SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

SUBSPACE CORRECTION METHODS FOR A CLASS OF NON-SMOOTH AND NON-ADDITIVE CONVEX VARIATIONAL PROBLEMS IN IMAGE PROCESSING

M. Hintermueller A. Langer

SFB-Report No. 2012-021 October 2012

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

SUBSPACE CORRECTION METHODS FOR A CLASS OF NON-SMOOTH AND NON-ADDITIVE CONVEX VARIATIONAL PROBLEMS IN IMAGE PROCESSING*

MICHAEL HINTERMÜLLER† AND ANDREAS LANGER‡

Abstract. The minimization of a functional composed of a non-smooth and non-additive regularization term and a combined L^1 and L^2 data-fidelity term is proposed. It is shown analytically and numerically that the new model has noticeable advantages over popular models in image processing tasks. For the numerical minimization of the new objective, subspace correction methods are introduced which guarantee the convergence and monotone decay of the associated energy along the iterates. Moreover, an estimate of the distance between the outcome of the subspace correction method and the global minimizer of the non-smooth objective is derived. This estimate and numerical experiments for image denoising, inpainting, and deblurring show that in practice the proposed subspace correction methods indeed converge to the global solution of the underlying minimization problem.

Key words. subspace correction, domain decomposition, total variation minimization, convex optimization, image restoration, combined L^1/L^2 data-fidelity, convergence analysis

AMS subject classifications. 68U10, 94A08, 49M27, 65K10, 90C06

1. Introduction. Subspace correction is a divide and conquer technique originally proposed for the numerical solution of partial differential equations. Algorithmically this is achieved by iteratively solving on each subspace an appropriately defined subproblem, which, in a variational setting, typically amounts to minimizing a smooth energy. For the overall algorithm, convergence, rate of convergence, and the independence of the rate of convergence from the mesh size of discretization are well-established.

For non-smooth problems, the resulting splitting algorithms still work fine as long as the energy splits additively with respect to the subspace decomposition. For such problems convergence and sometimes even rate of convergence are ensured; see e.g. [25, 46]. Moreover, for image deblurring problems preconditioning effects of a specific subspace correction algorithm for minimizing a non-smooth energy are shown in [48]. For non-smooth and *non-additive* energies, however, the research on subspace correction methods is far from being complete, and for some problem classes counterexamples do exist indicating failure of subspace correction; see e.g. [26, 49].

From a computational point of view, one of the appeals of subspace correction methods is given by the fact that parallel algorithms can be devised which exploit the capabilities of multiprocessor or multicore computer architectures. Main advantages of associated iterative solvers include (i) dimension reduction; (ii) enhancement of parallelism; (iii) localized treatment of complex and irregular geometries, singularities and anomalous regions; (iv) and sometimes reduction of the computational complexity of the underlying solution method. Among the important representatives of this algorithm class one finds the Jacobi method, the Gauss-Seidel method, point or block relaxation methods, multigrid methods, and domain decomposition methods. For further details on subspace correction and associated solvers we refer to [51].

In this paper we focus on subspace correction methods for a class of non-smooth and non-additive problems which arise in mathematical image processing. In this area the importance of devising such methods is clearly motivated by the continuous improvement of imaging hardware, which allows to increase resolutions or to acquire vast amounts of data. In the context of variational

*This work was supported by the Austrian Science Fund FWF through the START Project Y 305-N18 “Interfaces and Free Boundaries” and the SFB Project F32 04-N18 “Mathematical Optimization and Its Applications in Biomedical Sciences” as well as by the German Research Fund (DFG) through the Research Center MATHEON Project C28 and the SPP 1253 “Optimization with Partial Differential Equations”. M.H. also acknowledges support through a J. Tinsely Oden Fellowship at the Institute for Computational Engineering and Sciences (ICES) at UT Austin, Texas, USA.

†Department of Mathematics, Humboldt-University of Berlin, Unter den Linden 6, 10099 Berlin, Germany Email: hint@math.hu-berlin.de

‡Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria Email: andreas.langer@uni-graz.at

methods in image processing, this may lead to extremely large-scale problems which need to be processed routinely.

In image restoration, the non-smooth and non-additive total variation (TV), proposed in [42] for image denoising, plays a fundamental role as a regularization technique, since it preserves edges and discontinuities in images. In this context, one typically minimizes an energy that consists of a data-fidelity term, which enforces the consistency between the recovered and the measured image, and the total variation as the regularization term. The choice of the data term usually depends on the type of noise contained in the measured image data. In this vein, for images corrupted by *Gaussian noise* a quadratic L^2 data-fidelity term has been successfully used in first order methods, see e.g. [11, 12, 13, 16, 18, 19, 20, 21, 22, 29, 37, 40, 50, 54], as well as in second order methods, see e.g. [31]. In this approach, which we refer to as the L^2-TV model, the image u is recovered from the observed data g by solving

$$\min_u \alpha\|Tu-g\|^2_{L^2(\Omega)} + |Du|(\Omega), \tag{1.1}$$

where $\Omega \subset \mathbb{R}^2$ is an open bounded set with Lipschitz boundary, T is a bounded linear operator modelling the image-formation device (if the image is only corrupted by noise one sets $T = I$), and $\alpha > 0$ is a parameter. We recall, that for $u \in L^1(\Omega)$

$$V(u,\Omega) := \sup\left\{ \int_\Omega u \operatorname{div}\phi dx : \phi \in [C_c^1(\Omega)]^2, \|\phi\|_\infty \le 1 \right\}$$

is the variation of u. In the event that $V(u,\Omega) < \infty$ we denote $|Du|(\Omega) = V(u,\Omega)$ and call it the total variation of u in Ω; see [2] for more details. If $u \in W^{1,1}(\Omega)$, then $|Du|(\Omega) = \int_\Omega |\nabla u| dx$. The L^2-TV model usually fails in the presence of *salt-and-pepper noise*, where the noisy image g, throughout assumed to have a dynamic range of $c_{\min} \le g \le c_{\max}$, is given by

$$g(x) = \begin{cases} c_{\min} & \text{with probability } p_1 \in [0,1), \\ c_{\max} & \text{with probability } p_2 \in [0,1), \\ u(x) & \text{with probability } 1 - p_1 - p_2, \end{cases}$$

with $1 - p_1 - p_2 > 0$ [14]. Here, $p_1 + p_2$ defines the noise level. Recently a non-smooth L^1 data-fidelity term was suggested in [1], which treats impulse noise (e.g. salt-and-pepper noise) more successfully than a quadratic L^2 data term [38, 39, 23], i.e., instead of (1.1) one considers

$$\min_u \alpha\|Tu-g\|_{L^1(\Omega)} + |Du|(\Omega),$$

which we call the L^1-TV model.

In the case of simultaneous Gaussian and salt-and-pepper noise the choice of the data-fidelity is unclear, and the literature on this subject appears rather scarce. In order to accommodate such situations, a two-phase reconstruction approach is suggested in [9]. In fact, in the first phase (most of) the outliers are detected and in the second phase a variational functional consisting of a Mumford-Shah regularizer, which renders the problem non-convex, is minimized. In contrast to this development we tackle the problem of removing simultaneous Gaussian and salt-and-pepper noise by optimizing a convex functional with a total variation regularizer and a combination of a quadratic L^2-term and a non-smooth L^1-term. It turns out in our numerical experiments that such a combined data-fidelity term well suits the restoration task; see Figure 5.2 below. Analytically we show by means of an explicit example that the minimization of the newly proposed functional has noticeable advantages over the standard functionals, L^2-TV or L^1-TV model. Algorithmically, we adapt the approach in [4], which was originally proposed for solving the L^1-TV model only, to our case of a combined data-fidelity term.

As all of the aforementioned solvers for TV-minimization are confined to small and medium scale problems only, we propose and analyze subspace correction, domain decomposition, and coordinate descent methods as these are fundamental for reducing the overall problem to a finite number of subproblems with each of them of a size manageable for the above TV solvers.

Recently, in [26, 27, 28] non-overlapping and overlapping domain decomposition strategies were introduced for solving the L^2-TV problem. In this context, the major difficulty lies in the correct treatment of the interfaces of the domain decomposition patches, i.e. the preservation of crossing discontinuities and the correct matching where the solution is continuous. We emphasize that well-known approaches as those in [10, 17, 43, 44] are not directly applicable to the non-smooth and non-additive L^2-TV problem. In [27, 28] the convex objective under some linear constraint, ensuring the correct treatment of the internal interfaces, was iteratively minimized on each subdomain. While in these two papers an implementation guaranteeing convergence and monotonic decay of the objective energy is provided, convergence to the global minimizer of the L^2-TV problem *cannot* be ensured, in general. For one-dimensional problems, in [27] a proof is presented which establishes convergence of the overlapping domain decomposition algorithm to the global solution. We note that although this proof is carried out for any finite dimensional space, it is not yet clear how to prove convergence to the expected minimizer without further (and practically possibly critical) assumptions on the overlapping region for higher dimensions ($d > 1$).

In [26] a wavelet decomposition method is presented with similar properties as the aforementioned non-overlapping domain decomposition methods. In that paper, an additional condition is imposed which allows to establish global optimality of the limit point obtained by the domain decomposition method. Unfortunately, despite the good practical behavior of the method, this condition cannot be ensured to hold in general, as counterexamples have shown. Thus, unfortunately, with the aforementioned condition one can only check a posteriori whether the algorithm found the global minimizer or whether it failed to do so. Moreover, no error estimates are available.

In the present paper we generalize the subspace correction strategy to more general functionals, which consist of a non-smooth and non-additive regularization term and a weighted combination of an L^1-term and a quadratic L^2-term; see (2.1) below. In this setting, the L^2-TV model considered in [26, 27, 28] and the L^1-TV model are special instances. Note that the methods in [26, 27, 28] differ from our approach. In fact, in [26, 27, 28] each subspace minimization problem is approximated by a surrogate functional minimization, while we are minimizing on each subspace the exact subspace minimization problem. Thus, a different convergence analysis is required. Similarly to the domain decomposition methods in [27, 28] we are able to show that our subspace correction methods for the newly introduced functional are guaranteed to converge and to monotonically decrease the energy. In addition, we are able to establish an estimate of the distance of the limit point obtained from the subspace correction method to the true global minimizer. With the help of this estimate, we show in our numerical experiments that the sequence generated by our proposed algorithm indeed numerically converges to the expected minimizer of the objective functional.

The rest of the paper is organized as follows: In Section 2 we state the problem of interest and motivate the choice of the objective functional by an illustrative example. Moreover, we introduce our alternating and parallel subspace correction methods in a Banach space setting and state some convergence properties. Main notations and basic definitions are provided in Section 3. In Section 4 we describe the problem in a discrete setting and show optimality properties of our subspace correction methods which allow us to estimate the distance of a limit point obtained by subspace correction to the minimizer of the total energy. In Section 5 we present our subspace correction methods for the special cases of overlapping and non-overlapping domain decomposition, respectively. Details on the implementation of the solvers for the global minimization problem as well as for the domain decomposition methods are provided. Finally we show sequential and parallel numerical experiments for total variation minimization.

2. Problem Statement.

We are interested in solving the following minimization problem

$$\min_{u \in L^2(\Omega)} J_{\alpha_1,\alpha_2}(u) := \alpha_1 \|T_1 u - g_1\|_{L^1(\Omega)} + \alpha_2 \|T_2 u - g_2\|^2_{L^2(\Omega)} + \varphi(|Du|)(\Omega), \tag{2.1}$$

where $T_i : L^2(\Omega) \to L^2(\Omega)$ is a bounded linear operator, $g_i \in L^2(\Omega)$ is a given datum, $\alpha_i \geq 0$ for $i = 1, 2$, with $\alpha_1 + \alpha_2 > 0$, $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, and $\varphi(|\cdot|)$ is a convex function of measures representing regularization.

In what follows we assume that J_{α_1,α_2} is bounded from below and coercive, i.e., $\{J_{\alpha_1,\alpha_2} \leq$

$C\} = \{u \in L^2(\Omega) : J_{\alpha_1,\alpha_2}(u) \leq C\}$ is bounded in $L^2(\Omega)$ for all constants $C > 0$, in order to guarantee that problem (2.1) has solutions. Moreover we assume that

(A_φ) $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, nondecreasing in $\mathbb{R}^+$ with
- (i) $\varphi(0) = 0$.
- (ii) $cz - b \leq \varphi(z) \leq cz + b$, for all $z \in \mathbb{R}^+$ for some constant $c > 0$ and $b \geq 0$.

Note that for the particular example $\varphi(t) = t$, the third term in (2.1) becomes the well-known total variation of u in Ω and we call then (2.1) the L^1-L^2-TV model. Other functions which fulfill assumption (A_φ) are $\varphi(t) = \sqrt{1+t^2} - 1$ (the function of minimal surfaces) and $\varphi(t) = \log\cosh t$ [47].

In order to motivate our proposed model (2.1), we use a simple and illustrative example in 2D, where $\varphi(t) = t$, which we compare with the L^1-TV model, i.e., when $\alpha_2 = 0$ in (2.1), and with the L^2-TV model, i.e., when $\alpha_1 = 0$ in (2.1).

EXAMPLE 2.1. *Let the observed image $g_1 = g_2$ be the characteristic function $1_{B_r(0)}$ of a disk $B_r(0)$ centered at the origin with radius $r > 0$. We are interested in the explicit solution of the problem in (2.1) when $\Omega = \mathbb{R}^2$ and $\varphi(t) = t$ for the following three different cases: (i) $\alpha_1 = 0, \alpha_2 > 0$ (L^2-TV), (ii) $\alpha_1 > 0, \alpha_2 = 0$ (L^1-TV), (iii) $\alpha_1 > 0, \alpha_2 > 0$ (L^1-L^2-TV), when the operator $T_1 = T_2 = I$ is the identity operator, respectively.*

For the first two cases we recall the solutions found in [15, 36].

(i) For $\alpha_1 = 0, \alpha_2 > 0$ the unique minimizer u_{0,α_2} is given by

$$u_{0,\alpha_2} = \begin{cases} 0 & \text{if } 0 \leq r < \frac{1}{\alpha_2}, \\ \left(1 - \frac{1}{\alpha_2 r}\right) 1_{B_r(0)} & \text{if } r \geq \frac{1}{\alpha_2}. \end{cases}$$

(ii) For $\alpha_1 > 0, \alpha_2 = 0$ a minimizer $u_{\alpha_1,0}$ is given by

$$u_{\alpha_1,0} \in \begin{cases} \{0\} & \text{if } 0 \leq r < \frac{2}{\alpha_1}, \\ \{c1_{B_r(0)} : c \in [0,1]\} & \text{if } r = \frac{2}{\alpha_1}, \\ \{1_{B_r(0)}\} & \text{if } r > \frac{2}{\alpha_1}. \end{cases}$$

(iii) For $\alpha_1, \alpha_2 > 0$ one can reason that every minimizer has to be of the form $c1_{B_r(0)}$ for $c \in [0,1]$. Therefore we just need to minimize the function

$$J_{\alpha_1,\alpha_2}(c1_{B_r(0)}) = \alpha_1 \pi r^2 |1-c| + \alpha_2 \pi r^2 (1-c)^2 + 2\pi r c$$

over $c \in [0,1]$. Then the optimality condition for c is given by

$$-\alpha_1 \pi r^2 - 2\alpha_2 \pi r^2 (1-c) + 2\pi r = 0,$$

which is equivalent to

$$c = \frac{2\alpha_2 + \alpha_1}{2\alpha_2} - \frac{1}{\alpha_2 r}.$$

Hence, the unique minimizer is given by

$$u_{\alpha_1,\alpha_2} = \begin{cases} 0 & \text{if } 0 \leq r < \frac{2}{2\alpha_2+\alpha_1}, \\ \left(\frac{2\alpha_2+\alpha_1}{2\alpha_2} - \frac{1}{\alpha_2 r}\right) 1_{B_r(0)} & \text{if } \frac{2}{2\alpha_2+\alpha_1} \leq r \leq \frac{2}{\alpha_1}, \\ 1_{B_r(0)} & \text{if } r > \frac{2}{\alpha_1}. \end{cases}$$

From this example we clearly see the difference between the L^2-TV model and the L^1-TV model. When the L^1-fidelity is used, then the solution is constant except at a special value ($r = \frac{2}{\alpha_1}$) where it undergoes a sudden transition. When in addition to the L^1-fidelity also the L^2-term is present then the solution is constant except in an interval ($\frac{2}{2\alpha_2+\alpha_1} \leq r \leq \frac{2}{\alpha_1}$) where it experiences a smooth transition. On the contrary, when only the L^2-fidelity plus TV-term is used then the solution is only constant for $0 \leq r < \frac{2}{\alpha_2}$ and hyperbolically increasing otherwise.

The differences between the L^2-TV model and the L^1-TV model result in the following observation: Fix $\alpha_1 = \alpha_2 = \alpha > 0$ and set, as in Example 2.1, $g_1 = g_2 = 1_{B_r(0)}$ and $T_1 = T_2 = I$. Then the solution $u_{0,\alpha}$ of the L^2-TV model is identically 0 if $r < \frac{1}{\alpha}$. This is clearly an advantage over the L^1-TV model, where $u_{\alpha,0} = 0$ if $r < \frac{2}{\alpha}$, since smaller features can be maintained with the L^2-TV model. On the contrary, the L^2-TV model is not able to preserve the original features perfectly (except if $\alpha = \infty$) but only obtains them with a loss of energy, i.e., $u_{0,\alpha} = \left(1 - \frac{1}{\alpha r}\right) 1_{B_r(0)}$ if $r \geq \frac{1}{\alpha}$. This is different for the L^1-TV model, where we have that $u_{\alpha,0} = 1_{B_r(0)}$ if $r > \frac{2}{\alpha}$ and hence features can be perfectly preserved. This is naturally a clear advantage of the latter model.

For the combined L^1-L^2-TV model, we observe that $u_{\alpha,\alpha} = 0$ if $0 \leq r < \frac{2}{3\alpha}$ and hence even smaller features as with the L^2-TV model can be maintained. But still we are able to preserve original features perfectly as in the L^1-TV model. Moreover the transition between the just mentioned constant states is smooth, which is clearly a property coming from the L^2-term, which renders the solution unique.

