frac{\rho^j \eta}{2\alpha K}$ ] );
  endfor
   $\eta := 2\rho^K \eta / \theta$ ;
  u := COARSE[u, (1 -  $\theta$ ) $\eta$ ];
enddo

```

One can also modify this algorithm with a variable *descend parameter*  $\alpha_\eta$  (instead of the prescribed  $\alpha_{opt}$ ), according to, e.g., a classical *steepest-descent* (SD) criterion, see [35] for more details. The optimality benchmark of **SOLVE** was mostly limited to best approximations with respect to wavelet bases. Unfortunately, the efficient applicability of these methods is spoiled by the crucial problem of constructing well-conditioned boundary adapted bases especially on domains with complicated geometry or manifolds [37–39]. The wavelet bases constructed so far exhibit relatively high condition numbers or limited smoothness. The patching used to construct global smooth wavelets by domain decomposition techniques appears complicated and, in most cases, makes the conditioning worse. The global smoothness of the basis, when implementing adaptive schemes in [23, 24], is a necessary condition for getting compressibility (i.e., finitely banded approximations) of (infinite) discretization matrices  $L$ , especially for high order operators. This bottleneck has led to generalizations of the wavelet approach. These generalizations are based on *frames*, i.e., stable, redundant, non-orthogonal expansions [21, 35, 41, 60, 90], which are simpler to construct, hence potentially more attractive to practitioners.

## 2.2 The frame discretization

Frame construction is usually implemented by Overlapping Domain Decompositions (ODD) so no patching at the interfaces is needed to obtain global smoothness. Moreover, the use of frames, due to their intrinsic redundancy, tends to improve the conditioning (meant as the ratio between the largest and the smallest nonzero eigenvalue) of the corresponding discretization matrices and do not spoil their compressibility. Let us recall shortly the general setting.

Let  $\Omega$  be a Lipschitz domain in  $\mathbb{R}^n$ , possibly with re-entrant corners.  $H$  is a Hilbert space of smooth functions with the following embeddings  $H \subset L_2(\Omega) \subset H'$ . It is also useful to introduce shortly the notations for sequence spaces. For a countable index set  $\Lambda$ ,  $0 < p < \infty$ , we define the spaces  $\ell_p(\Lambda) = \{c = (c_\lambda)_{\lambda \in \Lambda} : \sum_{\lambda \in \Lambda} |c_\lambda|^p < \infty\}$  endowed with the natural (quasi-)norm, and for  $p = \infty$  we invoke the usual modification. The operator  $\mathcal{L} : H \rightarrow H'$  is, e.g., linear, and elliptic. The Laplace operator acting on  $H = H_0^1(\Omega)$  is an appropriate example:

$$\begin{aligned} -\Delta u &= f & \text{in } \Omega, \\ u &= 0 & \text{on } \partial\Omega. \end{aligned} \tag{3}$$

The task to perform is to solve adaptively  $\mathcal{L}u = f$ , for  $f \in H'$ , when  $u$  has limited Sobolev smoothness, whereas high Besov regularity [28, 32]. In particular, the assumption is that the solution  $u^*$  has *compressible expansions*  $u^* = \sum_\lambda u_\lambda^* \psi_\lambda$  (i.e., with a small number of relatively large coefficients  $u_\lambda$  in absolute value) with respect to a frame  $\Psi = \{\psi_\lambda\}$  (e.g., wavelets, shearlets) [21, 35, 41, 60, 90] constructed, e.g., on overlapping domain decompositions. In particular we choose a frame  $\Psi = \{\psi_\lambda\}_{\lambda \in \Lambda}$  for  $H$ , i.e.,  $\|f\|_{H'} \approx \|\langle f, \Psi \rangle\|_{\ell_2(\Lambda)}$ , and  $F^* : \ell_2(\Lambda) \rightarrow H : c \mapsto c^T \Psi := \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda$ ,  $\tilde{F} : H \rightarrow \ell_2(\Lambda) : u \mapsto \langle u, \tilde{\Psi} \rangle$  are bounded operators,  $\tilde{\Psi}$  is a dual frame. With the frame we discretize the problem as follows:

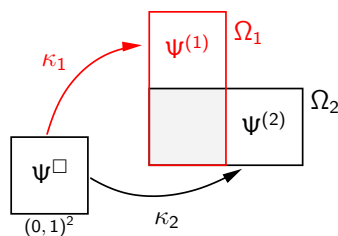
$$\mathcal{L}u^* = f \Rightarrow Lu^* = f,$$

$$\begin{cases} L := \langle \mathcal{L}\Psi, \Psi \rangle, \\ f := \langle f, \Psi \rangle. \end{cases}$$

For simplicity and with the hope not to create confusion in this informal discussion, we used here  $u^*$  and  $f$  both to indicate the function and the frame coefficients. The brackets  $\langle \cdot, \cdot \rangle$  denote the duality between  $H'$  and  $H$ .

The construction of the frame by means of an overlapping domain decomposition can be schematically described as follows:

For  $H = H_0^t(\Omega)$ , we fix a smooth reference Riesz basis  $\Psi^\square \subset H_0^t(\square)$ ,  $\square := (0, 1)^n$ , typically a wavelet basis with complementary Dirichlet boundary conditions. Given a suitable overlapping decomposition  $\Omega = \sum_{i=1}^K \Omega_i$ , and  $\kappa_i : \square \rightarrow \Omega_i$ ,  $C^m$ -diffeomorphisms,  $m \geq t$ , an appropriate lifting yields *aggregated frames* given by  $\Psi = \bigcup_{i=1}^K \Psi_i$ , see also [35, 63, 90].



Certainly, an overlapping domain decomposition generates regions of the domain where the side effect of the redundancy is that solutions are no longer uniquely representable by the global frame system. At first sight, it may seem that redundancy contradicts the minimality requirement on the amount of information being used to approximate the solution. Often accurate simulations already require processing a huge amount of data. How can one attempt such computations if the degrees of freedom are also made redundant? A figurative answer to this question is the so-called “dictionary example”: The larger and richer is my dictionary the *shorter* are the phrases I compose. The use of the proper terminology avoids long circumlocutions for describing an object. Of course, the key point is the capability of choosing the right terminology. Back to mathematical terms, the combination of adaptivity (i.e., the capability of choosing the right terminology) and redundancy (i.e., the richness or non-uniqueness of representations) can give rise to compressed and accurate approximations.

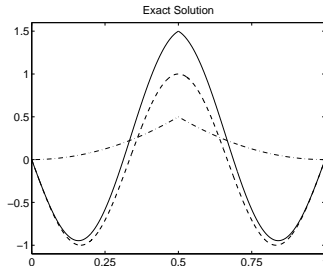
### 2.3 Optimality of adaptive frame approximations

Our main contribution in this setting was directed to the difficult estimate of the trade-off between compressibility and adaptivity effort. While for bases the best  $N$ -term approximation is the optimal benchmark, for frames this concept is not well-defined, due to the non-uniqueness of the expansion. Nevertheless, for frames constructed as a union of bases of the *same* nature (e.g., wavelet bases), as we do by using overlapping domain decompositions, we have to expect that the performances cannot be much better than using a unique global basis. Hence, in these cases, again the best  $N$ -term approximation with respect to a certain specific representation of the solution constructed via suitable partitions of unity does represent a good benchmark. The core of our work in this direction consisted in the proof of the optimality of compressive algorithms based on wavelet frame discretizations, despite the redundancy [34, 33, 35, 90].

### 2.4 Numerical experiments

In order to illustrate the mentioned theoretical results, we would like to present a few concrete numerical examples. For the discretization we use aggregated wavelet frames on suitable overlapping domain decompositions, as the union of local wavelet bases lifted to the subdomains. As such local bases we use

piecewise linear and piecewise quadratic wavelets with complementary boundary conditions from [37], with *order of polynomial exactness*  $d = 2$  and with  $\tilde{d} = 2$  *vanishing moments*. In particular, we impose here homogenous boundary conditions on the primal wavelets and free boundary conditions on the duals. We test the algorithms on both 1D and 2D Poisson problems.



**Fig. 2.** Exact solution (solid line) for the one-dimensional example being the sum of the dashed and dash-dotted functions.

