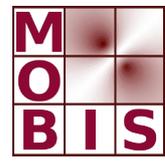




SpezialForschungsBereich F 32



Karl-Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz



Non-overlapping Domain Decomposition Methods for Dual Total Variation Based Image Denoising

M. Hintermüller A. Langer

SFB-Report No. 2013-014

October 2013

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**



Non-overlapping Domain Decomposition Methods for Dual Total Variation Based Image Denoising

Michael Hintermüller · Andreas Langer

Abstract In this paper non-overlapping domain decomposition methods for the pre-dual total variation minimization problem are introduced. Both parallel and sequential approaches are proposed for these methods for which convergence to a minimizer of the original problem is established. The associated subproblems are solved by a semi-smooth Newton method. Several numerical experiments are presented, which show the successful application of the sequential and parallel algorithm for image denoising.

Keywords Domain decomposition · Subspace correction · Total bounded variation · Pre-dual · Convex optimization · Convergence analysis · Image reconstruction

1 Introduction

In the context of image denoising, Rudin, Osher and Fatemi [31] proposed to use the total variation as a regularization technique. In this approach, the recovery of an unknown image is based on solving the constrained minimization problem

$$\begin{aligned} \min \int_{\Omega} |Du| \quad \text{over } u \in BV(\Omega) \\ \text{subject to (s.t.) } \int_{\Omega} u dx = \int_{\Omega} f dx \quad \text{and} \quad \int_{\Omega} (u - f)^2 dx = \sigma^2, \end{aligned} \tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is a simply connected domain with Lipschitz boundary $\partial\Omega$, $f \in L^2(\Omega)$ is the observed image, and σ is the standard deviation of Gaussian white noise with zero mean. Moreover, $BV(\Omega)$ denotes the *space of functions of bounded variation* [1, 15] with $u \in BV(\Omega)$ if and only if $u \in L^1(\Omega)$ and the total variation

$$\int_{\Omega} |Du| = \sup \left\{ \int_{\Omega} u \operatorname{div} v : v \in (C_c^1(\Omega))^2, \|v\|_{\infty} \leq 1 \right\}$$

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”.

M. Hintermüller

Department of Mathematics, Humboldt-University of Berlin, Unter den Linden 6, 10099 Berlin, Germany

Tel.: +49-30-2093-2668

Fax: +49-30-2093-5859

E-mail: hint@math.hu-berlin.de

A. Langer

Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010 Graz, Austria

E-mail: andreas.langer@uni-graz.at

is finite. Here, $C_c^1(\Omega)$ denotes the space of C^1 -functions with compact support in Ω and $\|\varphi\|_\infty = \sup_x \sqrt{\sum_i \varphi_i^2(x)}$. The space $BV(\Omega)$ endowed with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + \int_\Omega |Du|$ is a Banach space [20].

It is well-established that the non-smooth and non-additive total variation preserves edges and discontinuities in images [6,11], and many efficient numerical methods have been proposed for solving the total variation minimization problem (1); see for example [2,5,7–9,12–14,25,28,29,35]. A common strategy to solve (1) is to minimize 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, i.e., one solves

$$\min \frac{1}{2} \int_\Omega |u - f|^2 dx + \alpha \int_\Omega |Du| \quad \text{over } u \in BV(\Omega) \quad (2)$$

with $\alpha > 0$.

Due to the continuous improvement of hardware, the dimensionality of images and measurements in general is increasing, resulting in big data sets that need to be processed. While existing state-of-the-art numerical algorithms for the solution of (2) perform well on small- and medium-scale problems, none of these are capable of addressing extremely large problems in realistic CPU-time. In such a situation subspace correction, domain decomposition and block-coordinate descent methods are fundamental as they allow for splitting the computational workload and solving a sequence of smaller problems, possibly distributed on multiple cores. We emphasize, however, that well-known approaches as those in [4,10,32,33] are not directly applicable to the problem in (2) due to the non-smoothness and non-additivity of the total variation regularization.

Recently in [16–19,22,23,27] domain decomposition strategies were introduced for minimizing objectives including the total variation. In this context, the major difficulty lies in the correct treatment of the interfaces of the domain patches, i.e. the preservation of crossing discontinuities and the correct matching where the solution is continuous. While in these papers an implementation guaranteeing convergence and monotonic decay of the objective energy is provided, convergence to the global minimizer of the underlying problem cannot be ensured, in general. For one-dimensional problems, in [18] a proof is presented which establishes convergence of an overlapping domain decomposition algorithm to the global solution. However, it is not clear yet how to extend this proof to any finite dimensional space without introducing additional (possibly restrictive) assumptions.

In [16], for a wavelet decomposition method an additional condition is invoked which allows to establish global optimality of a limit point obtained by the 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, with the aforementioned condition one can only check a posteriori whether the algorithm found the global minimizer or whether it failed to do so.

Additionally to the convergence and monotonic decay of the energy the authors derived in [22,23] an estimate of the distance of the limit point obtained from domain decomposition methods to the true global minimizer. With the help of this estimate, it is demonstrated by numerical experiments that the sequence generated by the domain decomposition algorithm indeed approaches the global minimizer of the objective functional. However, a proof of convergence of such a domain decomposition method to the minimizer of the global problem seems elusive. Here the main difficulty lies in the presence of the non-additive and non-smooth total variation term.

In order to tackle the difficulties due to the minimization of a non-smooth and non-additive objective over a non-reflexive Banach space in (2), we consider the pre-dual problem of (2); see [21,24] for the derivation of the latter. In fact, the pre-dual reads:

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div} \mathbf{p} + f\|_{L^2(\Omega)}^2 \quad \text{over } \mathbf{p} \in H_0(\operatorname{div}, \Omega) \\ \text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}(x) \leq \alpha \mathbf{1} \text{ for almost every } x \in \Omega, \end{aligned} \quad (3)$$

where $H_0(\operatorname{div}, \Omega) := \{\mathbf{v} \in \mathbb{L}^2(\Omega) : \operatorname{div} \mathbf{v} \in L^2(\Omega), \mathbf{v} \cdot \mathbf{n} = 0 \text{ on } \partial\Omega\}$ and \mathbf{n} is the outward unit normal on $\partial\Omega$.

The smooth objective of L^2 -type and the pointwise constraints turn out to be more amenable to domain decomposition methods than the structure of (2). In fact, in the present paper we introduce non-overlapping domain decomposition methods for the problem in (3) and show their convergence to a global minimizer in a discrete setting. In [34] a convergence theory for coordinate gradient descent methods was established for a class of minimization problems. In that paper, convergence to a global minimizer could be proven only under the assumption of strict convexity of the objective. This latter assumption, however, does not hold for the problem in (3). Moreover, the convergence analysis presented in [34] is not directly adaptable to our problem. Hence for our proposed domain decomposition methods a different convergence analysis is needed.

In our numerical experiments we show that our domain decomposition approach for the dual problem (3) converges faster than the domain decomposition approach in [22] for the primal problem (2). Hence the proposed splitting method is not only superior from a theoretical point of view, since we can show convergence to the global solution, but also practically. These reasons make the domain decomposition algorithm for the (pre-)dual problem applicable to 3D imaging problems, where a large number of data have to be processed.

The rest of the paper is organized as follows: In Section 2 we present the main notations used throughout the rest of the paper. In Section 3 we introduce the sequential and parallel non-overlapping domain decomposition method and show its convergence to a minimizer of the global problem. Details on the numerical implementation of the solvers for the proposed domain decomposition methods are described in Section 4. Finally in Section 5 we show sequential and parallel numerical experiments for image denoising. We compare the computational performance of the proposed domain decomposition algorithm of the dual problem with the domain decomposition algorithm for the problem in (2) introduced in [22]. Moreover, we show the successful application of the proposed algorithm for 3D medical data.

2 Notations

In the rest of the paper we work in a finite dimensional setting by considering a finite regular (pixel) mesh as a discretization of a rectangular domain Ω . For simplicity, we take $\Omega := [0, 1]^2$, but an extension to rectangular domains is straightforward. Further we approximate functions \mathbf{p} by discrete functions, again denoted by \mathbf{p} . Since we are mainly interested in imaging problems, it is sufficient to introduce our main notations for a two-dimensional space only.

In our discrete setting, Ω is a mesh in \mathbb{R}^2 of size $N \times N$, where $N \in \mathbb{N}$ with mesh size $h = x_{i,j} - x_{i+1,j} = x_{i,j} - x_{i,j+1}$ for $x_{i,j} \in \Omega$. The considered “function spaces” are $\mathcal{H} = \mathbb{R}^{N \times N}$ and $V = \mathcal{H}^2$, with corresponding norm

$$\|u\|_{\mathcal{H}} = \|u\|_{\ell^2(\Omega)} = \left(h^2 \sum_{x \in \Omega} |u(x)|^2 \right)^{1/2}, \quad \text{for all } u \in \mathcal{H},$$

and

$$\|\mathbf{p}\|_V = \left(h^2 \sum_{x \in \Omega} |p^1(x)|^2 + |p^2(x)|^2 \right)^{1/2}, \quad \text{for all } \mathbf{p} = (p^1, p^2) \in V,$$

respectively. We define the scalar product of $u, v \in \mathcal{H}$ and of $\mathbf{p}, \mathbf{q} \in V$, respectively, by

$$\langle u, v \rangle_{\mathcal{H}} = h^2 \sum_{x \in \Omega} u(x)v(x), \quad \text{and} \quad \langle \mathbf{p}, \mathbf{q} \rangle_V = h^2 \sum_{x \in \Omega} \langle \mathbf{p}(x), \mathbf{q}(x) \rangle_{\mathbb{R}^2}$$

with $\langle y, z \rangle_{\mathbb{R}^2} = \sum_{j=1}^2 y_j z_j$ for every $y = (y_1, y_2) \in \mathbb{R}^2$ and $z = (z_1, z_2) \in \mathbb{R}^2$.

The discrete gradient $\nabla_{\Omega} u$ is denoted by $(\nabla_{\Omega} u)(x) = ((\nabla_{\Omega} u)^1(x), (\nabla_{\Omega} u)^2(x))$ with

$$(\nabla_{\Omega} u)^1(x_{i,j}) = \frac{1}{h} \cdot \begin{cases} u(x_{i+1,j}) - u(x_{i,j}) & \text{if } i < N, \\ 0 & \text{if } i = N, \end{cases}$$

and

$$(\nabla_{\Omega} u)^2(x_{i,j}) = \frac{1}{h} \cdot \begin{cases} u(x_{i,j+1}) - u(x_{i,j}) & \text{if } j < N, \\ 0 & \text{if } j = N, \end{cases}$$

for all $i = 1, \dots, N$ and $j = 1, \dots, N$. For an operator T we denote by T^* its adjoint (or transpose). Further we introduce the *discrete divergence* $\operatorname{div}_{\Omega} : V \rightarrow \mathcal{H}$ as $\operatorname{div}_{\Omega} := -\nabla_{\Omega}^*$ (∇_{Ω}^* is the adjoint of the gradient ∇_{Ω}), in analogy to the continuous setting. In our case, the discrete divergence operator is explicitly given by

$$\begin{aligned} (\operatorname{div}_{\Omega} \mathbf{p})(x_{i,j}) &= \frac{1}{h} \begin{cases} p^1(x_{i,j}) - p^1(x_{i-1,j}) & \text{if } 1 < i < N, \\ p^1(x_{i,j}) & \text{if } i = 1, \\ -p^1(x_{i-1,j}) & \text{if } i = N, \end{cases} \\ &+ \frac{1}{h} \begin{cases} p^2(x_{i,j}) - p^2(x_{i,j-1}) & \text{if } 1 < j < N, \\ p^2(x_{i,j}) & \text{if } j = 1, \\ -p^2(x_{i,j-1}) & \text{if } j = N, \end{cases} \end{aligned}$$

for every $\mathbf{p} = (p^1, p^2) \in V$. Note, that with these definitions of the gradient and divergence we have that $\langle \mathbf{p}, \nabla_{\Omega} u \rangle_V = \langle -\operatorname{div}_{\Omega} \mathbf{p}, u \rangle_{\mathcal{H}}$. In the sequel we also use the notations $\nabla := \nabla_{\Omega}$ and $\operatorname{div} := \operatorname{div}_{\Omega}$ when it is clear from the context that all of Ω is considered.

We will often use the symbols $\mathbf{1}$ and $\mathbf{0}$ to indicate the constant vectors with entry values 1 and 0, respectively.