2.1. Subspace Correction Approach. As in [26, 27, 28] for the L^2-TV model, we seek to minimize J_{α_1,α_2} by decomposing $L^2(\Omega)$ into two subspaces V_1 and V_2 such that $L^2(\Omega) = V_1 + V_2$. With this splitting we aim to solve (2.1) by the following alternating algorithm:

Choose an initial $u^{(0)} =: \tilde{u}_1^{(0)} + \tilde{u}_2^{(0)} \in V_1 + V_2$, for example, $u^{(0)} = 0$, and iterate

$$\begin{cases} u_1^{(n+1)} = \arg\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + \tilde{u}_2^{(n)}), \\ u_2^{(n+1)} = \arg\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + u_2), \\ u^{(n+1)} := u_1^{(n+1)} + u_2^{(n+1)}, \\ \tilde{u}_1^{(n+1)} = \chi_1 \cdot u^{(n+1)}, \\ \tilde{u}_2^{(n+1)} = \chi_2 \cdot u^{(n+1)}, \end{cases} \tag{2.2}$$

where $\chi_1, \chi_2 \in L^\infty(\Omega)$ have the properties (a) $\chi_1 + \chi_2 = 1$ and (b) $\chi_i \in V_i$ for $i = 1, 2$. Let $\kappa := \max\{\|\chi_1\|_\infty, \|\chi_2\|_\infty\} < \infty$. From these assumptions on χ_i we obtain that if the V_i's are orthogonal, i.e., $L^2(\Omega) = V_1 \oplus V_2$, then $\tilde{u}_i^{(n)} = u_i^{(n)}$ for all $n \in \mathbb{N}$ and, hence, in this case there is no need to introduce the variables $\tilde{u}_i^{(n)}$, cf. with (5.2) below. The parallel version of the algorithm in (2.2) reads as follows:

Choose an initial $u^{(0)} =: \tilde{u}_1^{(0)} + \tilde{u}_2^{(0)} \in V_1 + V_2$, for example, $u^{(0)} = 0$, and iterate

$$\begin{cases} u_1^{(n+1)} = \arg\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + \tilde{u}_2^{(n)}), \\ u_2^{(n+1)} = \arg\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n)} + u_2), \\ u^{(n+1)} := \frac{u_1^{(n+1)} + u_2^{(n+1)} + u^{(n)}}{2}, \\ \tilde{u}_1^{(n+1)} = \chi_1 \cdot u^{(n+1)}, \\ \tilde{u}_2^{(n+1)} = \chi_2 \cdot u^{(n+1)}. \end{cases} \tag{2.3}$$

We define the orthogonal complement of V_i in $L^2(\Omega)$ by V_i^c, i.e., $L^2(\Omega) = V_i \oplus V_i^c$ and we define by π_{V_i} the corresponding orthogonal projection onto V_i. Moreover, we define the domain of a functional $\mathcal{J} : L^2(\Omega) \to \bar{\mathbb{R}}$ as the set $\mathrm{Dom}(\mathcal{J}) = \{v \in L^2(\Omega) : \mathcal{J}(v) \neq \infty\}$.

Note that the subspace minimization problems in (2.2) and (2.3) can be written as constrained optimization problems of the form

$$\min_{v \in L^2(\Omega)} J_{\alpha_1,\alpha_2}(v) \quad \text{subject to (s.t.) } Av = b,$$

where $A : L^2(\Omega) \to L^2(\Omega)$ is a linear and continuous operator on $L^2(\Omega)$ and $b \in L^2(\Omega)$. In particular, we have

$$\min_{v \in L^2(\Omega)} J_{\alpha_1,\alpha_2}(v + b) \quad \text{s.t. } \pi_{V_i^c} v = 0,$$

or equivalently

$$\min_{v\in L^2(\Omega)} J_{\alpha_1,\alpha_2}(v) \quad \text{s.t. } \pi_{V_i^c}(v) = \pi_{V_i^c}(b), \tag{2.4}$$

where $b = u_1^{(n+1)}$ for the second minimization problem in (2.2) and $b = \tilde{u}_j^{(n)}$ for the first minimization problem in (2.2) and the minimization problems in (2.3) for $i = 1, 2$ and $j \in \{1, 2\} \setminus \{i\}$.

For any *attainable* $b \in V_j$, i.e., there exists an $u \in \mathrm{Dom}(J_{\alpha_1,\alpha_2})$ such that $\pi_{V_i^c}(u) = \pi_{V_i^c}(b)$, we observe that $\{u \in L^2(\Omega) : \pi_{V_i^c}(u) = \pi_{V_i^c}(b), J_{\alpha_1,\alpha_2}(u) \le C\} \subset \{J_{\alpha_1,\alpha_2} \le C\}$ for all $C > 0$, $i = 1, 2$, and $j \in \{1, 2\} \setminus \{i\}$. Hence the former set is bounded by the coercivity assumption and thus (2.4) has a solution, since every $u_i^{(n)}$ and $\tilde{u}_i^{(n)}$ generated by the algorithm in (2.2) and (2.3) is attainable.

Proposition 2.2. *The algorithms in* (2.2) *and* (2.3) *produce a sequence* $(u^{(n)})_n$ *in* $L^2(\Omega)$ *with the following properties:*

(i) $J_{\alpha_1,\alpha_2}(u^{(n)}) \ge J_{\alpha_1,\alpha_2}(u^{(n+1)})$ *for all* $n \in \mathbb{N}$*;*

(ii) The sequence $(u^{(n)})_n$ *has subsequences that weakly converge in* $L^2(\Omega)$ *and* $BV(\Omega)$*.*

Proof. First let us show (i) for the algorithm in (2.2). Observe that

$$J_{\alpha_1,\alpha_2}(u^{(n)}) = J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n)}+\tilde{u}_2^{(n)}) \ge J_{\alpha_1,\alpha_2}(u_1^{(n+1)}+\tilde{u}_2^{(n)}) \ge J_{\alpha_1,\alpha_2}(u_1^{(n+1)}+u_2^{(n+1)}) = J_{\alpha_1,\alpha_2}(u^{(n+1)}),$$

which proves the assertion.
To show (i) for the algorithm in (2.3) we consider first that

$$J_{\alpha_1,\alpha_2}(u^{(n)}) \ge \frac{1}{2}\left(J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + \tilde{u}_2^{(n)}) + J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n)} + u_2^{(n+1)})\right).$$

Moreover by convexity we obtain

$$J_{\alpha_1,\alpha_2}\left(\frac{u_1^{(n+1)} + u_2^{(n+1)} + \tilde{u}_1^{(n)} + \tilde{u}_2^{(n)}}{2}\right) \le \frac{1}{2}\left(J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + \tilde{u}_2^{(n)}) + J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n)} + u_2^{(n+1)})\right)$$

and hence $J_{\alpha_1,\alpha_2}(u^{(n)}) \ge J_{\alpha_1,\alpha_2}(u^{(n+1)})$.
From the above considerations we infer that $J_{\alpha_1,\alpha_2}(u^{(0)}) \ge J_{\alpha_1,\alpha_2}(u^{(n)})$ for all $n \in \mathbb{N}$. By the coercivity condition on J_{α_1,α_2}, $(u^{(n)})_n$ is uniformly bounded in $L^2(\Omega)$ and hence there exists a weakly convergent subsequence. Moreover, due to the presence of $\varphi(|Du|)$ in J_{α_1,α_2} and $\alpha_1+\alpha_2 > 0$ we obtain that $(u^{(n)})_n$ is bounded in $BV(\Omega)$. The compact embedding $BV(\Omega) \hookrightarrow L^q(\Omega)$, $q < \frac{d}{d-1}$, implies that a subsequence $(u^{(n_k)})_k$ converges in $L^q(\Omega)$ to a limit $u^{(\infty)} \in L^2(\Omega)$. By [3, Prop. 10.1.1] we even have that $u^{(\infty)} \in BV(\Omega)$, $\liminf_{n\to\infty} \varphi(|Du^{(n_k)}|)(\Omega) \ge \varphi(|Du^{(\infty)}|)(\Omega)$, and $u^{(n_k)}$ weakly converges to $u^{(\infty)}$ in $BV(\Omega)$, which concludes the proof. ∎

Remark 2.3. *Since the sequence* $(J_{\alpha_1,\alpha_2}(u^{(n)}))_n$ *is monotonically decreasing and bounded from below, it is also convergent.*

Proposition 2.4. *The sequences* $(u_i^{(n)})_n$ *and* $(\tilde{u}_i^{(n)})_n$ *for* $i = 1, 2$ *generated by the algorithm in* (2.2) *or* (2.3) *are bounded in* $L^2(\Omega)$ *and hence have weak accumulation points* $u_i^{(\infty)} \in L^2(\Omega)$ *and* $\tilde{u}_i^{(\infty)} \in L^2(\Omega)$*, respectively.*

Proof. The boundedness of $(u^{(n)})_n$ implies the boundedness of $(\tilde{u}_i^{(n)})_n$, since

$$\|\tilde{u}_i^{(n)}\|_{L^2(\Omega)} = \|\chi_i u^{(n)}\|_{L^2(\Omega)} \le \kappa \|u^{(n)}\|_{L^2(\Omega)} \le C < \infty \quad \text{for } i = 1, 2. \tag{2.5}$$

By the definition of $u_1^{(n+1)}$ and by the coercivity assumption on J_{α_1,α_2} we have that $\left(u_1^{(n+1)} + \tilde{u}_2^{(n)}\right)_n$ is bounded in $L^2(\Omega)$, i.e., there exists a constant $C > 0$ such that $\|u_1^{(n+1)} + \tilde{u}_2^{(n)}\|_{L^2(\Omega)} \le C$ for all $n \in \mathbb{N}$. Since $(\tilde{u}_2^{(n)})_n$ is bounded in $L^2(\Omega)$ by (2.5), the triangle inequality yields

$$\|u_1^{(n+1)}\|_{L^2(\Omega)} - \|\tilde{u}_2^{(n)}\|_{L^2(\Omega)} \le \|u_1^{(n+1)} + \tilde{u}_2^{(n)}\|_{L^2(\Omega)} \le C.$$

Hence $(u_1^{(n)})_n$ is bounded in $L^2(\Omega)$. By similar arguments we get the $L^2(\Omega)$-boundedness of $(u_2^{(n)})_n$. Consequently $(\tilde{u}_i^{(n)})_n$ and $(u_i^{(n)})_n$ have a weakly convergent subsequence with limits $\tilde{u}_i^{(\infty)}$ and $u_i^{(\infty)}$, respectively. ∎

3. Notations and Basic Definitions. In the rest of the paper we eventually work on a finite dimensional space by considering a finite regular mesh as a discretization of Ω. Essentially we approximate functions u by discrete functions, again denoted by u, and denote their gradient by ∇u. Thus we consider instead of the continuous functional (2.1) its discrete approximation, for ease again denoted by J_{α_1,α_2} in (4.1) below. Note that the discrete approximation (4.1) Γ-converges to the continuous functional (2.1), see [7, 34], and has the same singular nature as the continuous problem. Although in our applications we are mainly interested in imaging problems, i.e., two-dimensional problems, we keep the notation more general for any d-dimensional space.

In our discrete setting we define the *discrete d-orthotope* $\Omega = \{x_1^1 < \ldots < x_{N_1}^1\} \times \ldots \times \{x_1^d < \ldots < x_{N_d}^d\} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, and the considered "function spaces" are $\mathcal{H} = \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_d}$, where $N_i \in \mathbb{N}$ for $i = 1, \ldots, d$. For $u \in \mathcal{H}$ we write $u = u(x) = u(x_{i_1}^1, \ldots, x_{i_d}^d)$, where $i_k \in \{1, \ldots, N_k\}$ and $x \in \Omega$. Let $h = x_{i_k+1}^k - x_{i_k}^k$ be the equidistant step-size for all $k = 1, \ldots, d$. We define the scalar product of $u, v \in \mathcal{H}$ by

$$\langle u, v\rangle_{\mathcal{H}} = h^d \sum_{x\in\Omega} u(x)v(x)$$

and the scalar product of $p, q \in \mathcal{H}^d$ by

$$\langle p, q\rangle_{\mathcal{H}^d} = h^d \sum_{x\in\Omega} \langle p(x), q(x)\rangle_{\mathbb{R}^d}$$

with $\langle y, z\rangle_{\mathbb{R}^d} = \sum_{j=1}^d y_j z_j$ for every $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ and $z = (z_1, \ldots, z_d) \in \mathbb{R}^d$. In what follows we consider different norms. In particular we use

$$\|u\|_{\ell^p(\Omega)} = \left(h^d \sum_{x\in\Omega} |u(x)|^p\right)^{1/p}, \quad 1 \le p < \infty,$$

and

$$\|u\|_{\ell^\infty(\Omega)} = \sup_{x\in\Omega} |u(x)|.$$

Sometimes we do not specify the norm, i.e., we just write $\|\cdot\|$, which indicates that any norm can be taken.

The discrete gradient ∇u is denoted by $(\nabla u)(x) = ((\nabla u)^1(x), \ldots, (\nabla u)^d(x))$ with

$$(\nabla u)^j(x) = \frac{1}{h} \cdot \begin{cases} u(x_{i_1}^1, \ldots, x_{i_j+1}^j, \ldots, x_{i_d}^d) - u(x_{i_1}^1, \ldots, x_{i_j}^j, \ldots, x_{i_d}^d) & \text{if } i_j < N_j, \\ 0 & \text{if } i_j = N_j, \end{cases}$$

for all $j = 1, \ldots, d$ and for all $x \in \Omega$. Let $\varphi : \mathbb{R} \to \mathbb{R}$, which we define for $\omega \in \mathcal{H}^d$ by

$$\varphi(|\omega|)(\Omega) := h^d \sum_{x\in\Omega} \varphi(|\omega(x)|),$$

where $|y| = \sqrt{y_1^2 + \ldots + y_d^2}$. In particular we define the *total variation* of u by setting $\varphi(t) = t$ and $\omega = \nabla u$, i.e.,

$$|\nabla u|(\Omega) := h^d \sum_{x\in\Omega} |\nabla u(x)|.$$

For an operator T we denote by T^* its adjoint. Further we introduce the *discrete divergence* $\mathrm{div} : \mathcal{H}^d \to \mathcal{H}$ defined by $\mathrm{div} = -\nabla^*$ (∇^* is the adjoint of the gradient ∇), in analogy to the continuous setting. In our case, the discrete divergence operator is explicitly given by

$$(\operatorname{div} p)(x) = \begin{cases} \frac{1}{h}(p^1(x^1_{i_1},\dots,x^d_{i_d}) - p^1(x^1_{i_1-1},\dots,x^d_{i_d})) & \text{if } 1 < i_1 < N_1, \\ p^1(x^1_{i_1},\dots,x^d_{i_d}) & \text{if } i_1 = 1, \\ -p^1(x^1_{i_1-1},\dots,x^d_{i_d}) & \text{if } i_1 = N_1, \end{cases}$$
$$+ \dots + \begin{cases} \frac{1}{h}(p^d(x^1_{i_1},\dots,x^d_{i_d}) - p^d(x^1_{i_1},\dots,x^d_{i_d-1})) & \text{if } 1 < i_d < N_d, \\ p^d(x^1_{i_1},\dots,x^d_{i_d}) & \text{if } i_d = 1, \\ -p^d(x^1_{i_1},\dots,x^d_{i_d-1}) & \text{if } i_d = N_d, \end{cases}$$

for every $p = (p^1,\dots,p^d) \in \mathcal{H}^d$ and for all $x \in \Omega$. (Note that if the discrete domains Ω are not discrete d-orthotopes, then the definitions of the gradient and divergence operators have to be adjusted accordingly.) We will often use the symbol 1 to indicate the constant vector with entry values 1 and 1_D to indicate the characteristic function of the domain $D \subset \Omega$.

For a convex functional $\mathcal{J} : \mathcal{H} \to \bar{\mathbb{R}}$, we define the *subdifferential* of $\mathcal{J}$ at $v \in \mathcal{H}$ as the set valued mapping

$$\partial \mathcal{J}(v) := \begin{cases} \emptyset & \text{if } \mathcal{J}(v) = \infty, \\ \{v^* \in \mathcal{H} : \langle v^*, u - v\rangle_{\mathcal{H}} + \mathcal{J}(v) \le \mathcal{J}(u) \quad \forall u \in \mathcal{H}\} & \text{otherwise.} \end{cases}$$

It is clear from this definition that $0 \in \partial \mathcal{J}(v)$ if and only if v is a minimizer of $\mathcal{J}$. Since we deal with different spaces, namely $\mathcal{H}$, V_i, it is useful to sometimes distinguish in which space the subdifferential is defined by imposing a subscript $\partial_{V_i}\mathcal{J}$ for the subdifferential considered on the space V_i.

4. Properties of Subspace Correction Methods. In what follows we consider the discrete functional

$$J_{\alpha_1,\alpha_2}(u) = \alpha_1 \|T_1 u - g_1\|_{\ell^1(\Omega)} + \alpha_2 \|T_2 u - g_2\|^2_{\ell^2(\Omega)} + \varphi(|\nabla u|)(\Omega), \tag{4.1}$$

where $T_i : \mathcal{H} \to \mathcal{H}$ is a bounded linear operator, $g_i \in \mathcal{H}$ is a given datum, and $\alpha_i \ge 0$ for $i = 1, 2$ with $\alpha_1 + \alpha_2 > 0$. Moreover, we assume that φ fulfills the assumption (A_φ) and that J_{α_1,α_2} is again bounded from below and coercive.

In the sequential and parallel algorithm in (2.2) and (2.3) we denote the difference between the current subspace minimizer $u_i^{(n+1)}$ and the initial value $\tilde{u}_i^{(n)}$ by $s^{(n+\frac{i}{2})}$, i.e.,

$$s^{(n+\frac{i}{2})} := u_i^{(n+1)} - \tilde{u}_i^{(n)}, \quad \text{for } i = 1, 2. \tag{4.2}$$

It is easy to see that in the sequential domain decomposition algorithm

$$s^{(n+\frac{i}{2})} = \arg\min_s J_{\alpha_1,\alpha_2}(u^{(n)} + (i-1)s^{(n+\frac{i-1}{2})} + s) \quad \text{s.t. } \pi_{V_i^c} s = 0$$

and in the parallel version

$$s^{(n+\frac{i}{2})} = \arg\min_s J_{\alpha_1,\alpha_2}(u^{(n)} + s) \quad \text{s.t. } \pi_{V_i^c} s = 0$$

for $i = 1, 2$. Moreover, for $v, s \in \mathcal{H}$ a quadratic Taylor expansion yields

$$\min_s J_{\alpha_1,\alpha_2}(v+s) = \min_s 2\alpha_2 \langle T_2 s, T_2 v - g_2\rangle_{\mathcal{H}} + \alpha_2 \|T_2 s\|^2_{\ell^2(\Omega)} + \alpha_1 \|T_1(v+s) - g_1\|_{\ell^1(\Omega)} + \varphi(|\nabla(v+s)|)(\Omega).$$

Then the following lemma can be proven similarly to Lemma 1 of [46]. For its statement we define the quantities $v(i)$, $i = 1, 2$, as follows: For the sequential domain decomposition algorithm in (2.2) choose $v(i) = u^{(n)}$ if $i = 1$ and $v(i) = u^{(n)} + s^{(n+\frac{1}{2})}$ for $i = 2$, while for the parallel domain decomposition algorithm in (2.3) $v(i) = u^{(n)}$ for $i = 1, 2$.

