

Convergence Rate of Overlapping Domain Decomposition Methods for the Rudin–Osher–Fatemi Model Based on a Dual Formulation*

Huibin Chang[†], Xue-Cheng Tai[‡], Li-Lian Wang[§], and Danping Yang[¶]

Abstract. This paper is concerned with overlapping domain decomposition methods (DDMs), based on successive subspace correction (SSC) and parallel subspace correction (PSC), for the Rudin–Osher–Fatemi (ROF) model in image restoration. In contrast to recent attempts, we work with a dual formulation of the ROF model, where one significant difficulty resides in the decomposition of the global constraint of the dual variable. We introduce a stable “unity decomposition” using a set of “partition of unity functions,” which naturally leads to overlapping DDMs based on the dual formulation. The main objective of this paper is to rigorously analyze the convergence of the SSC and PSC algorithms and derive the rate of convergence $O(n^{-1/2})$, where n is the number of iterations. Moreover, we characterize the explicit dependence of the convergence rate on the subdomain overlapping size and other important parameters. To the best of our knowledge, such a convergence rate has not yet been claimed for domain decomposition related algorithms for the ROF model.

Key words. Rudin–Osher–Fatemi model, dual formulation, overlapping domain decomposition method, successive subspace correction, parallel subspace correction, convergence rate

AMS subject classifications. 68U10, 65M55, 74S20

DOI. 10.1137/140965016

1. Introduction. The Rudin–Osher–Fatemi (ROF) model introduced by Rudin, Osher, and Fatemi [33] is one of the fundamental variational models for image processing. It is known that the ROF model restores a clear image from a noise image g on a domain Ω in $\mathbb{R}^2$ by solving the minimization problem

$$(1.1) \quad \min_{u \in BV(\Omega)} \left\{ \lambda TV(u) + \frac{1}{2} \|u - g\|_{L^2(\Omega)}^2 \right\},$$

*Received by the editors April 14, 2014; accepted for publication (in revised form) December 8, 2014; published electronically March 10, 2015.

<http://www.siam.org/journals/siims/8-1/96501.html>

[†]School of Mathematical Sciences, Tianjin Normal University, Tianjin, 300387, People's Republic of China (changhuiyin@gmail.com). The research of this author was partially supported by Natural Science Foundation of China grants 11426165, U1304610, and 11401170, PHD Programme 52XB1304 of Tianjin Normal University, and the Science and Technology Planning Project of Henan Province (132102210185).

[‡]Department of Mathematics, University of Bergen, 5020, Bergen, Norway (tai@mi.uib.no).

[§]Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore (lilian@ntu.edu.sg). The research of this author was partially supported by Singapore MOE AcRF Tier 1 grant RG 15/12, MOE AcRF Tier 2 grant MOE 2013-T2-1-095 (ARC 44/13), and Singapore A*STAR-SERC-PSF grant 122-PSF-007.

[¶]Department of Mathematics and Shanghai Key Laboratory of Pure Mathematics and Mathematical Practice, East China Normal University, Shanghai, 200241, People's Republic of China (dpyang@math.ecnu.edu.cn). The research of this author was partially supported by Natural Science Foundation of China grants 11071080 and 11171113.

where $\lambda > 0$, $BV(\Omega)$ is the space of functions of bounded variation, and the total variation of u is defined by (cf. [1])

$$(1.2) \quad TV(u) := \sup_{\mathbf{p} \in \mathcal{K}} \int_{\Omega} u \operatorname{div} \mathbf{p} \, d\mathbf{x} \quad \text{with}$$

$$(1.3) \quad \mathcal{K} := \{\mathbf{p} = (p_1, p_2) \in (C_0^1(\Omega))^2 : |\mathbf{p}| := (p_1^2 + p_2^2)^{1/2} \leq 1\}.$$

Here, u is referred to as the primal variable, while $\mathbf{p}$ is dubbed as the dual variable.

Over the last two decades, many methods [34] have been proposed to solve the ROF model. In general, they can be classified into three categories based on the nature of manipulating the primal and/or dual variables.

- *Primal approaches.* Along this line, the reliable methods typically include the gradient descent method (see, e.g., [29, 33]) and the lagged diffusivity fixed-point iteration (see, e.g., [1, 44, 45, 46, 47]). They can produce good restoration of images but usually converge slowly. The fast algorithm based on the Bregman iteration (cf. [19, 32, 48, 49]) and the augmented Lagrangian technique (cf. [18, 42, 50]) has become the method of choice, thanks to the use of the variable-splitting technique and the availability of the fast Fourier transform for solving the subproblems. The equivalence of the Bregman iteration and augmented Lagrangian method was revealed in [42, 50]. For other algorithms such as the graph cuts method, additive operator splitting (AOS), and the multigrid method, we refer the reader to [53] and the references therein.
- *Dual approaches.* The important representative of dual methods is known as Chambolle's dual algorithm [6]. It essentially applies the Karush–Kuhn–Tucker (KKT) condition (cf. [5]) to (1.1)–(1.2), leading to the dual problem

$$(1.4) \quad \inf_{\mathbf{p} \in \mathcal{K}} \left\{ D(\mathbf{p}) := \int_{\Omega} (\lambda \operatorname{div} \mathbf{p} - g)^2 \, d\mathbf{x} \right\}$$

and then updating the primal variable by

$$(1.5) \quad u = g - \lambda \operatorname{div} \mathbf{p}.$$

- *Primal-dual approaches.* Such approaches were developed in [2, 3, 24] and later were intensively applied to image processing [7, 13, 54]. The work [13] generalized the primal-dual hybrid gradient (PDHG) algorithm [54] and introduced a fairly general framework for primal-dual algorithms. A first-order convergent primal-dual approach was developed and analyzed in [7], and the technique for acceleration to a second-order scheme was proposed.