For a convex functional $\mathcal{J} : \mathcal{H} \rightarrow \bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$, we define the *subdifferential* of \mathcal{J} at $v \in \mathcal{H}$ as the set valued mapping $\partial \mathcal{J} : \mathcal{H} \rightarrow 2^{\mathcal{H}}$

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

Whenever $w \in \mathcal{H}$ is fixed, $\partial \mathcal{J}(\cdot + w)(v)$, for $v \in \mathcal{H}$, denotes the evaluation of $\partial \mathcal{J}$ at $v + w$. 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 have to handle different spaces, namely \mathcal{H} , V , it is sometimes useful to distinguish the spaces in the definition of the subdifferential by imposing a subscript $\partial_{\mathcal{H}} \mathcal{J}$ for the subdifferential considered on the space \mathcal{H} , for instance.

3 Non-overlapping Domain Decomposition of the (Pre-)Dual Problem

We consider the splitting of Ω into two disjoint subdomains Ω_1 and Ω_2 such that $\Omega_2 = \Omega \setminus \Omega_1$ and $\Omega = \Omega_1 \cup \Omega_2$. For consistency with the definitions of the gradient and divergence operator, we assume that the subdomains Ω_i , $i = 1, 2$, as well as Ω are discrete rectangles. We stress that the shape of the domains as well as the decomposition into two subdomains represents by no means a restriction, but rather simplifies the presentation. Associated with this decomposition we define $V_i := \{\mathbf{v} \in V : \mathbf{v}(x) = \mathbf{0} \text{ for } x \in \Omega \setminus \Omega_i\}$ such that $V = V_1 \oplus V_2$ and by π_{V_i} we denote the corresponding orthogonal projection onto V_i . With this splitting we aim to solve the discrete problem

$$\begin{aligned} \min & \frac{1}{2} \|\operatorname{div} \mathbf{p} + f\|_{\mathcal{H}}^2 \quad \text{over } \mathbf{p} \in V \\ \text{s.t.} & -\alpha \mathbf{1} \leq \mathbf{p}(x) \leq \alpha \mathbf{1} \text{ for all } x \in \Omega, \end{aligned} \tag{4}$$

by the following iterative scheme:

Algorithm 1

- Initialize: $\mathbf{p}^{(0)} := \mathbf{p}_1^{(0)} + \mathbf{p}_2^{(0)} \in V_1 \oplus V_2$ with $-\alpha\mathbf{1} \leq \mathbf{p}^{(0)}(x) \leq \alpha\mathbf{1}$ for all $x \in \Omega$.
- Iterate: For $k = 0, 1, \dots$ do

$$\begin{cases} \mathbf{p}_1^{(k+1)} \in \arg \min_{\mathbf{p}_1 \in V_1} \left\{ \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \mathbf{p}_2^{(k)}) + f\|_{\mathcal{H}}^2 : -\alpha\mathbf{1} \leq \mathbf{p}_1(x) + \mathbf{p}_2^{(k)}(x) \leq \alpha\mathbf{1} \right\}, \\ \mathbf{p}_2^{(k+1)} \in \arg \min_{\mathbf{p}_2 \in V_2} \left\{ \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1^{(k+1)} + \mathbf{p}_2) + f\|_{\mathcal{H}}^2 : -\alpha\mathbf{1} \leq \mathbf{p}_1^{(k+1)}(x) + \mathbf{p}_2(x) \leq \alpha\mathbf{1} \right\}, \\ \mathbf{p}^{(k+1)} := \mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k+1)}. \end{cases}$$

For the ease of notation, from now on we use $-\alpha\mathbf{1} \leq \mathbf{p}(x) \leq \alpha\mathbf{1}$ instead of $-\alpha\mathbf{1} \leq \mathbf{p}(x) \leq \alpha\mathbf{1}$ for all $x \in \Omega$.

Several basic properties of Algorithm 1 are summarized next.

Proposition 1 Let $J(\mathbf{p}) := \frac{1}{2} \|\operatorname{div} \mathbf{p} + f\|_{\mathcal{H}}^2$ and $\mathbf{p}^{(k+\frac{1}{2})} := \mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k)}$. Then Algorithm 1 generates sequences $(\mathbf{p}^{(k)})_k \subset V$ and $(\mathbf{p}_i^{(k)})_k \subset V_i$ for $i = 1, 2$ with the following properties:

- (i) $J(\mathbf{p}^{(k)}) \geq J(\mathbf{p}^{(k+\frac{1}{2})}) \geq J(\mathbf{p}^{(k+1)})$ for all $k \in \mathbb{N}$;
- (ii) The sequences $(J(\mathbf{p}^{(k)}))_k$ and $(J(\mathbf{p}^{(k+\frac{1}{2})}))_k$ converge;
- (iii) $\lim_{k \rightarrow \infty} (J(\mathbf{p}^{(k+\frac{1}{2})}) - J(\mathbf{p}^{(k)})) = 0$;
- (iv) The sequences $(\mathbf{p}^{(k)})_k$ and $(\tilde{\mathbf{p}}^{(k)})_k$ are bounded in V and hence have subsequences that converge in V ;
- (v) The sequences $(\mathbf{p}_i^{(k)})_k$ for $i = 1, 2$ are bounded in V .

Proof (i) The assertion follows from the optimality properties of $\mathbf{p}_1^{(k)}$ and $\mathbf{p}_2^{(k)}$.

(ii)+(iii) The statements follow directly from (i) and the fact that $J(\cdot)$ is bounded from below.

(iv) Since $-\alpha\mathbf{1} \leq \mathbf{p}_1^{(k)}(x) + \mathbf{p}_2^{(k)}(x) \leq \alpha\mathbf{1}$ and $-\alpha\mathbf{1} \leq \mathbf{p}_1^{(k+1)}(x) + \mathbf{p}_2^{(k)}(x) \leq \alpha\mathbf{1}$ for all $x \in \Omega$ it follows that

$$\|\mathbf{p}^{(k)}\|_V \leq \alpha\sqrt{2} \text{ and } \|\mathbf{p}^{(k+\frac{1}{2})}\|_V \leq \alpha\sqrt{2} \text{ for all } k \in \mathbb{N}.$$

(v) Since $\mathbf{p}_i^{(k)} \in V_i$ it follows from the boundedness of $(\mathbf{p}^{(k)})_k$, see (iv), that

$$2\alpha^2 \geq \|\mathbf{p}^{(k)}\|_V^2 = \|\mathbf{p}_1^{(k)} + \mathbf{p}_2^{(k)}\|_V^2 = \|\mathbf{p}_1^{(k)}\|_V^2 + \|\mathbf{p}_2^{(k)}\|_V^2$$

and hence $(\mathbf{p}_i^{(k)})_k$ is bounded in V for $i = 1, 2$.

Proposition 2 For $i = 1, 2$ let $(\mathbf{p}_i^{(k)})_k$ be generated by Algorithm 1 and $(\mathbf{p}_i^{(k_l)})_l$ be a convergent subsequence with limit point $\mathbf{p}_i^{(\infty,0)}$. Then there exists an accumulation point $\mathbf{p}_i^{(\infty,1)}$ of the sequence $(\mathbf{p}_i^{(k_l+1)})_l$ such that

(i) $\mathbf{p}_1^{(\infty,0)}$ and $\mathbf{p}_1^{(\infty,1)}$ are minimizers of

$$\begin{aligned} & \min \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \mathbf{p}_2^{(\infty,0)}) + f\|_{\mathcal{H}}^2 \text{ over } \mathbf{p}_1 \in V_1 \\ & \text{s.t. } -\alpha\mathbf{1} \leq \mathbf{p}_1(x) + \mathbf{p}_2^{(\infty,0)}(x) \leq \alpha\mathbf{1}, \end{aligned}$$

and
(ii) $\mathbf{p}_2^{(\infty,0)}$ and $\mathbf{p}_2^{(\infty,1)}$ are minimizers of

$$\begin{aligned} & \min \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1^{(\infty,1)} + \mathbf{p}_2) + f\|_{\mathcal{H}}^2 \text{ over } \mathbf{p}_2 \in V_2 \\ & \text{s.t. } -\alpha\mathbf{1} \leq \mathbf{p}_1(x)^{(\infty,1)} + \mathbf{p}_2(x) \leq \alpha\mathbf{1}, \end{aligned}$$

Proof Set $F(\mathbf{p}) := \frac{1}{2} \|\operatorname{div}(\mathbf{p}) + f\|_{\mathcal{H}}^2 + I_{[-\alpha, \alpha]}(\mathbf{p})$, where

$$I_{[-\alpha, \alpha]}(\mathbf{p}) = \begin{cases} 0 & \text{if } -\alpha \mathbf{1} \leq \mathbf{p}(x) \leq \alpha \mathbf{1} \text{ for all } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases} \quad (5)$$

Then the minimization problem in V_1 is equivalent to

$$\min F(\mathbf{p}_1 + \mathbf{p}_2^{(k)}) \quad \text{over } \mathbf{p}_1 \in V_1.$$

Note that F is convex, lower semicontinuous, bounded from below, and monotonically decreasing with respect to the iterations, i.e., $F(\mathbf{p}^{(k)}) \geq F(\mathbf{p}^{(k+1)})$ for all $k \in \mathbb{N}$.

(i) Since $(\mathbf{p}_2^{(k)})_k$ is bounded, there exists a convergent subsequence $(\mathbf{p}_2^{(k_l)})_l$ with limit $\mathbf{p}_2^{(\infty, 0)}$. Further $(\mathbf{p}_1^{(k_l+1)})_l$ is bounded and hence there exists a convergent subsequence $(\mathbf{p}_1^{(k_{l_j}+1)})_j$ with limit $\mathbf{p}_1^{(\infty, 1)}$.

By the minimality property of $\mathbf{p}_1^{(k_{l_j}+1)}$ we have that $0 \in \partial_{V_1} F(\cdot + \mathbf{p}_2^{(k_{l_j})})(\mathbf{p}_1^{(k_{l_j}+1)})$. Then by [30, Theorem 24.4, p. 233] we obtain that $0 \in \partial_{V_1} F(\cdot + \mathbf{p}_2^{(\infty, 0)})(\mathbf{p}_1^{(\infty, 1)})$ and hence by convexity $\mathbf{p}_1^{(\infty, 1)} \in \arg \min_{\mathbf{p}_1 \in V_1} F(\mathbf{p}_1 + \mathbf{p}_2^{(\infty, 0)})$.

Analogously we get that $0 \in \partial_{V_2} F(\mathbf{p}_1^{(\infty, 1)} + \cdot)(\mathbf{p}_2^{(\infty, 1)})$, i.e., $\mathbf{p}_2^{(\infty, 1)} = \arg \min_{\mathbf{p}_2 \in V_2} F(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2)$. Moreover, by the monotone decrease of the energy we have that

$$F(\mathbf{p}^{(k)}) \geq F(\mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k)}) \geq F(\mathbf{p}^{(k+1)}) \quad \text{for all } k \in \mathbb{N}$$

and hence

$$F(\mathbf{p}^{(k)}) - F(\mathbf{p}^{(k+1)}) \geq F(\mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k)}) - F(\mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k+1)}) \geq 0.$$

Replacing (k) and $(k+1)$ by (k_{l_j}) and $(k_{l_j}+1)$, respectively, we get

$$0 \leq F(\mathbf{p}_1^{(k_{l_j}+1)} + \mathbf{p}_2^{(k_{l_j})}) - F(\mathbf{p}_1^{(k_{l_j}+1)} + \mathbf{p}_2^{(k_{l_j}+1)}) \leq F(\mathbf{p}_1^{(k_{l_j}+1)} + \mathbf{p}_2^{(k_{l_j})}) - F(\mathbf{p}_1^{(k_{l_j}+1)} + \mathbf{p}_2^{(k_{l_j+1})}).$$

Since $(\mathbf{p}_1^{(k_{l_j}+1)})_j$ is bounded, there exists a subsequence $(\mathbf{p}_1^{(k_{l_{j_m}+1})})_m$ with limit $\mathbf{p}_1^{(\infty, 0)}$ and we have

$$F(\mathbf{p}_1^{(k_{l_{j_m}+1})} + \mathbf{p}_2^{(k_{l_{j_m}})}) - F(\mathbf{p}_1^{(k_{l_{j_m}+1})} + \mathbf{p}_2^{(k_{l_{j_m}+1})}) \geq 0.$$

Proposition 1(ii) implies $\lim_{k \rightarrow \infty} [F(\mathbf{p}^{(k)}) - F(\mathbf{p}^{(k+r)})] = 0$ for all $r \in \mathbb{N}$. Consequently

$$\begin{aligned} 0 &\leq \lim_{m \rightarrow \infty} \left[F(\mathbf{p}_1^{(k_{l_{j_m}+1})} + \mathbf{p}_2^{(k_{l_{j_m}})}) - F(\mathbf{p}_1^{(k_{l_{j_m}+1})} + \mathbf{p}_2^{(k_{l_{j_m}+1})}) \right] \\ &\leq \lim_{m \rightarrow \infty} \left[F(\mathbf{p}_1^{(k_{l_{j_m}})} + \mathbf{p}_2^{(k_{l_{j_m}})}) - F(\mathbf{p}_1^{(k_{l_{j_m}+1})} + \mathbf{p}_2^{(k_{l_{j_m}+1})}) \right] = 0. \end{aligned}$$

Further

$$\begin{aligned} \lim_{m \rightarrow \infty} \left[F(\mathbf{p}_1^{(k_{l_{j_m}+1})} + \mathbf{p}_2^{(k_{l_{j_m}})}) - F(\mathbf{p}_1^{(k_{l_{j_m}+1})} + \mathbf{p}_2^{(k_{l_{j_m}+1})}) \right] \\ = F(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) - F(\mathbf{p}_1^{(\infty, 0)} + \mathbf{p}_2^{(\infty, 0)}). \end{aligned}$$

This implies

$$F(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) - F(\mathbf{p}_1^{(\infty, 0)} + \mathbf{p}_2^{(\infty, 0)}) = 0.$$

Since $\mathbf{p}_1^{(\infty, 1)}$ is a solution of

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \mathbf{p}_2^{(\infty, 0)}) + f\|_{\mathcal{H}}^2 \quad \text{over } \mathbf{p}_1 \in V_1 \\ \text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}_1(x) + \mathbf{p}_2^{(\infty, 0)}(x) \leq \alpha \mathbf{1}, \end{aligned}$$

so is $\mathbf{p}_1^{(\infty, 0)}$.