## 2.5 Poisson Equation on the Interval

We consider the variational formulation of the following problem of second order on the interval  $\Omega = (0, 1)$ , i.e.,  $n = 1$ , with homogenous boundary conditions

$$-u'' = f \quad \text{on } \Omega, \quad u(0) = u(1) = 0. \quad (4)$$

The right-hand side  $f$  is given as the functional defined by  $f(v) := 4v(\frac{1}{2}) + \int_0^1 g(x)v(x)dx$ , where

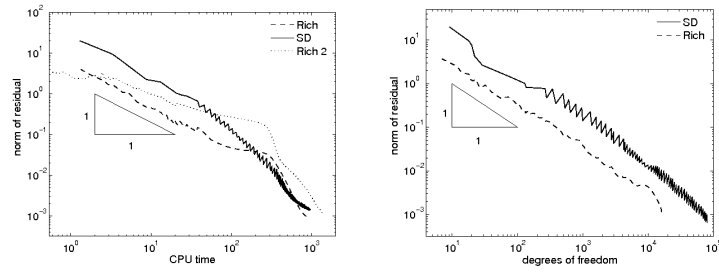
$$g(x) = -9\pi^2 \sin(3\pi x) - 4.$$

The solution is consequently given by

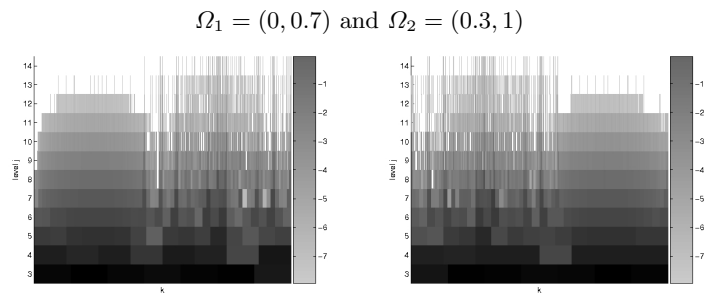
$$u(x) = -\sin(3\pi x) + \begin{cases} 2x^2 & , x \in [0, \frac{1}{2}) \\ 2(1-x)^2 & , x \in [\frac{1}{2}, 1] \end{cases},$$

see Figure 2. As an overlapping domain decomposition we choose  $\Omega = \Omega_1 \cup \Omega_2$ , where  $\Omega_1 = (0, 0.7)$  and  $\Omega_2 = (0.3, 1)$ . Associated to this decomposition we construct our aggregated wavelet frames just as the union of the local bases. It is shown in [34, 33, 90] that such a system is a (Gelfand) frame for  $H_0^t(\Omega)$  and that it can provide a suitable characterization of Besov spaces in terms of wavelet coefficients.

On the one hand, the solution  $u$  is contained in  $H_0^{s+1}(\Omega)$  only for  $s < \frac{1}{2}$ . This means that linear Galerkin methods can only converge with limited order. On the other hand, it can be shown that  $u \in B_\tau^s(L_\tau(\Omega))$  for any  $s > 0$ ,  $1/\tau = s - 1/2$  [28,



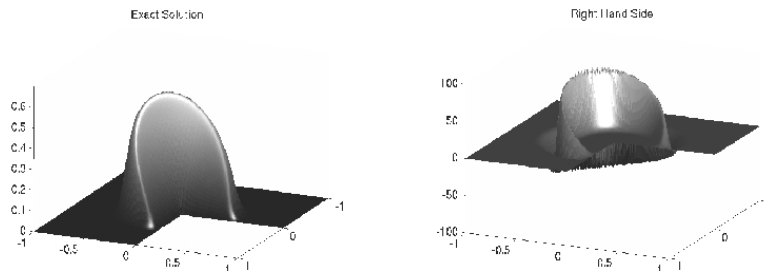
**Fig. 3.** *Left:* Convergence history of **SOLVE** with respect to CPU time. Tests for **SOLVE** with different damping parameters are shown. *Right:* Convergence history with respect to the support size of the iterands.



**Fig. 4.** Distribution of active wavelet frame elements in  $\Omega_1$  and  $\Omega_2$ .

32], so that the wavelet frame coefficients  $u_\lambda$  associated with  $u$  define a sequence in  $\ell_\tau^w$  for any  $s < \frac{d-t}{n}$  (see [48, 90]). This ensures that the choice of wavelets with suitable order  $d$  can allow for any order of convergence in adaptive schemes like that presented in this paper, in the sense that the error is  $O(N^{-s})$  where  $N$  is the number of unknowns. Due to our choice of piecewise linear wavelets with order  $d = 2$ , the optimal rate of convergence is expected to be  $s = \frac{d-t}{n} = 1$ . We show that the numerical experiments confirm this expected rate.

We tested the adaptive wavelet algorithm **SOLVE** with adaptive choice of the parameter  $\alpha_\eta$  according to the steepest descent (SD) rule, and with parameter  $\alpha_{opt} \approx 0.52$ ,  $\theta = 2/7$ ,  $K = 83$ , and with initial  $\eta = 64.8861$ . The numerical results in Figure 3 illustrate the optimal computational complexity of the two implementations. In particular, we show that **SOLVE** with steepest-descent parameter rule outperforms a suboptimal choice of the damping parameter ( $\alpha^* = 0.2 \leq \alpha_{opt} \approx 0.52$  in this specific test). In practice, the wrong guess of the damping parameter can even spoil convergence and/or optimality. Finally, Figure 4 illustrates the distribution of the active wavelet frame elements used by the steepest descent scheme, each of them corresponding to a colored rectangle. The two overlapping subintervals are shown separately. For both patches one observes that the adaptive scheme detects the singularity of the solution. The chosen frame elements arrange in a tree-like structure with large coefficients around the singularity, while on the smooth parts the coefficients are uniformly distributed, and along a fixed level they are of similar size here.



**Fig. 5.** Exact solution (left) and right-hand side for the two-dimensional Poisson equation in an  $L$ -shaped domain.

## 2.6 Poisson Equation on the $L$ -shaped Domain

We consider the model problem of the variational formulation of Poisson's equation in two spatial dimensions:

$$-\Delta u = f \text{ in } \Omega, \quad u|_{\partial\Omega} = 0. \quad (5)$$

The problem will be chosen in such a way that the application of *adaptive* algorithms pays off most, as is the case for domains with reentrant corners. Here,

the reentrant corners themselves lead to singular parts in the solutions, forcing them to have a limited Sobolev regularity, even for smooth right-hand sides  $f$ . We use

$$\mathcal{S}(r, \theta) := \zeta(r)r^{2/3} \sin\left(\frac{2}{3}\theta\right),$$

as the exact solution, which is shown together with the corresponding right-hand side in Figure 5. It is well-known that  $\mathcal{S} \in H^s(\Omega)$  for  $s < 5/3$  only, but it is contained in every Besov space  $B_\tau^s(L_\tau(\Omega))$ , where  $s > 0$ ,  $1/\tau = (s - 1)/2 + 1/2$  (see [28]). As has been previously noted, the convergence rate of a uniform refinement strategy is determined by the Sobolev regularity of the solution, while in the context of adaptive schemes it depends on the Besov regularity (cf. [29]). In particular, considering piecewise quadratic approximation, the best possible convergence rate in the  $H^1(\Omega)$ -norm for uniform refinement strategies is  $\mathcal{O}(N^{-(\frac{5}{3}-1)/2})$ , with  $N$  being the number of unknowns, whereas our adaptive frame scheme gives the optimal rate  $\mathcal{O}(N^{-1})$ .

For our numerical experiments, we use an aggregated wavelet frame. With  $\Omega_1 = (-1, 0) \times (-1, 1)$ ,  $\Omega_2 = (-1, 1) \times (-1, 0)$ , and  $\square = (0, 1)^2$ , let  $\kappa_i$  be affine bijections between  $\square$  and  $\Omega_i$  ( $i = 1, 2$ ). For  $\Psi^\square$  being a piecewise quadratic wavelet basis for  $H_0^1(\square)$ , where  $d = 3$  and  $\tilde{d} = 5$ , we set  $\Psi = \cup_{i=1}^2 \kappa_i(\Psi^\square)$ .

In Figure 6 we show some of the approximations and the corresponding point-wise differences to the exact solution produced by our steepest descent scheme using piecewise quadratic frame elements. The numerical results in Figure 7 illustrate the optimal convergence of the scheme with different descent parameter rules. Figure 8 shows similar results by using improved bases [33] with  $d = \tilde{d} = 3$ .