LEMMA 4.1. *Let* $P(u) = \alpha_1 \|T_1 u - g_1\|_{\ell^1(\Omega)} + \varphi(|\nabla u|)(\Omega)$. *For any* $v(i) \in \mathcal{H}$ *chosen according to the underlying algorithm, let* $s = s^{(n+\frac{i}{2})}$ *for* $i = 1, 2$. *Then*

$$J_{\alpha_1,\alpha_2}(v(i)+s) = J_{\alpha_1,\alpha_2}(v(i)) + 2\alpha_2 \langle T_2 s, T_2 v(i) - g_2\rangle_{\mathcal{H}} + \alpha_2 \|T_2 s\|^2_{\ell^2(\Omega)} + P(v(i)+s) - P(v(i))$$

and

$$2\alpha_2 \langle T_2 s, T_2 v(i) - g_2\rangle_{\mathcal{H}} + P(v(i)+s) - P(v(i)) \leq -2\alpha_2 \|T_2 s\|^2_{\ell^2(\Omega)}.$$

REMARK 4.2.
With $\alpha_2 > 0$, $v(i)$ *as above and* $s = s^{(n+\frac{i}{2})}$, *a direct consequence of Lemma 4.1 is that*

$$J_{\alpha_1,\alpha_2}(v(i)+s) - J_{\alpha_1,\alpha_2}(v(i)) \leq -\alpha_2 \|T_2 s\|^2_{\ell^2(\Omega)}, \tag{4.3}$$

where $\alpha_2 \|T_2 s\|^2_{\ell^2(\Omega)} > 0$ *whenever* $s \notin \ker T_2$. *Note that the above descent property holds in particular when* $T_2^* T_2$ *is invertible and* $\|s\| \neq 0$.

PROPOSITION 4.3. *Assume that* $T_2^* T_2$ *is invertible and* $\alpha_2 > 0$. *Let the sequence* $(u^{(n)})_n$ *be generated by the algorithm in* (2.2) *or* (2.3) *and let* $s^{(n+\frac{i}{2})}$ *be defined as in* (4.2). *Then we have the following statements:*

(i) $\|s^{(n+\frac{i}{2})}\| \to 0$ *for* $n \to \infty$,

(ii) $\|u^{(n+1)} - u^{(n)}\| \to 0$ *for* $n \to \infty$.

Proof. We begin by showing these statements for the sequential algorithm in (2.2). By the minimality property of $s^{(n+\frac{i}{2})}$ we have that whenever $\|s^{(n+\frac{i}{2})}\| \neq 0$ then

$$J_{\alpha_1,\alpha_2}(u^{(n)} + (i-1)s^{(n+\frac{i-1}{2})} + s^{(n+\frac{i}{2})}) < J_{\alpha_1,\alpha_2}(u^{(n)} + (i-1)s^{(n+\frac{i-1}{2})})$$

for $i = 1, 2$. Note that $u^{(n+1)} = u^{(n)} + s^{(n+\frac{1}{2})} + s^{(n+1)}$. From Remark 2.3 we know that J_{α_1,α_2} is convergent and hence by the above observation we obtain

$$J_{\alpha_1,\alpha_2}(u^{(n)} + (i-1)s^{(n+\frac{i-1}{2})} + s^{(n+\frac{i}{2})}) - J_{\alpha_1,\alpha_2}(u^{(n)} + (i-1)s^{(n+\frac{i-1}{2})}) \to 0 \quad for\ n \to \infty.$$

By (4.3) it follows then that $\|s^{(n+\frac{i}{2})}\| \to 0$ for $n \to \infty$ and for $i = 1, 2$, which proves (i). Since $u^{(n+1)} = u^{(n)} + s^{(n+\frac{1}{2})} + s^{(n+1)}$, (ii) immediately follows.

For the parallel algorithm in (2.3) we obtain by the minimality property of $s^{(n+\frac{i}{2})}$ that

$$J_{\alpha_1,\alpha_2}(u^{(n)} + s^{(n+\frac{i}{2})}) < J_{\alpha_1,\alpha_2}(u^{(n)}) \tag{4.4}$$

for $i = 1, 2$ whenever $\|s^{(n+\frac{i}{2})}\| \neq 0$. Hence by convexity and definition of $u^{(n+1)}$ in (2.3) we get

$$2J_{\alpha_1,\alpha_2}(u^{(n)}) > J_{\alpha_1,\alpha_2}(u^{(n)} + s^{(n+\frac{1}{2})}) + J_{\alpha_1,\alpha_2}(u^{(n)} + s^{(n+1)}) \geq 2J_{\alpha_1,\alpha_2}(u^{(n+1)}).$$

From (4.4), the convergence of J_{α_1,α_2}, and the previous inequalities we obtain

$$\underbrace{J_{\alpha_1,\alpha_2}(u^{(n)}) - J_{\alpha_1,\alpha_2}(u^{(n)} + s^{(n+\frac{1}{2})})}_{\geq 0} + \underbrace{J_{\alpha_1,\alpha_2}(u^{(n)}) - J_{\alpha_1,\alpha_2}(u^{(n)} + s^{(n+1)})}_{\geq 0}$$
$$\leq 2(J_{\alpha_1,\alpha_2}(u^{(n)}) - J_{\alpha_1,\alpha_2}(u^{(n+1)})) \to 0 \text{ for } n \to \infty.$$

By (4.3) we eventually have that $\|s^{(n+\frac{i}{2})}\| \to 0$ for $n \to \infty$ and for $i = 1, 2$. The second statement follows from $u^{(n+1)} = \frac{u_1^{(n+1)} + u_2^{(n+1)} + u^{(n)}}{2}$, since

$$\|u^{(n+1)} - u^{(n)}\| = \left\| \frac{u_1^{(n+1)} + u_2^{(n+1)} - u^{(n)}}{2} \right\| = \left\| \frac{s^{(n+\frac{1}{2})} + s^{(n+1)}}{2} \right\|.$$

□

REMARK 4.4. *From the proof of Proposition 4.3 we find that if* $T_2^* T_2$ *is invertible and* $\alpha_2 > 0$, *then we can replace the non-increase of the energy* J_{α_1,α_2} *in Proposition 2.2 (i) by a strict monotone decrease unless* $u^{(n+1)} = u^{(n)}$.

4.1. A Convergence Estimate. For proving convergence results of the algorithms in (2.2) and (2.3) we use a characterization of solutions of the minimization problem (2.1) similar to [47, Proposition 4.1] in the continuous setting for $\alpha_1 = 0$ and adapted in [27, Proposition 5.2] to the discrete case.

4.1.1. Characterization of Solutions. Following [24, Def. 4.1, p. 17] the *conjugate (or Legendre transform)* of a convex function $\phi : V \to \mathbb{R}$, with V a vector space with topological dual V^* and duality pairing $\langle \cdot, \cdot \rangle$, is defined by

$$\phi^*(u^*) = \sup_{u \in V} \{\langle u, u^* \rangle - \phi(u)\}.$$

This notion of a convex conjugate is useful when characterizing the solution of the minimization of (4.1) for $\alpha_1, \alpha_2 > 0$.

PROPOSITION 4.5. *Let $\zeta, u \in \mathcal{H}$. If the assumption (A_φ) holds true, then $\zeta \in \partial J_{\alpha_1,\alpha_2}(u)$ if and only if there exists $M = (M_0, M_1, M_2) \in \mathcal{H}^d \times \mathcal{H} \times \mathcal{H}$, and a constant $c_1 \geq 0$ with $|M_0(x)| \leq c_1$, $|M_1(x)| \leq \alpha_1$ for all $x \in \Omega$ such that*

$$\varphi(|(\nabla u)(x)|) + \langle M_0(x), \nabla u(x) \rangle_{\mathbb{R}^d} + \varphi_1^* (|M_0(x)|) = 0, \quad \text{for all } x \in \Omega, \tag{4.5}$$

$$M_2(x) = -2\alpha_2 (T_2 u - g_2)(x), \quad \text{for all } x \in \Omega, \tag{4.6}$$

$$\alpha_1 |(T_1 u - g_1)(x)| + M_1(x)((T_1 u)(x) - g_1(x)) = 0, \quad \text{for all } x \in \Omega, \tag{4.7}$$

$$T_1^* M_1 + T_2^* M_2 - \operatorname{div} M_0 + \zeta = 0, \tag{4.8}$$

where φ_1^* is the conjugate function of φ_1 defined by $\varphi_1(t) = \varphi(|t|)$ for $t \in \mathbb{R}$. If, additionally, φ is differentiable and $|(\nabla u)(x)| \neq 0$ for $x \in \Omega$, then we can compute M_0 as

$$M_0(x) = -\frac{\varphi'(|(\nabla u)(x)|)}{|(\nabla u)(x)|} (\nabla u)(x). \tag{4.9}$$

The proof of this proposition is deferred to the Appendix.

For $\alpha_2 = 0$ the minimization problem associated with the objective in (4.1) simplifies to

$$\min_{u \in \mathcal{H}} J_{\alpha_1,0}(u) = \alpha_1 \|(T_1 u - g_1)\|_{\ell^1(\Omega)} + \varphi(|\nabla u|)(\Omega), \tag{4.10}$$

and the system (4.5)-(4.8) reduces to (4.11)-(4.13) below.

COROLLARY 4.6. *Let $\zeta, u \in \mathcal{H}$. If the assumption (A_φ) holds true, then $\zeta \in \partial J_{\alpha_1,0}(u)$ if and only if there exists $M = (M_0, M_1) \in \mathcal{H}^d \times \mathcal{H}$, and a constant $c_1 \geq 0$ with $|M_0(x)| \leq c_1$, $|M_1(x)| \leq \alpha_1$ for all $x \in \Omega$ such that*

$$\varphi(|(\nabla u)(x)|) + \langle M_0(x), \nabla u(x) \rangle_{\mathbb{R}^d} + \varphi_1^* (|M_0(x)|) = 0, \quad \text{for all } x \in \Omega, \tag{4.11}$$

$$\alpha_1 |(T_1 u - g_1)(x)| + M_1(x)((T_1 u)(x) - g_1(x)) = 0, \quad \text{for all } x \in \Omega, \tag{4.12}$$

$$T_1^* M_1 - \operatorname{div} M_0 + \zeta = 0, \tag{4.13}$$

where φ_1^* is the conjugate function of φ_1 defined by $\varphi_1(t) = \varphi(|t|)$ for $t \in \mathbb{R}$.

4.1.2. Optimality Properties. By [32, Theorem 2.1.4, p. 305], the optimality condition for the subspace minimization problem in V_i, cf. (2.4), i.e.,

$$\xi_i^{(n+1)} \in \arg\min_{\xi_i \in \mathcal{H}} \{J_{\alpha_1,\alpha_2}(\xi_i) \ : \ \pi_{V_i^c} \xi_i = \pi_{V_i^c} b\}, \tag{4.14}$$

is

$$0 \in \partial J_{\alpha_1,\alpha_2}(\xi_i^{(n+1)}) + \eta_i^{(n+1)},$$

where $\eta_i^{(n+1)} \in \operatorname{Range}(\pi_{V_i^c}^*) \simeq V_i^c$ and $b \in V_j$ as in (2.4) for $i = 1, 2$ and $j \in \{1, 2\} \setminus \{i\}$. Note that indeed $\xi_1^{(n+1)}$ is optimal if and only if $u_1^{(n+1)} = \xi_1^{(n+1)} - \tilde{u}_2^{(n)}$ is optimal. Moreover, $\xi_2^{(n+1)}$ is optimal for the algorithm in (2.2) if and only if $u_2^{(n+1)} = \xi_2^{(n+1)} - u_1^{(n+1)}$ is optimal, and $\xi_2^{(n+1)}$ is optimal for the algorithm in (2.3) if and only if $u_2^{(n+1)} = \xi_2^{(n+1)} - \tilde{u}_1^{(n)}$ is optimal.

The following result is a consequence of Proposition 4.5 and Corollary 4.6. It relies on the fact that $\partial J_{\alpha_1,\alpha_2}(\xi)$ is compact for any $\xi \in \mathcal{H}$; see [6].

COROLLARY 4.7. *Let $(\xi^{(n)})_n \subset \mathcal{H}$ be bounded and $\eta^{(n)} \in \partial J_{\alpha_1,\alpha_2}(\xi^{(n)})$ for all $n \in \mathbb{N}$. Then $(\eta^{(n)})_n$ is bounded.*

Proof. Set $P(\xi^{(n)}) := \alpha_1 \|T_1\xi^{(n)} - g_1\|_{\ell^1(\Omega)} + \varphi(|\nabla\xi^{(n)}|)(\Omega)$. Then we have

$$\eta^{(n)} \in 2\alpha_2 T_2^*(T_2\xi^{(n)} - g_2) + \partial P(\xi^{(n)}).$$

Since T_2 is a bounded operator and $(\xi^{(n)})_n$ is bounded we are left with showing that the set $\partial P(\xi^{(n)})$ is bounded for all n. By Corollary 4.6 we have that

$$\begin{aligned}\partial P(\xi^{(n)}) = \{\, &\operatorname{div} M_0 - T_1^* M_1 \in \mathcal{H} : \|M_0\|_\infty \le c_1, \|M_1\|_\infty \le \alpha_1,\\ &\varphi(|(\nabla\xi^{(n)})(x)|) + \langle M_0(x), \nabla\xi^{(n)}(x)\rangle_{\mathbb{R}^d} + \varphi_1^*(|M_0(x)|) = 0,\\ &\alpha_1|(T_1\xi^{(n)} - g_1)(x)| + M_1(x)((T_1\xi^{(n)})(x) - g_1(x)) = 0 \text{ for all } x \in \Omega\}.\end{aligned}$$

Since c_1 and α_1 do not depend on n, the sequence of sets $(\partial P(\xi^{(n)}))_n$ is uniformly bounded and, hence, $(\eta^{(n)})_n$ is bounded as well. □

PROPOSITION 4.8. *Let $u_i^{(\infty)}$ and $\tilde{u}_i^{(\infty)}$ be accumulation points of the sequences $(u_i^{(n)})_n$ and $(\tilde{u}_i^{(n)})_n$, $i = 1, 2$, generated by the algorithms in* (2.2) *or* (2.3)*. Then the limits $u_1^{(\infty)}$ and $\tilde{u}_1^{(\infty)}$ are minimizers of $\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + \tilde{u}_2^{(\infty)})$. Further we have for the algorithm in* (2.2) *that $u_2^{(\infty)}$ and $\tilde{u}_2^{(\infty)}$ are minimizers of $\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(u_1^{(\infty)} + u_2)$ and for the algorithm in* (2.3) *that $u_2^{(\infty)}$ and $\tilde{u}_2^{(\infty)}$ are minimizers of $\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)} + u_2)$.*

Proof. Let us show the assertion first for the algorithm in (2.2). Along a convergent subsequence, which - without loss of generality - we denote again by $(u_i^{(n)})_n$ for $i = 1, 2$, we have by the minimality property of $u_1^{(n+1)}$ that $0 \in \partial_{V_1} J_{\alpha_1,\alpha_2}(\cdot + \tilde{u}_2^{(n)})(u_1^{(n+1)})$. By [41, Theorem 24.4, p 233] we obtain that $0 \in \partial_{V_1} J_{\alpha_1,\alpha_2}(\cdot + \tilde{u}_2^{(\infty)})(u_1^{(\infty)})$ and hence $u_1^{(\infty)} \in \arg\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + \tilde{u}_2^{(\infty)})$. We get by a similar argument that $0 \in \partial_{V_2} J_{\alpha_1,\alpha_2}(u_1^{(\infty)} + \cdot)(u_2^{(\infty)})$, i.e., $u_2^{(\infty)} = \arg\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(u_1^{(\infty)} + u_2)$.

Moreover, by the monotone decrease of the energy (see Proposition 2.2) we have that

$$J_{\alpha_1,\alpha_2}(u^{(n)}) \ge J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + \tilde{u}_2^{(n)}) \ge J_{\alpha_1,\alpha_2}(u^{(n+1)}) \quad \text{for all } n \in \mathbb{N} \tag{4.15}$$

and hence

$$J_{\alpha_1,\alpha_2}(u^{(n)}) - J_{\alpha_1,\alpha_2}(u^{(n+1)}) \ge J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + \tilde{u}_2^{(n)}) - J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n+1)} + \tilde{u}_2^{(n+1)}) \ge 0$$

as well as

$$J_{\alpha_1,\alpha_2}(u^{(n)}) - J_{\alpha_1,\alpha_2}(u^{(n+1)}) \ge J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + \tilde{u}_2^{(n)}) - J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + u_2^{(n+1)}) \ge 0.$$

Since J_{α_1,α_2} is bounded from below, we obtain $\lim_{n\to\infty}\left[J_{\alpha_1,\alpha_2}(u^{(n)}) - J_{\alpha_1,\alpha_2}(u^{(n+1)})\right] = 0$. Consequently

$$\begin{aligned}0 &= \lim_{n\to\infty}\left[J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + \tilde{u}_2^{(n)}) - J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n+1)} + \tilde{u}_2^{(n+1)})\right]\\ &= J_{\alpha_1,\alpha_2}(u_1^{(\infty)} + \tilde{u}_2^{(\infty)}) - J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)} + \tilde{u}_2^{(\infty)})\end{aligned} \tag{4.16}$$

as well as

$$\begin{aligned}0 &= \lim_{n\to\infty}\left[J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + \tilde{u}_2^{(n)}) - J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + u_2^{(n+1)})\right]\\ &= J_{\alpha_1,\alpha_2}(u_1^{(\infty)} + \tilde{u}_2^{(\infty)}) - J_{\alpha_1,\alpha_2}(u_1^{(\infty)} + u_2^{(\infty)}).\end{aligned} \tag{4.17}$$

From (4.16) we see that since $u_1^{(\infty)}$ is a solution of $\arg\min_{u_1\in V_1} J_{\alpha_1,\alpha_2}(u_1+\tilde{u}_2^{(\infty)})$, so is $\tilde{u}_1^{(\infty)}$. By an analogous argument we obtain from (4.17) that $\tilde{u}_2^{(\infty)} = \arg\min_{u_2\in V_2} J_{\alpha_1,\alpha_2}(u_1^{(\infty)}+u_2)$.