(ii) By similar arguments one proves the second assertion.

Next we show the convergence of Algorithm 1 to a minimizer of the (pre-)dual problem (4).

Theorem 1 *The sequence $(\mathbf{p}^{(k)})_k$ generated by Algorithm 1 has an accumulation point $\mathbf{p}^{(\infty)} \in V$ which is a solution of the problem in (4).*

Proof From the optimality of $\mathbf{p}_1^{(k+1)}$ and $\mathbf{p}_2^{(k+1)}$ we get

$$\begin{aligned} 0 &\in -\pi_{V_1} \nabla(\operatorname{div}(\mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k)}) + f) + \partial_{V_1} I_{[-\alpha, \alpha]}(\cdot + \mathbf{p}_2^{(k)})(\mathbf{p}_1^{(k+1)}) \\ 0 &\in -\pi_{V_2} \nabla(\operatorname{div}(\mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k+1)}) + f) + \partial_{V_2} I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(k+1)} + \cdot)(\mathbf{p}_2^{(k+1)}), \end{aligned} \quad (6)$$

where $I_{[-\alpha, \alpha]}(\cdot)$ is defined as in (5). By similar arguments as in Proposition 2 we extract suitable subsequences and due to their boundedness (Proposition 1), the hierarchy of sequences $(\mathbf{p}_2^{(k_l)})_l$, $(\mathbf{p}_1^{(k_{l_j}+1)})_j$, $(\mathbf{p}_2^{(k_{l_{j_m}}+1)})_m$ converges to limit points $\mathbf{p}_2^{(\infty, 0)}$, $\mathbf{p}_1^{(\infty, 1)}$, $\mathbf{p}_2^{(\infty, 1)}$, respectively. Note that the divergence of a bounded sequence is again bounded, i.e., the boundedness of $(\mathbf{p}^{(k)})_k \subset V$ implies the boundedness of $(\operatorname{div} \mathbf{p}^{(k)})_k$. Hence, we may replace k by $k_{l_{j_m}}$ in (6). By taking the limit $m \rightarrow \infty$ and using [30, Thm. 24.4, p. 233] we obtain

$$\begin{aligned} 0 &\in -\pi_{V_1} \nabla(\operatorname{div}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) + f) + \partial_{V_1} I_{[-\alpha, \alpha]}(\cdot + \mathbf{p}_2^{(\infty, 0)})(\mathbf{p}_1^{(\infty, 1)}) \\ 0 &\in -\pi_{V_2} \nabla(\operatorname{div}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 1)}) + f) + \partial_{V_2} I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \cdot)(\mathbf{p}_2^{(\infty, 1)}). \end{aligned}$$

By Proposition 2 we can replace $\mathbf{p}_2^{(\infty, 1)}$ by $\mathbf{p}_2^{(\infty, 0)}$ in the last inclusion. Let us define

$$\begin{aligned} \xi_1 &:= \pi_{V_1} \nabla(\operatorname{div}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) + f) \in \partial_{V_1} I_{[-\alpha, \alpha]}(\cdot + \mathbf{p}_2^{(\infty, 0)})(\mathbf{p}_1^{(\infty, 1)}) \\ \xi_2 &:= \pi_{V_2} \nabla(\operatorname{div}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) + f) \in \partial_{V_2} I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \cdot)(\mathbf{p}_2^{(\infty, 0)}). \end{aligned}$$

Note that $\xi := \xi_1 + \xi_2 = \nabla(\operatorname{div}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) + f)$. By the definition of the subdifferential we get

$$\begin{aligned} \langle \xi_1, \mathbf{y}_1 - \mathbf{p}_1^{(\infty, 1)} \rangle_V + I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) &\leq I_{[-\alpha, \alpha]}(\mathbf{y}_1 + \mathbf{p}_2^{(\infty, 0)}) \quad \text{for all } \mathbf{y}_1 \in V_1, \\ \langle \xi_2, \mathbf{y}_2 - \mathbf{p}_2^{(\infty, 0)} \rangle_V + I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) &\leq I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{y}_2) \quad \text{for all } \mathbf{y}_2 \in V_2. \end{aligned}$$

Summing up the last two inequalities yields

$$\begin{aligned} \langle \xi, \mathbf{y}_1 + \mathbf{y}_2 - (\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) \rangle_V + I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) \\ \leq I_{[-\alpha, \alpha]}(\mathbf{y}_1 + \mathbf{p}_2^{(\infty, 0)}) + I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{y}_2) - I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) \quad \text{for all } \mathbf{y}_i \in V_i. \end{aligned}$$

Since Ω_1 and Ω_2 are disjoint we have that

$$I_{[-\alpha, \alpha]}(\mathbf{p}_1 + \mathbf{p}_2) = I_{[-\alpha, \alpha]}(\mathbf{p}_1) + I_{[-\alpha, \alpha]}(\mathbf{p}_2) \quad (7)$$

for all $\mathbf{p}_i \in V_i$, $i = 1, 2$. By using this property we deduce

$$I_{[-\alpha, \alpha]}(\mathbf{y}_1 + \mathbf{p}_2^{(\infty, 0)}) + I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{y}_2) - I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) = I_{[-\alpha, \alpha]}(\mathbf{y}_1 + \mathbf{y}_2)$$

for all $\mathbf{y}_i \in V_i$, $i = 1, 2$. Hence, for $\mathbf{y} := \mathbf{y}_1 + \mathbf{y}_2$, we have

$$\langle \xi, \mathbf{y} - (\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) \rangle_V + I_{[-\alpha, \alpha]}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) \leq I_{[-\alpha, \alpha]}(\mathbf{y}) \quad \text{for all } \mathbf{y} \in V.$$

and consequently

$$0 \in -\nabla(\operatorname{div}(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}) + f) + \partial_V I_{[-\alpha, \alpha]}(\cdot)(\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}).$$

On the other hand, $\mathbf{p}^* \in V$ is a solution of the global problem (4) if and only if

$$0 \in -\nabla(\operatorname{div} \mathbf{p}^* + f) + \partial_V I_{[-\alpha, \alpha]}(\cdot)(\mathbf{p}^*).$$

Hence $\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 0)}$ is a minimizer of the global problem. By Proposition 1(iii) and since $0 \leq J(\mathbf{p}_1^{(k_{l_{j_m}}+1)} + \mathbf{p}_2^{(k_{l_{j_m}})}) - J(\mathbf{p}_1^{(k_{l_{j_m}}+1)} + \mathbf{p}_2^{(k_{l_{j_m}}+1)})$ it follows that $\mathbf{p}_1^{(\infty, 1)} + \mathbf{p}_2^{(\infty, 1)}$ is also a minimizer of (4).

Remark 1 (a) In the above proof we rely on the additivity of the function $I_{[-\alpha, \alpha]}(\cdot)$ with respect to a disjoint decomposition. Actually, Theorem 1 can be extended to more general problems of the type

$$\frac{1}{2} \|T(\mathbf{p}) + g\|_{\mathcal{H}}^2 + P(\mathbf{p}), \quad (8)$$

where $g \in \mathcal{H}$ is a given datum, $P : V \rightarrow \mathbb{R}$ a convex, lower semicontinuous function with the splitting property (7) and $T : V \rightarrow \mathcal{H}$ is a bounded linear operator.

Due to this property, our method can also handle deconvolution-type problems, i.e., (2) is replaced by

$$\min \frac{1}{2} \int_{\Omega} |Ku - f|^2 dx + \alpha \int_{\Omega} |Du| + \frac{\kappa}{2} \int_{\Omega} |u|^2 dx \quad \text{over } u \in BV(\Omega)$$

where K is a continuous linear operator from $L^2(\Omega)$ to $L^2(\Omega)$ modelling, e.g., convolution and $\kappa \geq 0$. We note that $\kappa > 0$ is useful, whenever K^*K is not invertible. The (pre-)dual in the discrete setting becomes

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div} \mathbf{p} + K^*f\|_M^2 \quad \text{over } \mathbf{p} \in V \\ \text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}(x) \leq \alpha \mathbf{1}, \end{aligned}$$

where $\|w\|_M^2 = w^T M w$ for $M = (\kappa I + K^*K)^{-1}$. Here, in a slight misuse of notation, we denote the discrete convolution operator also by K . Considering $T := M^{\frac{1}{2}} \operatorname{div}$ and $g := M^{\frac{1}{2}} K^*f$ in (8), the deconvolution problem fits into our framework.

- (b) As the considered problem (4) is convex but not strictly convex, our convergence analysis differs from the one in [34], where the strict convexity assumption is essential.

3.1 A Parallel Version and its Convergence

The parallel version of Algorithm 1 reads as follows:

Algorithm 2

- Initialize: $\mathbf{p}^{(0)} := \tilde{\mathbf{p}}_1^{(0)} + \tilde{\mathbf{p}}_2^{(0)} \in V_1 \oplus V_2$ with $-\alpha \mathbf{1} \leq \mathbf{p}^{(0)}(x) \leq \alpha \mathbf{1}$ for all $x \in \Omega$;
- Iterate: For $k = 0, 1, \dots$ do

$$\left\{ \begin{array}{l} \mathbf{p}_1^{(k+1)} \in \arg \min_{\mathbf{p}_1 \in V_1} \left\{ \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(k)}) + f\|_{\mathcal{H}}^2 : -\alpha \mathbf{1} \leq \mathbf{p}_1(x) + \tilde{\mathbf{p}}_2^{(k)}(x) \leq \alpha \mathbf{1} \right\}, \\ \mathbf{p}_2^{(k+1)} \in \arg \min_{\mathbf{p}_2 \in V_2} \left\{ \frac{1}{2} \|\operatorname{div}(\tilde{\mathbf{p}}_1^{(k)} + \mathbf{p}_2) + f\|_{\mathcal{H}}^2 : -\alpha \mathbf{1} \leq \mathbf{p}_2(x) + \tilde{\mathbf{p}}_1^{(k)}(x) \leq \alpha \mathbf{1} \right\}, \\ \mathbf{p}^{(k+1)} = \frac{\mathbf{p}_1^{(k+1)} + \mathbf{p}_2^{(k+1)} + \mathbf{p}^{(k)}}{2}, \\ \tilde{\mathbf{p}}_1^{(k+1)} := \pi_{V_1} \mathbf{p}^{(k+1)}, \\ \tilde{\mathbf{p}}_2^{(k+1)} := \pi_{V_2} \mathbf{p}^{(k+1)}. \end{array} \right.$$

We state similar convergence results as for the sequential algorithm and provide their proofs for the sake of completeness.

