Domain Decomposition Methods for Linear Inverse Problems with Sparsity Constraints

Massimo Fornasier
Program in Applied and Computational Mathematics
Princeton University
Fine Hall, Washington Road
Princeton NJ 08544-1000, U.S.A.
email: mfornasi@math.princeton.edu

Abstract

Quantities of interest appearing in concrete applications often possess sparse expansions with respect to a preassigned frame. Recently, there were introduced sparsity measures which are typically constructed on the basis of weighted ℓ_1 norms of frame coefficients. One can model the reconstruction of a sparse vector from noisy linear measurements as the minimization of the functional defined by the sum of the discrepancy with respect to the data and the weighted ℓ_1-norm of suitable frame coefficients. Thresholded Landweber iterations were proposed for the solution of the variational problem. Despite of its simplicity which makes it very attractive to users, this algorithm converges slowly. In this paper we investigate methods to accelerate significantly the convergence. We introduce and analyze sequential and parallel iterative algorithms based on alternating subspace corrections for the solution of the linear inverse problem with sparsity constraints. We prove their norm convergence to minimizers of the functional. We compare the computational cost and the behavior of these new algorithms with respect to the thresholded Landweber iterations.

Keywords: linear inverse problems, sparsity, thresholded Landweber iterations, domain decomposition methods
AMS subject classification: 15A29, 65M55, 65K10, 90C25, 52A41, 49M30

1 Introduction: Linear Inverse Problems with Sparsity Constraints

Often in applications the quantity of interest is not given explicitly, but only indirect observations are furnished by measurements. Although complex phenomena are often governed by nonlinear rules, still the assumption of linear dependence of the observations on the quantity of interest covers many interesting problems and surprisingly works well also for certain nonlinear situations. In this paper we are concerned with linear inverse problems which are mathematically described as follows.

Let $\mathcal{K}$ and $\mathcal{H}$ be (separable) Hilbert spaces and $A : \mathcal{K} \to \mathcal{H}$ a bounded linear operator. Assume we are given (observations) data $g \in \mathcal{H}$,

$$g = Af.$$

Then our goal consists in reconstructing the (unknown) element $f \in \mathcal{K}$. We are interested in particular in the situation when the corresponding linear mapping from the vector f to the vector g is not invertible or ill-conditioned. Moreover, we may assume that the data g are corrupted by noise. Thus, in order to deal with our reconstruction problem a *regularization* is required [22].

Of course, with incomplete data (i.e., few measurements) and noise disturbance, it is impossible to recover f without imposing further constraints which help to shape the solution of the problem. Therefore our main assumption throughout this paper is that f is *sparse* with respect to a pre-assigned frame or basis (for the Hilbert space $\mathcal{K}$) [8]. Our aim is to model the sparsity constraint within a regularization term. Let us clarify what we do mean with sparsity.

We assume that we have given a suitable frame $\{\psi_\lambda : \lambda \in \Lambda\} \subset \mathcal{K}$ indexed by a countable set Λ. This means that there exist constants $c_1, c_2 > 0$ such that

$$c_1 \|f\|_{\mathcal{K}}^2 \le \sum_{\lambda \in \Lambda} |\langle f, \psi_\lambda \rangle|^2 \le c_2 \|f\|_{\mathcal{K}}^2 \qquad \text{for all } f \in \mathcal{K}. \tag{1.1}$$

Orthonormal bases are particular examples of frames. Frames allow for a (stable) series expansion of any $f \in \mathcal{K}$ of the form

$$f = Fu := \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda \tag{1.2}$$

where $u = (u_\lambda)_{\lambda \in \Lambda} \in \ell_2(\Lambda)$. The linear operator $F : \ell_2(\Lambda) \to \mathcal{K}$ is called the *synthesis map* in frame theory. It is bounded due to the frame inequality (1.1). In contrast to orthonormal bases, the coefficients u_λ need not be unique, in general. For more information on frames and their differences with respect to bases we refer to [8].

For f to be sparse with respect to the frame $\{\psi_\lambda\}$, we mean that f can be well-approximated by a series of the form (1.2) with only a small number of non-vanishing coefficients u_λ. Sparsity also means that only few information is conveyed by f. It is reasonable to expect that only few measurements, although incomplete to identify an arbitrary element in $\mathcal{K}$, might be sufficient to characterize and to reconstruct f.

It is now established, see for instance [4, 3, 18], that sparsity can be modelled as the sequence u being contained in $\ell_1(\Lambda)$. Indeed, the minimization of the $\ell_1(\Lambda)$ norm promotes that only few entries are non-zero.

On the basis of these considerations, several authors, e.g., [19, 23, 38, 39, 42], proposed independently the regularized functional

$$\mathcal{J}(u) = \|g - Tu\|_{\mathcal{H}}^2 + \tau \|u\|_1 = \|g - Tu\|_{\mathcal{H}}^2 + \tau \sum_{\lambda \in \Lambda} |u_\lambda|, \tag{1.3}$$

which has to be minimized with respect to the vector of coefficients $u = (u_\lambda)_{\lambda\in\Lambda}$. Here we have introduced the operator $T = A \circ F : \ell_2(\Lambda) \to \mathcal{H}$, which combines the frame synthesis map with the original model A. The ℓ_1 norm in this functional clearly represents the regularization term. Once the minimizer $u = (u_\lambda)$ is determined we obtain a reconstruction of the vectors of interest by means of $f = Fu = \sum_\lambda u_\lambda \psi_\lambda$. An *iterative thresholding algorithm* can be taken to perform the minimization with respect to u: Pick an initial $u^{(0)} \in \ell_2(\Lambda)$ ($u^{(0)} = 0$ is a good choice) and iterate

$$u^{(n+1)} = \mathbb{S}_\tau(u^{(n)} + T^*(g - Tu^{(n)})), \quad n \geq 0. \tag{1.4}$$

where $\mathbb{S}_\tau$ is the so-called *soft-thresholding operator*, which acts componentwise $\mathbb{S}_\tau v = (S_\tau v_\lambda)_{\lambda\in\Lambda}$ and defined by

$$S_\tau(x) = \begin{cases} x - \operatorname{sign}(x)\frac{\tau}{2}, & |x| > \frac{\tau}{2} \\ 0, & \text{otherwise.} \end{cases}$$

In [14] the algorithm in (1.4) was analyzed and the authors proved that it converges strongly to a minimizer u^* of the functional $\mathcal{J}$. The proof of this result is based on the application of the Opial's fixed point theorem [33] which implies the weak convergence, and on specific properties of the thresholding operator which allow to turn the weak convergence into strong. Besides the elegant mathematics needed for the convergence proof, one of the major advantages of this algorithm is its simplicity, also in terms of implementation. Indeed thresholding methods combined with wavelets have been often presented, e.g., in image processing, as a possible good alternative to total variation minimization [5] which requires instead the solution of a degenerate partial differential equation. See [17] for a recent comparison of these two methods. Unfortunately, no rate of convergence is ensured for the algorithm in (1.4). In practice, the algorithm converges relatively fast for few very initial iterations, but after this short transition, it starts dramatically to slow down. These effects are very well documented in the paper [15], see also [31] for a discussion in applications. In particular, in [15] an alternative approach is proposed towards *projected gradient methods* where the iteration (1.4) is substituted with

$$u^{(n+1)} = \mathbb{P}_R(u^{(n)} + \beta^{(n)} T^*(g - Tu^{(n)})), \quad n \geq 0. \tag{1.5}$$

where $\mathbb{P}_R$ is the projection onto the ℓ_1-ball of radius $R > 0$, and $\beta^{(n)} > 0$ are suitable descent parameters. Again, this latter algorithm converges strongly to a minimizer of $\mathcal{J}$ where $\tau = \tau(R)$ is chosen according to R. Indeed, in this case the convergence is much faster in practice. Nevertheless, as soon as the dimension of the problem is very large the computation of the projection $\mathbb{P}_R$ and of an optimal $\beta^{(n)}$ may result again computationally demanding. Since no convergence rate can again be theoretically ensured for this second algorithm, it is difficult to estimate the trade-off between computational cost and fast convergence. A further alternative is the introduction of a quadratic term for $\varepsilon > 0$

$$\mathcal{J}_\varepsilon(u) \; = \; \|g - Tu\|_{\mathcal{H}}^2 + \tau\|u\|_1 + \varepsilon\|u\|_2^2. \tag{1.6}$$

The minimizer u^ε of this functional can be computed by the following iterations:

$$u^{(n+1)} = \frac{1}{1+\varepsilon}\mathbb{S}_\tau(u^{(n)} + T^*(g - Tu^{(n)})), \quad n \geq 0. \tag{1.7}$$

In this case and for $\|T\| < 1$, the function $v \to \frac{1}{1+\varepsilon}\mathbb{S}_\tau(v + T^*(g - Tv))$ is a contraction, hence the iteration converges linearly to the unique minimizer of $\mathcal{J}_\varepsilon$. By Γ-convergence, see for instance [27] for precise statements, one can show that there exist sequences of minimizers u^ε which converge to a minimizer u^* of $\mathcal{J}$. Unfortunately it is not possible to assess the rate of convergence of this latter approximation. In the papers [26, 27] a very general family of iterative thresholding algorithms is analyzed for joint sparse and vector valued recovery, and their convergence properties are also discussed. Generalizations to nonlinear inverse problems appear in [35, 36, 41].

We emphasize the enormous impact of inverse problems with sparsity constraints in applications such as geophysics and image processing, e.g., brain and astronomical imaging [16, 21, 25, 31, 38, 39]. Moreover, it is also worth to stress the strong relations between iterations as in (1.4) and adaptive schemes for the solution of linear and nonlinear PDE as proposed in [9, 10, 12, 13, 40].

In this paper we want to study a new acceleration method of the basic iterative thresholding algorithm (1.4) by alternating subspace corrections determined by a suitable decomposition of the label set Λ. As we shall discuss in detail in this paper, the benefit from this approach is twofold:

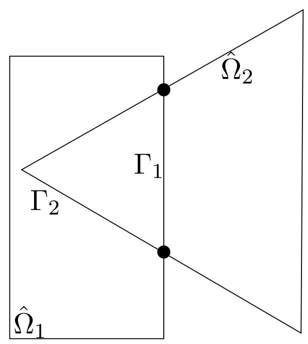
1. Instead of solving one large problem with many iteration steps, we can solve approximatively several smaller subproblems, which might lead to an acceleration of convergence and a reduction of the overall computational effort.

2. The subproblems do not need more sophisticated algorithms, simply reproduce at smaller dimension the original problem, and they can be solved in parallel.

The paper is organized as follows. In Section 2 we recall the main concepts related to domain decomposition methods for the solution of linear problems. In Section 3 we borrow these concepts for the sake of the minimization of the functional $\mathcal{J}$. We illustrate how to split the problem into two lower dimensional problems and we propose an associated sequential algorithm based again on iterative thresholding and alternating subspace corrections. We prove that the weak accumulation points of the sequence produced by this algorithm are minimizers of $\mathcal{J}$. We prove also norm convergence in $\ell_2(\Lambda)$ of subsequences. In case of unique minimizer, the whole sequence will be convergent (not only subsequences.) In Section 4 we modify the algorithm in order to be parallelizable. We prove similar results as for the sequential algorithm. Section 5 discusses the computational cost of the new algorithms and compares them with the thresholded Landweber iteration (1.4). In Section 6 we illustrate the extension of the decomposition to more than two subspaces and some further variations of the proposed algorithms.