## 2.7 Towards domain decomposition methods

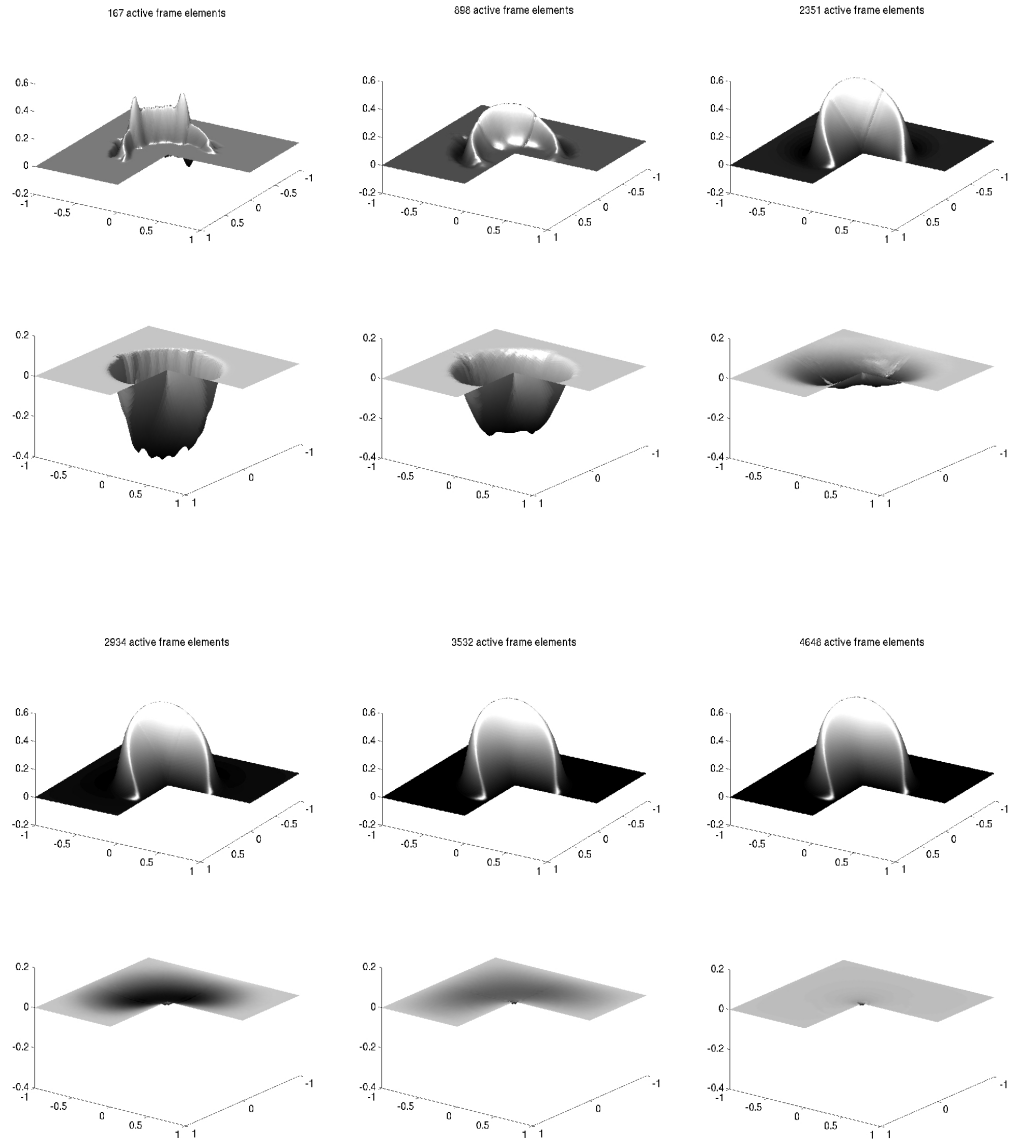
Figure 4 is very useful in order to highlight a fundamental behavior of **SOLVE**. In the overlapping region nearly the same wavelet coefficients related to both patches are simultaneously activated. This means that these algorithms are unable to eliminate the frame redundancy and to approximate the sparsest representation of the solution. This problem can be solved by realizing a coarsening that does not just threshold coefficients but really promotes very sparse representations. However, such a procedure needs to implement a basis pursuit strategy which can be computationally expensive (see the next section). A more promising direction is an adaptive implementation of a Schwarz alternating method, as proposed, e.g., in [84, 85],

**Algorithm 1.** *Multiplicative Schwarz iteration*

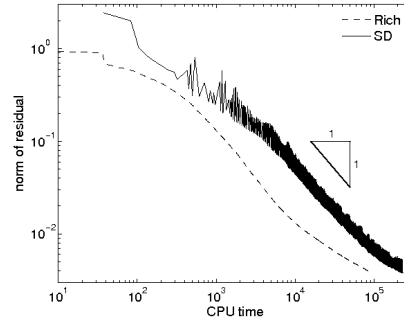
```

for  $k = 1, \dots,$ 
   $u_0^{k-1} = u^{k-1}$ 
  for  $i = 1, \dots, n$ 
     $u_i^{k-1} = u_{i-1}^{k-1} + Q_i^T \tilde{L}_i^{-1} Q_i (f - Lu_{i-1}^{k-1})$ 
  endfor
   $u^k = u_n^{k-1}$ 
endfor

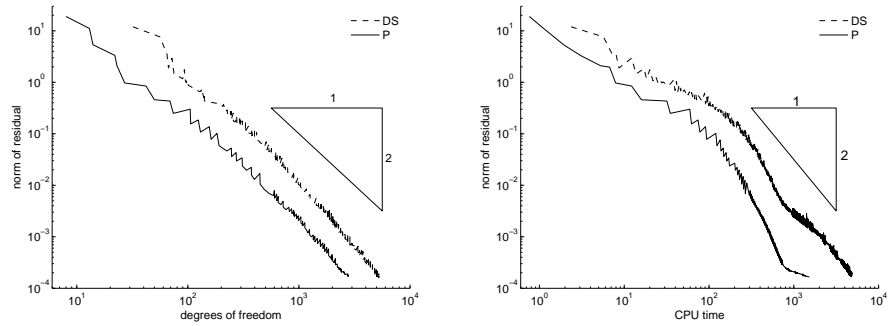
```



**Fig. 6.** Approximations and corresponding pointwise errors produced by the adaptive steepest descent algorithm, using piecewise quadratic frame elements. *Upper part:* Approximations with 167, 898, and 2351 frame elements. *Lower part:* Approximations with 2934, 3532, and 4648 frame elements.



**Fig. 7.** Convergence histories of **SOLVE** and **CDD2SOLVE** with respect to CPU time using piecewise quadratic frame elements ( $d = 3, \tilde{d} = 5$ ).



**Fig. 8.** Convergence histories of the adaptive steepest descent method with respect to the support size of the iterates (*left column*) or CPU time (*right column*) for  $d = \tilde{d} = 3$ . The algorithm has been tested with aggregated frames based on interval bases from [3] (solid line) and [37] (dashed line).

**Algorithm 2.** *Additive Schwarz iteration*

**for**  $k = 1, \dots,$   
 $u^k = u^{k-1} + \alpha \sum_{i=1}^n Q_i^T \tilde{L}_i^{-1} Q_i (f - Lu^{k-1})$   
**endfor**

where  $0 < \alpha \leq 1$  is a suitable damping parameter and  $Q_i$  are suitable projections onto coordinate subspaces and  $\tilde{L}_i^{-1}$  is an approximation of a local inverse of the operator  $L$ .

A recent exploration of this idea [94] confirms that an adaptive domain decomposition strategy successfully promotes sparser solutions, hence it is computationally far more advantageous, and it is also proven to be optimal. In particular, the wavelet coefficients activated in one patch contribute to the datum of the problem to be solved on the other patch, and are not going to re-activated again there. Presently, this is the best approach to adaptive numerical solution of elliptic PDEs by means of wavelet discretizations.

### 3 Sparse Recovery and Variational Problems

#### 3.1 $\ell_1$ -minimization and iterative soft-thresholding algorithms

The minimization of the functional

$$\mathcal{J}(u) := \|Lu - f\|_{\ell_2(A)}^2 + 2\gamma \|u\|_{\ell_1(A)}, \quad \gamma > 0 \tag{6}$$

proved to be an extremely efficient alternative to the well-known Tikhonov regularization [61], whenever

$$Lu = f,$$

is an ill-posed problem on  $\ell_2(A)$  and the solution  $u$  is expected to be a vector with a moderate number of nonzero entries. Indeed, the imposition of the  $\ell_1$ -constraint does promote a sparse solution. The restriction to problems on  $\ell_2(A)$  is just formal, this can always be achieved by discretization via a suitable frame for the underlying Hilbert space, as already shown in the previous section. The use of the  $\ell_1$ -norm as a sparsity-promoting functional can be found first in reflection seismology and in deconvolution of seismic traces [22, 87, 96]. In the last decade more understanding of the deep motivations why  $\ell_1$ -minimization tends to promote sparse recovery was developed. Rigorous results began to appear in the late-1980's, with Donoho and Stark [57] and Donoho and Logan [56]. Applications of  $\ell_1$ -minimization in statistical estimation began in the mid-1990's with the introduction of LASSO, which stands for Least Absolute Shrinkage and Selection Operator, cf. [97] for details<sup>1</sup>. In signal processing, Basis Pursuit [20] was proposed in compression applications for extracting the sparsest signal

<sup>1</sup> <http://www-stat.stanford.edu/~tibs/lasso.html>

representation from highly overcomplete frames. From these early steps the applications and understanding of compressive  $\ell_1$ -minimization have continued to increase dramatically. It is now hard to trace all the relevant results and applications [8, 13, 10, 52, 53], and it is beyond the scope of this short summary<sup>2</sup>. In fact,  $\ell_1$ -minimization has been so surprisingly effective in several applications, that Candès, Wakin, and Boyd call it the “modern least squares” in [14]. We thus clearly need efficient algorithms for the minimization of  $\mathcal{J}$ .