Proposition 3 Let $J(\mathbf{p}) := \frac{1}{2} \|\operatorname{div} \mathbf{p} + f\|_{\mathcal{H}}^2$. Then Algorithm 2 generates sequences $(\mathbf{p}^{(k)})_k \subset V$, $(\mathbf{p}_i^{(k)})_k \subset V_i$ and $(\tilde{\mathbf{p}}_i^{(k)})_k \subset V_i$ with the following properties:

- (i) $J(\mathbf{p}^{(k)}) \geq J(\mathbf{p}^{(k+1)})$ for all $k \in \mathbb{N}$;
- (ii) The sequence $(J(\mathbf{p}^{(k)}))_k$ converges;
- (iii) The sequence $(\mathbf{p}^{(k)})_k$ is bounded in V and hence has a subsequence that converges in V ;
- (iv) The sequences $(\mathbf{p}_i^{(k)})_k$ and $(\tilde{\mathbf{p}}_i^{(k)})_k$ for $i = 1, 2$ are bounded in V .

Proof (i) The assertion follows from the optimality properties of $\mathbf{p}_1^{(k)}$ and $\mathbf{p}_2^{(k)}$.

- (ii) The statement follows directly from (i) and the fact that $(J(\mathbf{p}^{(k)}))_k$ is bounded from below.
 (iii) Since $-\alpha \mathbf{1} \leq \mathbf{p}_1^{(k+1)}(x) + \tilde{\mathbf{p}}_2^{(k)}(x) \leq \alpha \mathbf{1}$ and $-\alpha \mathbf{1} \leq \tilde{\mathbf{p}}_1^{(k)}(x) + \mathbf{p}_2^{(k+1)}(x) \leq \alpha \mathbf{1}$ we get that

$$-\alpha \mathbf{1} \leq \frac{\mathbf{p}_1^{(k+1)}(x) + \mathbf{p}_2^{(k+1)}(x) + \mathbf{p}^{(k)}(x)}{2} \leq \alpha \mathbf{1}$$

for all $x \in \Omega$. From the definition of $\mathbf{p}^{(k+1)}$, see Algorithm 2, we deduce $-\alpha \mathbf{1} \leq \mathbf{p}^{(k+1)}(x) \leq \alpha \mathbf{1}$ for all $x \in \Omega$ and hence $\|\mathbf{p}^{(k)}\|_V \leq \alpha\sqrt{2}$.

- (iv) Since $\tilde{\mathbf{p}}_i^{(k)}, \mathbf{p}_i^{(k)} \in V_i$ it follows from the boundedness of $(\mathbf{p}^{(k)})_k$ and the inequality $(a+b)^{\frac{1}{2}} \geq \frac{1}{\sqrt{2}}(a^{\frac{1}{2}} + b^{\frac{1}{2}})$ for $a, b \geq 0$, that

$$\alpha\sqrt{2} \geq \|\mathbf{p}^{(k)}\|_V = \|\tilde{\mathbf{p}}_1^{(k)} + \tilde{\mathbf{p}}_2^{(k)}\|_V \geq \frac{1}{\sqrt{2}}(\|\tilde{\mathbf{p}}_1^{(k)}\|_V + \|\tilde{\mathbf{p}}_2^{(k)}\|_V)$$

and

$$\begin{aligned} 2\alpha\sqrt{2} &\geq 2\|\mathbf{p}^{(k)}\|_V = \|\mathbf{p}_1^{(k)} + \mathbf{p}_2^{(k)} + \mathbf{p}^{(k-1)}\|_V \geq \|\mathbf{p}_1^{(k)} + \mathbf{p}_2^{(k)}\|_V - \|\mathbf{p}^{(k-1)}\|_V \\ &\geq \frac{1}{\sqrt{2}}(\|\mathbf{p}_1^{(k)}\|_V + \|\mathbf{p}_2^{(k)}\|_V) - \|\mathbf{p}^{(k-1)}\|_V \end{aligned}$$

and hence $(\tilde{\mathbf{p}}_i^{(k)})_k$ and $(\mathbf{p}_i^{(k)})_k$ are bounded in V for $i = 1, 2$.

Proposition 4 For $i = 1, 2$ let $(\mathbf{p}_i^{(k)})_k$ be generated by Algorithm 2 and $(\mathbf{p}_i^{(k_l)})_l$ be a convergent subsequence with limit point $\mathbf{p}_i^{(\infty, 0)}$. Then there exists an accumulation point $\mathbf{p}_i^{(\infty, 1)}$ of the sequence $(\mathbf{p}_i^{(k_l+1)})_l$ such that

- (i) $\tilde{\mathbf{p}}_1^{(\infty, 0)}$ and $\mathbf{p}_1^{(\infty, 1)}$ are minimizers of

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(\infty, 0)}) + f\|_{\mathcal{H}}^2 \quad \text{over } \mathbf{p}_1 \in V_1 \\ \text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}_1(x) + \tilde{\mathbf{p}}_2^{(\infty, 0)}(x) \leq \alpha \mathbf{1}, \end{aligned}$$

and

- (ii) $\tilde{\mathbf{p}}_2^{(\infty, 0)}$ and $\mathbf{p}_2^{(\infty, 1)}$ are minimizers of

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div}(\tilde{\mathbf{p}}_1^{(\infty, 0)} + \mathbf{p}_2) + f\|_{\mathcal{H}}^2 \quad \text{over } \mathbf{p}_2 \in V_2 \\ \text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}_2(x) + \tilde{\mathbf{p}}_1^{(\infty, 0)}(x) \leq \alpha \mathbf{1}, \end{aligned}$$

Proof Set $F(\mathbf{p}) := \frac{1}{2} \|\operatorname{div}(\mathbf{p}) + f\|_{\mathcal{H}}^2 + I_{[-\alpha, \alpha]}(\mathbf{p})$ with $I_{[-\alpha, \alpha]}$ defined as in (5). Then the minimization problem in V_1 is equivalent to $\min_{\mathbf{p}_1 \in V_1} F(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(k)})$. Note that F is convex, lower semicontinuous, bounded from below, and monotonically decreasing with respect to the iterations, i.e., $F(\mathbf{p}^{(k)}) \geq F(\mathbf{p}^{(k+1)})$ for all $k \in \mathbb{N}$.

Since $(\tilde{\mathbf{p}}_2^{(k)})_k$ is bounded, there exists a convergent subsequence $(\tilde{\mathbf{p}}_2^{(k_l)})_l$ with limit $\tilde{\mathbf{p}}_2^{(\infty, 0)}$. Further $(\mathbf{p}_1^{(k_l+1)})_l$ is bounded and hence there exists a convergent subsequence $(\mathbf{p}_1^{(k_{l_j}+1)})_j$ with limit $\mathbf{p}_1^{(\infty, 1)}$.

By the minimality property of $\mathbf{p}_1^{(k_{l_j}+1)}$ we have that $0 \in \partial_{V_1} F(\cdot + \tilde{\mathbf{p}}_2^{(k_{l_j})})(\mathbf{p}_1^{(k_{l_j}+1)})$. Then by [30, Theorem 24.4, p. 233] we obtain that $0 \in \partial_{V_1} F(\cdot + \tilde{\mathbf{p}}_2^{(\infty, 0)})(\mathbf{p}_1^{(\infty, 1)})$ and hence by convexity $\mathbf{p}_1^{(\infty, 1)} \in \arg \min_{\mathbf{p}_1 \in V_1} F(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(\infty, 0)})$.

Analogously we get that $0 \in \partial_{V_2} F(\tilde{\mathbf{p}}_1^{(\infty, 0)} + \cdot)(\mathbf{p}_2^{(\infty, 1)})$, i.e., $\mathbf{p}_2^{(\infty, 1)} \in \arg \min_{\mathbf{p}_2 \in V_2} F(\tilde{\mathbf{p}}_1^{(\infty, 0)} + \mathbf{p}_2)$. Moreover, by the monotone decrease of the energy we have that

$$2F(\mathbf{p}^{(k)}) \geq F(\mathbf{p}_1^{(k+1)} + \tilde{\mathbf{p}}_2^{(k)}) + F(\tilde{\mathbf{p}}_1^{(k)} + \mathbf{p}_2^{(k+1)}) \geq 2F(\mathbf{p}^{(k+1)}) \quad \text{for all } k \in \mathbb{N}$$

and hence

$$2(F(\mathbf{p}^{(k)}) - F(\mathbf{p}^{(k+1)})) \geq F(\mathbf{p}_1^{(k+1)} + \tilde{\mathbf{p}}_2^{(k)}) + F(\tilde{\mathbf{p}}_1^{(k)} + \mathbf{p}_2^{(k+1)}) - 2F(\mathbf{p}^{(k+1)}) \geq 0.$$

The sequence $(\tilde{\mathbf{p}}_1^{(k_{l_j})})_l$ is bounded and hence there exists a convergent subsequence $(\tilde{\mathbf{p}}_1^{(k_{l_{j_m}})})_m$ with limit $\tilde{\mathbf{p}}_1^{(\infty,0)}$. Moreover, $(\mathbf{p}_2^{(k_{l_{j_m}})})_m$ is bounded and has a convergent subsequence $(\mathbf{p}_2^{(k_{l_{j_{m_n}}})})_n$ with limit $\mathbf{p}_2^{(\infty,0)}$. Thus we have

$$\lim_{n \rightarrow \infty} \left[F(\mathbf{p}_1^{(k_{l_{j_{m_n}}+1})} + \tilde{\mathbf{p}}_2^{(k_{l_{j_{m_n}}})}) + F(\tilde{\mathbf{p}}_1^{(k_{l_{j_{m_n}}})} + \mathbf{p}_2^{(k_{l_{j_{m_n}}+1})}) - 2F(\mathbf{p}^{(k_{l_{j_{m_n}+1}})}) \right] = 0.$$

Since $F(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(\infty,0)}) \geq F(\mathbf{p}_1^{(\infty,1)} + \tilde{\mathbf{p}}_2^{(\infty,0)})$ for all $\mathbf{p}_1 \in V_1$ and $F(\tilde{\mathbf{p}}_1^{(\infty,0)} + \mathbf{p}_2) \geq F(\tilde{\mathbf{p}}_1^{(\infty,0)} + \mathbf{p}_2^{(\infty,1)})$ for all $\mathbf{p}_2 \in V_2$ we deduce that

$$F(\tilde{\mathbf{p}}_1^{(\infty,0)} + \tilde{\mathbf{p}}_2^{(\infty,0)}) = F(\mathbf{p}_1^{(\infty,1)} + \tilde{\mathbf{p}}_2^{(\infty,0)}) \text{ and } F(\tilde{\mathbf{p}}_1^{(\infty,0)} + \mathbf{p}_2^{(\infty,1)}) = F(\tilde{\mathbf{p}}_1^{(\infty,0)} + \mathbf{p}_2^{(\infty,1)})$$

and hence

$$\tilde{\mathbf{p}}_1^{(\infty,0)} \in \arg \min_{\mathbf{p}_1 \in V_1} F(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(\infty,0)}) \text{ and } \tilde{\mathbf{p}}_2^{(\infty,0)} \in \arg \min_{\mathbf{p}_2 \in V_2} F(\tilde{\mathbf{p}}_1^{(\infty,0)} + \mathbf{p}_2),$$

which concludes the proof.

Theorem 2 *The sequence $(\mathbf{p}^{(k)})_k$ generated by Algorithm 2 has an accumulation point $\mathbf{p}^{(\infty)} \in V$ which is a solution of the problem in (4).*

Proof The proof can be carried out analogously to the proof of Theorem 1. However, we present an alternative proof.

By similar arguments as in Proposition 4 we extract suitable subsequences and due to their boundedness (Proposition 3), the hierarchy of sequences $(\tilde{\mathbf{p}}_2^{(k_l)})_l$, $(\mathbf{p}_1^{(k_{l_j+1})})_j$, $(\tilde{\mathbf{p}}_1^{(k_{l_{j_m}})})_m$ $(\mathbf{p}_2^{(k_{l_{j_{m_n}}+1})})_n$ converges to limit points $\mathbf{p}_2^{(\infty,0)}$, $\mathbf{p}_1^{(\infty,1)}$, $\tilde{\mathbf{p}}_1^{(\infty,0)}$ $\mathbf{p}_2^{(\infty,1)}$, respectively. Hence the system

$$\begin{aligned} \mathbf{p}_1^{(k_{l_{j_{m_n}}+1})} &\in \arg \min_{\mathbf{p}_1 \in V_1} \left\{ \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(k_{l_{j_{m_n}}})}) + f\|_{\mathcal{H}}^2 : -\alpha \mathbf{1} \leq \mathbf{p}_1(x) + \tilde{\mathbf{p}}_2^{(k_{l_{j_{m_n}}})}(x) \leq \alpha \mathbf{1} \right\}, \\ \mathbf{p}_2^{(k_{l_{j_{m_n}}+1})} &\in \arg \min_{\mathbf{p}_2 \in V_2} \left\{ \frac{1}{2} \|\operatorname{div}(\tilde{\mathbf{p}}_1^{(k_{l_{j_{m_n}}})} + \mathbf{p}_2) + f\|_{\mathcal{H}}^2 : -\alpha \mathbf{1} \leq \mathbf{p}_2(x) + \tilde{\mathbf{p}}_1^{(k_{l_{j_{m_n}}})}(x) \leq \alpha \mathbf{1} \right\}, \end{aligned}$$

converges by [30, Theorem 24.4, p. 233] to

$$\begin{aligned} \mathbf{p}_1^{(\infty,1)} &\in \arg \min_{\mathbf{p}_1 \in V_1} \left\{ \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(\infty,0)}) + f\|_{\mathcal{H}}^2 : -\alpha \mathbf{1} \leq \mathbf{p}_1(x) + \tilde{\mathbf{p}}_2^{(\infty,0)}(x) \leq \alpha \mathbf{1} \right\}, \\ \mathbf{p}_2^{(\infty,1)} &\in \arg \min_{\mathbf{p}_2 \in V_2} \left\{ \frac{1}{2} \|\operatorname{div}(\tilde{\mathbf{p}}_1^{(\infty,0)} + \mathbf{p}_2) + f\|_{\mathcal{H}}^2 : -\alpha \mathbf{1} \leq \mathbf{p}_2(x) + \tilde{\mathbf{p}}_1^{(\infty,0)}(x) \leq \alpha \mathbf{1} \right\}, \end{aligned}$$

for $n \rightarrow \infty$.