2 Domain Decomposition Methods

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



$\Omega = \hat{\Omega}_1 \cup \hat{\Omega}_2,\ \hat{\Omega}_1 \cap \hat{\Omega}_2 \neq \emptyset$

Figure 1: An overlapping subdomain decomposition.

subdomains [30, 2, 43, 7, 28, 34, 44, 29, 1, 32]. 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. The mentioned techniques can often be applied directly to the partial differential equation, but of course to apply them to the discretizations of the problem is also of major interest. In this paper we deal with frame discretizations and the domain decomposition method will be applied on the space of the frame coefficients. Domain decomposition methods, together with other known iterative methods for symmetric positive definite problems, such as multigrid methods, Jacobi and Gauss-Seidel iterations, and multilevel nodal basis preconditioners, can be viewed as *subspace correction methods*, see [43]. In this paper, two types of domain decomposition based iterative schemes for a frame discretized inverse problem will be considered. We discuss a *successive* subspace correction method (inspired by the so-called multiplicative Schwarz iteration) as well as a *parallel* subspace correction method (inspired by the so-called additive Schwarz iteration).

To introduce the approach we want to develop in this paper, we recall the main ideas of the most classical and well-known example of domain decomposition method, i.e., the *multiplicative Schwarz alternating algorithm.* Consider the second order self-adjoint elliptic problem

$$Lu = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega. \tag{2.1}$$

For the moment, let us restrict the discussion to the case of a decomposition of a domain $\Omega \subset \mathbb{R}^2$ into two overlapping subdomains, i.e., $\Omega = \hat{\Omega}_1 \cup \hat{\Omega}_2$, see Figure 1. Starting with an initial guess $u^{(0)}$, the multiplicative Schwarz alternating algorithm to solve (2.1) generates a sequence of approximations $u^{(1)},\ u^{(2)}, \dots$ by solving the following two local problems:

$$\left\{\begin{array}{lll} Lu_1^{(k+1)} & = f, & \text{in } \hat{\Omega}_1, \\ u_1^{(k+1)} & = u^{(k)}{}_{|\Gamma_1}, & \text{on } \Gamma_1, \\ u_1^{(k+1)} & = 0, & \text{on } \partial\hat{\Omega}_1 \setminus \Gamma_1, \end{array}\right. \quad \text{and} \quad \left\{\begin{array}{lll} Lu_2^{(k+1)} & = f, & \text{in } \hat{\Omega}_2, \\ u_2^{(k+1)} & = u_1^{(k+1)}{}_{|\Gamma_2}, & \text{on } \Gamma_2, \\ u_2^{(k+1)} & = 0, & \text{on } \partial\hat{\Omega}_2 \setminus \Gamma_2. \end{array}\right. \tag{2.2}$$

The next iterate $u^{(k+1)}$ is then defined by

$$u^{(k+1)}(x) = \begin{cases} u_2^{(k+1)}(x), & \text{if } x \in \hat{\Omega}_2 \\ u_1^{(k+1)}(x), & \text{if } x \in \Omega \setminus \hat{\Omega}_2. \end{cases} \tag{2.3}$$

By Stampacchia's Theorem, the variational formulation of (2.2) reads as follows: Let $u^0 \in H_0^1(\Omega)$. For $k = 0, 1, \dots$ compute

$$\begin{aligned} &u^{(k+1/2)} := u^{(k)} + u_1^{(k+1/2)}, \text{ where } u_1^{(k+1/2)} \text{ satisfies} \\ &u_1^{(k+1/2)} = \arg\min_{u_1 \in H_0^1(\hat{\Omega}_1)} J(u_1, u^{(k)}) \\ \\ &u^{(k+1)} := u^{(k+1/2)} + u_2^{(k+1/2)}, \text{ where } u_2^{(k+1/2)} \text{ satisfies} \\ &u_2^{(k+1/2)} = \arg\min_{u_2 \in H_0^1(\hat{\Omega}_2)} J(u_2, u^{(k+1/2)}) \end{aligned}$$

with $J(v, u) := \frac{1}{2} a(v, v) - (\langle f, v \rangle - a(u, v))$, $a(v, u) := \langle Lv, u \rangle$, as usual, being the corresponding bilinear form.

Inspired by this variational formulation of the classical Schwartz alternating algorithm, we propose a minimization of the functional in (1.3) by alternating minimizations of local problems restricted to suitable subspaces. Similar techniques of alternating minimizations of functionals with auxiliary variables appear also, e.g., in [6, 11, 24, 26, 27].

3 Domain Decompositions Adapted to Inverse Problems

In this section we introduce a sequential domain decomposition method for the linear inverse problem with sparsity constraints modelled by (1.3). The goal is to join the simplicity of the iterative approach (1.4) with a dimension reduction technique provided by a decomposition which will improve the convergence and the complexity of the algorithm without increasing the sophistication of the algorithm.

Before starting our discussion let us briefly introduce some of the spaces we will use in the following. For some countable index set Λ we denote by $\ell_p = \ell_p(\Lambda)$, $1 \le p \le \infty$ the space of real sequences $u = (u_\lambda)_{\lambda \in \Lambda}$ with norm

$$\|u\|_p \;=\; \|u\|_{\ell_p(\Lambda)} \;:=\; \left(\sum_{\lambda \in \Lambda} |u_\lambda|^p \right)^{1/p}, \quad 1 \le p < \infty$$

and $\|u\|_\infty := \sup_{\lambda \in \Lambda} |u_\lambda|$ as usual.

For simplicity, we start by decomposing the "domain" of the sequences Λ into two disjoint sets Λ_1, Λ_2 so that $\Lambda = \Lambda_1 \cup \Lambda_2$. The extension to decompositions into multiple subsets $\Lambda = \Lambda_1 \cup \dots \cup \Lambda_{\mathcal{N}}$ follows from an analysis similar to the basic case $\mathcal{N} = 2$, and we discuss this issue in Section 6. Associated to a decomposition $\mathcal{C} = \{\Lambda_1, \Lambda_2\}$ we define the *extension operators* $E_i : \ell_2(\Lambda_i) \to \ell_2(\Lambda)$, $(E_i v)_\lambda = v_\lambda$, if $\lambda \in \Lambda_i$, $(E_i v)_\lambda = 0$, otherwise, $i = 1, 2$. The adjoint operator, which we call the *restriction operator*, is denoted by $R_i := E_i^*$. With these operators we may define the functional $J(u_1, u_2)$, $J : \ell_2(\Lambda_1) \times \ell_2(\Lambda_2) \to \mathbb{R}$, given by

$$\mathcal{J}(u_1, u_2) := \mathcal{J}(E_1 u_1 + E_2 u_2).$$

For the sequence u_i we use the notation $u_{\lambda,i}$ in order to denote its components. In analogy to the Schwartz multiplicative algorithm, we want to formulate and to analyze the following algorithm:

$$\begin{cases} u_1^{(n+1)} = \arg\min_{v_1 \in \ell_2(\Lambda_1)} \mathcal{J}(v_1, u_2^{(n)}) \\ u_2^{(n+1)} = \arg\min_{v_2 \in \ell_2(\Lambda_2)} \mathcal{J}(u_1^{(n+1)}, v_2) \\ u^{(n+1)} := E_1 u_1^{(n+1)} + E_2 u_2^{(n+1)}. \end{cases} \tag{3.1}$$

Let us observe that $\|E_1 u_1 + E_2 u_2\|_{\ell_1(\Lambda)} := \|u_1\|_{\ell_1(\Lambda_1)} + \|u_2\|_{\ell_1(\Lambda_2)}$, hence

$$\arg \min_{v_1 \in \ell_2(\Lambda_1)} \mathcal{J}(v_1, u_2^{(n)}) = \arg \min_{v_1 \in \ell_2(\Lambda_1)} \|(g - TE_2 u_2^{(n)}) - TE_1 v_1\|_{\mathcal{H}}^2 + \tau \|v_1\|_1.$$

A similar formulation holds for $\arg\min_{v_2 \in \ell_2(\Lambda_2)} \mathcal{J}(u_1^{(n+1)}, v_2)$. This means that the solution of the local problems on Λ_i is of the *same* kind as the original problem $\arg\min_{u \in \ell_2(\Lambda)} \mathcal{J}(u)$, but the dimension for each has been reduced. Unfortunately the functionals $\mathcal{J}(u, u_2^{(n)})$ and $\mathcal{J}(u_1^{(n+1)}, v)$ do not need to have a unique minimizer. Therefore the formulation as in (3.1) is not in principle well defined. In the following we will consider a particular choice of the minimizers and in particular we will implement the algorithm in (1.4) in order to solve each local problem. This choice leads to the following algorithm:

$$\begin{cases} \begin{cases} u_1^{(n+1,0)} = u_1^{(n,L)} \\ u_1^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_1^{(n+1,\ell)} + R_1 T^*((g - TE_2 u_2^{(n,M)}) - TE_1 u_1^{(n+1,\ell)}) \right) \quad \ell = 0, \ldots, L-1 \end{cases} \\ \begin{cases} u_2^{(n+1,0)} = u_2^{(n,M)} \\ u_2^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_2^{(n+1,\ell)} + R_2 T^*((g - TE_1 u_1^{(n+1,L)}) - TE_2 u_2^{(n+1,\ell)}) \right) \quad \ell = 0, \ldots, M-1 \end{cases} \\ u^{(n+1)} := E_1 u_1^{(n+1,L)} + E_2 u_2^{(n+1,M)}. \end{cases} \tag{3.2}$$

Of course, for $L = M = \infty$ the previous algorithm realizes a particular instance of (3.1). However, in practice we will never execute an infinite number of inner iterations and therefore it is important to analyze the convergence of the algorithm when $L, M \in \mathbb{N}$ are finite. Moreover, as we will discuss in Section 5, the computational cost of the whole algorithm and its convergence rate depends on the choice of L and M. It is not convenient to choose them too large.

At this point the question is whether algorithm (3.2) really converges to a minimizer of the original functional $\mathcal{J}$. This is the scope of the following sections.

3.1 Weak convergence of the sequential DD algorithm

A main tool in the analysis of non-smooth functionals and their minima is the concept of subdifferential. Recall that for a convex functional F on some Banach space V its subdifferential $\partial F(x)$ at a point $x \in V$ with $F(x) < \infty$ is defined as the set

$$\partial F(x) = \{x^* \in V^*, x^*(z-x) + F(x) \le F(z) \text{ for all } z \in V\},$$

where V^* denotes the dual space of V. It is obvious from this definition that $0 \in \partial F(x)$ if and only if x is a minimizer of F.

Example 1. Let $V = \ell_1(\Lambda)$ and $F(x) := \|x\|_1$ is the ℓ_1 norm. We have

$$\partial \| \cdot \|_1(x) = \{\xi \in \ell_\infty(\Lambda) : \xi_\lambda \in \partial |\cdot|(x_\lambda), \lambda \in \Lambda\} \tag{3.3}$$

where $\partial|\cdot|(z) = \{\text{sign}(z)\}$ if $z \neq 0$ and $\partial|\cdot|(0) = [-1, 1]$.