Although the proof for the algorithm in (2.3) is similar, for the sake of completeness we provide the details here. By similar arguments as above one shows that $u_1^{(\infty)} \in \arg\min_{u_1\in V_1} J_{\alpha_1,\alpha_2}(u_1+\tilde{u}_2^{(\infty)})$ and $u_2^{(\infty)} = \arg\min_{u_2\in V_2} J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)}+u_2)$. Hence, we have

$$\begin{aligned} J_{\alpha_1,\alpha_2}(u_1^{(\infty)}+\tilde{u}_2^{(\infty)}) &\le J_{\alpha_1,\alpha_2}(u_1+\tilde{u}_2^{(\infty)}) \quad \forall\, u_1\in V_1,\\ J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)}+u_2^{(\infty)}) &\le J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)}+u_2) \quad \forall\, u_2\in V_2. \end{aligned} \tag{4.18}$$

Moreover by the monotone decrease of the energy J_{α_1,α_2} (see Proposition 2.2) we have

$$J_{\alpha_1,\alpha_2}(u^{(n)}) \ge \frac{1}{2}\left(J_{\alpha_1,\alpha_2}(u_1^{(n+1)}+\tilde{u}_2^{(n)}) + J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n)}+u_2^{(n+1)})\right) \ge J_{\alpha_1,\alpha_2}(u^{(n+1)})$$

and hence

$$2(J_{\alpha_1,\alpha_2}(u^{(n)})-J_{\alpha_1,\alpha_2}(u^{(n+1)})) \ge J_{\alpha_1,\alpha_2}(u_1^{(n+1)}+\tilde{u}_2^{(n)})+J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n)}+u_2^{(n+1)})-2J_{\alpha_1,\alpha_2}(u^{(n+1)}) \ge 0.$$

Since J_{α_1,α_2} is bounded below, we obtain $\lim_{n\to\infty}\left[J_{\alpha_1,\alpha_2}(u^{(n)})-J_{\alpha_1,\alpha_2}(u^{(n+1)})\right]=0$ and hence

$$\lim_{n\to\infty}\left[J_{\alpha_1,\alpha_2}(u_1^{(n+1)}+\tilde{u}_2^{(n)}) + J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n)}+u_2^{(n+1)}) - 2J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(n+1)}+\tilde{u}_2^{(n+1)})\right] = 0.$$

From (4.18) it follows that

$$J_{\alpha_1,\alpha_2}(u_1^{(\infty)}+\tilde{u}_2^{(\infty)})-J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)}+\tilde{u}_2^{(\infty)}) = 0 \quad \text{and} \quad J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)}+u_2^{(\infty)})-J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)}+\tilde{u}_2^{(\infty)}) = 0$$

and hence

$$\tilde{u}_1^{(\infty)} \in \arg\min_{u_1\in V_1} J_{\alpha_1,\alpha_2}(u_1+\tilde{u}_2^{(\infty)}) \quad \text{and} \quad \tilde{u}_2^{(\infty)} \in \arg\min_{u_2\in V_2} J_{\alpha_1,\alpha_2}(\tilde{u}_1^{(\infty)}+u_2),$$

which concludes the proof. □

REMARK 4.9.

(i) If V_1 and V_2 are disjoint spaces, then the algorithm in (2.2) *generates sequences $(u_i^{(n)})_n$ and $(\tilde{u}_i^{(n)})_n$ with $u_i^{(n)} = \tilde{u}_i^{(n)}$, and the algorithm in* (2.3) *generates sequences $(u_i^{(n)})_n$ and $(\tilde{u}_i^{(n)})_n$ with $\tilde{u}_i^{(n+1)} = \frac{1}{2}(u_i^{(n+1)}+u_i^{(n)})$, $\tilde{u}_i^{(\infty)} = u_i^{(\infty)}$, and $u^{(\infty)} = u_1^{(\infty)}+u_2^{(\infty)}$.*

(ii) In general, however, the algorithms in (2.2) *and* (2.3)*, respectively, generate sequences $(u_i^{(n)})_n$ and $(\tilde{u}_i^{(n)})_n$ with $u_i^{(n)} \ne \tilde{u}_i^{(n)}$, $i=1,2$. This relation is still valid in the limit case, i.e., for $n\to\infty$ we have $u_i^{(\infty)} \ne \tilde{u}_i^{(\infty)}$ (unless V_1 and V_2 are disjoint), although $u_i^{(\infty)}$ and $\tilde{u}_i^{(\infty)}$ are minimizers of the same minimization problem; see Proposition 2.4. This behavior can be attributed to the fact that the function J_{α_1,α_2} is not strictly convex and, thus, has in general more than one minimizer.*

THEOREM 4.10. *Assume that $\alpha_2 > 0$ and u^* is a minimizer of J_{α_1,α_2}. Let $u^{(\infty)}$ be an accumulation point of the sequence $(u^{(n)})_n$ generated by the algorithm in* (2.2) *or* (2.3)*. Then we have that either*

1. *$u^{(\infty)}$ is a minimizer of J_{α_1,α_2}, or*
2. *there exists a constant $\varepsilon > 0$ such that*

$$\|u^{(\infty)}-u^*\|_{\ell^2(\Omega)} \le \frac{\|\hat{\eta}\|_{\ell^2(\Omega)}}{\varepsilon},$$

where $\hat{\eta} \in \arg\min_{\eta\in\bigcup_{i=1}^2(\partial J_{\alpha_1,\alpha_2}(u^{(\infty)})\cap V_i^c)} \|\eta\|_{\ell^2(\Omega)}$. In particular, if $T_2^*T_2$ is positive definite with smallest Eigenvalue $\sigma > 0$, then we can specify the constant ε by $\varepsilon = \alpha_2\sigma$.

Proof. Let $\xi_i^{(n)} \in \arg\min_{\xi_i \in \mathcal{H}}\{J_{\alpha_1,\alpha_2}(\xi_i) \ : \ \pi_{V_i^c}\xi_i = \pi_{V_i^c} b\}$ with $b \in V_j$ as in (2.4) for $i = 1, 2$ and $j \in \{1, 2\} \setminus \{i\}$. Then, by [32, Theorem 2.1.4, p. 305] we have that $0 \in \partial J_{\alpha_1,\alpha_2}(\xi_i^{(n)}) + \eta_i^{(n)}$ for $\eta_i \in V_i^c$ and $i = 1, 2$. In other words, for example, for $i = 1$ we have that $\xi_1^{(n)} - \tilde{u}_2^{(n-1)} =: u_1^{(n)} \in \arg\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + \tilde{u}_2^{(n-1)})$. Since $\xi_i^{(n)}$ is bounded, also $\partial J_{\alpha_1,\alpha_2}(\xi_i^{(n)})$ is bounded and, cf. [41, Theorem 24.4, p 233], there exists $\eta_i^{(\infty)} \in V_i^c$ such that $0 \in \partial J_{\alpha_1,\alpha_2}(\xi_i^{(\infty)}) + \eta_i^{(\infty)}$. That is, $u_i^{(\infty)}$ is optimal in V_i for $i = 1, 2$. From Proposition 4.8 we get that if $u_i^{(\infty)}$ is optimal then also $\tilde{u}_i^{(\infty)}$ is optimal. Hence for $i = 1$ we have $u^{(\infty)} - \tilde{u}_2^{(\infty)} = \tilde{u}_1^{(\infty)} \in \arg\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + \tilde{u}_2^{(\infty)})$, which means that there exists $\hat{\eta}_1 \in V_1^c$ such that $0 \in \partial J_{\alpha_1,\alpha_2}(u^{(\infty)}) + \hat{\eta}_1$. Similarly we get for $i = 2$ that there exists $\hat{\eta}_2 \in V_2^c$ such that $0 \in \partial J_{\alpha_1,\alpha_2}(u^{(\infty)}) + \hat{\eta}_2$. By the definition of the subdifferential we obtain

$$J_{\alpha_1,\alpha_2}(u^{(\infty)}) \leq J_{\alpha_1,\alpha_2}(v) + \langle \hat{\eta}, u^{(\infty)} - v\rangle_{\mathcal{H}} \leq J_{\alpha_1,\alpha_2}(v) + \|\hat{\eta}\|_{\ell^2(\Omega)}\|v - u^{(\infty)}\|_{\ell^2(\Omega)} \tag{4.19}$$

for all $v \in \mathcal{H}$, where $\hat{\eta} = \arg\min_{\eta \in \bigcup_{i=1}^2 \left(\partial J_{\alpha_1,\alpha_2}(u^{(\infty)}) \cap V_i^c\right)} \|\eta\|_{\ell^2(\Omega)}$.

Let $u^* \in \arg\min_{u \in \mathcal{H}} J_{\alpha_1,\alpha_2}(u)$. Then the optimality of u^* yields that the directional derivative of J_{α_1,α_2} at u^* in any direction $s \in \mathcal{H}$ is non-negative, i.e., $J'_{\alpha_1,\alpha_2}(u^*; s) \geq 0$. Set $P(\xi^{(n)}) := \alpha_1\|T_1\xi^{(n)} - g_1\|_{\ell^1(\Omega)} + \varphi(|\nabla\xi^{(n)}|)(\Omega)$. Then by using Taylor's expansion we have for $s \in \mathcal{H}$ that

$$\begin{aligned} J_{\alpha_1,\alpha_2}(u^* + s) &= \alpha_2\|T_2u^* - g_2\|^2_{\ell^2(\Omega)} + \langle s, 2\alpha_2T_2^*(T_2u^* - g_2)\rangle_{\mathcal{H}} + \alpha_2\|T_2s\|^2_{\ell^2(\Omega)} + P(u^* + s) \\ &= J_{\alpha_1,\alpha_2}(u^*) + \langle s, 2\alpha_2T_2^*(T_2u^* - g_2)\rangle_{\mathcal{H}} + P(u^* + s) - P(u^*) + \alpha_2\|T_2s\|^2_{\ell^2(\Omega)}. \end{aligned}$$

By using $P(u^* + s) - P(u^*) \geq P'(u^*; s)$, which easily follows from the convexity of P, we obtain that

$$J_{\alpha_1,\alpha_2}(u^* + s) \geq J_{\alpha_1,\alpha_2}(u^*) + \langle s, 2\alpha_2T_2^*(T_2u^* - g_2)\rangle_{\mathcal{H}} + P'(u^*; s) + \alpha_2\|T_2s\|^2_{\ell^2(\Omega)}. \tag{4.20}$$

Since $J'_{\alpha_1,\alpha_2}(u^*; s) = \langle s, 2\alpha_2T_2^*(T_2u^* - g_2)\rangle_{\mathcal{H}} + P'(u^*; s) \geq 0$ and $\alpha_2\|T_2s\|^2_{\ell^2(\Omega)} \geq 0$ there exists a constant $\rho \geq 0$ such that $J_{\alpha_1,\alpha_2}(u^* + s) = J_{\alpha_1,\alpha_2}(u^*) + \rho$.

1. If $\rho = 0$ for $s := u^{(\infty)} - u^*$, then it immediately follows that $u^* + s = u^{(\infty)}$ is a minimizer of J_{α_1,α_2}.
2. If $\rho > 0$ for $s := u^{(\infty)} - u^*$, then, since $\|s\|_{\ell_2(\Omega)} \leq \beta < +\infty$, there exists an $\varepsilon > 0$ such that $\varepsilon\beta^2 \leq \rho$. Hence we obtain

$$J_{\alpha_1,\alpha_2}(u^* + s) \geq J_{\alpha_1,\alpha_2}(u^*) + \varepsilon\|u^{(\infty)} - u^*\|^2_{\ell_2(\Omega)} \tag{4.21}$$

Setting $v = u^*$ in (4.19) and using the resulting inequality in (4.21) yields

$$J_{\alpha_1,\alpha_2}(u^*) + \varepsilon\|u^{(\infty)} - u^*\|^2_{\ell^2(\Omega)} \leq J_{\alpha_1,\alpha_2}(u^{(\infty)}) \leq J_{\alpha_1,\alpha_2}(u^*) + \|\hat{\eta}\|_{\ell^2(\Omega)}\|u^* - u^{(\infty)}\|_{\ell^2(\Omega)}$$

and consequently

$$\|u^{(\infty)} - u^*\|_{\ell^2(\Omega)} \leq \frac{\|\hat{\eta}\|_{\ell^2(\Omega)}}{\varepsilon}.$$

If additionally $T_2^*T_2$ is symmetric positive definite with smallest Eigenvalue $\sigma > 0$, i.e., $\|T_2s\|^2_{\ell^2(\Omega)} \geq \sigma\|s\|^2_{\ell^2(\Omega)}$, we get from (4.20) that $J_{\alpha_1,\alpha_2}(u^* + s) \geq J_{\alpha_1,\alpha_2}(u^*) + \alpha_2\sigma\|u^{(\infty)} - u^*\|^2_{\ell_2(\Omega)}$. If we compare the last inequality with (4.21) we observe that now $\varepsilon = \alpha_2\sigma$, which finishes the proof.

□

We have the following immediate consequence of Theorem 4.10.

COROLLARY 4.11. *Let the assumptions of Theorem 4.10 hold true. If $\|\eta_i^{(n_\ell)}\| \to 0$ for $\ell \to \infty$ along a suitable subsequence $(n_\ell)_\ell$ for at least one $i \in \{1, 2\}$, then any accumulation point of the sequence $(u^{(n)})_n$ generated by the algorithm in (2.2) or (2.3) is a minimizer of J_{α_1,α_2}.*

4.2. A Modified Sequential Subspace Correction Method. Note that the algorithm in (2.2) is a special case of a coordinate descent method, where the spaces V_i are chosen in a cyclic manner; see [45, 46] for more details on different rules for choosing the subspaces. In the sense of coordinate descent methods our algorithm in (2.2) can be written as follows.

CD-Algorithm: Choose $u^{(0)} \in \mathcal{H}$ and iterate for $n = 0, 1, 2, \ldots$
1) choose a non-empty space $V_n \subset \mathcal{H}$;
2) compute $s^{(n)} = s_{T_2^* T_2}(u^{(n)}, V_n) \in \arg\min_s\{J_{\alpha_1,\alpha_2}(u^{(n)} + s) \text{ s.t. } \pi_{V_n^c} s = 0\}$;
3) set $u^{(n+1)} = u^{(n)} + s^{(n)}$.

In step 2, the subscript $T_2^* T_2$ refers to the Hessian of the smooth part of J_{α_1,α_2}. Compared to [46] here we choose the step size in the update of $u^{(n+1)}$ to be 1, which is justified by (4.3) for $\alpha_2 > 0$. As already mentioned, there exist several different ways of choosing V_n in each iteration. We suggest to select V_n such that

$$\|s_D(u^{(n)}, V_n)\|_{\ell^2(\Omega)} \geq \nu \|s_D(u^{(n)}, \mathcal{H})\|_{\ell^2(\Omega)}, \tag{4.22}$$

where $0 < \nu \leq 1$ and D is positive definite diagonal matrix. Here, the subscript D indicates that $T_2^* T_2$, the Hessian of the smooth part of J_{α_1,α_2}, is replaced by D. This rule is called the *Gauss-Southwell-r* rule, which also allows the choice $V_n = \mathcal{H}$. With this selection rule of the subspaces we are able to establish global convergence. The proof follows from [46].

THEOREM 4.12. *Assume $2\alpha_2 \|T_2\|^2 \geq \underline{\lambda} > 0$. Let $(u^{(n)})_n$, $(s^{(n)})_n$ be sequences generated by the CD-Algorithm. If (V_n) is chosen by the Gauss-Southwell-r rule with $D = \operatorname{diag}(T_2^* T_2)$ and $\|\operatorname{diag}(T_2^* T_2)\| \geq \underline{\delta} > 0$, then every cluster point of $(u^{(n)})_n$ is a minimizer of J_{α_1,α_2}.*

5. Application Domain Decomposition. The results of the previous sections are valid for any splitting of the function space $\mathcal{H}$. We concentrate now on decompositions which split the spatial domain into two subdomains. But let us emphasize that a generalization to a splitting into more domains is straightforward.

5.1. Overlapping Domain Decomposition. In this section we focus on an overlapping domain decomposition method. Thus we want to minimize (4.1) by decomposing Ω into two overlapping subdomains Ω_1 and Ω_2 such that $\Omega = \Omega_1 \cup \Omega_2$ and $\Omega_1 \cap \Omega_2 \neq \emptyset$. By Γ_1 we denote the interface between Ω_1 and $\Omega_2 \setminus \Omega_1$ and by Γ_2 the interface between Ω_2 and $\Omega_1 \setminus \Omega_2$. For consistency with the definitions of the gradient and divergence operators, we assume that the subdomains Ω_i as well as Ω are discrete d-orthotopes. We stress that this is by no means a restriction, but simplifies the presentation. Associated to the splitting of Ω we define $V_i = \{u \in \mathcal{H} : \operatorname{supp}(u) \subset \Omega_i\}$. One aims now to minimize J_{α_1,α_2} by the alternating algorithm in (2.2) or the parallel algorithm in (2.3). Note that the respective subspace minimization problems are actually constrained optimization problems of the type (2.4). In particular, for the alternating algorithm we have in V_1

$$\min_{v \in \mathcal{H}} J_{\alpha_1,\alpha_2}(v) \quad \text{s.t. } \pi_{V_1^c} v = \pi_{V_1^c} \tilde{u}_2^{(n)}, \tag{5.1}$$

and in V_2 we have

$$\min_{v \in \mathcal{H}} J_{\alpha_1,\alpha_2}(v) \quad \text{s.t. } \pi_{V_2^c} v = \pi_{V_2^c} u_1^{(n+1)},$$

while for the parallel algorithm the minimization in V_1 is again (5.1) and in V_2 it changes to

$$\min_{v \in \mathcal{H}} J_{\alpha_1,\alpha_2}(v) \quad \text{s.t. } \pi_{V_2^c} v = \pi_{V_2^c} \tilde{u}_1^{(n)}.$$

5.2. Non-Overlapping Domain Decomposition. In the non-overlapping domain decomposition method we want to minimize (4.1) by decomposing Ω into two non-overlapping subdomains Ω_1 and Ω_2 such that $\Omega = \Omega_1 \cup \Omega_2$ and $\Omega_1 = \Omega \setminus \Omega_2$. For consistency with the definitions of the gradient and divergence operator, we again assume that the subdomains Ω_i as well as Ω are discrete d-orthotopes. Associated to the splitting of Ω we define by $V_i = \{u \in \mathcal{H} : \operatorname{supp}(u) \subset \Omega_i\}$

the function space of the subdomain Ω_i. Then we minimize J_{α_1,α_2} either by the parallel algorithm in (2.3) or by the alternating algorithm in (2.2), which specifies to:

Choose an initial $u^{(0)} =: u_1^{(0)} + u_2^{(0)} \in V_1 \oplus V_2$, for example, $u^{(0)} = 0$, and iterate

$$\begin{cases} u_1^{(n+1)} = \arg\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + u_2^{(n)}), \\ u_2^{(n+1)} = \arg\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + u_2), \\ u^{(n+1)} := u_1^{(n+1)} + u_2^{(n+1)}. \end{cases} \tag{5.2}$$

The subspace optimization problems for the alternating version are

$$\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + u_2^{(n)}) = \min_{u_1 \in V_1} \alpha_1 \|T_1 u_1 - (g_1 - T_1 u_2^{(n)})\|_{\ell^1(\Omega)} + \alpha_2 \|T_2 u_1 - (g_2 - T_2 u_2^{(n)})\|^2_{\ell^2(\Omega)} + \varphi(|\nabla(u_1 + u_2^{(n)})|)(\Omega)$$

in V_1 and

$$\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + u_2) = \min_{u_2 \in V_2} \alpha_1 \|T_1 u_2 - (g_1 - T_1 u_1^{(n+1)})\|_{\ell^1(\Omega)} + \alpha_2 \|T_2 u_2 - (g_2 - T_2 u_1^{(n+1)})\|^2_{\ell^2(\Omega)} + \varphi(|\nabla(u_1^{(n+1)} + u_2)|)(\Omega)$$

in V_2. Upon adjusting notation, for the parallel algorithm the subspace minimization problems look similar.