By Proposition 4 we have that $\mathbf{p}_2^{(\infty,1)}$ and $\tilde{\mathbf{p}}_2^{(\infty,0)}$ are minimizers of

$$\begin{aligned} &\min \frac{1}{2} \|\operatorname{div}(\tilde{\mathbf{p}}_1^{(\infty,0)} + \mathbf{p}_2) + f\|_{\mathcal{H}}^2 \text{ over } \mathbf{p}_2 \in V_2 \\ &\text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}_2(x) + \tilde{\mathbf{p}}_1^{(\infty,0)}(x) \leq \alpha \mathbf{1}, \end{aligned}$$

and $\mathbf{p}_1^{(\infty,1)}$ and $\tilde{\mathbf{p}}_1^{(\infty,0)}$ are minimizers of

$$\begin{aligned} &\min \frac{1}{2} \|\operatorname{div}(\mathbf{p}_1 + \tilde{\mathbf{p}}_2^{(\infty,0)}) + f\|_{\mathcal{H}}^2 \text{ over } \mathbf{p}_1 \in V_1 \\ &\text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}_1(x) + \tilde{\mathbf{p}}_2^{(\infty,0)}(x) \leq \alpha \mathbf{1}, \end{aligned}$$

Hence we get the following first order conditions for $\tilde{\mathbf{p}}_1^{(\infty,0)} \in V_1$ and $\tilde{\mathbf{p}}_2^{(\infty,0)} \in V_2$:

$$\left\langle \operatorname{div}(\tilde{\mathbf{p}}_1^{(\infty,0)} + \tilde{\mathbf{p}}_2^{(\infty,0)}) + f, \operatorname{div} \mathbf{q}_1 - \operatorname{div} \tilde{\mathbf{p}}_1^{(\infty,0)} \right\rangle_{\mathcal{H}} \geq 0$$

for all $\mathbf{q}_1 \in V_1$, $-\alpha \mathbf{1} \leq \mathbf{q}_1(x) + \tilde{\mathbf{p}}_2^{(\infty,0)}(x) \leq \alpha \mathbf{1}$ for all $x \in \Omega$.

$$\left\langle \operatorname{div}(\tilde{\mathbf{p}}_1^{(\infty,0)} + \tilde{\mathbf{p}}_2^{(\infty,0)}) + f, \operatorname{div} \mathbf{q}_2 - \operatorname{div} \tilde{\mathbf{p}}_2^{(\infty,0)} \right\rangle_{\mathcal{H}} \geq 0$$

for all $\mathbf{q}_2 \in V_2$, $-\alpha \mathbf{1} \leq \mathbf{q}_2(x) + \tilde{\mathbf{p}}_1^{(\infty,0)}(x) \leq \alpha \mathbf{1}$ for all $x \in \Omega$.

Then by adding up the last two inequalities we obtain

$$\left\langle \operatorname{div}(\tilde{\mathbf{p}}_1^{(\infty,0)} + \tilde{\mathbf{p}}_2^{(\infty,0)}) + f, \operatorname{div}(\mathbf{q}_1 + \mathbf{q}_2) - \operatorname{div}(\tilde{\mathbf{p}}_1^{(\infty,0)} + \tilde{\mathbf{p}}_2^{(\infty,0)}) \right\rangle_{\mathcal{H}} \geq 0$$

for all $\mathbf{q}_i \in V_i$ with $-\alpha \mathbf{1} \leq \mathbf{q}_1(x) + \tilde{\mathbf{p}}_2^{(\infty,0)}(x) \leq \alpha \mathbf{1}$ for all $x \in \Omega$ and $-\alpha \mathbf{1} \leq \mathbf{q}_2(x) + \tilde{\mathbf{p}}_1^{(\infty,0)}(x) \leq \alpha \mathbf{1}$ for all $x \in \Omega$ for $i = 1, 2$. Since Ω_1 and Ω_2 are disjoint and $-\alpha \mathbf{1} \leq \mathbf{q}_1(x) + \tilde{\mathbf{p}}_2^{(\infty,0)}(x) \leq \alpha \mathbf{1}$ as well as $-\alpha \mathbf{1} \leq \mathbf{q}_2(x) + \tilde{\mathbf{p}}_1^{(\infty,0)}(x) \leq \alpha \mathbf{1}$ it follows that $-\alpha \mathbf{1} \leq \mathbf{q}_1(x) + \mathbf{q}_2(x) \leq \alpha \mathbf{1}$ and $-\alpha \mathbf{1} \leq \tilde{\mathbf{p}}_1^{(\infty,0)}(x) + \tilde{\mathbf{p}}_2^{(\infty,0)}(x) \leq \alpha \mathbf{1}$ for all $x \in \Omega$. Hence $\mathbf{p}^{(\infty)} = \tilde{\mathbf{p}}_1^{(\infty,0)} + \tilde{\mathbf{p}}_2^{(\infty,0)} \in V$ is a solution of the global problem in (4).

4 Implementation

For technical reasons, we will require a certain splitting property of the discrete divergence operator with respect to a disjoint decomposition of the domain Ω , i.e.,

$$\begin{aligned} \sum_{x \in \Omega} \operatorname{div}_{\Omega}(\mathbf{p}_1 + \mathbf{p}_2)(x) &= \left(\sum_{x \in \Omega_1 \cup \hat{\Omega}_2} \operatorname{div}_{\Omega_1 \cup \hat{\Omega}_2}(\mathbf{p}_1 + \mathbf{p}_2)(x) \right) + d_1(\mathbf{p}_2) \\ \sum_{x \in \Omega} \operatorname{div}_{\Omega}(\mathbf{p}_1 + \mathbf{p}_2)(x) &= \left(\sum_{x \in \Omega_2 \cup \hat{\Omega}_1} \operatorname{div}_{\Omega_2 \cup \hat{\Omega}_1}(\mathbf{p}_1 + \mathbf{p}_2)(x) \right) + d_2(\mathbf{p}_1) \end{aligned}$$

where d_i is a suitable function only depending on \mathbf{p}_{i^c} , $i^c \in \{1, 2\} \setminus \{i\}$, and $\hat{\Omega}_i \subset \Omega_i$ is a small stripe around the interface between Ω_1 and Ω_2 for $i = 1, 2$. For appropriately chosen $\hat{\Omega}_i$, $i = 1, 2$, this splitting property always holds.

Let us consider, for example, the subspace minimization in V_1 , i.e.,

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div}_{\Omega}(\mathbf{p}_1 + \mathbf{p}_2^{(k)}) + f\|_{\mathcal{H}}^2 \quad \text{over } \mathbf{p}_1 \in V_1 \\ \text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}_1(x) \leq \alpha \mathbf{1} \text{ for all } x \in \Omega_1, \end{aligned}$$

which can be written as

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div}_{\Omega_1 \cup \hat{\Omega}_2}(\mathbf{p}_1 + \mathbf{p}_2^{(k)}) + f + d_1(\mathbf{p}_2^{(k)})\|_{\ell^2(\Omega_1 \cup \hat{\Omega}_2)}^2 \quad \text{over } \mathbf{p}_1 \in V_1 \\ \text{s.t. } -\alpha \mathbf{1} \leq \mathbf{p}_1(x) \leq \alpha \mathbf{1} \text{ for all } x \in \Omega_1, \end{aligned} \tag{9}$$

where we used the above splitting property of the divergence operator.

We define $\hat{V}_2 := \{\mathbf{v} \in V : \mathbf{v}(x) = \mathbf{0} \text{ for } x \in \Omega \setminus \hat{\Omega}_2\}$ and rewrite the constrained minimization problem (9) in the following form

$$\begin{aligned} \min \frac{1}{2} \|\operatorname{div}_{\Omega_1 \cup \hat{\Omega}_2}(\boldsymbol{\xi}_1) + f + d_1(\mathbf{p}_2^{(k)})\|_{\ell^2(\Omega_1 \cup \hat{\Omega}_2)}^2 \quad \text{over } \boldsymbol{\xi}_1 \in V_1 \oplus \hat{V}_2 \\ \text{s.t. } \pi_{\hat{V}_2} \boldsymbol{\xi}_1 = \pi_{\hat{V}_2} \mathbf{p}_2^{(k)} \quad \text{and } -\alpha \mathbf{1} \leq \boldsymbol{\xi}_1(x) \leq \alpha \mathbf{1} \text{ for all } x \in \Omega_1 \cup \hat{\Omega}_2. \end{aligned}$$

Note that for $i = 1, 2$ and $i^c = \{1, 2\} \setminus \{i\}$ indeed ξ_i is optimal if and only if $\mathbf{p}_i = \xi_i - \mathbf{p}_{i^c}$ is optimal. In order to solve this constrained minimization problem we use the *augmented Lagrangian method* [3, 26]:

Algorithm 3

- Choose $\xi_1^{(0)} (= \pi_{V_1 \oplus \widehat{V}_2} \mathbf{p}^{(k)}) \in V_1, \mathbf{q}^{(0)} (= \mathbf{0}) \in V_1$, and $\mu > 0$
- Iterate: For $n = 0, 1, \dots$

$$\xi_1^{(n+1)} \in \arg \min_{\xi_1 \in V_1 \oplus \widehat{V}_2} \left\{ \frac{1}{2} \|\operatorname{div}_{\Omega_1 \cup \widehat{\Omega}_2}(\xi_1) + f + d_1(\mathbf{p}_2^{(k)})\|_{\ell^2(\Omega_1 \cup \widehat{\Omega}_2)}^2 - \langle \mathbf{q}^{(n)}, \xi_1 - \xi_1^{(n)} \rangle_V \right. \\ \left. + \frac{\mu}{2} \|\pi_{\widehat{V}_2}(\xi_1 - \mathbf{p}_2^{(k)})\|_{\ell^2(\Omega_1 \cup \widehat{\Omega}_2)}^2 : -\alpha \mathbf{1} \leq \xi_1(x) \leq \alpha \mathbf{1} \right\}$$

$$\mathbf{q}^{(n+1)} := \mathbf{q}^{(n)} - \mu \pi_{\widehat{V}_2}^* \pi_{\widehat{V}_2}(\xi_1^{(n+1)} - \mathbf{p}_2^{(k)})$$