It will turn out to be useful to us the auxiliary functional

$$\mathcal{J}^S(u,a) := \|g - Tu\|_{\mathcal{H}}^2 + \tau\|u\|_1 + \|u-a\|_2^2 - \|Tu - Ta\|_{\mathcal{H}}^2. \tag{3.4}$$

A direct calculation shows

$$\mathcal{J}^S(u,a) = \|(a + T^*(g - Ta)) - u\|_2^2 + \|u\|_1 - \|a + T^*(g-Ta)\|_2^2 + \|g\|_{\mathcal{H}}^2 - \|Ta\|_{\mathcal{H}}^2 + \|a\|_2^2.$$

In the following we assume that $\|T\| < 1$. This condition can be always achieved by suitable rescaling of T and g. Observe that

$$\|u-a\|_2^2 - \|Tu - Ta\|_{\mathcal{H}}^2 \ge C\|u-a\|_2^2, \tag{3.5}$$

for $C = (1 - \|T\|^2) > 0$. Hence

$$\mathcal{J}(u) = \mathcal{J}^S(u,u) \le \mathcal{J}^S(u,a), \tag{3.6}$$

and

$$\mathcal{J}^S(u,a) - \mathcal{J}^S(u,u) \ge C\|u-a\|_2^2. \tag{3.7}$$

In particular, $\mathcal{J}^S$ is strictly convex with respect to u and it has a unique minimizer with respect to u once a is fixed. By observing that $\partial(\|T \cdot - g\|_{\mathcal{H}}^2)(u) = \{2T^*(Tu - g)\}$ (see [26, Lemma 3.2]) and by an application of [20, Proposition 5.2] combined with the example above, we obtain the following characterizations of the subdifferentials of $\mathcal{J}$ and $\mathcal{J}^S$.

Lemma 3.1. *i) The subdifferential of $\mathcal{J}$ at u is given by*

$$\begin{aligned}\partial\mathcal{J}(u) &= 2T^*(Tu-g) + \tau\partial\|\cdot\|_1(u)\\ &= \{\xi \in \ell_\infty(\Lambda) : \xi_\lambda \in [2T^*(Tu-g)]_\lambda + \tau\partial|\cdot|(u_\lambda)\}.\end{aligned}$$

ii) The subdifferential of $\mathcal{J}^S$ with respect to the sole component u is given by

$$\begin{aligned}\partial_u\mathcal{J}^S(u,a) &= -2(a + T^*(g - Ta)) + 2u + \tau\partial\|\cdot\|_1(u)\\ &= \{\xi \in \ell_\infty(\Lambda) : \xi_\lambda \in [-2(a + T^*(g-Ta))]_\lambda + 2u_\lambda + \tau\partial|\cdot|(u_\lambda)\}.\end{aligned}$$

Since $\mathbb{S}_\tau(z)$ is the unique solution of the subdifferential inclusion $0 \in 2(u-z)+\tau\partial\|\cdot\|_1(u)$, see for instance [26] and [14, Proposition 2.1], from Lemma 3.1 ii) we obtain immediately

$$\arg\min_{u\in\ell_2(\Lambda)} \mathcal{J}^S(u,a) = \mathbb{S}_\tau(a+T^*(g-Ta)).$$

In light of this result we can reformulate the algorithm in (3.2) by

$$\left\{\begin{array}{l} \left\{\begin{array}{l} u_1^{(n+1,0)} = u_1^{(n,L)} \\ u_1^{(n+1,\ell+1)} = \arg\min_{u_1\in\ell_2(\Lambda_1)} \mathcal{J}^S(E_1u_1+E_2u_2^{(n,M)}, E_1u_1^{(n+1,\ell)}+E_2u_2^{(n,M)}) \quad \ell=0,\dots,L-1 \end{array}\right. \\ \left\{\begin{array}{l} u_2^{(n+1,0)} = u_2^{(n,M)} \\ u_2^{(n+1,\ell+1)} = \arg\min_{u_2\in\ell_2(\Lambda_2)} \mathcal{J}^S(E_1u_1^{(n+1,L)}+E_2u_2, E_1u_1^{(n+1,L)}+E_2u_2^{(n+1,\ell)}) \quad \ell=0,\dots,M-1 \end{array}\right. \\ u^{(n+1)} := E_1u_1^{(n+1,L)}+E_2u_2^{(n+1,M)}. \end{array}\right. \tag{3.8}$$

Before we actually start proving the weak convergence of the algorithm in (3.8) we recall the following definition [37].

Definition 1. Let V be a topological space and $\mathcal{A} = (A_n)_{n\in\mathbb{N}}$ a sequence of subsets of V. The subset $A \subseteq V$ is called the *limit of the sequence* $\mathcal{A}$, and we write $A = \lim_n A_n$, if

$$A = \{a \in V : \exists a_n \in A_n, a = \lim_n a_n\}.$$

The following observation will be useful for us, see, e.g., [37, Proposition 8.7].

Lemma 3.2. *Assume that Γ is a convex function on $\mathbb{R}^M$ and $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^M$ a convergent sequence with limit x such that $\Gamma(x_n), \Gamma(x) < \infty$. Then the subdifferentials satisfy*

$$\lim_{n\to\infty} \partial\Gamma(x_n) \subseteq \partial\Gamma(x).$$

In other words, the subdifferential $\partial\Gamma$ of a convex function is an outer semicontinuous set-valued function*.*

Theorem 3.3 (Weak convergence)**.** *The algorithm in* (3.8) *produces a sequence $(u^{(n)})_{n\in\mathbb{N}}$ in $\ell_2(\Lambda)$ whose weak accumulation points are minimizers of the functional $\mathcal{J}$. In particular, the set of the weak accumulation points is non-empty and if $u^{(\infty)}$ is a weak accumulation point then*

$$u^{(\infty)} = \mathbb{S}_\tau(u^{(\infty)}+T^*(g-Tu^{(\infty)})).$$

Proof. Let us first observe that by (3.6)

$$\begin{aligned} \mathcal{J}(u^{(n)}) = \mathcal{J}^S(u^{(n)},u^{(n)}) &= \mathcal{J}^S(E_1u_1^{(n,L)}+E_2u_2^{(n,M)}, E_1u_1^{(n,L)}+E_2u_2^{(n,M)}) \\ &= \mathcal{J}^S(E_1u_1^{(n,L)}+E_2u_2^{(n,M)}, E_1u_1^{(n+1,0)}+E_2u_2^{(n,M)}). \end{aligned}$$

By definition of $u_1^{(n+1,1)}$ and its minimal properties we have

$$\mathcal{J}^S(E_1u_1^{(n,L)}+E_2u_2^{(n,M)}, E_1u_1^{(n+1,0)}+E_2u_2^{(n,M)}) \geq \mathcal{J}^S(E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)}, E_1u_1^{(n+1,0)}+E_2u_2^{(n,M)}).$$

Again, an application of (3.6) gives

$$\mathcal{J}^S(E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,0)}+E_2u_2^{(n,M)}) \geq \mathcal{J}^S(E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)})$$

Putting in line these inequalities we obtain

$$\mathcal{J}(u^{(n)}) \geq \mathcal{J}^S(E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)}).$$

In particular, from (3.7) we have

$$\mathcal{J}(u^{(n)}) - \mathcal{J}^S(E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)}) \geq C\|u_1^{(n+1,1)}-u_1^{(n+1,0)}\|^2_{\ell_2(\Lambda_1)}.$$

By induction we obtain

$$\begin{aligned}\mathcal{J}(u^{(n)}) &\geq \mathcal{J}^S(E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,1)}+E_2u_2^{(n,M)}) \geq \dots \\ &\geq \mathcal{J}^S(E_1u_1^{(n+1,L)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,L)}+E_2u_2^{(n,M)}),\end{aligned}$$

and

$$\mathcal{J}(u^{(n)}) - \mathcal{J}^S(E_1u_1^{(n+1,L)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,L)}+E_2u_2^{(n,M)}) \geq C\sum_{\ell=0}^{L-1}\|u_1^{(n+1,\ell+1)}-u_1^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_1)}.$$

By definition of $u_2^{(n+1,1)}$ and its minimal properties we have

$$\begin{aligned}&\mathcal{J}^S(E_1u_1^{(n+1,L)}+E_2u_2^{(n,M)},E_1u_1^{(n+1,L)}+E_2u_2^{(n,M)}) \\ \geq\ &\mathcal{J}^S(E_1u_1^{(n+1,L)}+E_2u_2^{(n+1,1)},E_1u_1^{(n+1,L)}+E_2u_2^{(n+1,0)}).\end{aligned}$$

By similar arguments as above we find

$$\mathcal{J}(u^{(n)}) \geq \mathcal{J}^S(E_1u_1^{(n+1,L)}+E_2u_2^{(n+1,M)},E_1u_1^{(n+1,L)}+E_2u_2^{(n+1,M)}) = \mathcal{J}(u^{(n+1)}), \tag{3.9}$$

and

$$\mathcal{J}(u^{(n)}) - \mathcal{J}(u^{(n+1)}) \geq C\left(\sum_{\ell=0}^{L-1}\|u_1^{(n+1,\ell+1)}-u_1^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_1)} + \sum_{m=0}^{M-1}\|u_2^{(n+1,m+1)}-u_2^{(n+1,m)}\|^2_{\ell_2(\Lambda_2)}\right). \tag{3.10}$$

From (3.9) we have $\mathcal{J}(u^{(0)}) \geq \mathcal{J}(u^{(n)}) \geq \tau\|u^{(n)}\|_{\ell_1(\Lambda)} \geq \tau\|u^{(n)}\|_{\ell_2(\Lambda)}$. This means that $(u^{(n)})_{n\in\mathbb{N}}$ is uniformly bounded in $\ell_2(\Lambda)$, hence there exists a weakly convergent subsequence $(u^{(n_j)})_{j\in\mathbb{N}}$. Let us denote $u^{(\infty)}$ the weak limit of the subsequence. For simplicity, we rename such subsequence by $(u^{(n)})_{n\in\mathbb{N}}$. Moreover, since the sequence $(\mathcal{J}(u^{(n)}))_{n\in\mathbb{N}}$ is monotonically decreasing and bounded from below by 0, it is also convergent. From (3.10) and the latter convergence we deduce

$$\left(\sum_{\ell=0}^{L-1}\|u_1^{(n+1,\ell+1)}-u_1^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_1)} + \sum_{m=0}^{M-1}\|u_2^{(n+1,m+1)}-u_2^{(n+1,m)}\|^2_{\ell_2(\Lambda_2)}\right) \to 0, \quad n\to\infty. \tag{3.11}$$

In particular, by the standard inequality $(a^2+b^2) \geq \frac{1}{2}(a+b)^2$ for $a, b > 0$ and the triangle inequality, we have also

$$\|u^{(n)} - u^{(n+1)}\|_{\ell_2(\Lambda)} \to 0, \quad n \to \infty. \tag{3.12}$$

We would like now to show that

$$0 \in \lim_{n\to\infty} \partial\mathcal{J}(u^{(n)}) \subset \partial\mathcal{J}(u^{(\infty)}).$$

To this end, and in light of Lemma 3.1, we reason componentwise. By definition of $u_1^{(n+1,L)}$ we have

$$0 \in [-2(u_1^{(n+1,L-1)} + R_1 T^*((g - TE_2 u_2^{(n,M)}) - TE_1 u_1^{(n+1,L-1)}))]_\lambda + 2u_{\lambda,1}^{(n+1,L)} + \tau\partial|\cdot|(u_{\lambda,1}^{(n+1,L)}), \tag{3.13}$$

and by definition of $u_2^{(n+1,M)}$ we have

$$0 \in [-2(u_2^{(n+1,M-1)} + R_2 T^*((g - TE_1 u_1^{(n+1,L)}) - TE_2 u_2^{(n+1,M-1)}))]_\lambda + 2u_{\lambda,2}^{(n+1,M)} + \tau\partial|\cdot|(u_{\lambda,2}^{(n+1,M)}). \tag{3.14}$$