An iterative thresholding algorithm, realized by Richardson-Landweber steps followed by a *soft-thresholding*  $H_\gamma$  (Figure 1), was proposed for this task [27, 42, 46, 88, 97],

$$u^{(n+1)} = H_\gamma \left( u^{(n)} + L^*(f - Lu^{(n)}) \right), \quad n \in \mathbb{N}, \quad (7)$$

with a *fixed* threshold parameter  $\gamma > 0$ .

Unfortunately, despite its simplicity which makes it very attractive to users, this algorithm does not perform very well, as it has been recently verified systematically, e.g., in [77] where a comparison with several other alternative methods [4, 44, 62, 76] is carefully provided. However, since the mentioned methods are formulated as a sequential algorithm, none of them is able to address in real-time, or at least in an acceptable computational time, extremely large problems, such as 4D imaging (spatial plus temporal dimensions) from functional magnetic-resonance in nuclear medical imaging, astronomical imaging or global terrestrial seismic tomography. For this reason, and parallel to the development of very successful approaches for well-posed problems [94], a “domain decomposition” algorithm was proposed in [64] together with its parallelization, and generalizations to general subspace corrections appear in [69].

We briefly illustrate below the general setting of the results in [44, 64, 69].

### 3.2 Accelerated projected gradient methods

A concrete recipe to identify adaptively *good* shape parameters  $\gamma = (\gamma_n)_n$  in iterative thresholding algorithms is by construction of a suitable convex set  $K$  and a projection map  $P_K$  such that  $P_K(u) = H_\gamma(u)$  for an adaptive  $\gamma := \gamma(K, u)$ . In the case of the specific functional (6) the set  $K$  is an  $\ell_1$ -ball, while  $H_{\gamma_n}$  is the soft-thresholding operator. This leads to the following accelerated projected gradient/steepest descent iteration, that turns out to be in several situations [77] much faster than (7):

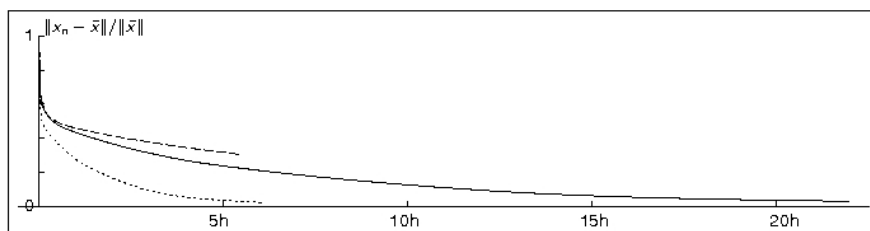
$$\begin{cases} u^{(n+1)} = P_K(u^{(n)} + L_n^*(f - L_n u^{(n)})), \\ L_n \rightarrow L, \end{cases} \quad (8)$$

Note that, by definition of  $P_K$ , this iteration corresponds to a compressive algorithm where  $H_{\gamma_n}$  is adapted at each iteration.

<sup>2</sup> The reader can also find a sufficiently comprehensive collection of the ongoing recent developments at the web-site <http://www.compressedensing.com/>.

**Theorem 1.** *The sequence  $(u^{(n)})_{n \in \mathbb{N}}$  as defined in (8), where the operator approximation  $L_n^* L_n$  is possibly chosen according to a suitable steepest-descent criterion, converges in norm to a minimizer of  $\mathcal{D}(u) = \|Lu - f\|_{\ell_2(\Lambda)}^2$  on  $K$ .*

For the theoretical analysis and a discussions on the performances of this algorithm we refer the reader to [44]. Let us just include an example inspired by a



**Fig. 9.** The different convergence rates of the thresholded Landweber algorithm (solid line), the projected Landweber algorithm (dashed line), where no adaptive descent parameters are used, and the projected steepest descent algorithm (dotted line), for the third example. The projected steepest descent algorithm converges about four times faster than the thresholded Landweber iteration. The projected Landweber iteration does better at first (not visible in this plot), but loses with respect to iterative thresholding afterwards. The horizontal axis has time (in hours), the vertical axis displays the relative error.

real-life application in geoscience [78], in particular an application in seismic tomography based on earthquake data. The solution space consists of the wavelet coefficients of a 2D seismic velocity perturbation. There are 8192 degrees of freedom. In this particular case the number of data is 1848. Hence the matrix  $L$  has 1848 rows and 8192 columns. We apply the different methods to the same noisy data that are used in [78] and measure the time to convergence up to a specified relative error (see Figure 9). This example illustrates the slow convergence of the thresholded Landweber algorithm (7), and the improvements made by a projected steepest descent iteration (8).

### 3.3 Domain decomposition methods

A different strategy in order to deal with very large problems is based on suitable domain decomposition methods, as analyzed in [64]. By splitting disjointly the index set  $\Lambda = \Lambda_1 \cup \Lambda_2$ , alternating (sequential and parallel) algorithms can be

formulated:

$$\begin{cases} u_{\Lambda_1}^{(n+1,\ell+1)} = H_\gamma(u_{\Lambda_1}^{(n+1,\ell)} + L_{n,\Lambda_1}^*((f - L_{n,\Lambda_2}u_{\Lambda_2}^{(n,M)}) - L_{n,\Lambda_1}u_{\Lambda_1}^{(n+1,\ell)})), \\ \ell = 0, \dots, L-1, \\ u_{\Lambda_2}^{(n+1,m+1)} = H_\gamma(u_{\Lambda_2}^{(n+1,m)} + L_{n,\Lambda_2}^*((f - L_{n,\Lambda_1}u_{\Lambda_1}^{(n+1,L)}) - L_{n,\Lambda_2}u_{\Lambda_2}^{(n+1,m)})), \\ m = 0, \dots, M-1. \\ u^{(n+1)} = u_{\Lambda_1}^{(n+1,L)} + u_{\Lambda_2}^{(n+1,M)}. \end{cases} \quad (9)$$

Here we denoted  $u_{\Lambda_i}$  any vector supported on  $\Lambda_i$ , and with  $L_{n,\Lambda_i}$  the submatrix of  $L_n$  where only the columns corresponding to entries in  $\Lambda_i$  are considered. The latter algorithm is derived as an instance of the following alternating minimization: Pick an initial  $u_{\Lambda_1}^{(0,L)} + u_{\Lambda_2}^{(0,M)} := u^{(0)} \in \ell_2(\Lambda)$ , for example  $u^{(0)} = 0$ , and iterate

$$\begin{cases} u_{\Lambda_1}^{(n+1,L)} \approx \arg \min_{\text{supp}(v_1) \subset \Lambda_1} \mathcal{J}(v_1 + u_{\Lambda_2}^{(n,M)}) \\ u_{\Lambda_2}^{(n+1,M)} \approx \arg \min_{\text{supp}(v_2) \subset \Lambda_2} \mathcal{J}(u_{\Lambda_1}^{(n+1,L)} + v_2) \\ u^{(n+1)} := u_{\Lambda_1}^{(n+1,L)} + u_{\Lambda_2}^{(n+1,M)}. \end{cases}$$

**Theorem 2.** *The algorithm in (9) produces a sequence  $(u^{(n)})_{n \in \mathbb{N}}$  in  $\ell_2(\Lambda)$  whose strong accumulation points are minimizers of the functional  $\mathcal{J}$ . In particular, the set of strong accumulation points is non-empty.*

### 3.4 Subspace correction methods

Actually, the functional (6) is the prototype model of more general problems, that we shall now describe. Let  $\mathcal{H}$  be a real separable Hilbert space. We are interested in the numerical minimization in  $\mathcal{H}$  of the general form of functionals

$$\mathcal{J}(u) := \|Lu - f\|_{\mathcal{H}}^2 + 2\gamma\psi(u), \quad (10)$$

where  $L \in \mathcal{L}(\mathcal{H})$  is a bounded linear operator,  $f \in \mathcal{H}$  is a datum, and  $\gamma > 0$  is a fixed constant. The function  $\psi : \mathcal{H} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  is a semi-norm for a suitable subspace  $\mathcal{H}^\psi$  of  $\mathcal{H}$ . An example of this setting is certainly (6). In particular, we investigate splittings into arbitrary orthogonal subspaces  $\mathcal{H} = V_1 \oplus V_2$  for which we may have

$$\psi(\pi_{V_1}(u) + \pi_{V_2}(v)) \neq \psi(\pi_{V_1}(u)) + \psi(\pi_{V_2}(v)), \quad u, v \in \mathcal{H},$$

where  $\pi_{V_i}$  is the orthogonal projection onto  $V_i$ . With this splitting we can address the minimization of  $\mathcal{J}$  by suitable instances of the following alternating algorithm: Pick an initial  $V_1 \oplus V_2 \ni u_1^{(0)} + u_2^{(0)} := u^{(0)} \in \mathcal{H}^\psi$ , for example  $u^{(0)} = 0$ , and iterate

$$\begin{cases} u_1^{(n+1)} \approx \arg \min_{v_1 \in V_1} \mathcal{J}(v_1 + u_2^{(n)}) \\ u_2^{(n+1)} \approx \arg \min_{v_2 \in V_2} \mathcal{J}(u_1^{(n+1)} + v_2) \\ u^{(n+1)} := u_1^{(n+1)} + u_2^{(n+1)}. \end{cases}$$

This algorithm is implemented by solving the subspace minimizations via an *oblique thresholding* iteration, which is defined by means of Lagrange multipliers.