The minimization problem in Algorithm 3 can be treated as proposed in [21, Section 3], where a family of approximating problems is considered. This method has the advantage of an image resolution independent convergence for fixed β, μ and it converges locally at a superlinear rate. More precisely, the minimization problem is replaced by a regularized problem, i.e., we consider

$$\min \frac{1}{2\beta} |\nabla \xi_1|^2 + \frac{1}{2} \|\operatorname{div}_{\Omega_1 \cup \widehat{\Omega}_2}(\xi_1) + f + d_1(\mathbf{p}_2^{(k)})\|_{\ell^2(\Omega_1 \cup \widehat{\Omega}_2)}^2 + \frac{\gamma}{2} \|P_{div} \xi_1\|_{\ell^2(\Omega_1 \cup \widehat{\Omega}_2)}^2 \\ + \frac{1}{2\beta} |\max(0, \beta(\xi_1 - \alpha \mathbf{1}))|^2 + \frac{1}{2\beta} |\min(0, \beta(\xi_1 + \alpha \mathbf{1}))|^2 - \langle \mathbf{q}^{(n)}, \xi_1 - \xi_1^{(n)} \rangle_V \quad (10) \\ + \frac{\mu}{2} \|\pi_{\widehat{V}_2}(\xi_1 - \mathbf{p}_2^{(k)})\|_{\ell^2(\Omega_1 \cup \widehat{\Omega}_2)}^2 \quad \text{over } \xi_1 \in V_1 \oplus \widehat{V}_2,$$

where $\beta > 0, \gamma \geq 0$, ∇ denotes the discrete vector gradient in $\Omega_1 \cup \widehat{\Omega}_2$, and P_{div} is the orthogonal projection in $V_1 \oplus \widehat{V}_2$ onto $\{\mathbf{v} \in V_1 \oplus \widehat{V}_2 : \operatorname{div}_{\Omega_1 \cup \widehat{\Omega}_2} \mathbf{v}(x) = 0 \text{ for all } x \in \Omega_1 \cup \widehat{\Omega}_2\}$. As argued in [21], the parameter γ has no noticeable effect on the result and thus can be set to 0, as it is done in our numerical experiments below. Let ξ_1^* denote the unique solution to (10). Then it satisfies the following optimality conditions

$$-\frac{1}{\beta} \Delta \xi_1^* - \nabla_{\Omega_1 \cup \widehat{\Omega}_2} (\operatorname{div}_{\Omega_1 \cup \widehat{\Omega}_2} \xi_1^* + f + d_1(\mathbf{p}_2^{(k)})) + \gamma P_{div} \xi_1^* - \mathbf{q}^{(n)} + \mu \pi_{\widehat{V}_2}^* \pi_{\widehat{V}_2}(\xi_1^* - \mathbf{p}_2^{(k)}) + \lambda = 0 \\ \lambda = \max(0, \beta(\xi_1^* - \alpha \mathbf{1})) + \min(0, \beta(\xi_1^* + \alpha \mathbf{1})),$$

where Δ is the discrete vector Laplacian in $\Omega_1 \cup \widehat{\Omega}_2$. This leads to the following algorithm for computing $\xi_1^{(n+1)}$, cf. [21, Algorithm B],

Algorithm 4

- Initialize: $\zeta_{(0)} = \xi_1^{(n)}$
- Iterate: Set $\ell = 0$. While stopping rule not satisfied do:
 - Set, for $j = 1, 2$,

$$\begin{aligned}\mathcal{A}_{\ell+1}^{+,j} &= \{x : \zeta_{(\ell)}^j(x) - \alpha > 0\}, \\ \mathcal{A}_{\ell+1}^{-,j} &= \{x : \zeta_{(\ell)}^j(x) + \alpha < 0\}, \\ \mathcal{I}_{\ell+1}^j &= (\Omega_1 \cup \widehat{\Omega}_2) \setminus (\mathcal{A}_{\ell+1}^{+,j} \cup \mathcal{A}_{\ell+1}^{-,j})\end{aligned}$$

- Solve for $\zeta_{(\ell+1)}$

$$\begin{aligned}\left(-\frac{1}{\beta}\Delta - \nabla_{\Omega_1 \cup \widehat{\Omega}_2} \operatorname{div}_{\Omega_1 \cup \widehat{\Omega}_2} + \gamma P_{div}\right) \zeta_{(\ell+1)} + \mu \pi_{V_2}^* \pi_{\widehat{V}_2}(\zeta_{(\ell+1)} - \mathbf{p}_2^{(k)}) + \\ \beta(\zeta_{(\ell+1)} - \alpha \mathbf{1}) \chi_{\mathcal{A}_{\ell+1}^+} + \beta(\zeta_{(\ell+1)} + \alpha \mathbf{1}) \chi_{\mathcal{A}_{\ell+1}^-} - \\ \nabla_{\Omega_1 \cup \widehat{\Omega}_2} \operatorname{div}_{\Omega_1 \cup \widehat{\Omega}_2}(f + d_1(\mathbf{p}_2^{(k)})) - \mathbf{q}^{(n)} = 0.\end{aligned}$$

- $\ell := \ell + 1$.
- Set $\xi_1^{(n+1)} = \zeta_{(\ell)}$.

In the above algorithm we used $\chi_{\mathcal{A}_{\ell+1}^\pm} = (\chi_{\mathcal{A}_{\ell+1}^\pm}^1, \chi_{\mathcal{A}_{\ell+1}^\pm}^2)$ with

$$\chi_{\mathcal{A}_{\ell+1}^\pm}^j = \begin{cases} 1 & \text{if } x \in \mathcal{A}_{\ell+1}^{\pm,j}, \\ 0 & \text{if } x \notin \mathcal{A}_{\ell+1}^{\pm,j}, \end{cases}$$

for $j = 1, 2$. The while-loop terminates when the residual of the respective linear system drops below a user-specified tolerance for the first time. This tolerance is given in iteration $\ell + 1$ by $\text{tol}_{\ell+1} = 0.1 \min(r_\ell^{1.25}, r_\ell)$, where r_ℓ denotes the residual in the ℓ -th iteration, as suggested in [21]. The linear system in Algorithm 4 is resolved by the conjugate gradient method (CG-method). The stopping tolerance for the CG-method in iteration ℓ is again the aforementioned tolerance tol_ℓ .

We conclude this section by mentioning that the implementation for the minimization in V_2 works analogously by simply adjusting the notations accordingly.

5 Numerical Examples

In this section we present numerical experiments for the proposed sequential and parallel algorithms for image denoising. The values of the parameters are chosen experimentally, i.e., we choose the values which give a good compromise between visual quality and computational time of the algorithm. Note that the dynamic range of all image data considered in this paper is $[0, 1]$. All the computations presented 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.

5.1 Denoising of 2D Image Data

In our numerical experiments for 2D image denoising we terminate Algorithm 1 and Algorithm 2 as soon as the residual of the global regularized problem, cf. (10), drops below a certain threshold $R^* > 0$.

We consider the “barbara” image of size 512×512 pixels, depicted in Figure 1, which is corrupted by Gaussian white noise with variance 0.03. The sequential and parallel domain decomposition algorithms are applied to this image by decomposing the image domain into $D = 2, 4, 8, 16$ subdomains. In particular, we either split the physical domain into stripes or windows as shown in Figure 1. In

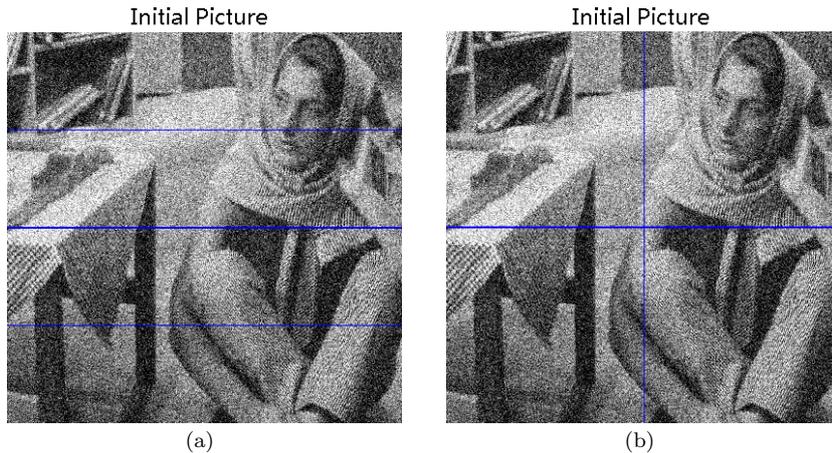


Fig. 1 Domain decomposition in 4 stripes, see (a), and 4 windows, see (b), of the "barbara" image corrupted by Gaussian noise with standard deviation 0.03.

# domains	sequential	parallel
$D = 2$ (stripe):	124 s / 6 it (20 s/it)	262 s / 16 it (16 s/it)
$D = 4$ (stripe):	145 s / 7 it (20 s/it)	231 s / 19 it (12 s/it)
$D = 4$ (window):	150 s / 7 it (21 s/it)	236 s / 22 it (10 s/it)
$D = 8$ (stripe):	173 s / 8 it (21 s/it)	253 s / 22 it (11 s/it)
$D = 16$ (stripe):	235 s / 10 it (23 s/it)	353 s / 26 it (13 s/it)
$D = 16$ (window):	180 s / 8 it (22 s/it)	302 s / 26 it (11 s/it)

Table 1 Restoration of the "barbara" image in Figure 1 of size 512×512 pixels: Computational performance (CPU time in seconds / the number of iterations and the average CPU time for one iteration in Algorithm 1 and Algorithm 2 respectively) for the sequential and parallel domain decomposition algorithms with $\alpha = 0.13$, $\beta = 7$, $\gamma = 0$, $\mu = 500$ for different numbers of subdomains ($D = 2, 4, 8, 16$). The algorithms are stopped as soon as the global residual reaches the significance level $2.1 * 10^{-4}$.

the algorithms we choose $\alpha = 0.13$, $\beta = 7$, $\gamma = 0$, and $\mu = 500$. The iterations are terminated as soon as the global residual reaches the significance value of $R^* = 2.1 * 10^{-4}$. We note that R^* was chosen empirically as the quality of the reconstruction was satisfactory and no noticeable change occurred for smaller values of R^* . For different numbers of splittings we show in Table 1 the required computational time and the number of required iterations. Moreover, in Table 1 we also display the approximate average time for one (outer) iteration of Algorithm 1 or Algorithm 2, respectively. From Table 1 we observe that the number of iterations is increasing with the number of domains. This phenomenon may be attributed to the increasing number of interfaces, where the correct solution can only be obtained during the iterations, while inside the domains the solution seems to be immediately achieved. Due to the fact that the image size is rather small, parallelization does not pay off, since the communication time between the processors adds to the CPU-time and more iterations are needed until convergence is reached. The latter fact comes from the averaging of the current iterate with the previous one, cf. Algorithm 2. Nevertheless we emphasize that for each single iteration significantly less computational time is needed with the parallel version than with the sequential one. We also note, that we ran experiments in which we modified the parallel algorithm in a way that the update is done without any averaging, i.e., $\mathbf{p}^{(k+1)} = \sum_{i=1}^D \mathbf{p}_i^{(k+1)}$. Additionally to the fact that analytically we cannot proof convergence with this modified update, we observed in our numerical experiments that convergence to a global minimizer cannot be guaranteed in general. Thus the averaging process in the update is not only theoretically but also practically necessary.

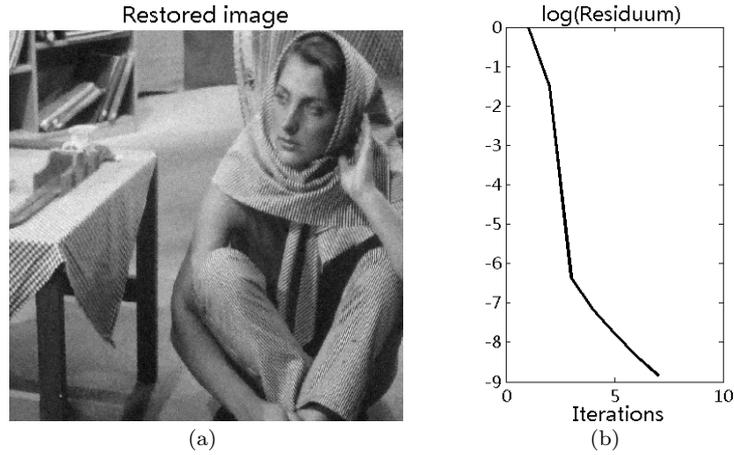


Fig. 2 In (a) we depict the restoration of the “barbara” image in Figure 1. For the “barbara” image of size 2024×2024 pixels the decrease of the logarithm of the global residual for the sequential domain decomposition with 4 subdomains (stripes) is plotted in (b).

# domains	sequential	parallel
$D = 2$ (stripe):	2736 s / 5 it (547 s/it)	4580 s / 13 it (352 s/it)
$D = 4$ (stripe):	2475 s / 6 it (412 s/it)	2898 s / 16 it (181 s/it)
$D = 4$ (window):	2209 s / 5 it (441 s/it)	3407 s / 17 it (200 s/it)
$D = 8$ (stripe):	2980 s / 7 it (425 s/it)	2292 s / 19 it (120 s/it)
$D = 16$ (stripe):	2852 s / 7 it (407 s/it)	2825 s / 22 it (128 s/it)
$D = 16$ (window):	2435 s / 6 it (405 s/it)	2902 s / 21 it (138 s/it)

Table 2 Restoration of the “barbara” image in Figure 1 rescaled to size 2048×2048 pixels: Computational performance (CPU time in seconds / the number of iterations and the average CPU time for one iteration in Algorithm 1 and Algorithm 2 respectively) for the sequential and parallel domain decomposition algorithms with $\alpha = 0.13$, $\beta = 7$, $\gamma = 0$, $\mu = 500$ for different numbers of subdomains ($D = 2, 4, 8, 16$). The algorithms are stopped as soon as the global residual reaches the significance level 2.1×10^{-4} .