Let us compute $\partial\mathcal{J}(u^{(n+1)})_\lambda$,

$$\partial\mathcal{J}(u^{(n+1)})_\lambda = [-2T^*(g - TE_1 u_1^{(n+1,L)} - TE_2 u_2^{(n+1,M)})]_\lambda + \tau\partial|\cdot|(u_{\lambda,i}^{(n+1,K)}), \tag{3.15}$$

where $\lambda \in \Lambda_i$ and $K = L, M$ for $i = 1, 2$ respectively. We would like to find a $\xi_\lambda^{(n+1)} \in \partial\mathcal{J}(u^{(n+1)})_\lambda$ such that $\xi_\lambda^{(n+1)} \to 0$ for $n \to \infty$. By (3.13) we have that for $\lambda \in \Lambda_1$

$$0 = [-2(u_1^{(n+1,L-1)} + R_1 T^*((g - TE_2 u_2^{(n,M)}) - TE_1 u_1^{(n+1,L-1)}))]_\lambda + 2u_{\lambda,1}^{(n+1,L)} + \tau\xi_{\lambda,1}^{(n+1)},$$

for a $\xi_{\lambda,1}^{(n+1)} \in \partial|\cdot|(u_{\lambda,1}^{(n+1,L)})$, and, by (3.14), for $\lambda \in \Lambda_2$

$$0 = [-2(u_2^{(n+1,M-1)} + R_2 T^*((g - TE_1 u_1^{(n+1,L)}) - TE_2 u_2^{(n+1,M-1)}))]_\lambda + 2u_{\lambda,2}^{(n+1,M)} + \tau\xi_{\lambda,2}^{(n+1)},$$

for a $\xi_{\lambda,2}^{(n+1)} \in \partial|\cdot|(u_{\lambda,2}^{(n+1,M)})$. Thus by subtracting zero from (3.15) as represented by the previous two formulas, we can choose

$$\xi_\lambda^{(n+1)} = 2(u_{\lambda,1}^{(n+1,L)} - u_{\lambda,1}^{(n+1,L-1)}) + [R_1 T^* TE_1(u_1^{(n+1,L)} - u_1^{(n+1,L-1)})]_\lambda + [R_1 T^* TE_2(u_2^{(n+1,M)} - u_1^{(n,M)})]_\lambda,$$

if $\lambda \in \Lambda_1$ and

$$\xi_\lambda^{(n+1)} = 2(u_{\lambda,2}^{(n+1,M)} - u_{\lambda,2}^{(n+1,M-1)}) + [R_2 T^* TE_1(u_2^{(n+1,M)} - u_1^{(n+1,M-1)})]_\lambda,$$

if $\lambda \in \Lambda_2$. For both these choices, from (3.11) and (3.12) we have $\xi_\lambda^{(n+1)} \to 0$ for $n \to \infty$. By continuity of T, weak convergence of $u^{(n)}$ (which implies componentwise convergence), and Lemma 3.2 we obtain

$$0 \in \lim_{n\to\infty} \partial\mathcal{J}(u^{(n)})_\lambda \subset \partial\mathcal{J}(u^{(\infty)})_\lambda, \quad \forall\lambda \in \Lambda.$$

It follows from Lemma 3.1 that $0 \in \partial \mathcal{J}(u^{(\infty)})$. By the properties of the subdifferential we have that $u^{(\infty)}$ is a minimizer of $\mathcal{J}$. Of course, the reasoning above hold for any weak convergent subsequence and therefore all weak accumulation points of the original sequence $(u^{(n)})_n$ are minimizers of $\mathcal{J}$.

Similarly, by taking now the limit for $n \to \infty$ in (3.13) and (3.14), and by using (3.11) we obtain

$$0 \in [-2(R_1 u^{(\infty)} + R_1 T^*((g - TE_2R_2u^{(\infty)}) - TE_1R_1u^{(\infty)}))]_\lambda + 2u_\lambda^{(\infty)} + \tau\partial|\cdot|(u_\lambda^{(\infty)}),$$

for $\lambda \in \Lambda_1$ and

$$0 \in [-2(R_2 u^{(\infty)} + R_2 T^*((g - TE_1R_1u^{(\infty)}) - TE_2R_2u^{(\infty)})]_\lambda + 2u_\lambda^{(\infty)} + \tau\partial|\cdot|(u_\lambda^{(\infty)}).$$

for $\lambda \in \Lambda_2$. In other words, we have

$$0 \in \partial_u \mathcal{J}^S(u^{(\infty)}, u^{(\infty)}).$$

An application of Lemma 3.1 and [14, Proposition 2.1] imply

$$u^{(\infty)} = \mathbb{S}_\tau(u^{(\infty)} + T^*(g - Tu^{(\infty)})).$$

■

REMARKS:

1. Because $u^{(\infty)} = \mathbb{S}_\tau(u^{(\infty)} + T^*(g - Tu^{(\infty)}))$, we could infer the minimality of $u^{(\infty)}$ by invoking [14, Proposition 3.10]. In the previous proof we wanted to present an alternative argument based on differential inclusions.
2. Since $(u^{(n)})_{n\in\mathbb{N}}$ is bounded and (3.11) holds, also $(u_i^{n,\ell})_{n,\ell}$ are bounded for $i = 1, 2$.

3.2 Strong convergence of the sequential DD algorithm

In this section we want to show that the convergence of a subsequence $(u^{n_j})_j$ to any accumulation point $u^{(\infty)}$ holds not only in the weak topology, but also in the Hilbert space $\ell_2(\Lambda)$ norm. Let us define

$$\begin{aligned} \eta^{(n+1)} &:= u_1^{(n+1,L)} - u_1^{(\infty)}, \qquad \eta^{(n+1/2)} := u_1^{(n+1,L-1)} - u_1^{(\infty)}, \\ \mu^{(n+1)} &:= u_2^{(n+1,M)} - u_2^{(\infty)}, \qquad \mu^{(n+1/2)} := u_2^{(n+1,M-1)} - u_2^{(\infty)}, \end{aligned}$$

where $u_i^{(\infty)} := R_i u^{(\infty)}$. From Theorem 3.3 we also have

$$u_i^{(\infty)} = \mathbb{S}_\tau(\underbrace{u_i^{(\infty)} + R_i T^*(g - TE_1u_1^{(\infty)} - TE_2u_2^{(\infty)})}_{:=h_i}), \quad i = 1, 2.$$

Let us also denote $h = E_1h_1 + E_2h_2$ and $\xi^{(n)} = E_1\eta^{(n+1/2)} + E_2\mu^{(n+1/2)}$.

For the proof of strong convergence we need the following technical lemmas.

Lemma 3.4 (Lemma 2.2 [14])**.** *The operator* $\mathbb{S}_\tau$ *is non-expansive, i.e.,* $\|\mathbb{S}_\tau(u) - \mathbb{S}_\tau(v)\|_2 \leq \|u - v\|_2$.

Lemma 3.5. $\|T\xi^{(n)}\|_{\mathcal{H}}^2 \to 0$ *for* $n \to \infty$.

Proof. Since

$$\begin{aligned}
\eta^{(n+1)} - \eta^{(n+1/2)} &= \mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}) - \mathbb{S}_\tau(h_1) - \eta^{(n+1/2)} \\
\mu^{(n+1)} - \mu^{(n+1/2)} &= \mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}) - \mathbb{S}_\tau(h_2) - \mu^{(n+1/2)},
\end{aligned}$$

and $\|\eta^{(n+1)} - \eta^{(n+1/2)}\|_{\ell_2(\Lambda_1)} = \|u_1^{(n+1,L)} - u_1^{(n+1,L-1)}\|_{\ell_2(\Lambda_1)} \to 0$, $\|\mu^{(n+1)} - \mu^{(n+1/2)}\|_{\ell_2(\Lambda_1)} = \|u_2^{(n+1,M)} - u_2^{(n+1,M-1)}\|_{\ell_2(\Lambda_2)} \to 0$ by (3.11), we have

$$\begin{aligned}
&\|\mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}) - \mathbb{S}_\tau(h_1) - \eta^{(n+1/2)}\|_{\ell_2(\Lambda_1)} \qquad (3.16) \\
\geq\; &\left| \|\mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}) - \mathbb{S}_\tau(h_1)\|_{\ell_2(\Lambda_1)} - \|\eta^{(n+1/2)}\|_{\ell_2(\Lambda_1)} \right| \to 0,
\end{aligned}$$

and

$$\begin{aligned}
&\|\mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}) - \mathbb{S}_\tau(h_2) - \mu^{(n+1/2)}\|_{\ell_2(\Lambda_2)} \qquad (3.17) \\
\geq\; &\left| \|\mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}) - \mathbb{S}_\tau(h_2)\|_{\ell_2(\Lambda_2)} - \|\mu^{(n+1/2)}\|_{\ell_2(\Lambda_2)} \right| \to 0.
\end{aligned}$$

By non-expansivity of $\mathbb{S}_\tau$ we have the estimates

$$\begin{aligned}
&\|\mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}) - \mathbb{S}_\tau(h_2)\|_{\ell_2(\Lambda_2)} \\
\leq\; &\|(I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}\|_{\ell_2(\Lambda_2)} \\
\leq\; &\|(I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1/2)}\|_{\ell_2(\Lambda_2)} + \underbrace{\|R_2T^*TE_1(\eta^{(n+1/2)}) - \eta^{(n+1)})\|_{\ell_2(\Lambda_2)}}_{:=\varepsilon^{(n)}}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\|\mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}) - \mathbb{S}_\tau(h_1)\|_{\ell_2(\Lambda_1)} \\
\leq\; &\|(I - R_1T^*TE_1)\mu^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}\|_{\ell_2(\Lambda_1)} \\
\leq\; &\|(I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n+1/2)}\|_{\ell_2(\Lambda_1)} + \underbrace{\|R_1T^*TE_2(\mu^{(n+1/2)} - \mu^{(n)})\|_{\ell_2(\Lambda_1)}}_{\delta^{(n)}}.
\end{aligned}$$

Moreover, combining the previous inequalities, we obtain also

$$\begin{aligned}
&\|\mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)} - \mathbb{S}_\tau(h_1)\|_{\ell_2(\Lambda_1)}^2 \\
+\; &\|\mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)} - \mathbb{S}_\tau(h_2)\|_{\ell_2(\Lambda_2)}^2 \\
\leq\; &\|(I - R_1T^*TE_1)\mu^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}\|_{\ell_2(\Lambda_1)}^2 + \|(I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}\|_{\ell_2(\Lambda_2)}^2 \\
\leq\; &\|\xi^{(n)}\|_{\ell_2(\Lambda)}^2 + ((\varepsilon^{(n)})^2 + (\delta^{(n)})^2 + C'(\varepsilon^{(n)} + \delta^{(n)})).
\end{aligned}$$