A very special situation occurs when $(T_1 u_2)(x) = 0$ for all $x \in \Omega$ and $(T_2 u_1)(x) = 0$ for all $x \in \Omega$, which is the case when $T_i = 1_{\Omega_i} \tilde{T}_i$ with $\tilde{T}_i : \mathcal{H} \to \mathcal{H}$ such that for all $v_j \in V_j$ we have $\tilde{T}_i v_j \in V_j$ for $j = 1,2$ and $i = 1,2$ (e.g., $\tilde{T}_i = I$ or $\tilde{T}_i = 1_{\Omega \setminus K}$ with $K \subset \Omega$). Then the above subspace minimization problems simplify to

$$\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + u_2^{(n)}) = \min_{u_1 \in V_1} \alpha_1 \|T_1 u_1 - g_1\|_{\ell^1(\Omega)} + \varphi(|\nabla(u_1 + u_2^{(n)})|)(\Omega) \tag{5.3}$$

in V_1 and

$$\min_{u_2 \in V_2} J_{\alpha_1,\alpha_2}(u_1^{(n+1)} + u_2) = \min_{u_2 \in V_2} \alpha_2 \|T_2 u_2 - g_2\|^2_{\ell^2(\Omega)} + \varphi(|\nabla(u_1^{(n+1)} + u_2)|)(\Omega) \tag{5.4}$$

in V_2.

The parallel version of the previous algorithm is the one in (2.3).

5.3. Numerical Implementation. In this section we propose an implementation of the domain decomposition algorithms in (2.2) and (2.3) when the domain is split into overlapping and non-overlapping subdomains for the particular case when $\varphi(|\nabla u|)(\Omega) = |\nabla u|(\Omega)$ (total variation of u in Ω). Specifically, the sequential non-overlapping domain decomposition algorithm can be written as in (5.2). In the sequel we assume that $\|T_i\| < 1$ for $i = 1,2$, which is not at all a restriction, since if a norm exceeds 1, then a proper rescaling of the problem reestablishes the desired setting. Our implementation works equally for the non-overlapping and overlapping domain decomposition algorithm. Therefore we will use notations which we introduced for an overlapping splitting but which are also meaningful for a non-overlapping decomposition. More precisely, in the non-overlapping case $\hat{\Omega}_i = \Omega_i \setminus \Omega_{\check{i}}$ turns out to be all of Ω_i for $i = 1,2$ and $\check{i} \in \{1,2\} \setminus \{i\}$, and for a non-overlapping decomposition we have that $\Gamma_1 = \Gamma_2$ is the interface between Ω_1 and Ω_2.

We already note here that the dynamic range of all image data considered below is $[c_{\min}, c_{\max}] := [0,1]$.

5.3.1. L^1-L^2-TV Minimization. In order to compute a minimizer of the global problem

$$\arg\min_{u \in \mathcal{H}} J_{\alpha_1,\alpha_2}(u), \tag{5.5}$$

where J_{α_1,α_2} is the functional in (4.1) with $\varphi(t) = t$, we suggest a simple algorithm, which is an adaptation of an algorithm that was originally proposed for L^1-TV minimization problems in [4]. The idea is to consider a regularization of J_{α_1,α_2}, i.e., the functional

$$\alpha_1 \|v\|_{\ell^1(\Omega)} + \frac{1}{2\gamma}\|T_1 u - g_1 - v\|^2_{\ell^2(\Omega)} + \alpha_2 \|T_2 u - g_2\|^2_{\ell^2(\Omega)} + |\nabla u|(\Omega), \tag{5.6}$$

where $\gamma > 0$ is small, so that we have $g_1 \approx T_1 u - v$. Actually for $\gamma \to 0$ (5.6) approaches the objective functional in (5.5). Now we minimize (5.6) with respect to u and v, which we perform in the following alternating way:

(1) For fixed u solve

$$\min_{v\in\mathcal{H}} \alpha_1 \|v\|_{\ell^1(\Omega)} + \frac{1}{2\gamma}\|T_1 u - g_1 - v\|^2_{\ell^2(\Omega)}. \tag{5.7}$$

The minimizer v^* of (5.7) can be easily computed via a soft-thresholding, i.e., $v^* = \mathrm{ST}(T_1\xi_1 - g_1, \gamma\alpha_1)$, where

$$\mathrm{ST}(g,\beta) = \begin{cases} g_{i,j} - \beta & \text{if } g_{i,j} > \beta, \\ 0 & \text{if } |g_{i,j}| \le \beta, \\ g_{i,j} + \beta & \text{if } g_{i,j} < -\beta. \end{cases} \tag{5.8}$$

(2) For fixed v solve

$$\min_{u\in\mathcal{H}} \frac{1}{2\gamma}\|T_1 u - g_1 - v\|^2_{\ell^2(\Omega)} + \alpha_2 \|T_2 u - g_2\|^2_{\ell^2(\Omega)} + |\nabla u|(\Omega). \tag{5.9}$$

In order to solve this minimization problem we introduce a *surrogate functional*. For this purpose, assume $a, u \in \mathcal{H}$ and define

$$\begin{aligned} S(u,a) :=& \frac{1}{2\gamma}\|T_1 u - g_1 - v\|^2_{\ell^2(\Omega)} + \alpha_2 \|T_2 u - g_2\|^2_{\ell^2(\Omega)} + |\nabla u|(\Omega) \\ &+ \frac{1}{2\gamma}\left(\|u-a\|^2_{\ell^2(\Omega)} - \|T_1(u-a)\|^2_{\ell^2(\Omega)}\right) + \alpha_2\left(\|u-a\|^2_{\ell^2(\Omega)} - \|T_2(u-a)\|^2_{\ell^2(\Omega)}\right) \\ =& \frac{1}{2\gamma}\|u - z_1\|^2_{\ell^2(\Omega)} + \alpha_2\|u - z_2\|^2_{\ell^2(\Omega)} + |\nabla u|(\Omega) + \psi, \end{aligned}$$

where $z_1 = a + T_1^*(g_1 + v - T_1 a)$, $z_2 = a + T_2^*(g_2 - T_2 a)$, and ψ is a function independent of u. Note that

$$\min_{u\in\mathcal{H}} S(u,a) \Leftrightarrow \min_{u\in\mathcal{H}} \left\| u - \frac{\gamma}{1+2\alpha_2\gamma}\left(\frac{1}{\gamma}z_1 + 2\alpha_2 z_2\right)\right\|^2_{\ell^2(\Omega)} + \frac{2\gamma}{1+2\alpha_2\gamma}|\nabla u|(\Omega), \tag{5.10}$$

which is a variant of the ROF-problem [42]. There exist several numerical methods for solving (5.10) efficiently; see for example [11, 16, 18, 21, 22, 29, 30, 31, 37, 40]. Hence an approximate solution of (5.9) can be computed by the following iterative algorithm: Initialize $u^{(0)} \in \mathcal{H}$ and iterate

$$u^{(\ell+1)} = \arg\min_{u\in\mathcal{H}} S(u, u^{(\ell)}) \quad \ell \ge 0. \tag{5.11}$$

If $T_1 = T_2 = I$, then we can solve (5.9) directly by means of one of the aforementioned solvers, since it becomes the ROF-problem, i.e.,

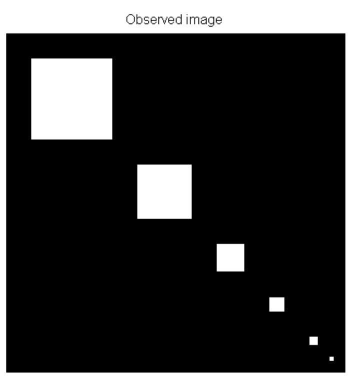
$$\min_{u\in\mathcal{H}} \left\| u - \frac{\gamma}{1+2\alpha_2\gamma}\left(\frac{1}{\gamma}(g_1+v) + 2\alpha_2 g_2\right)\right\|^2_{\ell^2(\Omega)} + \frac{2\gamma}{1+2\alpha_2\gamma}|\nabla u|(\Omega).$$

Numerical Examples. In Example 2.1 above we compute the exact solution of the minimization problem (2.1) with $\varphi(t) = t$. There we show that the newly proposed L^1-L^2-TV model better preserves the original signal than either the L^1-TV model or the L^2-TV model. In this section we support this result by numerical computations for different choices of α_1 and α_2 in (5.5) for a given image g $(= g_1 = g_2)$, which is specified below. As a comparison for the different restoration qualities of the image we use the PSNR (peak signal-to-noise ratio) given by

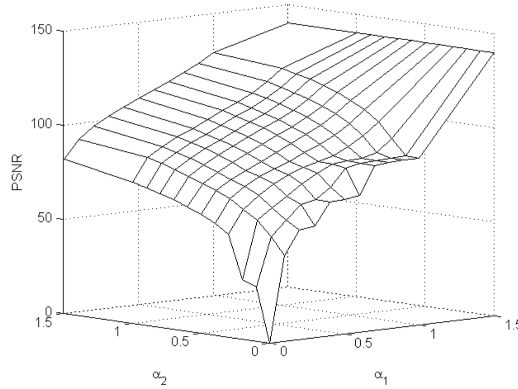
$$\mathrm{PSNR} = 20 \log \frac{1}{\|u_{org} - u^*\|},$$

where u_{org} denotes the original image before any corruption and u^* the restored image. In general we have that the higher the PSNR value is the "closer" is the reconstruction to the true image.

The chosen test image u_{org}, shown in Figure 5.1(a), consists of squares of various sizes. We are interested in selecting α_i, $i = 1, 2$, such that the original image u_{org} is preserved best. Therefore we compute the solution of (5.5) with $g = u_{org}$ and $T_i = I$, $i = 1, 2$, for $\alpha_1, \alpha_2 \in \{0, 0.1, 0.2, ..., 0.9, 1, 1.5\}$ and depict the obtained PSNR values in Figure 5.1(b). Note that for $\alpha_1 = \alpha_2 = 0$ we set the PSNR value to the default value of 0. For $\alpha_2 = 0$ and $\alpha_1 > 0$ we see the typical behavior of the L^1-TV model. In fact, depending on the size of α_1 different scales of the image features are preserved exactly. Typically, for decreasing α_1 features at smaller scales are suddenly "lost" in the reconstruction whereas other features are still recovered perfectly. Also the fading-away effect of image features at various scales depending on the decreasing choice of α_2 of the L^2-TV model can be seen clearly. However, Figure 5.1(b) shows that a combination of $\alpha_1 > 0, \alpha_2 > 0$ gives always a better restoration then setting one of the parameters to 0, respectively.



(a)

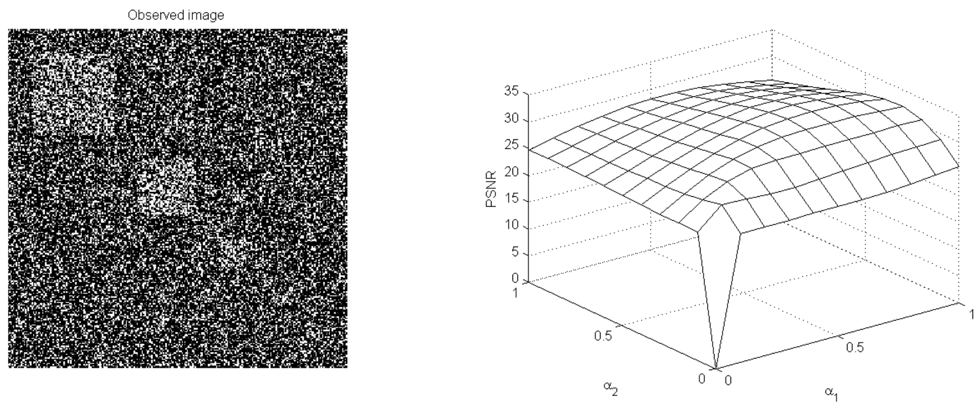


(b)

FIG. 5.1. *(a) Phantom image. (b) PSNR-values of the scale space generated by minimizing the L^1-L^2-TV model for different choices of the parameters α_1 and α_2.*

In the second experiment we corrupt the original image of Figure 5.1(a) by *Gaussian noise and salt-and-pepper noise*, i.e., g is now the image in Figure 5.2(a). Then we again compute the solution of (5.5) for $\alpha_1, \alpha_2 \in \{0, 0.1, 0.2, ..., 1\}$ and depict the obtained PSNR values in Figure (5.2)(b). The maximal PSNR value is reached for $\alpha_1 = 0.7$ and $\alpha_2 = 0.4$, which shows that for a combination of these two noise types the L^1-L^2-TV model outperforms the L^1-TV model as well as the L^2-TV model.

5.3.2. Implementation of the Domain Decomposition Algorithms. If we compute the minimizer of the functional (4.1) either via the sequential or parallel non-overlapping domain decomposition algorithm or via the sequential or parallel overlapping domain decomposition algorithm, then, on each subdomain, we have to solve a constrained optimization problem of the



(a) (b)

FIG. 5.2. *(a) Image of Figure 5.1 corrupted by Gaussian noise (with zero mean and variance $\sigma = 0.1$) and 75% salt-and-pepper noise (more precisely $p_1 = 0.5$ and $p_2 = 0.25$). (b) PSNR-values of the minimizer of the L^1-L^2-TV model for different choices of the parameters α_1 and α_2.*

type

$$\min_{\xi \in \mathcal{H}} J_{\alpha_1,\alpha_2}(\xi) \quad \text{s.t. } A\xi = b \tag{5.12}$$

where $A : \mathcal{H} \to \mathcal{H}$ is a linear operator or more precisely an orthogonal projection, i.e., $A = \pi_{V_i^c}$ for $i = 1, 2$. There exist several numerical methods, which efficiently solve (5.12). Instances are the *Augmented Lagrangian Method* [5, 33] or its variations known as Bregman iterations [40, 52, 53], because of their relation to the Bregman distance [8].

Note that the functional J_{α_1,α_2} is defined on all of Ω. We describe now how one may reduce the dimensionality of the subproblems and solve the resulting problems.

Subspace Minimization. We consider the minimization problem in Ω_1, which can be written as

$$\min_{u_1 \in V_1} J_{\alpha_1,\alpha_2}(u_1 + u_2) = \alpha_1 \|T_1(u_1 + u_2) - g_1\|_{\ell^1(\Omega)} + \alpha_2 \|T_2(u_1 + u_2) - g_2\|^2_{\ell^2(\Omega)} + |\nabla(u_1 + u_2)|(\Omega), \tag{5.13}$$

where $u_2 \in V_2$ is fixed. In order to compute a minimizer of (5.13) we use the algorithm described in Section 5.3.1 by noting that we minimize over $u_1 \in V_1$. That is, (i) we introduce a new variable $v = T_1(u_1 + u_2) - g_1$, (ii) we regularize the functional in (5.13) as in (5.6), (iii) we solve (5.7), and (iv) instead of (5.9) we minimize

$$\min_{u_1 \in V_1} \frac{1}{2\gamma} \|T_1 u_1 - (g_1 - T_1 u_2) - v\|^2_{\ell^2(\Omega)} + \alpha_2 \|T_2 u_1 - (g_2 - T_2 u_2)\|^2_{\ell^2(\Omega)} + |\nabla(u_1 + u_2)|(\Omega). \tag{5.14}$$

Similarly to (5.11), an approximate solution to (5.14) may be computed by the following iterative algorithm. Initialize $u_1^{(0)} \in V_1$ and iterate

$$u_1^{(\ell+1)} \in \arg \min_{u_1 \in V_1} S(u_1 + u_2, u_1^{(\ell)} + u_2) \quad \ell \geq 0. \tag{5.15}$$

Thanks to the splitting property of the total variation, i.e.,

$$|\nabla(u_1 + u_2)|(\Omega) = |\nabla(u_1 + u_2)|(\Omega_1 \cup \tilde{\Omega}_2) + f(u_2), \tag{5.16}$$

where f is a function independent of u_1 (see [2]), we can restrict (5.15) to $\Omega_1 \cup \tilde{\Omega}_2$, where $\tilde{\Omega}_2 \subset \Omega_2^c$ is a small neighborhood strip around the interface Γ_1. Hence the minimization problem in (5.15) is equivalent to

$$\min_{u_1\in V_1}\left\|u_1+u_2-\frac{\gamma}{1+2\alpha_2\gamma}\left(\frac{1}{\gamma}z_1+2\alpha_2 z_2\right)\right\|^2_{\ell^2(\Omega_1\cup\tilde\Omega_2)}+\frac{2\gamma}{1+2\alpha_2\gamma}|\nabla(u_1+u_2)|(\Omega_1\cup\tilde\Omega_2)$$

with $z_1=u_1^{(\ell)}+u_2+T_1^*(g_1+v-T_1u_1^{(\ell)}-T_1u_2)$ and $z_2=u_1^{(\ell)}+u_2+T_2^*(g_2-T_2u_1^{(\ell)}-T_2u_2)$. We compute a solution of this problem by solving the following constrained minimization problem

$$\begin{aligned}&\min_{\xi_1\in V_1\oplus\tilde V_2}\left\|\xi_1-\frac{\gamma}{1+2\alpha_2\gamma}\left(\frac{1}{\gamma}z_1+2\alpha_2 z_2\right)\right\|^2_{\ell^2(\Omega_1\cup\tilde\Omega_2)}+\frac{2\gamma}{1+2\alpha_2\gamma}|\nabla\xi_1|(\Omega_1\cup\tilde\Omega_2),\\ &\text{s.t. } \pi_{\tilde V_2}\xi_1=\pi_{\tilde V_2}u_2,\end{aligned}\tag{5.17}$$

where $\tilde V_2:=\{v\in\mathcal{H}:\mathrm{supp}(v)\subset\tilde\Omega_2\}$. Note that ξ_1 is optimal if and only if $u_1=\xi_1-\pi_{\tilde V_2}u_2\in V_1$ is optimal. We compute a minimizer of the problem in (5.17), which is of the form (5.12), by the Bregmanized Operator Splitting - Split Bregman Algorithm [35].

REMARK 5.1. *The minimization* (5.7) *can be achieved very efficiently on the whole domain Ω, since we only have to perform a soft-thresholding. On the other hand we could restrict the constrained L^2-TV minimization* (5.17) *to the domain $\Omega_1\cup\tilde\Omega_2$, i.e., on Ω_1 plus a small stripe around the interface. This is possible since we freed u_1 from the operators T_1 and T_2 and because of the splitting property of the total variation* (5.16).