Since an image of size 512×512 pixels is rather small, for the sake of comparison we rescaled the “barbara” image to a size of 2048×2048 pixels and denoised it by the sequential and parallel algorithm with the same parameters and stopping criterion as used above. The results of these experiments are documented in Table 2. For a decomposition into 4 stripes the restored image is plotted in Figure 2 (a), while in Figure 2 (b) the progress of the global residual over the iterations is depicted. Still, the sequential algorithm is faster than its parallel version, which can be explained as before. However, for a decomposition into 8 stripes the parallel domain decomposition algorithm outperforms the sequential one, although more than twice as many iterations are needed. Additionally we observe that the number of domains for which the algorithm gives the best performance clearly depends on the size of the considered image. While for the small size “barbara” image the best performance is achieved for a splitting into two domains for the sequential algorithm and into four domains for the parallel version, for the rescaled image four domains for the sequential and eight domains for the parallel algorithm converge fastest. We further note that the CPU-cost per iteration exhibits a stronger increase for the sequential algorithm when compared with the parallel one. This indicates that for yet larger image sizes a trade-off between the sequential and the parallel method will be reached rendering the parallel version more efficient for growing image data sizes. This claim is substantiated in Section 5.3 for the example of 3D medical imaging where the parallel method is indeed more efficient than the sequential method.

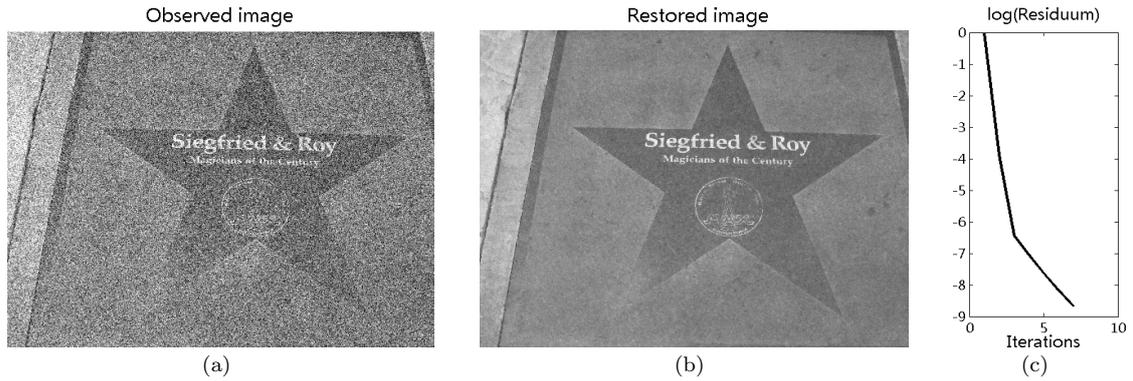


Fig. 3 Domain decomposition of an image (size 1920×2576 pixels) corrupted by Gaussian noise with variance 0.03, see (a). In (b) we show the restored image, whereby we used the sequential domain decomposition algorithm for 4 domains with the following parameters: $\alpha = 0.13$, $\beta = 7$, $\gamma = 0$, $\mu = 500$. In (c) we depict the progress of the logarithm of the global residual.

# domains	sequential	parallel
$D = 2$ (stripe):	2896 s / 5 it (579 s/it)	5374 s / 14 it (383 s/it)
$D = 4$ (stripe):	2675 s / 6 it (445 s/it)	3262 s / 18 it (181 s/it)
$D = 4$ (window):	2848 s / 6 it (474 s/it)	3535 s / 17 it (207 s/it)
$D = 8$ (stripe):	3114 s / 7 it (444 s/it)	3191 s / 21 it (151 s/it)
$D = 16$ (stripe):	3673 s / 8 it (459 s/it)	3786 s / 23 it (164 s/it)
$D = 16$ (window):	3267 s / 7 it (466 s/it)	3385 s / 22 it (153 s/it)

Table 3 Restoration of the image in Figure 3 of size 1920×2576 pixels: Computational performance (CPU time in seconds / the number of iterations and the average CPU time for one iteration in Algorithm 1 and Algorithm 2 respectively) for the sequential and parallel domain decomposition algorithms with $\alpha = 0.13$, $\beta = 7$, $\gamma = 0$, $\mu = 500$ for different numbers of subdomains ($D = 2, 4, 8, 16$). The algorithms are stopped as soon as the global residual reaches the significance level $2.1 \cdot 10^{-4}$.

We also tested the sequential and parallel domain decomposition algorithm for the image in Figure 3, which is of size 1920×2576 pixels corrupted by Gaussian white noise with variance 0.03, and for the image in Figure 4 of size 1280×1280 pixels corrupted by Gaussian noise with variance 0.1. The results are reported in Table 3 and Table 4. For the restoration of the image in Figure 3 we set the parameters in the algorithms to $\alpha = 0.13$, $\beta = 7$, $\gamma = 0$, $\mu = 500$, while for the one in Figure 4 we choose $\alpha = 0.25$, $\beta = 5$, $\gamma = 0$, $\mu = 500$. Upon inspection of the figures in the tables, similar conclusions as for the previous test examples can be drawn.

5.2 Comparison of the Dual and Primal Domain Decomposition Approach

In [22] domain decomposition methods directly applied to the minimization of the primal problem (2) are introduced, which we refer to as the primal domain decomposition approach. In fact, we only compare the respective sequential algorithms, i.e., Algorithm 1 with the algorithm in [22, (3.1)], since a similar behaviour can be observed from the parallel versions. The comparison is done with respect to CPU-time and iterations for the restoration of the “barbara” image of size 512×512 pixels corrupted by Gaussian white noise with variance 0.03. The physical domain is split into different numbers of stripes ($D = 2, 3, 4, 5, 6, 7, 8$) as depicted in Figure 1(a). For the restoration we choose $\alpha = 1/3$ in (2) and (3). Further, in Algorithm 1 we set $\beta = 70$, $\gamma = 0$, $\mu = 500$, while in the primal domain decomposition algorithm we set $\gamma = 0.01$, $\mu = 1$, and $\delta = 0.99$ as proposed in [22]. As a

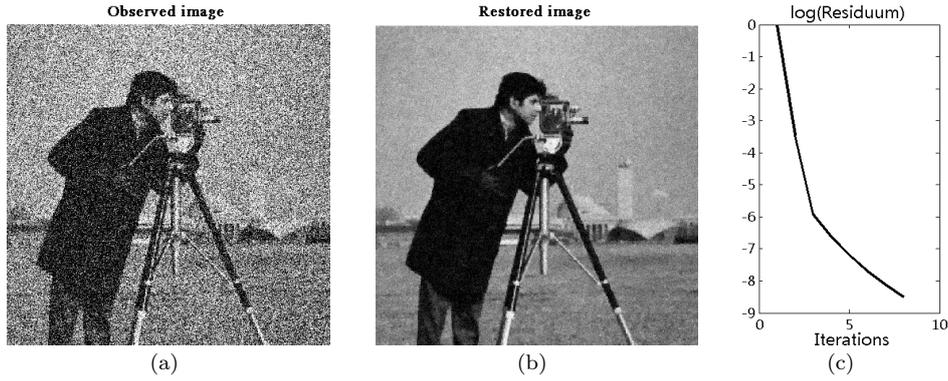


Fig. 4 Domain decomposition of an image (size 1280×1280 pixels) corrupted by Gaussian noise with variance 0.1, see (a). In (b) we show the restored image, whereby we used the sequential domain decomposition algorithm for 4 domains with the following parameters: $\alpha = 0.25$, $\beta = 5$, $\gamma = 0$, $\mu = 500$. In (c) we depict the progress of the logarithm of the global residual.

# domains	sequential	parallel
$D = 2$ (stripe):	1207 s / 6 it (201 s/it)	2705 s / 20 it (135 s/it)
$D = 4$ (stripe):	1108 s / 7 it (158 s/it)	1640 s / 24 it (68 s/it)
$D = 4$ (window):	1210 s / 7 it (172 s/it)	1801 s / 22 it (81 s/it)
$D = 8$ (stripe):	1326 s / 9 it (147 s/it)	1309 s / 27 it (48 s/it)
$D = 16$ (stripe):	1506 s / 10 it (150 s/it)	1682 s / 31 it (54 s/it)
$D = 16$ (window):	1199 s / 8 it (149 s/it)	1590 s / 28 it (56 s/it)

Table 4 Restoration of the image in Figure 4 of size 1280×1280 pixels: Computational performance (CPU time in seconds / the number of iterations and the average CPU time for one iteration in Algorithm 1 and Algorithm 2 respectively) for the sequential and parallel domain decomposition algorithms with $\alpha = 0.25$, $\beta = 5$, $\gamma = 0$, $\mu = 500$ for different numbers of subdomains ($D = 2, 4, 8, 16$). The algorithms are stopped as soon as the global residual reaches the significance level 2.1×10^{-4} .

stopping criterion we use the energy in (2) which is at hand for both approaches, since for a solution p^* of (3) and a minimizer u^* of (2) we have $u^* = \operatorname{div} p^* + f$; see [21]. For obtaining an appropriate stopping value for our tests we first solve the dual domain decomposition algorithm with a certain accuracy, i.e., we terminate the iterations as soon as the global residual reaches the significance value $R^* = 10^{-2}$. Then the observed primal energy is taken to be the stopping value, i.e., we stop the algorithms as soon as the primal energy drops below 0.0645388 for the first time in our numerical experiments. While the dual domain decomposition algorithm reaches the prescribed energy after a few iterations, the computation by the primal domain decomposition method takes many more iterations or does not even reach the desired value within 500 iterations; see Table 5. Since only a few iterations are needed for the dual algorithm, it converges much faster in CPU-time than the primal domain decomposition approach. This behaviour makes the dual splitting strategy superior for large-scale problems as they may occur in 3D imaging.

5.3 Reconstruction of 3D Computerized Tomography (CT) Data

We test Algorithm 1 and Algorithm 2 for denoising 3D medical data of $512 \times 512 \times 48$ voxels, i.e., the data consists of 48 slices with each slice containing 512×512 pixels; see Figure 5 for a vertical cut through the restoration. Solving this restoration task by minimizing (3) involves a huge computational effort. Therefore it is sensible to split the computation load into smaller pieces by using

# domains	dual	primal
$D = 2$:	233 s / 29 it (8 s/it)	344 s / 136 it (2.5 s/it)
$D = 3$:	168 s / 30 it (5.6 s/it)	3957 s / 500 it (7.9 s/it)
$D = 4$:	113 s / 29 it (3.8 s/it)	5199 s / 500 it (10.3 s/it)
$D = 5$:	97 s / 29 it (3.3 s/it)	6032 s / 500 it (12 s/it)
$D = 6$:	92 s / 29 it (3.1 s/it)	6725 s / 500 it (13.4 s/it)
$D = 7$:	92 s / 28 it (3.2 s/it)	6733 s / 500 it (13.4 s/it)
$D = 8$:	89 s / 28 it (3.1 s/it)	7265 s / 500 it (14.5 s/it)

Table 5 Comparison of the proposed dual domain decomposition Algorithm 1 and the primal domain decomposition algorithms introduced in [22] for the restoration of the image in Figure 1 of size 512×512 pixels: Computational performance (CPU time in seconds / the number of iterations and the average CPU time for one iteration) for different numbers of subdomains ($D = 2, 3, 4, 5, 6, 7, 8$). The algorithms are stopped as soon as the global primal energy in (2) drops below 0.0645388 or the number of iterations reaches 500 first.