The constant $C' > 0$ is due to the boundedness of $u^{(n,\ell)}$. Certainly, by (3.11), for every $\varepsilon > 0$ there exists n_0 such that for $n > n_0$ we have $(\varepsilon^{(n)})^2 + (\delta^{(n)})^2 + C'(\varepsilon^{(n)} + \delta^{(n)})) \leq \varepsilon$. Therefore, if

$$\begin{aligned} & \|\mathbb{S}_\tau(h_1 + (I - R_1 T^* T E_1)\eta^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)} - \mathbb{S}_\tau(h_1)\|^2_{\ell_2(\Lambda_1)} \\ + & \|\mathbb{S}_\tau(h_2 + (I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)} - \mathbb{S}_\tau(h_2)\|^2_{\ell_2(\Lambda_2)} \geq \|\xi^{(n)}\|^2_{\ell_2(\Lambda)}, \end{aligned}$$

then

$$\begin{aligned} 0 \leq & \|(I - R_1 T^* T E_1)\mu^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)}\|^2_{\ell_2(\Lambda_1)} \\ + & \|(I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)}\|^2_{\ell_2(\Lambda)} - \|\xi^{(n)}\|^2_{\ell_2(\Lambda)} \\ \leq & (\varepsilon^{(n)})^2 + (\delta^{(n)})^2 + C'(\varepsilon^{(n)} + \delta^{(n)})) \leq \varepsilon \end{aligned}$$

If, instead, we have

$$\begin{aligned} & \|\mathbb{S}_\tau(h_1 + (I - R_1 T^* T E_1)\eta^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)}) - \mathbb{S}_\tau(h_1)\|^2_{\ell_2(\Lambda_1)} \\ + & \|\mathbb{S}_\tau(h_2 + (I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)}) - \mathbb{S}_\tau(h_2)\|^2_{\ell_2(\Lambda_2)} < \|\xi^{(n)}\|^2_{\ell_2(\Lambda)}, \end{aligned}$$

then by (3.16) and (3.17)

$$\begin{aligned} \|\xi^{(n)}\|^2_{\ell_2(\Lambda)} - & \Big(\|(I - R_1 T^* T E_1)\mu^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)})\|^2_{\ell_2(\Lambda_1)} \\ + & \|(I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)})\|^2_{\ell_2(\Lambda_2)}\Big) \\ \leq & \|\xi^{(n)}\|^2_{\ell_2(\Lambda)} - \|\mathbb{S}_\tau(h_1 + (I - R_1 T^* T E_1)\eta^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)}) - \mathbb{S}_\tau(h_1)\|^2_{\ell_2(\Lambda_1)} \\ - & \|\mathbb{S}_\tau(h_2 + (I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)}) - \mathbb{S}_\tau(h_2)\|^2_{\ell_2(\Lambda_2)} \\ = & \Big|\|\xi^{(n)}\|^2_{\ell_2(\Lambda)} - \|\mathbb{S}_\tau(h_1 + (I - R_1 T^* T E_1)\eta^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)}) - \mathbb{S}_\tau(h_1)\|^2_{\ell_2(\Lambda_1)} \\ - & \|\mathbb{S}_\tau(h_2 + (I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)}) - \mathbb{S}_\tau(h_2)\|^2_{\ell_2(\Lambda_2)}\Big| \\ \leq & \Big|\|\eta^{(n+1/2)}\|^2_{\ell_2(\Lambda_1)} - \|\mathbb{S}_\tau(h_1 + (I - R_1 T^* T E_1)\eta^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)}) - \mathbb{S}_\tau(h_1)\|^2_{\ell_2(\Lambda_1)}\Big| \\ + & \Big|\|\mu^{(n+1/2)}\|^2_{\ell_2(\Lambda_2)} - \|\mathbb{S}_\tau(h_2 + (I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)}) - \mathbb{S}_\tau(h_2)\|^2_{\ell_2(\Lambda_2)}\Big| \leq \varepsilon \end{aligned}$$

for n large enough. This implies

$$\begin{aligned} \lim_{n\to\infty} \Big[\|\xi^{(n)}\|^2_{\ell_2(\Lambda)} - & \Big(\|(I - R_1 T^* T E_1)\mu^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)}\|^2_{\ell_2(\Lambda_1)} \\ + & \|(I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)}\|^2_{\ell_2(\Lambda_2)}\Big)\Big] = 0 \end{aligned}$$

Observe now that

$$\begin{aligned} & \|(I - R_1 T^* T E_1)\mu^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n)}\|^2_{\ell_2(\Lambda_1)} + \|(I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1)}\|^2_{\ell_2(\Lambda_2)} \\ \leq & (\|(I - R_1 T^* T E_1)\mu^{(n+1/2)} - R_1 T^* T E_2 \mu^{(n+1/2)}\|_{\ell_2(\Lambda_1)} + \delta^{(n)})^2 \\ + & (\|(I - R_2 T^* T E_2)\mu^{(n+1/2)} - R_2 T^* T E_1 \eta^{(n+1/2)}\|_{\ell_2(\Lambda_2)} + \varepsilon^{(n)})^2 \\ \leq & \|(I - T^* T)\xi^{(n)}\|^2_{\ell_2(\Lambda)} + \Big((\varepsilon^{(n)})^2 + (\delta^{(n)})^2 + 2C'(\varepsilon^{(n)} + \delta^{(n)})\Big), \end{aligned}$$

for a suitable constant $C' > 0$ as above. Therefore we have

$$\begin{aligned}
\|\xi^{(n)}\|^2_{\ell_2(\Lambda)} &- \Big(\|(I - R_1T^*TE_1)\mu^{(n+1/2)} - R_1T^*TE_2\mu^{(n)})\|^2_{\ell_2(\Lambda_1)} \\
&+ \|(I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)})\|^2_{\ell_2(\Lambda_2)}\Big) \\
&\geq \|\xi^{(n)}\|^2_{\ell_2(\Lambda)} - \|(I - T^*T)\xi^{(n)}\|^2_{\ell_2(\Lambda)} - \Big((\varepsilon^{(n)})^2 + (\delta^{(n)})^2 + 2C'(\varepsilon^{(n)} + \delta^{(n)})\Big) \\
&= 2\|T\xi^{(n)}\|^2_{\mathcal{H}} - \|T^*T\xi^{(n)}\|^2_{\ell_2(\Lambda)} - \Big((\varepsilon^{(n)})^2 + (\delta^{(n)})^2 + 2C'(\varepsilon^{(n)} + \delta^{(n)})\Big) \\
&\geq \|T\xi^{(n)}\|^2_{\mathcal{H}} - \Big((\varepsilon^{(n)})^2 + (\delta^{(n)})^2 + 2C'(\varepsilon^{(n)} + \delta^{(n)})\Big).
\end{aligned}$$

This implies $\|T\xi^{(n)}\|^2_{\mathcal{H}} \to 0$ for $n \to \infty$.

■

Lemma 3.6. *For $h = E_1h_1 + E_2h_2$, $\|\mathbb{S}_\tau(h + \xi^{(n)}) - \mathbb{S}_\tau(h) + \xi^{(n)}\|_{\ell_2(\Lambda)} \to 0$, for $n \to \infty$.*

Proof. We have

$$\begin{aligned}
&\mathbb{S}_\tau(h + \xi^{(n)} - T^*T\xi^{(n)}) \\
=\ & E_1\Big(\mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n+1/2)})\Big) \\
+\ & E_2\Big(\mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1/2)})\Big)
\end{aligned}$$

Therefore, we can write

$$\begin{aligned}
&\mathbb{S}_\tau(h + \xi^{(n)} - T^*T\xi^{(n)}) \\
=\ & E_1\Big[\mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}) \\
+\ & \mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n+1/2)}) \\
-\ & \mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)})\Big] \\
+\ & E_2\Big[\mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}) \\
+\ & \mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1/2)}) \\
-\ & \mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)})\Big].
\end{aligned}$$

By using the non-expansivity of $\mathbb{S}_\tau$, the boundedness of the operators E_i, R_i, T^*T, and the triangle inequality we obtain,

$$\begin{aligned}
&\|\mathbb{S}_\tau(h + \xi^{(n)}) - \mathbb{S}_\tau(h) + \xi^{(n)}\|^2_{\ell_2(\Lambda)} \\
\leq\ & C'\Big(\|\mathbb{S}_\tau(h + \xi^{(n)} - T^*T\xi^{(n)}) - \mathbb{S}_\tau(h) + \xi^{(n)}\|^2_{\ell_2(\Lambda)} + \|\mathbb{S}_\tau(h + \xi^{(n)}) - \mathbb{S}_\tau(h + \xi^{(n)} - T^*T\xi^{(n)})\|^2_{\ell_2(\Lambda)}\Big) \\
\leq\ & C''\Bigg(\underbrace{\|\mathbb{S}_\tau(h_1 + (I - R_1T^*TE_1)\eta^{(n+1/2)} - R_1T^*TE_2\mu^{(n)}) - \mathbb{S}_\tau(h_1) - \eta^{(n+1/2)}\|^2_{\ell_2(\Lambda)}}_{:=A^{(n)}}
\end{aligned}$$

$$
\begin{aligned}
&+ \underbrace{\|\mathbb{S}_\tau(h_2 + (I - R_2T^*TE_2)\mu^{(n+1/2)} - R_2T^*TE_1\eta^{(n+1)}) - \mathbb{S}_\tau(h_2) - \mu^{(n+1/2)}\|^2_{\ell_2(\Lambda)}}_{:=B^{(n)}} \\
&+ \underbrace{\|\mu^{(n+1/2)} - \mu^{(n)}\|^2_{\ell_2(\Lambda_2)} + \|\eta^{(n)} - \eta^{(n+1/2)}\|^2_{\ell_2(\Lambda_1)} + C'''\|\mu^{(n+1/2)} - \mu^{(n)}\|^2_{\ell_2(\Lambda_2)}\|\eta^{(n)} - \eta^{(n+1/2)}\|^2_{\ell_2(\Lambda_1)}}_{:=C^{(n)}} \\
&+ \left.\underbrace{\|T^*T\xi^{(n)}\|^2_{\ell_2(\Lambda)}}_{:=D^{(n)}}\right),
\end{aligned}
$$

the constant $C''' > 0$ due to the boundedness of $u^{(n,\ell)}$. The quantities $A^{(n)}, B^{(n)}$ vanish for $n \to \infty$ because of (3.16) and (3.17). The quantity $C^{(n)}$ vanishes for $n \to \infty$ because of (3.11), and $D^{(n)}$ vanishes $n \to \infty$ thanks to Lemma 3.5. ■

Lemma 3.7 (Lemma 3.18 [14])**.** *If for some $a \in \ell_2(\Lambda)$ and some sequence $(\xi^{(n)})_{n\in\mathbb{N}}$, $\text{w-lim}_{n\to\infty} \xi^{(n)} = 0$ and $\lim_{n\to\infty} \|\mathbb{S}_\tau(a+\xi^{(n)})-\mathbb{S}_\tau(a)-\xi^{(n)}\|_{\ell_2(\Lambda)} = 0$, then $\lim_{n\to\infty} \|\xi^{(n)}\|_{\ell_2(\Lambda)} = 0$.*

By combining the previous technical achievements, we can now state the strong convergence.