REMARK 5.2 (L^1-TV Minimization). *In the case when $\alpha_2=0$ and $\alpha_1>0$, i.e., the minimization problem in* (5.5) *becomes the L^1-TV model, each subspace minimization problem can be computed in the same way as described above. In fact, we first minimize* (5.7) *and then we solve the constrained minimization problem* (5.17)*, which simplifies to*

$$\min_{\xi_1\in V_1\oplus\tilde V_2}\|\xi_1-z_1\|^2_{\ell^2(\Omega_1\cup\tilde\Omega_2)}+2\gamma|\nabla\xi_1|(\Omega_1\cup\tilde\Omega_2)\quad s.t.\ \pi_{\tilde V_2}\xi_1=\pi_{\tilde V_2}u_2.$$

REMARK 5.3 (Denoising). *If $T_1=T_2=I$, then we do not need surrogate functionals and hence we do not have to perform the iterative algorithm* (5.15)*. Instead we restrict* (5.14) *directly to $\Omega_1\cup\tilde\Omega_2$ and solve the following constrained minimization problem*

$$\begin{aligned}&\min_{\xi_1\in V_1\oplus\tilde V_2}\left\|\xi_1-\frac{\gamma}{1+2\alpha_2\gamma}\left(\frac{1}{\gamma}(g_1+v)+2\alpha_2 g_2\right)\right\|^2_{\ell^2(\Omega_1\cup\tilde\Omega_2)}+\frac{2\gamma}{1+2\alpha_2\gamma}|\nabla\xi_1|(\Omega_1\cup\tilde\Omega_2)\\ &s.t.\ \pi_{\tilde V_2}\xi_1=\pi_{\tilde V_2}u_2.\end{aligned}\tag{5.18}$$

The minimization problem in Ω_2 can be computed in the same way by adjusting the notations accordingly.

A Special Case. The implementation of the special case $T_i=1_{\Omega_i}\tilde T_i$ and $g_i=1_{\Omega_i}\tilde g_i$, where $\tilde T_i:\mathcal{H}\to\mathcal{H}$ and $\tilde g_i\in\mathcal{H}$, for $i=1,2$, is considered next. Note that the case considered here is more general than the situation discussed in Section 5.2 on page 15. The minimization problem in (5.5) can be written as

$$\min_{u\in\mathcal{H}}\alpha_1\|\tilde T_1u-\tilde g_1\|_{\ell^1(\Omega_1)}+\alpha_2\|\tilde T_2u-\tilde g_2\|^2_{\ell^2(\Omega_2)}+|\nabla u|(\Omega).$$

When we solve this problem via one of the suggested domain decomposition methods, then on each subdomain we have to compute the minimizer of a constrained optimization problem. For example, in Ω_1 we have

$$\min_{u_1\in V_1}\{\alpha_1\|\tilde T_1(u_1+u_2)-\tilde g_1\|_{\ell^1(\Omega_1)}+\alpha_2\|\tilde T_2(u_1+u_2)-\tilde g_2\|^2_{\ell^2(\Omega_2)}+|\nabla(u_1+u_2)|(\Omega)\}.\tag{5.19}$$

A solution of this problem can be obtained as described in Section 5.3.2. However, numerically it turns out that there exists a more efficient way of minimizing the problem in Ω_1, while this strategy is not applicable for the minimization in Ω_2 due to the special structure of the minimization problems with the ℓ^1-norm defined only on Ω_1. We explain the main idea of this approach.

1. Free u_1 from the influence of $\tilde{T}_2$ (respectively T_2) by introducing a surrogate functional in a similar way as before, i.e., for $a \in V_1$

$$\begin{aligned} S(u_1+u_2,a) &:= \alpha_1\|\tilde{T}_1(u_1+u_2)-\tilde{g}_1\|_{\ell^1(\Omega_1)} + \alpha_2\|T_2(u_1+u_2)-g_2\|^2_{\ell^2(\Omega)} + |\nabla(u_1+u_2)|(\Omega) \\ &\quad + \alpha_2\left(\|u_1-a\|^2_{\ell^2(\Omega)} - \|T_2(u_1-a)\|^2_{\ell^2(\Omega)}\right) \\ &= \alpha_1\|\tilde{T}_1(u_1+u_2)-\tilde{g}_1\|_{\ell^1(\Omega_1)} + \alpha_2\|u_1-z_2\|^2_{\ell^2(\Omega)} + |\nabla(u_1+u_2)|(\Omega) + \psi, \end{aligned}$$

where $z_2 = a + T_2^*(g_2 - T_2u_2 - T_2a)$ and ψ is a function independent of u_1. Then compute an approximate solution of (5.19) by the following algorithm: Initialize $u_1^{(0)} \in V_1$ and iterate

$$u_1^{(\ell+1)} = \arg\min_{u_1\in V_1} S(u_1+u_2, u_1^{(\ell)}) \quad \ell \geq 0; \tag{5.20}$$

2. Thanks to (5.16), we can restrict the surrogate functional iteration to $\Omega_1 \cup \tilde{\Omega}_2$.
3. In each surrogate iteration (5.20) one has to solve a constrained minimization problem. Exemplarily we describe here the Bregmanized Operator Splitting of [53]. For this purpose, let $\mu, \delta > 0$ and initialize $\xi_1^{(0)} \in V_1 \oplus \tilde{V}_2$ and $b^{(0)} = b = u_2$. Then for $k = 0, 1, \ldots$ solve

$$\begin{aligned} \xi_1^{(k+1)} &= \arg\min_{\xi_1\in V_1\oplus\tilde{V}_2} S(\xi_1, u_1^{(\ell)}) + \frac{\mu}{\delta}\|\xi_1 - (\xi_1^{(k)} - \delta\pi^*_{\tilde{V}_2}(\pi_{\tilde{V}_2}\xi_1^{(k)} - b^{(k)}))\|^2_{\ell^2(\Omega_1\cup\tilde{\Omega}_2)} \\ b^{(k+1)} &= b^{(k)} - \pi_{\tilde{V}_2}\xi_1^{(k+1)} + b. \end{aligned} \tag{5.21}$$

4. Solve the minimization problem in (5.21) by the algorithm introduced in Section 5.3.1.

Remark 5.4. *Practically it seems that recomputing the Bregman update outside of the algorithm of Section 5.3.1 is preferable (than computing the update inside the algorithm of Section 5.3.1, as it is done in Section 5.3.2), as the resulting overall algorithm seems to converge faster according to our numerical practice.*

Remark 5.5. *In the case when $\tilde{T}_1 = \tilde{T}_2 = I$ and $\Omega_1 \cap \Omega_2 = \emptyset$, then on each domain we have to solve the following constrained minimization problems*

$$\min_{u_1\in V_1\oplus\tilde{V}_2} \alpha_1\|u_1-\tilde{g}_1\|_{\ell^1(\Omega_1\cup\tilde{\Omega}_2)} + |\nabla(u_1+u_2)|(\Omega_1\cup\tilde{\Omega}_2) \quad s.t.\ \pi_{\tilde{V}_2}u_1 = 0 \tag{5.22}$$

and

$$\min_{u_2\in V_2\oplus\tilde{V}_1} \alpha_2\|u_2-\tilde{g}_2\|^2_{\ell^2(\Omega_2\cup\tilde{\Omega}_1)} + |\nabla(u_1+u_2)|(\Omega_2\cup\tilde{\Omega}_1) \quad s.t.\ \pi_{\tilde{V}_1}u_2 = 0, \tag{5.23}$$

where $u_1 \in V_1$ is fixed in (5.23) *and $\tilde{\Omega}_1$ is defined analogously to $\tilde{\Omega}_2$. The subspace minimization problem* (5.23) *can be solved, for example, by the Bregmanized Operator Splitting - Split Bregman algorithm [35], while one may solve* (5.22) *as suggested in this section above starting at point 3.*

Due to the structure of the problem, for the minimization in Ω_2 this approach is not applicable and hence we suggest to use the strategy of Section 5.3.2.

5.4. Numerical Experiments. In the following we present numerical experiments for the proposed sequential and parallel algorithms. In particular, we show applications in image denoising, inpainting, and deblurring. The values of the parameters α_1 and α_2 in the objective functional (5.5) are chosen according to the application and experimentally, i.e., we choose the values which give a good compromise between visual quality and computational time of the algorithm. The optimal selection of α_1 and α_2 is an interesting research topic in its own right, but is beyond the scope of the present paper. However, it has been shown in several examples, see [9, 38, 39], that if only salt-and-pepper noise is present in an image then the L^1-TV model outperforms the L^2-TV model. In fact, the minimizer of the L^1-TV model ensures that for any $x \in \Omega$ we have $(T_1u)(x) = g_1(x)$ when $g_1(x)$ is not corrupted while the remaining points correspond to the regularization. The ℓ^2-norm cannot do that and is additionally less sensitive to small errors than the

ℓ^1-norm, which leads the data-fitting error of the ℓ^1-term to be smaller than of the ℓ^2-term. These facts motivate that we will use the pure L^1-TV model when only salt-and-pepper noise corrupted the image of interest, while when only Gaussian noise is present in the image then we are advised to use the pure L^2-TV model.

5.4.1. Numerical Results – Sequential Algorithms. In our numerical experiments we terminate our sequential algorithms (2.2) and (5.2) as soon as the norm of the difference of two successive iterates drops below a certain threshold. More precisely, we use as a stopping criterion

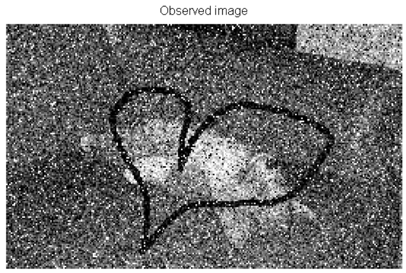
$$\|u^{(n)} - u^{(n+1)}\| < 10^{-6},$$

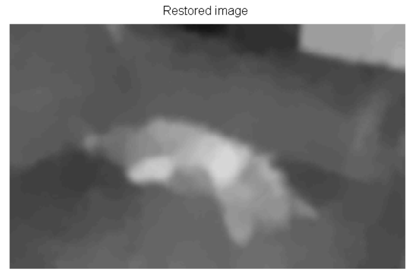
which seems suitable for our purposes, since it indicates that we are close to a solution. In fact, if our algorithm converges at least linearly, i.e., there exists an $\varepsilon \in (0,1)$ and an $m > 0$ such that for all $n \geq m$ we have $\|u^{(n+1)} - u^{(\infty)}\| \leq \varepsilon \|u^{(n)} - u^{(\infty)}\|$, our above stopping criterion ensures that the distance between our obtained result u and $u^{(\infty)}$ is $\|u - u^{(\infty)}\| \leq \frac{10^{-6}\varepsilon}{1-\varepsilon}$. Moreover, if we set $\alpha_2 > 0$, then we depict the minimal Lagrange multiplier $\eta^{(n)} := \min_i\{\|\eta_i^{(n)}\|_{\ell^2(\Omega)}\}$, which - according to Corollary 4.11 (see also Theorem 4.10) - indicates how close the computed solution is to the real global solution. In fact, by monitoring the progress of the minimal Lagrange multiplier during the iterations we are able to decide numerically whether the proposed domain decomposition algorithms converge indeed to the global solution; see Figure 5.3(c)-(d), Figure 5.5(c)-(d), and Figure 5.6(c).

We apply the overlapping and non-overlapping domain decomposition algorithm in (2.2) to the image shown in Figure 5.3(a) by decomposing the image domain into two overlapping or non-overlapping subdomains respectively. This image of size 167×270 pixels has partly lost data (black heart) while it is also corrupted by 10% of salt-and-pepper noise (i.e., 10% of the pixels are either flipped to black or white) and by Gaussian white noise with zero mean and variance 0.03. In this example the operators T_1 and T_2 act as $T_i u = 1_{\Omega \setminus K} u$ for $i = 1, 2$, where Ω denotes the image domain and $K \subset \Omega$ the set in which the original image content got lost. The parameters α_1 and α_2 are chosen to be 0.4, while $\gamma = 0.01$, $\mu = 1$, and $\delta = 0.99$. In Figure 5.3(b) we depict the result computed by the overlapping domain decomposition algorithm. Since α_2 is choosen to be positive we can check by plotting the progress of the minimal Lagrange multiplier whether the iterates computed by the domain decomposition algorithms indeed converge to the expected minimizer of the global functional. In fact, we see in Figure 5.3(c) and (d) that the norm of the minimial Lagrange multipliers converge to 0 and hence the algorithms generate sequences which converge to the global minimizer.

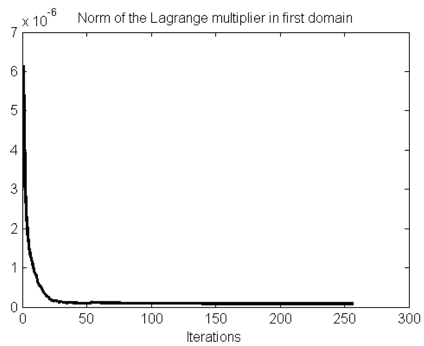
In the next example we present the successful application of a domain decomposition for the problem of pure L^1-TV minimization. Figure 5.4(a) shows the previously used image rescaled to size 334×540 pixels which is now corrupted by a Gaussian blur with kernel size 15×15 pixels and standard deviation 2 and in addition 2% salt-and-pepper noise. In order to restore the image we decompose the image domain into two non-overlapping subdomains and solve the resulting problems on the respective subdomains alternating by the non-overlapping domain decomposition algorithm (5.2). Since there is no Gaussian noise present, this is a typical example for L^1-TV minimization, i.e., we set $\alpha_2 = 0$ in (5.5). We choose $\alpha_1 = \frac{5}{3}$, $\gamma = 0.01$, $\mu = 1$, and $\delta = 0.99$ and obtain the image in Figure 5.4(b).

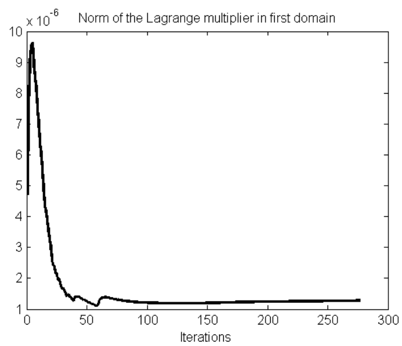
Further we illustrate the successful application of the non-overlapping domain decomposition algorithm (5.2) when both salt-and-pepper noise and Gaussian noise are present. In particular, we apply our algorithm to an image with a missing part which is corrupted by salt-and-pepper noise in the upper half while in the lower half only Gaussian noise is present. We are aware that this is a rather artificial example but very interesting from a numerical point of view. Note that since the total variation is non-local and hence non-additive it is not possible to obtain a correct global solution by just cutting the image into an upper and a lower part, and then computing the solutions separately and putting them together. However, since we are in the setting of the special situation of Section 5.2, we are able by using our non-overlapping domain decomposition algorithm in (5.2) to split the image into domains in which only one type of noise is present. Then we only





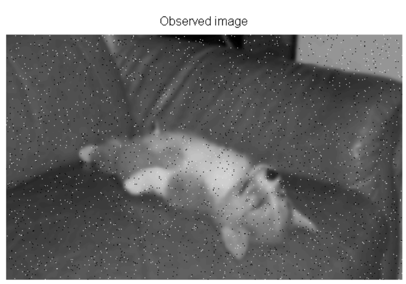
(a) (b)

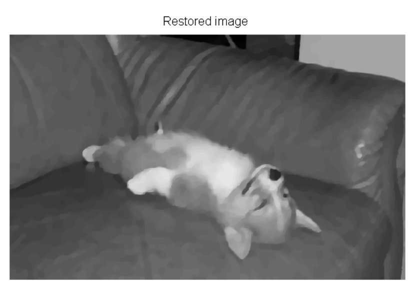




(c) (d)

Fig. 5.3. *Domain decomposition for L^1-L^2-TV minimization. Parameters: $\alpha_1 = \alpha_2 = 0.4$, $\gamma = 0.01$, $\mu = 1$, $\delta = 0.99$, and ROF-problem solved via Split Bregman with tolerance 10^{-3}. In (a) we show an image of size 167×270 pixels with a missing part (black heart) and corrupted by 10% salt-and-pepper noise and Gaussian noise with zero mean and variance 0.03. The restored image is shown in (b). In (c) we depict the progress of the minimal Lagrange multiplier $\eta^{(n)}$ obtained by the overlapping domain decomposition algorithm while in (d) we plot the one obtained by the non-overlapping domain decomposition.*

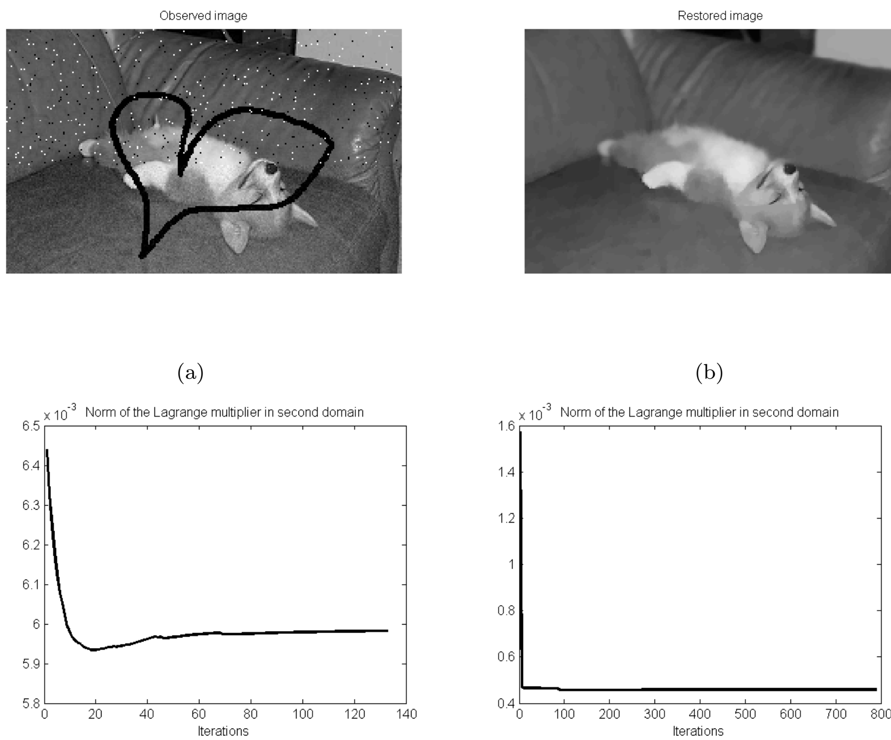




(a) (b)

Fig. 5.4. *Non-overlapping domain decomposition algorithm for L^1-TV minimization. Parameters: $\alpha_1 = 5/3$, $\gamma = 0.01$, $\mu = 1$, $\delta = 0.99$, and ROF-problem solved via Split Bregman with tolerance 10^{-4}. In (a) we show an image of size 334×540 pixels which is corrupted by a Gaussian blur (size 15×15; standard deviation 2) and 2% salt-and-pepper noise. The restored image is shown in (b).*

have to solve on each domain either a constrained L^1-TV minimization problem, cf. (5.3), or a constrained L^2-TV minimization problem, cf. (5.4). These problems are in general easier to solve than the original L^1-L^2-TV problem. Figure 5.5(a) is such an image (size 167×270 pixels), which we restore by the non-overlapping domain decomposition algorithm (5.2) with $\mu = 100$ in the upper half and $\mu = 1$ in the lower half, $\alpha_1 = \frac{5}{3}$, $\alpha_2 = \frac{50}{3}$, $\gamma = 0.01$, and $\delta = 0.99$. The computed result is shown in Figure 5.5(b). By depicting the norm of the minimal Lagrange multiplier $\eta^{(n)}$ we check additionally, whether the algorithm converges to the right solution. In Figure 5.5(c) we see the progress of the minimal Lagrange multiplier for the image in Figure 5.5(a) with size 167×270 pixels. By improving the image resolution yielding a three times finer grid, i.e., the image has now 501×810 pixels, we obtain a sequence $(\eta^{(n)})_n$ which converges to a significantly smaller norm; see Figure 5.5(d). If we keep increasing the image resolution we observe that the norm of the elements of the sequence $(\eta^{(n)})_n$ continuous to converge to smaller and smaller values. Hence, we conclude that the convergence of the domain decomposition algorithms may also depend on the step-size h.