# domains	parallel	sequential
$D = 2$:	9791 s / 6 it	9758 s / 4 it
$D = 4$:	7123 s / 7 it	8238 s / 4 it
$D = 8$:	5903 s / 7 it	7569 s / 4 it

Table 6 Comparison of the proposed sequential and parallel domain decomposition algorithms with $\alpha = 0.013$, $\beta = 7$, $\gamma = 0$, $\mu = 500$ for denoising 3D medical data of $512 \times 512 \times 48$ voxels. Computational performance (CPU time in seconds / the number of iterations) for different numbers of subdomains ($D = 2, 4, 8, 16$). The algorithms are stopped as soon as $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\| < 10^{-6}$.

a domain decomposition strategy. A perhaps too simple splitting approach would be, minimizing (3) on each slice separately, i.e., instead of solving the whole 3D problem we solve 48 2D problems. This strategy only finds minimizers of each slice but does not find the minimizer of the associated 3D problem. Moreover, it is clear that since no neighboring 2D-slices are taken into account, high variations in the vertical direction, i.e., the direction perpendicular to the 2D-slices, are allowed; see Figure 5 right column. A comparison of the solutions of this suboptimal strategy with our proposed domain decomposition algorithm applied to the genuine 3D image restoration problem shows that edges as well as uniform parts are much better preserved in the vertical direction, when dealing with 3D data rather than stacking 2D reconstructions; compare Figure 5. Hence, our proposed method is superior to the strategy which solves the problem slice by slice. This also illustrates that a careful treatment of the interfaces of domain patches is crucial.

Moreover, we compare the performance of Algorithm 1 with its parallel version, i.e., Algorithm 2, for denoising the 3D medical data of $512 \times 512 \times 48$ voxels, considered already above. For the parameters we choose $\alpha = 0.013$, $\beta = 7$, $\gamma = 0$, and $\mu = 500$ for different numbers of subdomains ($D = 2, 4, 8$). We terminate the algorithms 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 $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k+1)}\| < 10^{-6}$, which seems suitable for our purposes. 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 $k \geq m$ we have $\|\mathbf{p}^{(k)} - \mathbf{p}^{(k+1)}\| \leq \varepsilon \|\mathbf{p}^{(k)} - \mathbf{p}^{(\infty)}\|$, the above stopping criterion ensures that the distance between our obtained result \mathbf{p} and $\mathbf{p}^{(\infty)}$ is $\|\mathbf{p} - \mathbf{p}^{(\infty)}\| \leq \frac{10^{-6}\varepsilon}{1-\varepsilon}$. Note that in the parallel domain decomposition algorithm for the update of the iterate $\mathbf{p}^{(k+1)}$ an averaging between the previous iterate $\mathbf{p}^{(k)}$ and the current iterates $\mathbf{p}_i^{(k+1)}$, $i = 1, \dots, D$, has to be performed; cf. Algorithm 2. Hence, we expect that the parallel algorithm needs more iterations until successful termination. Additionally, the communication time of the processors adds to the overall computing time. Nevertheless it turns out that the parallel method converges faster in time than the sequential algorithm for a sufficiently large number of subdomains; see Table

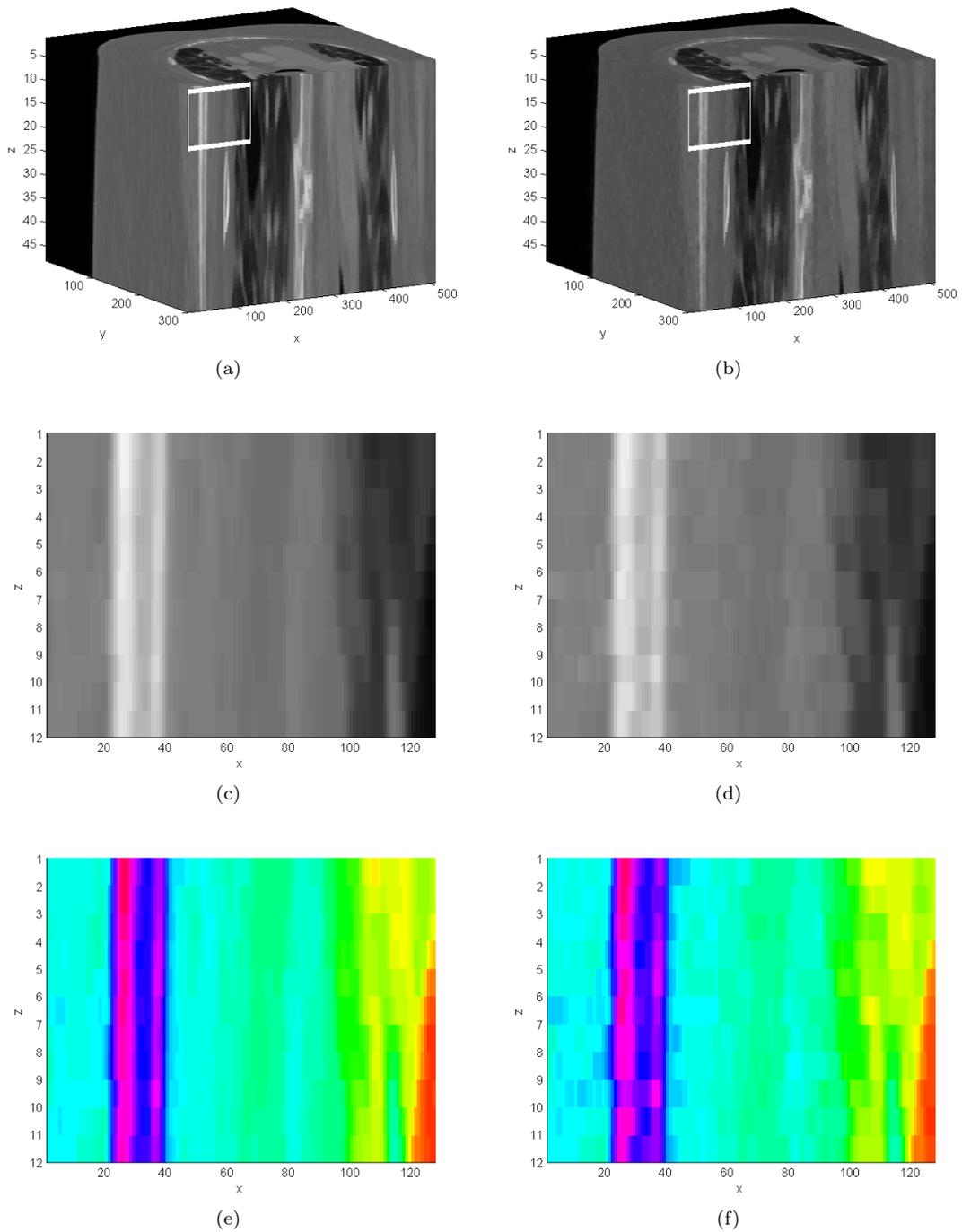


Fig. 5 Reconstruction of 3D CT data given by 48 slices. In the left column we show the reconstruction by the domain decomposition algorithm by splitting the data into 3D blocks, while the right column shows the reconstruction of a slice by slice reconstruction without taking the neighbor slices into account, i.e., solving 2D problems. In (c) and (d) we zoomed in on the in (a) and (b) highlighted area. In order to visualize the differences in the reconstruction of the two approaches we color in (e) and (f) the zoomed area.

6. In fact, the parallel method with a decomposition into 8 domains performs best. These results confirm our previous observation in Section 5.1, where we already anticipated that the parallel version is more efficient than the sequential method for large image data sizes, as it is the case for 3D image data sets.

References

1. Ambrosio, L., Fusco, N. and Pallara, D.: *Functions of bounded variation and free discontinuity problems.*, Oxford Mathematical Monographs. Oxford: Clarendon Press. xviii, 2000.
2. Aubert, G. and Kornprobst, P.: *Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variation*, Springer, 2002.
3. Bertsekas, D.: *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
4. Carstensen, C.: *Domain decomposition for a non-smooth convex minimization problem and its application to plasticity*, Numerical Linear Algebra with Applications, **4**, 177–190 (1998)
5. Chambolle, A.: *An algorithm for total variation minimization and applications*. J. Math. Imaging Vision **20**(1-2), 89–97 (2004)
6. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., and Pock, T.: *An Introduction to Total Variation for Image Analysis*. In M. Fornasier (ed.), Theoretical Foundations and Numerical Methods for Sparse Recovery, Radon Series on Computational and Applied Mathematics, Vol. 9, pp. 263-340, De Gruyter Verlag, Berlin, 2010.
7. Chambolle, A. and Darbon, J.: *On total variation minimization and surface evolution using parametric maximum flows*, Int. J. Comput. Vis. **84**, 288–307 (2009)
8. Chambolle, A. and Lions, P.-L.: *Image recovery via total variation minimization and related problems.*, Numer. Math. **76**(2), 167–188 (1997)
9. Chan, T. F., Golub, G. H., and Mulet, P.: *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput. **20**(6), 1964–1977 (1999)
10. Chan, T. F. and Mathew, T. P.: *Domain decomposition algorithms*, Acta Numerica, **3**, 61–143 (1994)
11. Chan, T. F. and Shen, J.: *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, 2005.
12. Combettes, P. L., and Wajs, V. R.: *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., **4**(4), 1168–1200 (2005)
13. Daubechies, I., Teschke, G., and Vese, L.: *Iteratively solving linear inverse problems under general convex constraints*, Inverse Probl. Imaging **1**(1), 29–46 (2007)
14. Dobson, D. and Vogel, C. R.: *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal. **34**(5), 1779–1791 (1997)
15. Evans, L. C. and Gariepy, R. F.: *Measure Theory and Fine Properties of Functions.*, CRC Press, 1992.
16. Fornasier, M., Kim, Y., Langer, A., and Schönlieb, C.-B.: *Wavelet Decomposition Method for L_2/TV -Image Deblurring*, SIAM J. Imaging Sciences, **5**, 857–885 (2012)
17. Fornasier, M., Langer, A., and Schönlieb, C.-B.: *Domain Decomposition Methods for Compressed Sensing*, Proc. Int. Conf. SampTA09, Marseilles, 2009. arXiv:0902.0124v1 [math.NA]
18. Fornasier, M., Langer, A., and Schönlieb, C.-B.: *A convergent overlapping domain decomposition method for total variation minimization*, Num. Math., **116**, 645–685 (2010).
19. Fornasier, M. and Schönlieb, C.-B.: *Subspace correction methods for total variation and ℓ_1 -minimization*, SIAM J. Numer. Anal., **47**, 3397–3428 (2009)
20. Giusti, E.: *Minimal Surfaces and Functions of Bounded Variation*. Birkhuser, Boston, 1984
21. Hintermüller, M. and Kunisch, K.: *Total Bounded Variation Regularization as a Bilaterally Constrained Optimization Problem*, SIAM J. Appl. Math., **64**(4), 1311–1333 (2004)
22. Hintermüller, M. and Langer, A.: *Subspace Correction Methods for a Class of Non-smooth and Non-additive Convex Variational Problems in Image Processing*, accepted by SIAM J. Imaging Sciences, 2013, 34 pp.
23. Hintermüller, M. and Langer, A.: *Surrogate Functional Based Subspace Correction Methods for Image Processing*, Proceedings of the 21st International Conference on Domain Decomposition Methods, Renne, 2012.
24. Hintermüller, M. and Rautenberg, C. N.: *On the density of classes of closed, convex sets in Sobolev spaces arising from pointwise constraints on function values, the gradient or the divergence*, IFB-Report No. 71 (09/2013) Institute of Mathematics and Scientific Computing, University of Graz.
25. Hintermüller, M. and Stadler, G.: *An Infeasible Primal-Dual Algorithm for Total Bounded Variation-based Inf-Convolution-Type Image Restoration*, SIAM J. Sci. Comput. **28**(1), 1–23 (2006)
26. Ito, K. and Kunisch, K.: *Lagrange Multiplier Approach to Variational Problems and Applications*, Series: Advances in Design and Control (No. 15) SIAM, 2008.
27. Langer, A., Osher, S., and Schönlieb, C.-B.: *Bregmanized Domain Decomposition for Image Restoration*, J. Sci. Comput., **54**, 549-576 (2013)
28. Nesterov, Y.: *Smooth minimization of non-smooth functions*. Mathematical Programming., Ser. A **103**, 127–152 (2005)
29. Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W.: *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul. **4**(2), 460–489 (2005)

30. Rockafellar, R. T.: *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
31. Rudin, L. I., Osher, S., and Fatemi, E.: *Nonlinear total variation based noise removal algorithms*, *Physica D* **60**(1-4), 259–268 (1992)
32. Tai, X.-C. and Tseng, P.: *Convergence rate analysis of an asynchronous space decomposition method for convex minimization*, *Math. Comp.*, **71**, 1105–1135 (2001)
33. Tai, X.-C. and Xu, J.: *Global convergence of subspace correction methods for convex optimization problems*, *Math. Comp.*, **71**, 105–124 (2002)
34. Tseng P., and Yun, S.: *A Coordinate Gradient Descent Method for Nonsmooth Separable Minimization*, *Math. Program., Ser. B* **117**, 387–423 (2009)
35. Vese, L.: *A study in the BV space of a denoising-deblurring variational problem.*, *Appl. Math. Optim.* **44**, 131–161 (2001)