The attribute “*oblique*” emphasizes the presence of a fixed additional subspace that contributes to the computation of the thresholded solution. In [69] we provide a detailed analysis of the convergence properties of this sequential algorithm and of its modification for parallel computation. Extensions to nonorthogonal decompositions  $\mathcal{H} = V_1 + V_2$  are reported in [65]<sup>3</sup>.

We want to mention here a special application of the results above towards domain decomposition methods for total variation minimization and the solution of its associated degenerate elliptic PDE’s. As we already shown in the previous section dedicated to the adaptive solution of well-posed elliptic problems, domain decomposition methods were introduced as techniques for solving partial differential equations based on a decomposition of the spatial domain of the problem into several subdomains. The initial equation restricted to the subdomains defines a sequence of new local problems. The main goal is to solve the initial equation via the solution of the local problems. This procedure induces a dimension reduction which is the major responsible of the success of such a method. Indeed, one of the principal motivations is the formulation of solvers which can be easily parallelized.

We can apply the theory and the algorithms described above to adapt domain decompositions to the minimization of functionals with total variation constraints. In this case the interesting solutions are usually discontinuous, e.g., along curves in 2D. These discontinuities may cross the interfaces of the domain decomposition patches. Hence, the crucial difficulty is the correct treatment of interfaces, with the preservation of crossing discontinuities and the correct matching where the solution is continuous instead. We consider the minimization of the functional  $\mathcal{J}$  in the following different setting: Let  $\Omega \subset \mathbb{R}^n$ , for  $n = 1, 2$ , be a bounded open set with Lipschitz boundary. We are interested in the case when  $\mathcal{H} = L^2(\Omega)$ ,  $\mathcal{H}^\psi = BV(\Omega)$  and  $\psi(u) = |Du|(\Omega)$ , the variation of  $u$ . Then a nonoverlapping domain decomposition  $\Omega = \Omega_1 \cup \Omega_2$  induces the space splitting into  $V_i := \{u \in L^2(\Omega) : \text{supp}(u) \subset \Omega_i\}$ ,  $i = 1, 2$ . For overlapping domain decomposition methods in this setting, we refer the reader to [65]. Hence, by means of the proposed alternating algorithm, we can address the minimization of the functional

$$\mathcal{J}(u) := \|Tu - g\|_{L^2(\Omega)}^2 + 2\alpha|Du|(\Omega).$$

It is important to mention that several numerical strategies to perform efficiently total variation minimization have been proposed in the literature. We list a few of the relevant ones, ordered by their chronological appearance:

(i) the approach of Chambolle and Lions [18] by re-weighted least squares, see also [43] for generalizations and refinements in the context of compressed sensing;

(ii) variational approximation via local quadratic functionals as in the work of Vese et al. [98, 3];

---

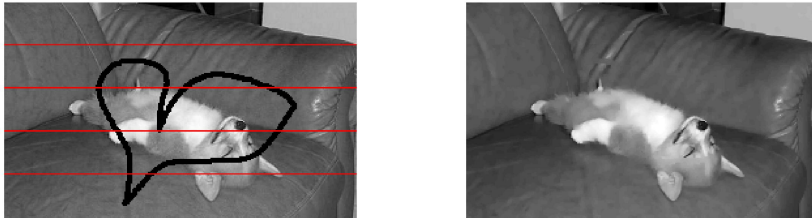
<sup>3</sup> Matlab software and the numerical experiments are provided at [http://homepage.univie.ac.at/carola.schoenlieb/webpage\\_tv/dode/tv\\_dode\\_numerics.htm](http://homepage.univie.ac.at/carola.schoenlieb/webpage_tv/dode/tv_dode_numerics.htm).

(iii) iterative thresholding algorithms based on projections onto convex sets as in the work of Chambolle [16] as well as in the work of Combettes-Wajs [27] and Daubechies et al. [46];

(iv) iterative minimization of the Bregman distance as in the work of Osher et al. [83];

(v) the approach proposed by Nesterov [82] and its modifications by Weiss et al. [99].

These approaches differ significantly, and it seems that the ones collected in the groups iv) and v) do show presently the best performances in practice. However, being formulated as sequential algorithms, none of the mentioned methods is able to address in real-time, or at least in an acceptable computational time, extremely large problems, and domain decomposition strategies are fundamental in such cases.



**Fig. 10.** An example of the solution of *total variation inpainting* (a nonlinear interpolation of an image with a missing part to be recovered) by means of the domain decomposition method [69]. The problem is split into 5 subdomains and it is correctly solved, providing a global minimization.

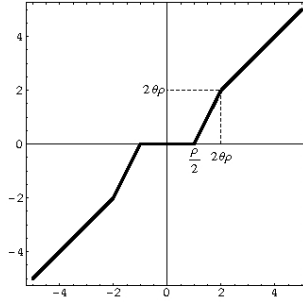
### 3.5 Joint sparsity and iterative *firm-thresholding* algorithms

Actually, total variation and  $\ell_1$ -constraints are of course not the only ways of promoting sparse solutions. In [67, 68] we investigated more general forms of functionals that not only generate different types of compressive algorithms, but also deal with vector-valued solutions with components coupled by identical sparsity patterns. *Joint sparsity* naturally occurs, for instance, in color images, where, e.g., the three color channels RGB can usually be well approximated by a jointly sparse wavelet or curvelet expansion since edges appear at the same locations throughout all channels. However, the range of applicability of this approach is not limited to color image restoration. Multimodal brain imaging, distributed networks and compressed sensing, and several other problems with

coupled vector valued solutions are fields where one can expect fruitful applications. In this more general context we are given a Hilbert space  $\mathcal{H}$ , and we assume that  $L$  is a linear operator mapping  $\ell_2(\Lambda)^M$  into  $\mathcal{H}$  where  $M \in \mathbb{N}$  indicates the number of vector components of the solution. We consider the functional

$$J(u, v) = J_{\theta, \rho, \omega}^{(q)}(u, v) := \|Lu - f\|_{\mathcal{H}}^2 + \sum_{\lambda \in \Lambda} v_\lambda \|u_\lambda\|_q + \sum_{\lambda \in \Lambda} \omega_\lambda \|u_\lambda\|_2^2 + \sum_{\lambda \in \Lambda} \theta_\lambda (\rho_\lambda - v_\lambda)^2, \quad (11)$$

where  $\|\cdot\|_q$  denotes the usual  $\ell_q$ -norm on  $\mathbb{R}^M$ ,  $q \in [1, \infty]$  and  $\theta = (\theta_\lambda)$ ,  $\omega = (\omega_\lambda)$ ,  $\rho = (\rho_\lambda)$  are suitable sequences of positive parameters. The variable  $u$  is assumed to be in  $\ell_2(\Lambda)^M$  and  $v_\lambda \geq 0$  for all  $\lambda \in \Lambda$ . Observe, that  $u_\lambda$  is a vector in  $\mathbb{R}^M$ .



**Fig. 11.** For  $M = 1$  and  $\omega = 0$ , the thresholding function  $H_{\theta, \rho, \omega}^{(q)}$  acts componentwise applying the firm-thresholding curve here depicted.