(a) (b)

(c) (d)

FIG. 5.5. *Non-overlapping domain decomposition algorithm for L^1-L^2-TV minimization. Parameters: $\alpha_1 = \frac{5}{3}$, $\alpha_2 = \frac{50}{3}$, $\gamma = 0.01$, $\mu = 100$ in Ω_1 and $\mu = 1$ in Ω_2, $\delta = 0.99$, and ROF-problem solved via Split Bregman with tolerance 10^{-4}. In (a) we show an image of size 167×270 pixels with a missing part (black heart) and corrupted by 2% salt-and-pepper noise and Gaussian noise with variance 0.001. The restored image is shown in (b). In (c) we depict the progress of the minimal Lagrange multiplier $\eta^{(n)}$ as well as in (d) for a three times finer grid.*

With respect to the Gauss-Southwell-r rule considered in Section 4.2, we observe in our previous L^1-L^2-TV minimization examples that inequality (4.22) is more likely to be satisfied with a fixed constant $\nu > 0$ for overlapping rather than non-overlapping domain decomposition. For instance, for the problem depicted in Figure 5.3 the Gauss-Southwell-r rule is satisfied with $\nu \leq 0.03837$ along the iteration in the overlapping case. For the non-overlapping decomposition one has $\nu \leq 0.000195$.

In the next section we show the successful application of our solvers when both Gaussian noise as well as salt-and-pepper noise are present simultaneously (and in a non-separated fashion) in an image; see the results depicted in Figure 5.6.

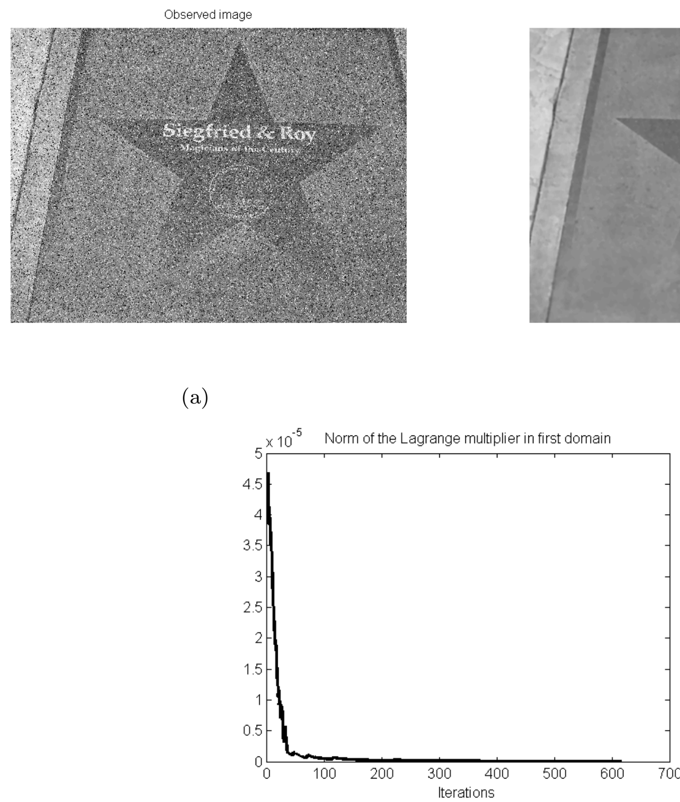
5.4.2. Numerical Results – Parallel Algorithms. Finally, we show the efficiency of the parallel algorithm in (2.3) for non-overlapping and overlapping domain decomposition and compare their numerical performance with the L^1-L^2-TV algorithm introduced in Section 5.3.1. Note that in the L^1-L^2-TV algorithm the problem is solved on all of Ω without any splitting into subdomains. In the domain decomposition algorithms we consider domain splittings into $D = 4, 8, 16, 32$ subdomains. Since we are comparing now the convergence speed of different algorithms the stopping criterion used before is no longer suitable. Now we stop the algorithms as soon as the energy J_{α_1,α_2} reaches a significance level J^*, i.e., when $J_{\alpha_1,\alpha_2}(u^{(n)}) \leq J^*$ for the first time. The level J^* is chosen visually, i.e., we once restore the image of interest until we observe a visually satisfying restoration and record the associated energy-value as J^*.

For our comparison let us consider the image in Figure 5.6, which is of size 1920×2576 pixels and corrupted by Gaussian noise with standard deviation 0.01 as well as by 10% salt-and-pepper noise on all Ω; see Figure 5.6(a). In the domain decomposition algorithms as well as in the L^1-L^2-TV algorithm we denoise this image by choosing $\alpha_1 = 0.5$, and $\alpha_2 = 0.4$. The computations are done in Matlab on a Linux cluster with 32 kernels, where each kernel has 2 processors and each processor 4 cores, i.e., on a computer with 256 cores, and the multithreading-option is activated such that all algorithms (including the L^1-L^2-TV algorithm without domain decomposition) take advantage of the parallel infrastructure offered by the hardware. For the domain decomposition algorithms we split the domain into non-overlapping or overlapping strips. The overlap is chosen to be a stripe of width 10 pixels, i.e., the overlap is of size 10×2576 pixels. For different numbers of splittings we show in Table 5.1 the required computational time and the number of iterations until the algorithms reach the significant energy of $J^* = 0.080041483485$ (see Figure 5.6(b) for the restored image). Note that the structure of the problems in the subdomains is different from the one of the global problem. More precisely, on each subdomain we have to solve constrained minimization problems, cf (2.4), which are structurally more difficult to solve than just minimizing an energy as for the global problem. Hence by domain decomposition on the one hand we reduce the dimensionality of the problem, but on the other hand we increase the complexity on each subdomain. Additionally, we also have to take the communication time of the processors into account. These facts add to the overall computing time. Therefore we cannot expect a very dramatic decrease in computational time once the number of subdomains gets large. Nevertheless, we see in Table 5.1 that the domain decomposition algorithms for splittings with $D = 4, 8, 16, 32$ are still much faster than without decomposition ($D = 1$). In this case, for a non-overlapping splitting into 8 domains the best performance is guaranteed, while for decomposing into 16 or more domains the algorithm already requires more time to reach its stopping criterion.

Splitting the image domain into larger subdomains, as it happens for an overlapping decomposition, one may expect an increase in computational time. This is not necessarily true, as the solution in the overlap is computed twice per iteration, which decreases the number of iterations. We even see in Table 5.1 that for a fixed number of subdomains the larger the overlapping region is the less iterations are performed. In our numerical experiments we observe that for an overlapping decomposition into 16 domains with overlaps of size 50×2576 pixels the domain decomposition algorithm performs best with respect to the number of iterations and computational time.

We also observe that with increasing the number of subdomains the number of iterations is decreasing. For a non-overlapping decomposition the number of iterations is only decreasing very slowly, while for overlapping decompositions the decay is more noticeable. For the overlapping splitting when doubling the number of domains we see from Table 5.1 that for a larger overlap the absolute reduction of the number of iterations is larger than for a smaller overlap, while the relative reduction is bigger for smaller overlap.

6. Conclusion. We have proposed a combined L^1/L^2-TV energy with total variation regularization which outperforms the pure L^1-TV or L^2-TV models as it preserves details better than the L^2-TV model and it does not suffer from a sudden loss of image features like the L^1-TV model. Moreover, it is superior (in PSNR) in restoration tasks where images are corrupted simultaneously



(a) (b)

(c)

FIG. 5.6. *Parallel domain decomposition for L^1-L^2-TV minimization of an image (size 1920×2576 pixels) corrupted by Gaussian noise with standard deviation 0.01 and 10% salt-and-pepper noise, see (a). In (b) we show the restored image, whereby we used the non-overlapping domain decomposition algorithm for 8 domains with the following parameters: $\alpha_1 = 0.5$, $\alpha_2 = 0.4$, $\gamma = 0.01$, $\mu = 10$, $\delta = 0.99$, and ROF-problem solved via Split Bregman with tolerance 10^{-4}. In (c) we depict the progress of the minimal Lagrange multiplier $\eta^{(n)}$.*

by Gaussian and salt-and-pepper noise.

For the numerical solution of the L^1/L^2-TV energy we have proposed and analyzed sequential and parallel subspace correction methods, which generate a convergent (sub)sequence of iterates and a monotone decrease of the energy. Moreover, we have shown that the distance between limit points and the global minimizer of the L^1/L^2-TV energy is bounded by the norm of the minimal Lagrange multiplier associated with involved subspace projection constraints. In our numerical experiments this norm appeared often very small or it might even tend to 0, indicating that convergence to the global optimum was obtained. However, in the rare cases where the norm of the minimal Lagrange multiplier did not tend to 0 we observed a resolution dependent effect reducing the multiplier norm under increasing resolution. This behavior may certainly motivate further research on the convergence of subspace correction methods for minimizing non-smooth and non-additive objectives. We have also shown that the parallel version pays off up to the number of subdomains where the communication between processors becomes dominant.

We also mention that the theoretical analysis of subspace correction methods for non-smooth and non-additive functionals is still far from being complete. In particular, in general Banach spaces there is not much known about such methods and their convergence to a global minimizer. Not even in a discrete setting for dimensions $d > 1$ this question has been answered yet without invoking (rather restrictive) assumptions.

Appendix A. Proof of Proposition 4.5.

# domains	non-overlapping alg.	overlapping alg. (overlap 10×2576 pixels)	overlapping alg. (overlap 50×2576 pixels)
$D = 1$ (L^1-L^2-TV alg.):	7882 s / 698 it		
$D = 4$:	5727 s / 617 it	5834 s / 607 it	5998 s / 561 it
$D = 8$:	5090 s / 618 it	5074 s / 596 it	5265 s / 499 it
$D = 16$:	5409 s / 588 it	5432 s / 560 it	5014 s / 371 it
$D = 32$:	6814 s / 586 it	6605 s / 501 it	5203 s / 242 it

TABLE 5.1

Denoising for the image in Figure 5.6: Computational performance (CPU time in seconds and the number of iterations) for the global L^1-L^2-TV algorithm and for the parallel domain decomposition algorithms with $\alpha_1 = 0.5$, $\alpha_2 = 0.4$ for different numbers of subdomains ($D = 4, 8, 16, 32$) and overlapping sizes.

It is clear that $\zeta \in \partial J_{\alpha_1,\alpha_2}(u)$ if and only if $u \in \operatorname{argmin}_{v\in\mathcal{H}}\{J_{\alpha_1,\alpha_2}(v) - \langle \zeta, v\rangle_{\mathcal{H}}\}$, and let us consider the following variational problem:

$$\inf_{v\in\mathcal{H}}\{J_{\alpha_1,\alpha_2}(v) - \langle\zeta, v\rangle_{\mathcal{H}}\} = \inf_{v\in\mathcal{H}}\{\alpha_1\|T_1 v - g_1\|_{\ell^1(\Omega)} + \alpha_2\|T_2 v - g_2\|^2_{\ell^2(\Omega)} + \varphi(|\nabla v|)(\Omega) - \langle\zeta, v\rangle_{\mathcal{H}}\}. \quad (\mathcal{P})$$

We denote such an infimum by $\inf(\mathcal{P})$. Now we compute $(\mathcal{P}^*)$, the dual of $(\mathcal{P})$. Let $\mathcal{F} : \mathcal{H} \to \mathbb{R}$, $\mathcal{G} : \mathcal{H}^d \times \mathcal{H} \times \mathcal{H} \to \mathbb{R}$, $\mathcal{G}_0 : \mathcal{H}^d \to \mathbb{R}$, $\mathcal{G}_1 : \mathcal{H} \to \mathbb{R}$, $\mathcal{G}_2 : \mathcal{H} \to \mathbb{R}$, such that

$$\begin{aligned}
\mathcal{F}(v) &= -\langle\zeta, v\rangle_{\mathcal{H}},\\
\mathcal{G}_0(w_0) &= \varphi(|w_0|)(\Omega),\\
\mathcal{G}_1(\bar{w}) &= \alpha_1\|w_1 - g_1\|_{\ell^1(\Omega)},\\
\mathcal{G}_2(\bar{w}) &= \alpha_2\|w_2 - g_2\|^2_{\ell^2(\Omega)},\\
\mathcal{G}(w) &= \mathcal{G}_0(w_0) + \mathcal{G}_1(w_1) + \mathcal{G}_2(w_2),
\end{aligned}$$

with $w = (w_0, w_1, w_2) \in \mathcal{H}^d \times \mathcal{H} \times \mathcal{H}$. Then the dual problem of $(\mathcal{P})$ is given by (cf. [24, p 60])

$$\sup_{p^*\in\mathcal{H}^d\times\mathcal{H}\times\mathcal{H}}\{-\mathcal{F}^*(\Lambda^* p^*) - \mathcal{G}^*(-p^*)\}, \quad (\mathcal{P}^*)$$

where $\Lambda : \mathcal{H} \to \mathcal{H}^d \times \mathcal{H} \times \mathcal{H}$ is defined by

$$\Lambda v = ((\nabla v)^1, \ldots, (\nabla v)^d, T_1 v, T_2 v)$$

and Λ^* is its adjoint. We denote the supremum in $(\mathcal{P}^*)$ by $\sup(\mathcal{P}^*)$. Using the definition of the conjugate function we compute $\mathcal{F}^*$ and $\mathcal{G}^*$. In particular, we have

$$\mathcal{F}^*(\Lambda^* p^*) = \sup_{v\in\mathcal{H}}\{\langle\Lambda^* p^*, v\rangle_{\mathcal{H}} - \mathcal{F}(v)\} = \sup_{v\in\mathcal{H}}\langle\Lambda^* p^* + \zeta, v\rangle_{\mathcal{H}} = \begin{cases} 0 & \text{if } \Lambda^* p^* + \zeta = 0,\\ \infty & \text{otherwise,}\end{cases}$$

where $p^* = (p_0^*, p_1^*, p_2^*)$, and due to the separability of $\mathcal{G}$ we find

$$\begin{aligned}
\mathcal{G}^*(p^*) &= \sup_{w\in\mathcal{H}^d\times\mathcal{H}\times\mathcal{H}}\{\langle p^*, w\rangle_{\mathcal{H}^d\times\mathcal{H}\times\mathcal{H}} - \mathcal{G}(w)\}\\
&= \sup_{w_0\in\mathcal{H}}\{\langle p_0^*, w_0\rangle_{\mathcal{H}^d} - \mathcal{G}_0(w_0)\} + \sup_{w_1\in\mathcal{H}}\{\langle p_1^*, w_1\rangle_{\mathcal{H}} - \mathcal{G}_1(w_1)\} + \sup_{w_2\in\mathcal{H}}\{\langle p_2^*, w_2\rangle_{\mathcal{H}} - \mathcal{G}_2(w_2)\}\\
&= \mathcal{G}_0^*(p_0^*) + \mathcal{G}_1^*(p_1^*) + \mathcal{G}_2^*(p_2^*).
\end{aligned}$$

We have that

$$\mathcal{G}_2^*(p_2^*) = \left\langle \frac{p_2^*}{4\alpha_2} + g_2, p_2^* \right\rangle_{\mathcal{H}},$$

$$\mathcal{G}_1^*(p_1^*) = \langle p_1^*, g_1 \rangle_{\mathcal{H}},$$

if $|p_1^*| \leq \alpha_1$, and (see [24])

$$\mathcal{G}_0^*(p_0{}^*) = \varphi_+^* \left(|p_0{}^*|\right)(\Omega)$$

if $|p_0^*(x)| \in \operatorname{Dom} \varphi_+^*$, where φ_+^* is the conjugate function of φ_+ defined by

$$\varphi_+(t) := \varphi(|t|) \quad \text{for } t \in \mathbb{R}.$$

Therefore we can write $(\mathcal{P}^*)$ in the following way

$$\sup_{p^* \in \mathcal{K}} \left\{ -\left\langle \frac{-p_2^*}{4\alpha_2} + g_2, -p_2^* \right\rangle_{\mathcal{H}} - \langle g_1, -p_1^* \rangle_{\mathcal{H}} - \varphi_+^* \left(|p_0^*|\right)(\Omega) \right\}, \tag{A.1}$$

where

$$\mathcal{K} = \left\{ p^* \in \mathcal{H}^d \times \mathcal{H} \times \mathcal{H} : |p_0{}^*(x)| \in \operatorname{Dom} \varphi_+^* \text{ and } |p_1^*(x)| \leq \alpha_1 \text{ for all } x \in \Omega, \Lambda^* p^* + \zeta = 0 \right\}.$$

The function φ_+ also fulfills assumption (A_φ)(ii) (i.e., there exists $c_1 > 0, b \geq 0$ such that $c_+ z - b \leq \varphi_+(z) \leq c_+ z + b$, for all $z \in \mathbb{R}^+$). The conjugate function of φ_+ is given by $\varphi_+^*(t) = \sup_{z \in \mathbb{R}} \{\langle t, z \rangle - \varphi_+(z)\}$. Using the previous inequalities and the fact that φ_+ is an even function (i.e., $\varphi_+(z) = \varphi_+(-z)$ for all $z \in \mathbb{R}$) we have

$$\sup_{z \in \mathbb{R}} \{\langle t, z \rangle - c_+|z| + b\} \geq \sup_{z \in \mathbb{R}} \{\langle t, z \rangle - \varphi_+(z)\} \geq \sup_{z \in \mathbb{R}} \{\langle t, z \rangle - c_1|z| - b\} = \begin{cases} -b & \text{if } |t| \leq c_1, \\ \infty & \text{else.} \end{cases} \tag{A.2}$$

In particular, one can see that $t \in \operatorname{Dom} \varphi_+^*$ if and only if $|t| \leq c_1$.