Theorem 3.8 (Strong convergence)**.** *The algorithm in (3.8) 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.*

Proof. Let $u^{(\infty)}$ be a weak accumulation point and let $(u^{(n_j)})_{j\in\mathbb{N}}$ be a subsequence weakly convergent to $u^{(\infty)}$. Let us denote the latter sequence $(u^{(n)})_{n\in\mathbb{N}}$ again. With the notation used in this section, by Theorem 3.3 and (3.11) we have that $\xi^{(n)} = E_1\eta^{(n+1/2)} + E_2\mu^{(n+1/2)}$ weakly converges to zero. By Lemma 3.6 we have $\lim_{n\to\infty} \|\mathbb{S}_\tau(h+\xi^{(n)})-\mathbb{S}_\tau(h)-\xi^{(n)}\|_{\ell_2(\Lambda)} = 0$. From Lemma 3.7 we conclude that $\xi^{(n)} = E_1\eta^{(n+1/2)} + E_2\mu^{(n+1/2)}$ converges to zero strongly. Again by (3.11) we have that $(u^{(n)})_{n\in\mathbb{N}}$ converges to $u^{(\infty)}$ strongly. ■

4 A Parallel Domain Decomposition Method

The most natural modification to (3.2) in order to obtain a parallelizable algorithm is to substitute the term $u^{(n+1,L)}$ with $R_1u^{(n)}$ in the second inner iterations. This makes the inner iterations on Λ_1 and Λ_2 mutually independent, hence executable by two processors at the same time. We obtain the following algorithm:

$$
\left\{
\begin{array}{l}
\left\{
\begin{array}{ll}
u_1^{(n+1,0)} = R_1u^{(n)} & \\
u_1^{(n+1,\ell+1)} = \mathbb{S}_\tau\left(u_1^{(n+1,\ell)} + R_1T^*((g - TE_2R_2u^{(n)}) - TE_1u_1^{(n+1,\ell)})\right) & \ell = 0,\ldots,L-1
\end{array}
\right. \\
\left\{
\begin{array}{ll}
u_2^{(n+1,0)} = R_2u^{(n)} & \\
u_2^{(n+1,\ell+1)} = \mathbb{S}_\tau\left(u_2^{(n+1,\ell)} + R_2T^*((g - TE_1R_1u^{(n)}) - TE_2u_2^{(n+1,\ell)})\right) & \ell = 0,\ldots,M-1
\end{array}
\right. \\
u^{(n+1)} := E_1u_1^{(n+1,L)} + E_2u_2^{(n+1,M)}.
\end{array}
\right.
\tag{4.1}
$$

The behavior of this algorithm is somehow bizzare. Indeed, the algorithm usually alternates between the two subsequences given by $u^{(2n)}$ and its consecutive iteration $u^{(2n+1)}$. These two sequences are complementary, in the sense that they encode independent patterns of the solution. During the iterations the two subsequences slowly approach to each other, merging the complementary features and shaping the final limit which usually coincides with the wanted minimal solution, see Fig. 2. Unfortunately, this "oscillatory behavior" makes impossible to prove monotonicity of the sequence $(\mathcal{J}(u^{(n)}))_{n\in\mathbb{N}}$ and we have no proof of convergence. However, since the subsequences are early indicating different features of the

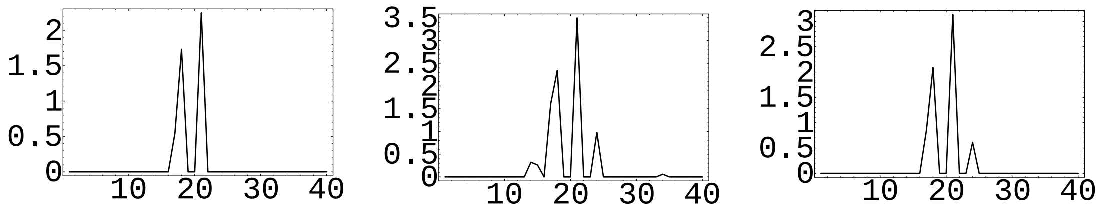


Figure 2: On the left we show $u^{(2n)}$, in the center $u^{(2n+1)}$, and on the right $u^{(\infty)}$. The two consecutive iterations contains different features which will appear in the solution.

final limit, we may modify the algorithm by substituting $u^{(n+1)} := E_1 u_1^{(n+1,L)} + E_2 u_2^{(n+1,M)}$ with $u^{(n+1)} := \frac{(E_1 u_1^{(n+1,L)} + E_2 u_2^{(n+1,M)}) + u^{(n)}}{2}$ that is the average of the current iteration and the previous one:

$$\begin{cases} \begin{cases} u_1^{(n+1,0)} = R_1 u^{(n)} \\ u_1^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_1^{(n+1,\ell)} + R_1 T^*((g - TE_2 R_2 u^{(n)}) - TE_1 u_1^{(n+1,\ell)}) \right) & \ell = 0, \ldots, L-1 \end{cases} \\ \begin{cases} u_2^{(n+1,0)} = R_2 u^{(n)} \\ u_2^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_2^{(n+1,\ell)} + R_2 T^*((g - TE_1 R_1 u^{(n)}) - TE_2 u_2^{(n+1,\ell)}) \right) & \ell = 0, \ldots, M-1 \end{cases} \\ u^{(n+1)} := \frac{(E_1 u_1^{(n+1,L)} + E_2 u_2^{(n+1,M)}) + u^{(n)}}{2}. \end{cases} \tag{4.2}$$

In the following we provide the convergence proof for the iterations in (4.2).

4.1 Weak convergence of the parallel DD algorithm

Theorem 4.1 (Weak convergence). *The algorithm in (4.2) produces a sequence $(u^{(n)})_{n\in\mathbb{N}}$ in $\ell_2(\Lambda)$ whose weak accumulation points are minimizers of the functional $\mathcal{J}$. In particular, the set of the weak accumulation points is non-empty and if $u^{(\infty)}$ is a weak accumulation point then*

$$u^{(\infty)} = \mathbb{S}_\tau(u^{(\infty)} + T^*(g - Tu^{(\infty)})).$$

Proof. By following the arguments in the proof of Theorem 3.3 we find

$$\mathcal{J}(u^{(n)}) - \mathcal{J}(E_1 u_1^{(n+1,L)} + E_2 R_2 u^{(n)}) \geq C \sum_{\ell=0}^{L-1} \|u_1^{(n+1,\ell+1)} - u_1^{(n+1,\ell)}\|_{\ell_2(\Lambda_1)}^2.$$

and

$$\mathcal{J}(u^{(n)}) - \mathcal{J}(E_1 R_1 u^{(n)} + E_2 u_2^{(n+1,M)}) \geq C \sum_{\ell=0}^{M-1} \|u_2^{(n+1,\ell+1)} - u_2^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_2)}.$$

By adding and halving the previous inequalities we obtain

$$\begin{aligned} & \mathcal{J}(u^{(n)}) - \frac{1}{2}\left(\mathcal{J}(E_1 u_1^{(n+1,L)} + E_2 R_2 u^{(n)}) + \mathcal{J}(E_1 R_1 u^{(n)} + E_2 u_2^{(n+1,M)})\right) \\ \geq \quad & \frac{C}{2}\left(\sum_{\ell=0}^{L-1} \|u_1^{(n+1,\ell+1)} - u_1^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_1)} + \sum_{\ell=0}^{M-1} \|u_2^{(n+1,\ell+1)} - u_2^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_2)}\right). \end{aligned}$$

By convexity we have

$$\begin{aligned} \left\|Tu^{(n+1)} - g\right\|^2_{\mathcal{H}} &= \left\|T\left(\frac{(E_1 u_1^{(n+1,L)} + E_2 u_2^{(n+1,M)}) + u^{(n)}}{2}\right) - g\right\|^2_{\mathcal{H}} \\ &\leq \frac{1}{2}\|T(E_1 u^{(n+1,L)} + E_2 R_2 u^{(n)}) - g\|^2_{\mathcal{H}} + \frac{1}{2}\|T(E_1 R_1 u^{(n)} + E_2 u^{(n+1,M)}) - g\|^2_{\mathcal{H}}. \end{aligned}$$

Moreover, by the triangle inequality we have

$$\begin{aligned} \|u^{(n+1)}\|_{\ell_1(\Lambda)} &\leq \frac{1}{2}\left(\|E_1 u_1^{(n+1,L)} + E_2 u_2^{(n+1,M)})\|_{\ell_1(\Lambda)} + \|u^{(n)}\|_{\ell_1(\Lambda)}\right) \\ &= \frac{1}{2}\left(\|E_1 u_1^{(n+1,L)} + E_2 R_2 u^{(n)})\|_{\ell_1(\Lambda)} + \|E_1 R_1 u^{(n)} + E_2 u_2^{(n+1,M)}\|_{\ell_1(\Lambda)}\right). \end{aligned}$$

By the last two inequalities we immediately show

$$\mathcal{J}(u^{(n+1)}) \leq \frac{1}{2}\left(\mathcal{J}(E_1 u_1^{(n+1,L)} + E_2 R_2 u^{(n)}) + \mathcal{J}(E_1 R_1 u^{(n)} + E_2 u_2^{(n+1,M)})\right),$$

hence

$$\mathcal{J}(u^{(n)}) - \mathcal{J}(u^{(n+1)}) \geq \frac{C}{2}\left(\sum_{\ell=0}^{L-1} \|u_1^{(n+1,\ell+1)} - u_1^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_1)} + \sum_{\ell=0}^{M-1} \|u_2^{(n+1,\ell+1)} - u_2^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_2)}\right). \tag{4.3}$$

We have $\mathcal{J}(u^{(0)}) \geq \mathcal{J}(u^{(n)}) \geq \tau\|u^{(n)}\|_{\ell_1(\Lambda)} \geq \tau\|u^{(n)}\|_{\ell_2(\Lambda)}$. This means that $(u^{(n)})_{n\in\mathbb{N}}$ is uniformly bounded in $\ell_2(\Lambda)$, hence there exists a weakly convergent subsequence $(u^{(n_j)})_{j\in\mathbb{N}}$. Let us denote $u^{(\infty)}$ the weak limit of the subsequence. For simplicity, we rename such subsequence by $(u^{(n)})_{n\in\mathbb{N}}$. Moreover, since the sequence $(\mathcal{J}(u^{(n)}))_{n\in\mathbb{N}}$ is monotonically decreasing and bounded from below by 0, it is also convergent. From (4.3) and the latter convergence we deduce

$$\left(\sum_{\ell=0}^{L-1} \|u_1^{(n+1,\ell+1)} - u_1^{(n+1,\ell)}\|^2_{\ell_2(\Lambda_1)} + \sum_{m=0}^{M-1} \|u_2^{(n+1,m+1)} - u_2^{(n+1,m)}\|^2_{\ell_2(\Lambda_2)}\right) \to 0, \quad n \to \infty. \tag{4.4}$$

In particular, by the standard inequality $(a^2+b^2) \geq \frac{1}{2}(a+b)^2$ for $a, b > 0$ and the triangle inequality, we have also

$$\begin{aligned}\sum_{\ell=0}^{L-1} \|u_1^{(n+1,\ell+1)} - u_1^{(n+1,\ell)}\|_{\ell_2(\Lambda_1)}^2 &\geq C'' \left(\sum_{\ell=0}^{L-1} \|u_1^{(n+1,\ell+1)} - u_1^{(n+1,\ell)}\|_{\ell_2(\Lambda_1)} \right)^2 \\ &\geq C'' \|E_1 u^{(n+1,L)} - E_1 R_1 u^{(n)}\|_{\ell_2(\Lambda)}^2 \\ &= C'' \|E_1 u^{(n+1,L)} + E_1 R_1 u^{(n)} - 2 E_1 R_1 u^{(n)}\|_{\ell_2(\Lambda)}^2 .\end{aligned}$$