We are interested in the joint minimizer  $(u^*, v^*)$  of this functional,  $u^*$  is then considered the optimal solution. The variable  $v$  is an auxiliary variable that plays the role of an indicator of the sparsity pattern. As argued in [67]  $J$  promotes *joint sparsity*, i.e.,  $u^* = (u^{*(\ell)})_{\ell=1}^M$  can be expected to be jointly sparse,  $\text{supp}(u^{*(\ell)}) \subset \Lambda_0$ , for all  $\ell = 1, \dots, M$ , and for a fixed  $\Lambda_0 \subset \Lambda$ ,  $\#\Lambda_0 < \infty$ . In [67] we proposed an iterative algorithm for computing the minimizer of  $J(u, v)$ . It consists of alternating a minimization with respect to  $u$  and  $v$ . More formally, for some initial choice  $v^{(0)}$ , for example  $v^{(0)} = (\rho_\lambda)_{\lambda \in \Lambda}$ , we define

$$\begin{aligned} u^{(n)} &:= \arg \min_{u \in \ell_2(\Lambda)^M} J(u, v^{(n-1)}), \\ v^{(n)} &:= \arg \min_{v \geq 0} J(u^{(n)}, v). \end{aligned} \quad (12)$$

Once again we see that an alternating minimization on several auxiliary variables plays a useful role in order to generate compressive algorithms. The minimizer  $v^{(n)}$  of  $J(u^{(n)}, v)$ , for a fixed  $u^{(n)}$ , can be computed explicitly by the formula

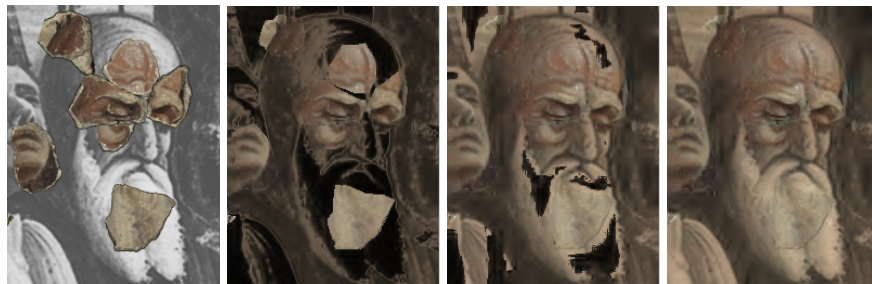
$$v_\lambda^{(n)} = \begin{cases} \rho_\lambda - \frac{1}{2\theta_\lambda} \|u_\lambda^{(n)}\|_q, & \|u_\lambda^{(n)}\|_q < 2\theta_\lambda \rho_\lambda, \\ 0, & \text{otherwise,} \end{cases} \quad \lambda \in \Lambda. \quad (13)$$

The minimization of  $J(u, v^{(n-1)})$  with respect to  $u$  and fixed  $v^{(n-1)}$  can be done by a thresholded Landweber iteration similar to the one analyzed in [42]. We showed in [67] that, for suitable choices of the parameters  $\theta, \rho, \omega$ , the algorithm (12) converges to the minimizer of the functional  $J$  and we show its linear convergence rate.

The functional  $J = J_{\theta, \rho, \omega}^{(q)}$  depends on several parameters. In the effort of clarifying their role in shaping the optimal solution  $u^*$  we discovered an intriguing relationship between our new functional (11) and hard-thresholding, and more generally to the so-called *firm-thresholding*. We proved that different choices of the parameters  $\theta = (\theta_\lambda), \omega = (\omega_\lambda)$  and  $\rho = (\rho_\lambda)$  do generate an entire family of thresholding algorithms, which perform the corresponding minimization by the iteration

$$u^{(n)} = H_{\theta, \rho, \omega}^{(q)}(u^{(n-1)} + L^*(f - Lu^{(n-1)})), \quad n \in \mathbb{N}. \quad (14)$$

Actually, for the simplest case  $M = 1$ , i.e., in the scalar situation, the triple  $\gamma = (\theta, \rho, \omega)$  very clearly defines a “shape parameter” since the form of the corresponding thresholding function  $H_\gamma := H_{\theta, \rho, \omega}^{(q)}$  (called *firm-thresholding*) changes accordingly, see Figure 11. In [68] we provided the proof of convergence of the latter algorithm to the minimizer  $u^*$ . Note that in this case, the auxiliary variable  $v$  does not play an explicit role anymore, although its presence implicitly contributes to the specific shape of the firm-thresholding function. By employing techniques of  $\Gamma$ -convergence and variational limits [40], we proved that the dependence of minimizers on the shape parameters  $\theta = (\theta_\lambda), \omega = (\omega_\lambda)$  is continuous.



**Fig. 12.** We illustrate a recolorization (three iterates of the joint-sparsity promoting algorithm (14)) of an image from a few color fragments and gray level information of the missing parts [66].

### 3.6 Compressive algorithms meet free-discontinuity problems

As there exists a natural correspondence between total variation and  $\ell_1$ -minimizations, there is also a corresponding *compressibility* in terms of derivatives to the one

promoted by firm-thresholding algorithms. As we discovered recently, functionals of the type (11) are discrete approximations of functionals modelling free-discontinuity problems. The terminology “free-discontinuity problems” was introduced by De Giorgi in the late-1980’s to indicate a class of variational problems that consist in the minimization of a functional, involving both volume and surface energies, depending on a closed set  $K$ , and a function  $u$  usually smooth outside of  $K$ : typically

- $K$  is not fixed a priori and it is an unknown of the problem;
- $K$  is not a boundary in general, but a free-surface inside the domain of the problem.

For a broad overview on free-discontinuity problems and their analysis, we refer to [2]. One of the best known examples of free-discontinuity problems is the one modelled by the Mumford-Shah functional [81] defined by

$$J(u, K) := \int_{\Omega \setminus K} [|\nabla u|^2 + \alpha(u - g)^2] dx + \beta \mathcal{H}_{n-1}(K \cap \Omega).$$

The set  $\Omega$  is a bounded open subset of  $\mathbb{R}^n$ ,  $\alpha, \beta > 0$  are fixed parameters and  $g \in L^\infty(\Omega)$ . Here  $\mathcal{H}_n$  denotes the  $n$ -dimensional Hausdorff measure. In this model, we seek for a function  $u \in W^{1,2}(\Omega \setminus K)$  that approximates the datum  $g$ , the function  $u$  is smooth out of the discontinuity set  $K$ . In visual analysis  $g$  is a given noisy image that we want to approximate by  $u$  which is instead smooth except for a rectifiable set  $K$ , the set  $K$  is also used in order to *segment* the image into connected components.

In fact, the Mumford-Shah functional is the continuous version of a previous discrete formulation of the image segmentation problem proposed by Geman and Geman in [71]; see also the work of Blake and Zisserman in [6]. Let us recall this discrete approach. For simplicity let  $n = 2$  (as for image processing problems),  $\Omega = [0, 1]^2$ , and let  $u_{i,j} = u(hi, hj)$ ,  $(i, j) \in \mathbb{Z}^2$  be a discrete function defined on  $\Omega_h := \Omega \cap h\mathbb{Z}^2$ , for  $h > 0$ . Define  $W_h(t) = \min\{t^2, \beta/h\}$  to be the truncated quadratic potential, and

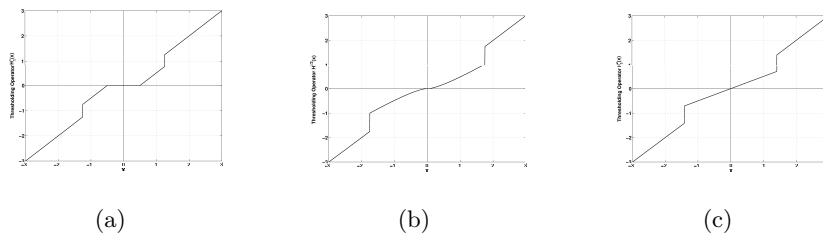
$$\begin{aligned} \mathcal{J}_{\sqrt{\beta/h}}(u) &:= h^2 \sum_{(hi, hj) \in \Omega_h} W_h \left( \frac{u_{i+1,j} - u_{i,j}}{h} \right) \\ &+ h^2 \sum_{(hi, hj) \in \Omega_h} W_h \left( \frac{u_{i,j+1} - u_{i,j}}{h} \right) \\ &+ \alpha h^2 \sum_{(hi, hj) \in \Omega_h} (u_{i,j} - g_{i,j})^2. \end{aligned}$$

Chambolle et al. [7, 19, 15] gave formal clarification as to how the discrete functional  $\mathcal{J}_{\sqrt{\beta/h}}$  approximates the continuous functional  $J$ . It has been pointed out in [70] that, by discretization of the Mumford-Shah functional by means of suitable finite elements, and then by expressing the problem in terms of sole

discrete derivatives, one can re-formulate the problem into a finite dimensional nonconvex and constrained optimization of the general type:

$$\begin{cases} \text{Minimize } \mathcal{J}_\gamma^p(u) = [\|Lu - f\|_{\ell_2^M}^2 + \sum_{i=1}^N \min\{|u_i|^p, \gamma^p\}] \\ \text{subject to } \mathcal{Q}u = 0. \end{cases} \quad (15)$$

where  $\mathcal{Q}$  is a suitable linear constraint,  $\gamma > 0$ , and  $1 \leq p \leq 2$ . In [70] we remarkably proved that these nonconvex functionals have always minimizers independently of the particular choice of  $L$ . Note that this result is not obvious since the functional is not convex, and for arbitrary choices of noninjective operators  $L$ , the problem does not seem *a priori* to be coercive. We also pointed out that the computation of these minimizers is an NP-hard problem [1].