From $\Lambda^* p^* + \zeta = 0$ we obtain

$$\begin{aligned} \langle \Lambda^* p^*, \omega \rangle_{\mathcal{H}} + \langle \zeta, \omega \rangle_{\mathcal{H}} &= \langle p^*, \Lambda \omega \rangle_{\mathcal{H}^{d+2}} + \langle \zeta, \omega \rangle_{\mathcal{H}} \\ &= \langle p_0^*, \nabla \omega \rangle_{\mathcal{H}^d} + \langle p_1^*, T_1 \omega \rangle_{\mathcal{H}} + \langle p_2^*, T_2 \omega \rangle_{\mathcal{H}} + \langle \zeta, \omega \rangle_{\mathcal{H}} = 0 \quad \text{for all } \omega \in \mathcal{H}. \end{aligned} \tag{A.3}$$

Then, since $\langle p_0{}^*, \nabla \omega \rangle_{\mathcal{H}^d} = \langle -\operatorname{div} p_0^*, \omega \rangle_{\mathcal{H}}$ (see Section 3), we have

$$T_1^* p_1^* + T_2^* p_2^* - \operatorname{div} p_0^* + \zeta = 0.$$

Hence we can write $\mathcal{K}$ in the following way

$$\begin{aligned} \mathcal{K} = \Big\{ p^* = (p_0^*, p_1^*, p_2^*) \in \mathcal{H}^d \times \mathcal{H} \times \mathcal{H} : &|p_0{}^*(x)| \leq c_1 \text{ and } |p_1^*(x)| \leq \alpha_1 \text{ for all } x \in \Omega, \\ & T_1^* p_1^* + T_2^* p_2^* - \operatorname{div} p_0^* + \zeta = 0 \Big\}. \end{aligned}$$

We now apply the duality results from [24, Theorem III.4.1], since the objective functional in $(\mathcal{P})$ is convex, continuous with respect to Λv in $\mathcal{H}^d \times \mathcal{H} \times \mathcal{H}$, and $\inf(\mathcal{P})$ is finite. Consequently, $\inf(\mathcal{P}) = \sup(\mathcal{P}^*) \in \mathbb{R}$ and $(\mathcal{P}^*)$ has a solution $M = (M_0, M_1, M_2) \in \mathcal{K}$.

Let us assume that u is a solution of $(\mathcal{P})$ and M is a solution of $(\mathcal{P}^*)$. From $\inf(\mathcal{P}) = \sup(\mathcal{P}^*)$ we get

$$\begin{aligned} \alpha_1 \|T_1 u - g_1\|_{\ell^1(\Omega)} &+ \alpha_2 \|T_2 u - g_2\|_{\ell^2(\Omega)}^2 + \varphi(|\nabla u|)(\Omega) - \langle \zeta, u \rangle_{\mathcal{H}} \\ &= -\left\langle \frac{-M_2}{4\alpha_2} + g_2, -M_2 \right\rangle_{\mathcal{H}} - \langle g_1, -M_1 \rangle_{\mathcal{H}} - \varphi_1^* \left(|M_0|\right)(\Omega), \end{aligned} \tag{A.4}$$

where $M=(M_0,M_1,M_2)\in\mathcal{H}^d\times\mathcal{H}\times\mathcal{H}$, $|M_0(x)|\le c_1$, $|M_1|\le\alpha_1$, and $T_1^*M_1+T_2^*M_2-\operatorname{div}M_0+\zeta=0$, which verifies (4.8). In particular, (A.3) and (A.4) yield

$$\begin{aligned}&\alpha_1\|T_1u-g_1\|_{\ell^1(\Omega)}+\alpha_2\|T_2u-g_2\|^2_{\ell^2(\Omega)}+\varphi(|\nabla u|)(\Omega)+\langle M_1,T_1u\rangle_{\mathcal{H}}+\langle M_2,T_2u\rangle_{\mathcal{H}}\\&\quad+\langle M_0,\nabla u\rangle_{\mathcal{H}^d}+\left\langle\frac{-M_2}{4\alpha_2}+g_2,-M_2\right\rangle_{\mathcal{H}}+\langle g_1,-M_1\rangle_{\mathcal{H}}+\varphi_1^*\left(|M_0|\right)(\Omega)=0.\end{aligned}\tag{A.5}$$

We rewrite (A.5) in the following form:

$$\begin{aligned}&\sum_{x\in\Omega}\alpha_1|(T_1u-g_1)(x)|+\sum_{x\in\Omega}\alpha_2|(T_2u-g_2)(x)|^2+\sum_{x\in\Omega}\varphi(|(\nabla u)(x)|)+\sum_{x\in\Omega}M_1(x)(T_1u)(x)\\&\quad+\sum_{x\in\Omega}M_2(x)(T_2u)(x)+\sum_{x\in\Omega}\langle M_0(x),\nabla u(x)\rangle_{\mathbb{R}^d}-\sum_{x\in\Omega}\left(\frac{-M_2(x)}{4\alpha_2}+g_2(x)\right)(M_2)(x)\\&\quad+\sum_{x\in\Omega}g_1(x)(-M_1)(x)+\sum_{x\in\Omega}\varphi_1^*\left(|M_0(x)|\right)=0.\end{aligned}\tag{A.6}$$

Now for the various terms in (A.6) we have:

1. $\alpha_1|(T_1u-g_1)(x)|+M_1(x)((T_1u)(x)-g_1(x))\ge 0$ since $|M_1(x)|\le\alpha_1$.
2. $\alpha_2|(T_2u-g_2)(x)|^2+M_2(x)(T_2u(x)-g_2(x))+\frac{M_2(x)^2}{4\alpha_2}=\left(\sqrt{\alpha_2}(T_2u-g_2)(x)+\frac{M_2(x)}{2\sqrt{\alpha_2}}\right)^2\ge 0$.
3. $\varphi(|(\nabla u)(x)|)+\langle M_0(x),\nabla u(x)\rangle_{\mathbb{R}^d}+\varphi_1^*\left(|M_0(x)|\right)\ge\varphi(|(\nabla u)(x)|)-\sum_{j=1}^d|M_0^j(x)||S^j|+\varphi_1^*\left(|M_0(x)|\right)\ge 0$ by the definition of φ_1^*, since

$$\varphi_1^*\left(|M_0(x)|\right)=\sup_{S=(S^1,\cdots,S^d)\in\mathbb{R}^d}\{\sum_{j=1}^d|M_0^j(x)||S^j|-\varphi(|S|)\}.$$

Hence, condition (A.6) reduces to

$$\varphi(|(\nabla u)(x)|)+\langle M_0(x),\nabla u(x)\rangle_{\mathbb{R}^d}+\varphi_1^*\left(|M_0(x)|\right)=0,\quad\text{for all } x\in\Omega,\tag{A.7}$$

$$M_2(x)=-2\alpha_2(T_2u-g_2)(x),\quad\text{for all } x\in\Omega,\tag{A.8}$$

$$\alpha_1|(T_1u-g_1)(x)|+M_1(x)((T_1u)(x)-g_1(x))=0,\quad\text{for all } x\in\Omega.\tag{A.9}$$

Conversely, if there exists $M=(M_0,M_1,M_2)\in\mathcal{H}^d\times\mathcal{H}\times\mathcal{H}$ with $|M_0(x)|\le c_1$ and $|M_1|\le\alpha_1$, which fulfills conditions (4.5) and (4.8), then it is clear from our previous considerations that equation (A.4) holds. Let us denote the functional on the left-hand side of (A.4) by

$$P(u):=\alpha_1\|T_1u-g_1\|_{\ell^1(\Omega)}+\alpha_2\|T_2u-g_2\|^2_{\ell^2(\Omega)}+\varphi(|\nabla u|)(\Omega)-\langle\zeta,u\rangle_{\mathcal{H}}$$

and the functional on the right-hand side of (A.4) by

$$P^*(M):=-\left\langle\frac{-M_2}{4\alpha_2}+g_2,-M_2\right\rangle_{\mathcal{H}}-\langle g_1,-M_1\rangle_{\mathcal{H}}-\varphi_1^*\left(|M_0|\right)(\Omega).$$

Hence $\inf P=\inf(\mathcal{P})$ and $\sup P^*=\sup(\mathcal{P}^*)$. Since P is convex, continuous with respect to Λu in $\mathcal{H}^d\times\mathcal{H}\times\mathcal{H}$, and $\inf(\mathcal{P})$ is finite we know from duality results [24, Theorem III.4.1] that $\inf(\mathcal{P})=\sup(\mathcal{P}^*)\in\mathbb{R}$. We assume that M is no solution of $(\mathcal{P}^*)$, i.e., $P^*(M)<\sup(\mathcal{P}^*)$, and u is no solution of $(\mathcal{P})$, i.e, $P(u)>\inf(\mathcal{P})$. Then we have that

$$P(u)>\inf(\mathcal{P})=\sup(\mathcal{P}^*)>P^*(M).$$

Thus (A.4) is valid if and only if M is a solution of $(\mathcal{P}^*)$ and u is a solution of $(\mathcal{P})$ which is equivalent to $\zeta\in\partial J_{\alpha_1,\alpha_2}(u)$.

If, additionally, φ is differentiable and $|(\nabla u)(x)|\ne 0$ for $x\in\Omega$, then $M_0(x)$ can be computed explicitly. In fact, from equation (A.7) (respectively (4.5)) we have

$$\varphi_1^*\left(|-M_0(x)|\right)=-\langle M_0(x),(\nabla u)(x)\rangle_{\mathbb{R}^d}-\varphi(|(\nabla u)(x)|).\tag{A.10}$$

From the definition of conjugate functions we have

$$\begin{aligned}
\varphi_1^*(|-M_0(x)|) &= \sup_{t\in\mathbb{R}} \{|-M_0(x)|t - \varphi_1(t)\} \\
&= \sup_{t\geq 0} \{|-M_0(x)|t - \varphi_1(t)\} \\
&= \sup_{t\geq 0} \sup_{\substack{S\in\mathbb{R}^d \\ |S|=t}} \{\langle -M_0(x), S\rangle_{\mathbb{R}^d} - \varphi_1(|S|)\} \\
&= \sup_{S\in\mathbb{R}^d} \{\langle -M_0(x), S\rangle_{\mathbb{R}^d} - \varphi(|S|)\}.
\end{aligned} \tag{A.11}$$

Now, if $|(\nabla u)(x)| \neq 0$ for $x \in \Omega$, then it follows from (A.10) that the supremum is taken in $S = |(\nabla u)(x)|$ and we have $\nabla_S(-\langle M_0(x), S\rangle_{\mathbb{R}^d} - \varphi(|S|)) = 0$ which implies

$$M_0^j(x) = -\frac{\varphi'(|(\nabla u)(x)|)}{|(\nabla u)(x)|}(\nabla u)^j(x) \quad j = 1, \ldots, d,$$

and verifies (4.9). This finishes the proof.

REFERENCES

[1] S. Aliney, *A property of the minimum vectors of a regularizing functional defined by means of absolute norm*, IEEE Transaction on Signal Processing, 45 (1997), pp. 913–917.

[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] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*, MPS-SIAM Series on Optimization, 2006.

[4] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, *Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection*, Int. J. Comp. Vis., 67 (2006), pp. 111–136.

[5] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982.

[6] D. Bertsekas, A. Nedić, and A. Ozdaglar *Convex Analysis and Optimization*, Athena Scientific, Belmont, Massachusetts, 2003.

[7] A. Braides, Γ-*Convergence for Beginners*, No.22 in Oxford Lecture Series in Mathematics and Its Applications. Oxford University Press, 2002.

[8] L. Bregman, *A relaxation method of finding a common point of convex sets and its application to the solution of convex programming problems.* J. Wych. Math. and Math.Phys., 7 (1967), pp. 200–217.

[9] J. F. Cai, R. Chan, and M. Nikolova, *Two phase methods for deblurring images corrupted by impulse plus Gaussian noise*, AIMS J. Inverse Prob. Imaging, 2 (2008), pp. 187–204.

[10] C. Carstensen, *Domain decomposition for a non-smooth convex minimization problem and its application to plasticity*, Numerical Linear Algebra with Applications, 4 (1998), pp. 177–190.

[11] A. Chambolle, *An algorithm for total variation minimization and applications.* J. Math. Imaging Vision , 20 (2004), pp. 89–97.

[12] A. Chambolle, J. Darbon, *On total variation minimization and surface evolution using parametric maximum flows*, Int. J. Comput. Vis., 84 (2009), pp. 288–307.

[13] A. Chambolle and P.-L. Lions: *Image recovery via total variation minimization and related problems.*, Numer. Math., 76 (1997), pp. 167–188.

[14] R. H. Chan, C.-W. Ho, and M. Nikolova, *Salt-and-Pepper Noise Removal by Median-Type Noise Detectors and Details-Preserving Regularization*, IEEE Transactions on Image Processing, 14 (2005), pp. 1479–1485.

[15] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

[16] T. F. Chan, G. H. Golub, and P. Mulet,: *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput. 20 (1999), pp. 1964–1977.

[17] T. F. Chan and T. P. Mathew, Domain decomposition algorithms, Acta Numerica, 3 (1994), pp. 61–143.

[18] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[19] J. Darbon and M. Sigelle, *A fast and exact algorithm for total variation minimization*, IbPRIA 2005, 3522 (2005), pp. 351–359.

[20] J. Darbon and M. Sigelle, *Image Restoration with Discrete Constrained Total Variation Part I: Fast and Exact Optimization*, J. Math. Imaging and Vision, 26 (2006), pp. 261–276.

[21] I. Daubechies, G. Teschke, and L. Vese, *Iteratively solving linear inverse problems under general convex constraints*, Inverse Probl. Imaging, 1 (2007), pp. 29–46.

[22] D. Dobson and C. R. Vogel, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal., 34 (1997), pp. 1779–1791.

[23] Y. Dong, M. Hintermüller, and M. Neri, *An Efficient Primal-Dual Method for L^1 TV Image Restoration*, SIAM J. Imaging Sciences, 2 (2009), pp. 1168–1189.
[24] I. Ekeland and R. Temam, *Convex analysis and variational problems. Translated by Minerva Translations, Ltd., London.*, Studies in Mathematics and its Applications. Vol. 1. Amsterdam - Oxford: North-Holland Publishing Company; New York: American Elsevier Publishing Company, Inc., 1976.
[25] M. Fornasier, *Domain decomposition methods for linear inverse problems with sparsity constraints*, Inverse Problems, **23** (2007), pp. 2505–2526.
[26] M. Fornasier, Y. Kim, A. Langer, and C.-B. Schönlieb, *Wavelet Decomposition Method for L_2/TV-Image Deblurring*, SIAM J. Imaging Sciences, 5 (2012), pp. 857–885.
[27] M. Fornasier, A. Langer, and C.-B. Schönlieb, *A convergent overlapping domain decomposition method for total variation minimization*, Num. Math., 116 (2010), pp. 645–685.
[28] M. Fornasier and C.-B. Schönlieb, *Subspace correction methods for total variation and ℓ_1-minimization*, SIAM J. Numer. Anal., 47 (2009), pp. 3397–3428.
[29] T. Goldstein and S. Osher, *The split bregman method for ℓ_1 regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
[30] M. Hintermüller and K. Kunisch, *Total bounded variation regularization as bilaterally constrained optimization problem*, SIAM J. Appl. Math., 64 (2004), pp. 1311-1333.
[31] M. Hintermüller and G. Stadler, *An Infeasible Primal-Dual Algorithm for Total Bounded Variation–Based Inf-Convolution-Type Image Restoration*, SIAM J. Sci. Comput., 28 (2006), pp. 1–23.
[32] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, Vol. 305 of Grundlehren der mathematischen Wissenschaften, Springer-Verlag: Berlin, 1996.
[33] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Series: Advances in Design and Control (No. 15) SIAM, 2008.
[34] A. Langer, *Subspace Correction and Domain Decomposition Methods for Total Variation Minimization*, Ph.D. thesis, Johannes Kepler University Linz, 2011.
[35] A. Langer, S. Osher, and C.-B. Schönlieb, *Bregmanized Domain Decomposition for Image Restoration*, to appear in J. Sci. Comput., DOI 10.1007/s10915-012-9603-x
[36] Y. Meyer, *Oscillating Pattern in Image Processing and Nonlinear Evolution Equations*, AMS University Lecture Series 22, AMS, Providence, RI, 2001.
[37] Y. Nesterov, *Smooth minimization of non-smooth functions.* Mathematical Programming., Ser. A, 103 (2005), pp. 127–152.
[38] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers,* SIAM Journal on Numerical Analysis, 40 (2002), pp. 965–994.
[39] M. Nikolova, *A variational approach to remove outliers and impulse noise,* Journal of Mathematical Imaging and Vision, 20 (2004), pp. 99–120.
[40] 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 (2005), pp. 460–489.
[41] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princton, NJ, 1970.
[42] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms.*, Physica D, 60 (1992), pp. 259–268.
[43] X.-C. Tai and P. Tseng, *Convergence rate analysis of an asynchronous space decomposition method for convex minimization*, Math. Comp., 71 (2001), pp. 1105–1135.
[44] X.-C. Tai and J. Xu, *Global convergence of subspace correction methods for convex optimization problems*, Math. Comp., 71 (2002), pp. 105–124.
[45] P. Tseng, *Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization,* Journal of Optimization Theory and Applications, 109 (2001), pp. 475–494.
[46] P. Tseng and S. Yun, *A Coordinate Gradient Descent Method for Nonsmooth Separable Minimization*, Math. Program., Ser. B 117 (2009), pp. 387–423.
[47] L. Vese, *A study in the BV space of a denoising-deblurring variational problem.*, Appl. Math. Optim., 44 (2001), pp. 131–161.
[48] C. Vonesch and M. Unser, *A fast multilevel algorithm for wavelet-regularized image restoration,* IEEE Transaction on Image Processing, 18 (2009), pp. 509–523.
[49] J. Warga, *Minimizing certain convex functions*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 588–593.
[50] P. Weiss, L. Blanc-Féraud, and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM J. Sci. Comput., 31 (2009), pp. 2047-2080.
[51] J. Xu, *The method of subspace corrections*, J. Comp. Appl. Math., 128 (2001), pp. 335–362.
[52] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing*, SIAM J. Imaging Sci., 1 (2008), pp. 143–168.
[53] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276.
[54] M. Zhu and T. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, UCLA, Center for Applied Math., CAM Reports 08-34, 2008.