Analogously we have

$$\sum_{\ell=0}^{M-1} \|u_2^{(n+1,\ell+1)} - u_2^{(n+1,\ell)}\|_{\ell_2(\Lambda_2)}^2 \geq C'' \|E_2 u^{(n+1,M)} + E_2 R_2 u^{(n)} - 2 E_2 R_2 u^{(n)}\|_{\ell_2(\Lambda)}^2 .$$

By denoting $C'' = \frac{1}{2} C'''$ we obtain

$$\begin{aligned}&\frac{C}{2} \left(\sum_{\ell=0}^{L-1} \|u_1^{(n+1,\ell+1)} - u_1^{(n+1,\ell)}\|_{\ell_2(\Lambda_1)}^2 + \sum_{\ell=0}^{M-1} \|u_2^{(n+1,\ell+1)} - u_2^{(n+1,\ell)}\|_{\ell_2(\Lambda_2)}^2 \right) \\ \geq\ & \frac{CC'''}{4} \|E_1 u^{(n+1,L)} + E_2 u^{(n+1,M)} + u^{(n)} - 2u^{(n)}\|_{\ell_2(\Lambda)}^2 \\ \geq\ & CC''' \|u^{(n+1)} - u^{(n)}\|_{\ell_2(\Lambda)}^2 .\end{aligned}$$

Therefore, we finally have

$$\|u^{(n)} - u^{(n+1)}\|_{\ell_2(\Lambda)} \to 0, \quad n \to \infty. \tag{4.5}$$

By definition of $u_1^{(n+1,L)}$ we have

$$0 \in [-2(u_1^{(n+1,L-1)} + R_1 T^*((g - T E_2 R_2 u^{(n)}) - T E_1 u_1^{(n+1,L-1)}))]_\lambda + 2 u_{\lambda,1}^{(n+1,L)} + \tau \partial |\cdot|(u_{\lambda,1}^{(n+1,L)}), \tag{4.6}$$

and by definition of $u_2^{(n+1,M)}$ we have

$$0 \in [-2(u_2^{(n+1,M-1)} + R_2 T^*((g - T E_1 R_1 u^{(n)}) - T E_2 u_2^{(n+1,M-1)}))]_\lambda + 2 u_{\lambda,2}^{(n+1,M)} + \tau \partial |\cdot|(u_{\lambda,2}^{(n+1,M)}). \tag{4.7}$$

Similarly to the argument used in the proof of Theorem 3.3, by taking now the limit for $n \to \infty$ in (4.6) and (4.7), and by using (4.4) we obtain

$$0 \in [-2(R_1 u^{(\infty)} + R_1 T^*((g - T E_2 R_2 u^{(\infty)}) - T E_1 R_1 u^{(\infty)}))]_\lambda + 2 u_\lambda^{(\infty)} + \tau \partial |\cdot|(u_\lambda^{(\infty)}),$$

for $\lambda \in \Lambda_1$ and

$$0 \in [-2(R_2 u^{(\infty)} + R_2 T^*((g - T E_1 R_1 u^{(\infty)}) - T E_2 R_2 u^{(\infty)})]_\lambda + 2 u_\lambda^{(\infty)} + \tau \partial |\cdot|(u_\lambda^{(\infty)}).$$

for $\lambda \in \Lambda_2$. In other words, we have

$$0 \in \partial_u \mathcal{J}^S(u^{(\infty)}, u^{(\infty)}).$$

An application of Lemma 3.1 and [14, Proposition 2.1] implies that

$$u^{(\infty)} = \mathbb{S}_\tau(u^{(\infty)} + T^*(g - T u^{(\infty)})).$$

We conclude the minimality of $u^{(\infty)}$ by an application of [14, Proposition 3.10]. ∎

4.2 Strong convergence of the parallel DD algorithm

By using the same notations as in Subsection 3.2, we can prove the convergence of the parallel domain decomposition algorithm (4.2).

Theorem 4.2 (Strong convergence)**.** *The algorithm in* (4.2) *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.*

Proof. Let $u^{(\infty)}$ be a weak accumulation point and let $(u^{(n_j)})_{j\in\mathbb{N}}$ be a subsequence weakly convergent to $u^{(\infty)}$. Let us denote the latter sequence $(u^{(n)})_{n\in\mathbb{N}}$ again. By Theorem 4.1 and (4.4) we have that $\xi^{(n)} = E_1\eta^{(n+1/2)} + E_2\mu^{(n+1/2)}$ weakly converges to zero. By substituting the use of (3.11) with the one of (4.4) whenever relevant and by substituting $\eta^{(n+1)}$ with $\eta^{(n)}$ in the proofs, one easily verifies that both Lemma 3.5 and Lemma 3.6 hold again. In particular, we have $\lim_{n\to\infty}\|\mathbb{S}_\tau(h+\xi^{(n)}) - \mathbb{S}_\tau(h) - \xi^{(n)}\|_{\ell_2(\Lambda)} = 0$. From Lemma 3.7 we conclude that $\xi^{(n)} = E_1\eta^{(n+1/2)} + E_2\mu^{(n+1/2)}$ converges to zero strongly. Again by (4.4) we have that $(u^{(n)})_{n\in\mathbb{N}}$ converges to $u^{(\infty)}$ strongly. ∎

REMARK: If $\mathcal{J}$ has a unique minimizer then necessarily the whole sequences $(u^{(n)})_{n\in\mathbb{N}}$ produced both by (3.2) and (4.2) converge in norm to it (and not only a subsequence.) Unfortunately, we could not prove the uniqueness of the accumulation point without this assumption, although numerical experiments support the conjecture that

1. the accumulation point is indeed unique;
2. it coincides with the *same* limit of the thresholded Landweber iterations.

A similar analysis can be provided for the minimization of $\mathcal{J}_\varepsilon(u) = \|g - Tu\|^2_{\mathcal{H}} + \tau\|u\|_1 + \varepsilon\|u\|_2^2$ via domain decompositions. In this case we have to consider in front of all the thresholding operations an additional scalar factor $\frac{1}{1+\varepsilon}$, giving, e.g., for the sequential algorithm, the following iterations

$$\begin{cases} \begin{cases} u_1^{(n+1,0)} = u_1^{(n,L)} \\ u_1^{(n+1,\ell+1)} = \frac{1}{1+\varepsilon}\mathbb{S}_\tau\left(u_1^{(n+1,\ell)} + R_1T^*((g - TE_2u_2^{(n,M)}) - TE_1u_1^{(n+1,\ell)})\right) & \ell = 0,\dots,L-1 \end{cases} \\ \begin{cases} u_2^{(n+1,0)} = u_2^{(n,M)} \\ u_2^{(n+1,\ell+1)} = \frac{1}{1+\varepsilon}\mathbb{S}_\tau\left(u_2^{(n+1,\ell)} + R_2T^*((g - TE_1u_1^{(n+1,L)}) - TE_2u_2^{(n+1,\ell)})\right) & \ell = 0,\dots,M-1 \end{cases} \\ u^{(n+1)} := E_1u_1^{(n+1,L)} + E_2u_2^{(n+1,M)}. \end{cases} \tag{4.8}$$

In this case, the functional $\mathcal{J}_\varepsilon$ is strictly convex, hence it has always a unique minimizer. Therefore, the whole sequences $(u^{(n)})_{n\in\mathbb{N}}$ produced, e.g., by (4.8) will converge to its minimizer.

5 On the Computational Cost

Let us assume that $\#\Lambda = N < \infty$ and that $\#\Lambda_1 = N/2$. For simplicity we assume that the matrix representing the operator T is full, so that the matrix-vector multiplication by

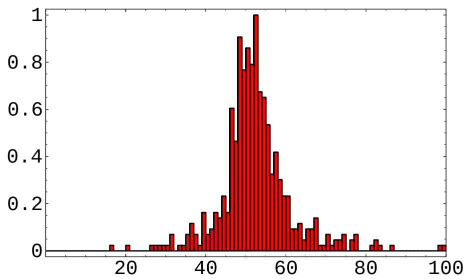


Figure 3: We assume $\mathcal{K} = \mathbb{R}^{40}$ and $\mathcal{H} = \mathbb{R}^{10}$, T is a 40×10 random matrix with $\|T\| < 1$, and $g \in \mathbb{R}^{10}$ is a random vector. We fix the regularization parameter $\tau = 0.1$. The figure shows the normalized frequency for multiple random trials of the percentage ratio between the number of operations required by the sequential domain decomposition method (3.2) in order to achieve an accuracy of 10^{-15} and the one required by the thresholded Landweber iteration (1.4). Here we have fixed $L = M = 8$. This experiment confirms that in most of the cases (i.e., with high probability) the computational cost due to (3.2) is half the one of (1.4).

the matrix T^*T costs $\mathcal{O}(N^2)$ algebraic operations. The computational cost of the original algorithm (1.4) is therefore $\mathcal{O}(N^2 \times n_{\max})$, where $n_{\max}$ is the number of iterations to achieve the desired accuracy. Here we have neglected the cost of $\mathbb{S}_\tau$ which in practice can be executed very rapidly (compared to the matrix-vector multiplication.) Let us now estimate the cost due to (3.2). For each outer iteration (indicated by the label n) we execute $L + M$ inner iterations (indicated by the label ℓ.) For each inner iteration we have to execute $\mathcal{O}\left(\frac{N}{2}\right)^2$ operations (due to the matrix-vector multiplications with halved dimension.) Therefore the total cost is given by $\mathcal{O}((L + M) \times \frac{N^2}{4} \times m_{\max})$, where $m_{\max}$ is the number of outer iterations to achieve the desired accuracy. In practice, we can verify experimentally (on random matrices T) that one can choose the parameters $L, M, m_{\max}$ so that $\frac{(L+M)\times m_{\max}}{4} \sim \frac{n_{\max}}{2}$ in order to achieve the *same* accuracy, see Figure 3. This means that by decomposing the problem as in (3.2) we can halve the computational cost. Note also that this operation does not imply any significant increasing of the complexity of the implementation. In particular, no parallelization is yet required. Indeed, algorithm (3.2) is fully sequential, i.e., it is implementable by a single processor.