**Fig. 13.** The component-wise discontinuous thresholding functions  $H_{(1,1)}$ ,  $H_{(3/2,1)}$ , and  $H_{(2,1)}$ , with parameters  $p \in \{1, 3/2, 2\}$ , respectively, and  $\gamma = 1$ .

Associated to the unconstrained minimization of  $\mathcal{J}_r^p$  we designed an iterative thresholding algorithm

$$u^{(n+1)} = H_{(p,\gamma)}(u^{(n)} + L^*(f - Lu^{(n)})), \quad (16)$$

where  $H_{(p,r)}$  is again a *thresholding function* which acts component-wise as depicted in Figure 13 for  $p \in \{1, 3/2, 2\}$ . This algorithm converges to fixed points  $\bar{u}$  of the iteration (16).

**Theorem 3.** *A fixed point  $\bar{u}$  of the iteration (16) is a local minimizer of the functional  $\mathcal{J}_\gamma^p$  defined in (15).*

More surprising is that global minimizers of  $\mathcal{J}_\gamma^p$  are also fixed points, as shown in the following theorem.

**Theorem 4.** *Any global minimizer  $u^*$  of  $\mathcal{J}_\gamma^p$  is a fixed point of the iteration (13).*

We reiterate that the computation of global minimizers is an NP-hard problem, and therefore the algorithm (16) is of particular importance because it furnishes a method to approximate local minimizer, but also it has the chance of finding a global one. We note that on a ball  $B(\bar{u}, \varepsilon(\gamma))$  around an equilibrium point  $\bar{u}$  of radius  $\varepsilon(\gamma) > 0$  sufficiently small, the functional  $\mathcal{J}_\gamma^p$  is convex; it is possible to show that  $\mathcal{J}_\gamma^p$  is in fact *strictly* convex whenever  $\bar{u} = u^*$  is a global minimizer. Hence a global minimizer is necessarily an isolated minimizer, whereas we cannot ensure the same property for local minimizers if  $L$  has a nontrivial null-space; in this case, local minimizers may form continuous sets. We conclude the following remark.

**Corollary 1.** *Minimizers of  $\mathcal{J}_\gamma^p$  are isolated.*

## 4 Conclusion and future perspectives

We gave an overview of a few significant instances of iterative compressive algorithms which show their versatility in several and diverse applications. We ranged through optimal adaptive solution of PDEs, inverse problems with sparsity constraints, and free-discontinuity problems. We wanted to emphasize the capability of suitable implementations of compressive algorithms, e.g., via domain decomposition and subspace correction methods, of performing effective dimensionality reductions, also in problems of very large dimension. We stressed the variational nature of such algorithms that usually dispose of an associated energy functional which is minimized during the iterations, and provides additional robustness to the iterative process.

We expect that future developments of numerical methods, for instance, for nonlinear PDEs [95], will continue to be influenced by such approaches: new tools to cope with the “curse of dimensionality”, further systematic developments of adaptivity in the presence of different scales, probabilistic algorithms, an increasing role for combinatorial aspects of the underlying algorithms, a few examples expected from future developments.

## Acknowledgment

The author acknowledges the support of the FWF project no. Y-432-N15 START-award “Sparse Approximation and Optimization in High Dimensions”.

## References

1. B. Alexeev and R. Ward, *On the complexity of free-discontinuity and  $\ell_0$  regularization in image processing*, preprint, (2009)

2. L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems.*, Oxford Mathematical Monographs. Oxford: Clarendon Press. xviii, 2000.
3. G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variation*, Springer, 2002.
4. 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.
5. P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates.*, Numer. Math. **97** (2004), no. 2, 219–268.
6. A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, 1987.
7. B. Bourdin and A. Chambolle. Implementation of an adaptive finite-element approximation of the Mumford-Shah functional. *Numer. Math.*, 85(4):609–646, 2000.
8. E. J. Candès, *Compressive sampling*, International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 1433–1452.
9. E. J. Candès, M. Wakin, and S. Boyd, *Enhancing sparsity by reweighted  $\ell_1$  minimization*, (Technical Report, California Institute of Technology).
10. 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.
11. E. Candès and T. Tao, *Near optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425.
12. 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.
13. E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4203–4215.
14. E. J. Candès, M. Wakin, and S. Boyd, *Enhancing sparsity by reweighted  $l_1$  minimization*, (Technical Report, California Institute of Technology).
15. A. Chambolle, *Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations.*, SIAM J. Appl. Math. **55** (1995), no. 3, 827–863.
16. A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision **20** (2004), no. 1-2, 89–97.
17. 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.
18. A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems.*, Numer. Math. **76** (1997), no. 2, 167–188.
19. A. Chambolle and G. Dal Maso, *Discrete approximation of the Mumford-Shah functional in dimension two*, M2AN Math. Model. Numer. Anal. **33** (1999), no. 4, 651–672.
20. S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. **1** (1998), 33–61, Reprint.
21. O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, 2003.
22. J. F. Claerbout and F. Muir, *Robust modeling with erratic data*, Geophysics **38** (1973), no. 5, 826–844.
23. A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations — Convergence rates*, Math. Comp. **70** (2001), 27–75.
24. A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods II: Beyond the elliptic case*, Found. Comput. Math. **2** (2002), no. 3, 203–245.

25. A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet schemes for nonlinear variational problems*, SIAM J. Numer. Anal. **41** (2003), no. 5, 1785–1823.
26. A. Cohen, M. Hoffmann, and M. Reiss, *Adaptive wavelet Galerkin methods for linear inverse problems.*, SIAM J. Numer. Anal. **42** (2004), no. 4, 1479–1501.
27. P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul. **4** (2005), no. 4, 1168–1200.
28. S. Dahlke, *Besov regularity for elliptic boundary value problems on polygonal domains*, Applied Mathematics Letters **12** (1999), 31–38.
29. S. Dahlke, W. Dahmen, and R. DeVore, *Nonlinear approximation and adaptive techniques for solving elliptic operator equations*, Multiscale Wavelet Methods for Partial Differential Equations (W. Dahmen, A. Kurdila, and P. Oswald, eds.), Academic Press, San Diego, 1997, pp. 237–283.
30. S. Dahlke, W. Dahmen, R. Hochmuth, and R. Schneider, *Stable multiscale bases and local error estimation for elliptic problems*, Appl. Numer. Math. **23** (1997), 21–48.
31. S. Dahlke, W. Dahmen, and K. Urban, *Adaptive wavelet methods for saddle point problems — Optimal convergence rates*, SIAM J. Numer. Anal. **40** (2002), no. 4, 1230–1262.
32. S. Dahlke and R. DeVore, *Besov regularity for elliptic boundary value problems*, Commun. Partial Differ. Equations **22** (1997), no. 1&2, 1–16.
33. S. Dahlke, M. Fornasier, M. Primbs, T. Raasch, and M. Werner, *Nonlinear and adaptive frame approximation schemes for elliptic pdes: Theory and numerical experiments*, to appear in Numerical Methods for Partial Differential Equations (2009).
34. S. Dahlke, M. Fornasier, and T. Raasch, *Adaptive frame methods for elliptic operator equations*, Adv. Comput. Math. **27** (2007), no. 1, 27–63.
35. 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.
36. S. Dahlke, R. Hochmuth, and K. Urban, *Adaptive wavelet methods for saddle point problems*, M2AN Math. Model. Numer. Anal. **34** (2000), 1003–1022.
37. W. Dahmen and R. Schneider. *Wavelets with complementary boundary conditions — Function spaces on the cube*. *Result. Math.*, 34(3–4):255–293, 1998.
38. W. Dahmen and R. Schneider. *Composite wavelet bases for operator equations*. *Math. Comp.*, 68:1533–1567, 1999.
39. W. Dahmen and R. Schneider. *Wavelets on manifolds I. Construction and domain decomposition*. *SIAM J. Math. Anal.*, 31:184–230, 1999.
40. G. Dal Maso, *An Introduction to  $\Gamma$ -Convergence*, Birkhäuser, Boston, 1993.
41. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.
42. 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.
43. I. Daubechies, R. DeVore, M. Fornasier, and S. Güntürk, *Iteratively re-weighted least squares minimization for sparse recovery*, to appear in Commun. Pure Appl. Math. (2009), arXiv:0807.0575
44. 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.
45. I. Daubechies and G. Teschke, *Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising.*, Appl. Comput. Harmon. Anal. **19** (2005), no. 1, 1–16.
46. I. Daubechies, G. Teschke, and L. Vese, *Iteratively solving linear inverse problems under general convex constraints*, Inverse Problems and Imaging **1** (2007), 29–46.

47. L. Daudet and B. Torr sani, *Hybrid representations for audiophonic signal encoding*, Signal Processing **82** (2002), no. 11, 1595–1617.
48. R. DeVore, *Nonlinear approximation*, Acta Numerica **7** (1998), 51–150.
49. D. L. Donoho, *Superresolution via sparsity constraints.*, SIAM J. Math. Anal. **23** (1992), no. 5, 1309–1331.
50. D. L. Donoho, *De-noising by soft-thresholding.*, IEEE Trans. Inf. Theory **41** (1995), no. 3, 613–627.
51. D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition.*, Appl. Comput. Harmon. Anal. **2** (1995), no. 2, 101–126.
52. D. L. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory **52** (2006), no. 4, 1289–1306.
53. D. L. Donoho, *High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension*, Discrete Comput. Geom. **35** (2006), no. 4, 617–652.
54. D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81** (1994), no. 3, 425–455.
55. D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, *Wavelet shrinkage: asymptopia?*, J. Roy. Statist. Soc. Ser. B **57** (1995), no. 2, 301–369.
56. D. L. Donoho and B. F. Logan, *Signal recovery and the large sieve*, SIAM J. Appl. Math. **52** (1992), no. 2, 557–591.
57. D. L. Donoho and P. B. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. **49** (1989), no. 3, 906–931.
58. D. L. Donoho and J. Tanner, *Sparse nonnegative solutions of underdetermined linear equations by linear programming*, Proc. Nat. Acad. Sci. **102** (2005), no. 27, 9446–9451.
59. D. L. Donoho and J. Tanner, *Counting faces of randomly-projected polytopes when the projection radically lowers dimension*, preprint (2006).
60. R. J. Duffin and A. C. Schaeffer, *A class of nonharmonic Fourier series*, Trans. Amer. Math. Soc. **72** (1952), 341–366.
61. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems.*, Mathematics and its Applications (Dordrecht). 375. Dordrecht: Kluwer Academic Publishers., 1996.
62. 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.
63. M. Fornasier, *Quasi-orthogonal decompositions of structured frames*, J. Math. Anal. Appl. **289** (2004), no. 1, 180–199.
64. M. Fornasier, *Domain decomposition methods for linear inverse problems with sparsity constraints*, Inverse Problems **23** (2007), 2505–2526.
65. M. Fornasier, A. Langer, and C. B. Sch nlieb, *Domain decomposition methods for compressed sensing*, Proceedings of SampTA09, Marseilles, (2009).
66. M. Fornasier, R. Ramlau, and G. Teschke, *The application of joint sparsity and total variation minimization algorithms to a real-life art restoration problem*, to appear in Adv. Comput. Math., (2009)
67. M. Fornasier and H. Rauhut, *Recovery algorithms for vector valued data with joint sparsity constraints*, SIAM J. Numer. Anal. **46** (2008), no. 2, 577–613.
68. M. Fornasier and H. Rauhut, *Iterative thresholding algorithms*, Appl. Comput. Harmon. Anal. Appl. Comput. Harmon. Anal., **25** (2008), no. 2, 187–208.
69. M. Fornasier and C. B. Sch nlieb, *Subspace correction methods for total variation and  $\ell_1$ -minimization*, to appear in SIAM J. Numer. Anal. (2009).
70. M. Fornasier and R. Ward, *Iterative thresholding meets free-discontinuity problems*, arXiv:0901.2605 (2009).

71. S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.*, IEEE Trans. Pattern Anal. Mach. Intell **6** (1984), 721–741.
72. A. C. Gilbert, M. Strauss, J. A. Tropp, and R. Vershynin, *One sketch for all: Fast algorithms for compressed sensing*, preprint (2006).
73. A. C. Gilbert and J. A. Tropp, *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Trans. Inform. Theory, **53** (2007), no. 12, 4655–4666.
74. R. Gribonval and M. Nielsen, *Highly sparse representations from dictionaries are unique and independent of the sparseness measure*, Appl. Comput. Harmon. Anal. **22** (2007), no. 3, 335–355.
75. K. Gröchenig, *Foundations of Time–Frequency Analysis*, Birkhäuser, 2000.
76. S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, *A method for large-scale  $\ell_1$ -regularized least squares problems with applications in signal processing and statistics*, IEEE Journal on Selected Topics in Signal Process. (2007).
77. I. Loris, *On the performance of algorithms for the minimization of  $\ell_1$ -penalized functionals*, Inverse Problems **25** (2009), no. 3, 035008.
78. I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen, *Tomographic inversion using  $\ell_1$ -norm regularization of wavelet coefficients*, Geophys. J. Int. **170** (2007), 359–370.
79. S. Mallat, *A Wavelet Tour of Signal Processing. 2nd Ed.*, San Diego, CA: Academic Press., 1999.
80. P. Morin, R. H. Nochetto, and K. G. Siebert, *Convergence of adaptive finite element methods*, SIAM Rev. **44** (2002), 631–658.
81. D. Mumford and K. Shah, *Optimal approximation by piecewise smooth functions and associated variational problems*, Communications on Pure and Applied Mathematics **42** (1989), 577–684.
82. Y. Nesterov, *Smooth minimization of non-smooth functions*. Mathematic Programming, Ser. A, **103** (2005), 127–152.
83. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An Iterative Regularization Method for Total Variation-Based Image Restoration*, Multiscale Model. Simul. **4**, no. 2 (2005) 460–489.
84. P. Oswald, *Frames and space splittings in Hilbert spaces*, Survey lectures on multilevel schemes for elliptic problems in Sobolev spaces. <http://www.faculty.iu-bremen.de/poswald/bonn1.pdf>, 1997.
85. P. Oswald, *Multilevel frames and Riesz bases in Sobolev spaces*, Survey lectures on multilevel schemes for elliptic problems in Sobolev spaces <http://www.faculty.iu-bremen.de/poswald/bonn2.pdf>, 1997.
86. L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms.*, Physica D **60** (1992), no. 1-4, 259–268.
87. F. Santosa and W. W. Symes, *Linear inversion of band-limited reflection seismograms*, SIAM J. Sci. Stat. Comput. **7** (1986), no. 4, 1307–1330.
88. J.-L. Starck, E. J. Candès, and D. L. Donoho, *Astronomical image representation by curvelet transform*, Astronomy and Astrophysics **298** (2003), 785–800.
89. J.-L. Starck, M. K. Nguyen, and F. Murtagh, *Wavelets and curvelets for image deconvolution: a combined approach*, Signal Proc. **83** (2003), 2279–2283.
90. R. Stevenson, *Adaptive solution of operator equations using wavelet frames*, SIAM J. Numer. Anal. **41** (2003), no. 3, 1074–1100.
91. R. Stevenson, *An optimal adaptive finite element method*, SIAM J. Numer. Anal., **42** (2005), no. 5, 2188–2217.
92. R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. **7** (2007), no. 2, 245–269.

93. R. Stevenson and M. Werner, *Computation of differential operators in aggregated wavelet frame coordinates*, IMA J. Numer. Anal. 28(2), 354-381 (2008)
94. R. Stevenson and M. Werner, *A multiplicative Schwarz adaptive wavelet method for elliptic boundary value problems*, Math. Comput. **78** (2009), no. 266, 619-644.
95. E. Tadmor, *Numerical methods for nonlinear partial differential equations*, preprint (2009).
96. H. L. Taylor, S. C. Banks, and J. F. McCoy, *Deconvolution with the  $\ell_1$  norm*, Geophysics **44** (1979), no. 1, 39-52.
97. R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B **58** (1996), no. 1, 267-288.
98. L. Vese, *A study in the BV space of a denoising-deblurring variational problem.*, Appl. Math. Optim. **44** (2001), 131-161.
99. P. Weiss, L. Blanc-Féraud, and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, to appear in SIAM J. Sci. Comput., (2009).
100. J. Zou, A. C. Gilbert, M. Strauss, and I. Daubechies, *Theoretical and experimental analysis of a randomized algorithm for sparse Fourier transform analysis*, J. Comput. Phys. **211** (2005), 572-595.