We illustrate the characteristic dynamics of the thresholded Landweber iteration in Fig. 4 (on the top-left) by plotting the trajectory of the iterations $(\|u^{(n)}\|_1, \|Tu^{(n)} - g\|_{\mathcal{H}})$: Indeed, while the algorithm initially converges relatively fast, then it overshoots the limit value of $\|u^{(n)}\|_1$ and takes very long to re-correct back. We have to imagine that, starting from $u^{(0)} = 0$, the algorithm generate a path $\{u^{(n)}\}_{n\in\mathbb{N}}$ which is initially fully contained in the ℓ_1-ball $B_R := \{u \in \ell_2(\Lambda) : \|u\|_1 \leq R\}$, with $R := \|u^{(\infty)}(\tau)\|_1$. Then it gets out of the ball to come back to it only at the limit, typically on a vertex or on an edge of the ball (which correspond with regions where several components are indeed zero.) It is this "tail" which requires most of the computational cost. A similar dynamics is realized by the sequential domain decomposition method, but it is visible that the subspace corrections due

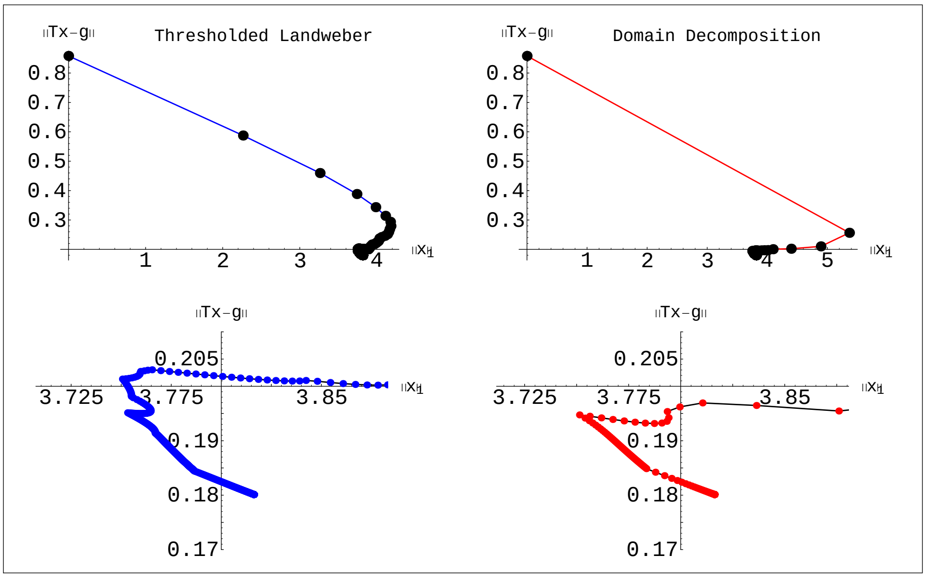


Figure 4: Dynamics of the iterations $(\|u^{(n)}\|_1, \|Tu^{(n)} - g\|_{\mathcal{H}})$ of the thresholded Landweber iterations (on the left) and of the sequential domain decomposition algorithm (on the right). On the bottom row we compare the final iterations ("the tail".)

to the inner iterations are indeed accelerating the convergence. Such acceleration becomes very relevant on the "tail", where the thresholded Landweber iteration is very slow, so that much fewer steps are needed to get to convergence. This acceleration compensates the effort due to few lower dimensional subspace corrections. This will not be true anymore for L and M too large and the trade-off between acceleration and computational cost has to be considered. Indeed, this is a rather common issue in domain decomposition methods. A theoretical *a priori* estimation of this trade-off is far from being achieved and a very interesting open problem.

6 Variations on a Theme

In this section we make explicit the generalization of the subspace correction algorithms to multiple decompositions. We split now the index set Λ into multiple disjoint sets $\Lambda_1, \Lambda_2, \dots \Lambda_{\mathcal{N}}$ so that $\Lambda = \bigcup_{i=1}^{\mathcal{N}} \Lambda_i$.

Associated to a decomposition $\mathcal{C} = \{\Lambda_1, \Lambda_2, \dots \Lambda_{\mathcal{N}}\}$ we define the *extension operators* $E_i : \ell_2(\Lambda_i) \to \ell_2(\Lambda)$, $(E_i v)_\lambda = v_\lambda$, if $\lambda \in \Lambda_i$, $(E_i v)_\lambda = 0$, otherwise, $i = 1, 2, \dots, \mathcal{N}$. Again we denote R_i the adjoint of E_i. For a sequence of natural numbers $L_1, \dots, L_{\mathcal{N}}$ we define the sequential iterations

$$
\begin{cases}
\begin{cases}
u_1^{(n+1,0)} = u_1^{(n,L_1)} \\
u_1^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_1^{(n+1,\ell)} + R_1 T^*((g - \sum_{i=2}^{\mathcal{N}} TE_i u_i^{(n,L_i)}) - TE_1 u_1^{(n+1,\ell)}) \right) \\
\ell = 0, \dots, L_1 - 1
\end{cases} \\
\dots \\
\begin{cases}
u_{\mathcal{N}}^{(n+1,0)} = u_{\mathcal{N}}^{(n,L_{\mathcal{N}})} \\
u_{\mathcal{N}}^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_{\mathcal{N}}^{(n+1,\ell)} + R_{\mathcal{N}} T^*((g - \sum_{i=1}^{\mathcal{N}-1} TE_i u_i^{(n+1,L_i)}) - TE_{\mathcal{N}} u_{\mathcal{N}}^{(n+1,\ell)}) \right) \\
\ell = 0, \dots, L_{\mathcal{N}} - 1
\end{cases} \\
u^{(n+1)} := \sum_{i=1}^{\mathcal{N}} E_i u_i^{(n+1,L_i)}.
\end{cases}
\tag{6.1}
$$

The analysis of this algorithm follows from a straightforward generalization of the case $\mathcal{N} = 2$ and the proofs are essentially identical. For the parallel version again we have to take into account a suitable average of two consecutive iterations together with the number $\mathcal{N}$ of patches. We obtain the following parallel algorithm

$$
\begin{cases}
\begin{cases}
u_1^{(n+1,0)} = u_1^{(n,L_1)} \\
u_1^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_1^{(n+1,\ell)} + R_1 T^*((g - \sum_{i=2}^{\mathcal{N}} TE_i R_i u^{(n)}) - TE_1 u_1^{(n+1,\ell)}) \right) \\
\ell = 0, \dots, L_1 - 1
\end{cases} \\
\dots \\
\begin{cases}
u_{\mathcal{N}}^{(n+1,0)} = u_{\mathcal{N}}^{(n,L_{\mathcal{N}})} \\
u_{\mathcal{N}}^{(n+1,\ell+1)} = \mathbb{S}_\tau \left(u_{\mathcal{N}}^{(n+1,\ell)} + R_{\mathcal{N}} T^*((g - \sum_{i=1}^{\mathcal{N}-1} TE_i R_i u^{(n)}) - TE_{\mathcal{N}} u_{\mathcal{N}}^{(n+1,\ell)}) \right) \\
\ell = 0, \dots, L_{\mathcal{N}} - 1
\end{cases} \\
u^{(n+1)} := \frac{\sum_{i=1}^{\mathcal{N}} E_i u_i^{(n+1,L_i)} + (\mathcal{N}-1) u^{(n)}}{\mathcal{N}}.
\end{cases}
\tag{6.2}
$$

With this modification the proof of convergence again follows from the approach considered for the case $\mathcal{N} = 2$.

Acknowledgment

The author acknowledges the financial support provided by the European Union's Human Potential Programme under contract MOIF-CT-2006-039438. He also thanks the Program in Applied and Computational Mathematics, Princeton University, for the hospitality during the preparation of this work.

References

[1] H. H. Bauschke, F. Deutsch, H. Hundal, and S-H. Park, *Accelerating the convergence of the method of alternating projections*, Trans. Americ. Math. Soc. **355** (2003), no. 9, 3433–3461.

[2] J. H. Bramble, J. E. Pasciak, J. Wang, and J. Xu, *Convergence estimates for product iterative methods with applications to domain decomposition*, Math. Comp. **57** (1991), no. 195, 1–21.

[3] 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.

[4] E. J. Candès and T. Tao, *Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?*, IEEE Trans. Inf. Theory **52** (2006), no. 12, 5406–5425.

[5] 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.

[6] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems.*, Numer. Math. **76** (1997), no. 2, 167–188.

[7] T. F. Chan and T. P. Mathew, *Domain decomposition algorithms*, Acta Numerica (1994), 61–143.

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

[9] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods II: Beyond the elliptic case*, Found. Comput. Math. **2** (2002), no. 3, 203–245.

[10] ———, *Adaptive methods for nonlinear variational problems*, SIAM J. Numer. Anal. **41** (2003), no. 5, 1785–1823.

[11] L. D. Cohen, *Auxiliary variables and two-step iterative algorithms in computer vision problems*, J. Math. Imaging Vision **6** (1996), no. 1, 59–83.

[12] S. Dahlke, M. Fornasier, and T. Raasch, *Adaptive frame methods for elliptic operator equations*, Adv. Comput. Math. (2007), doi: 10.1007/s10444-005-7501-6.

[13] 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. (2007), doi:10.1093/imanum/drl035.

[14] 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.

[15] I. Daubechies, M. Fornasier, and I. Loris, *Accelaration of the projected gradient method for linear inverse problems with sparsity constraints*, preprint (2007).

[16] 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.

[17] I. Daubechies, G. Teschke, and L. Vese, *Iteratively solving linear inverse problems under general convex constraints*, Inverse Probl. Imaging **1** (2007), no. 1, 29–46.

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

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

[20] I. Ekeland and R. Témam, *Convex analysis and variational problems*, SIAM, 1999.

[21] M. Elad, J.-L. Starck, P. Querre, and D.L. Donoho, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Appl. Comput. Harmon. Anal. **19** (2005), 340–358.

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

[23] 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.

[24] M. Fornasier and R. March, *Restoration of color images by vector valued BV functions and variational calculus*, preprint (2006).

[25] M. Fornasier and F. Pitolli, *Adatpive iterative thresholding algorithms for magnetoenceophalography (MEG)*, preprint (2007).

[26] M. Fornasier and H. Rauhut, *Recovery algorithms for vector valued data with joint sparsity constraints*, preprint (2006).

[27] ______, *Iterative thresholding algorithms*, preprint (2007).

[28] U. Langer, *Lecture notes on domian decomposition methods*, Department for Numerical Mathematics and Optimaization, Institute of Mathematics, Johannes Kepler University Linz (Austria).

[29] Y-J. Lee, J. Xu, and L. Zikatanov, *Successive subspace correction method for singular system of equations*, Fourteenth International Conference on Domain Decomposition Methods (I. Herrera, D. E. Keyes, O. B. Widlund, and R. Yates, eds.), UNAM, 2003, pp. 315–321.

[30] P. L. Lions, *On the Schwarz alternating method*, Proc. First Internat. Sympos. on Domain Decomposition Methods for Partial Differential Equations (R. Glowinski, G. H. Golub, G. A. Meurant, and J. Périaux, eds.), SIAM, Philadelphia, PA, 1988.

[31] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen, *Tomographic inversion using ℓ_1-norm regularization of wavelet coefficients*, preprint (2006).

[32] R. Nabben and D.B. Szyld, *Schwarz iterations for symmetric positive semidefinite problems*, Research Report 05-11-03, Department of Mathematics, Temple University, 2005.

[33] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc. **73** (1967), 591–597.

[34] A. Quarteroni and A. Valli, *Domain decomposition methods for partial differential equations*, Numerical Mathematics and Scientific Computation, The Clarendon Press Oxford University Press, New York, 1999, Oxford Science Publications.

[35] R. Ramlau and G. Teschke, *Tikhonov replacement functionals for iteratively solving nonlinear operator equations.*, Inverse Probl. **21** (2005), no. 5, 1571–1592.

[36] R. Ramlau and G. Teschke, *A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints*, Numer. Math. **104** (2006), no. 2, 177–203.

[37] R.T. Rockafellar and R.J.B. Wets, *Variational analysis*, Grundlehren der Mathematischen Wissenschaften, vol. 317, Springer-Verlag, Berlin, 1998.

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

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

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

[41] G. Teschke, *Multi-frame representations in linear inverse problems with mixed multi-constraints*, Appl. Comput. Harm. Anal. **22** (2007), no. 1, 43–60.

[42] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B **58** (1996), no. 1, 267–288.

[43] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Rev. **34** (1992), no. 4, 581–613.

[44] J. Xu and L. Zikatanov, *The method of alternating projections and the method of subspsce corrections in Hilbert space*, Report AM223, Department of Mathematics, Pann State University, 2000.