As the important numerical mean for PDEs, the domain decomposition method (DDM) has two merits (cf. [8, 11, 20, 35, 40, 51]): (i) it can break down the problem into a sequence of subproblems of much smaller scale, so faster and better-conditioned solvers can be constructed; and (ii) it allows for parallel computations. In general, there are two different routes to implement DDMs. The first is to work with the Euler–Lagrangian equation associated with the variational problem, which might lead to a linear or nonlinear PDE. In practice, various techniques including two-level methods, multigrid methods, and preconditioning techniques

are available for the linear equations (see, e.g., [11, 26, 27, 28, 35, 51]), but it is significantly challenging to solve the nonlinear equations. The second is to use the parallel subspace correction (PSC) and successive subspace correction (SSC) methods [36, 38, 39, 41] to decompose the variational problems into subproblems defined on subdomains. It can be applied to more general and complex variational problems, where the energy of the cost functional is guaranteed to decrease monotonically.

There has been much recent interest in developing DDMs for image processing. An overlapping DDM with Dirichlet transmission conditions was developed in [14] for image denoising using Gaussian curvature. The thesis [31] provided an efficient parallel algorithm based on a DDM in primal-dual formulation for the ROF model. The PSC and SSC algorithms were integrated with graph cuts for image restoration and segmentation in [12, 37, 53]. Such a combination led to appealing methods for image processing. We remark that in [53], a coarse mesh correction was added to the proposed domain decomposition scheme which was usually difficult to implement for nonlinear problems. Xu, Chang, and Qin [52] applied such a DDM to image deblurring, where a special treatment for the global blur operator was considered. Chang et al. [9] considered the DDMs for the nonlocal total variation based image restoration problems. Some interesting algorithms built upon domain decomposition techniques were proposed by [15, 16, 17, 22, 23, 25]. Instead of solving the subproblem of SSC or PSC directly, an approximation of the subproblem using the surrogate functional (or iterative proximity map) was adopted in [15, 16, 17], and the approximate subproblem was then solved by oblique thresholding. Hintermüller and Langer [22] proposed and analyzed subspace correction methods for a class of L^1 - L^2 -TV models. Some efficient techniques for such a subspace correction idea were introduced in [25]. Very recently, Hintermüller and Langer [23] studied a nonoverlapping DDM based on a dual formulation for the ROF model with anisotropic total variation.

In contrast to most existing works, we concern ourselves with the overlapping DDM for a dual formulation of (1.4). Indeed, the dual approach, e.g., Chambolle's algorithm, has proven to be efficient for solving the ROF model. However, for images of large size, it is advantageous to employ domain decomposition techniques for the vector-valued dual variable. However, one difficulty lies in the decomposition of the convex set $\mathcal{K}$ with the constraint $|\mathbf{p}| \leq 1$. Note that in [23], the constraints for two components of $\mathbf{p}$ are $|p_1|, |p_2| \leq 1$, which can be decomposed by following [36]. A second difficulty resides in the fact that the cost functional of the dual formulation is neither strictly convex nor strongly convex. Indeed, the framework in [36] cannot be applied to this context.

The paper is organized as follows. We describe the overlapping DDMs in section 2. In section 3, we conduct the convergence analysis of the proposed algorithms. We present in section 4 the numerical experiments to support the theoretical results. We conclude the paper with some final remarks.

2. Overlapping domain decomposition algorithms. In this section, we describe the overlapping domain decomposition algorithms for (1.4). We start with introducing some notations. Define

$$(2.1) \quad H(\operatorname{div}; \Omega) = \{\mathbf{p} \in (L^2(\Omega))^2 : \operatorname{div} \mathbf{p} \in L^2(\Omega)\},$$

which is equipped with the graph norm

$$\|\mathbf{p}\|_{H(\operatorname{div};\Omega)} = \left(\|\mathbf{p}\|_{(L^2(\Omega))^2}^2 + \|\operatorname{div} \mathbf{p}\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}}.$$

Introduce the following subspace of $H(\operatorname{div};\Omega)$:

$$(2.2) \quad H_0(\operatorname{div};\Omega) = \text{closure of } (C_0^\infty(\Omega))^2 \text{ in the } H(\operatorname{div};\Omega) \text{ graph norm.}$$

Recall that if Ω is a bounded Lipschitz domain in $\mathbb{R}^2$, then

$$(2.3) \quad H_0(\operatorname{div};\Omega) = \{\mathbf{p} \in H(\operatorname{div};\Omega) : \mathbf{p} \cdot \mathbf{n}|_{\partial\Omega} = 0\}$$

(see, e.g., [30, Thm. 3.25]), where $\mathbf{n}$ is the outer unit vector normal to $\partial\Omega$. In view of (2.3), we introduce

$$(2.4) \quad K := \{\mathbf{p} \in H_0(\operatorname{div};\Omega) : |\mathbf{p}| \leq 1\}$$

and consider the alternative formulation of (1.4):

$$(2.5) \quad \min_{\mathbf{p} \in K} \left\{ D(\mathbf{p}) := \int_{\Omega} (\lambda \operatorname{div} \mathbf{p} - g)^2 dx \right\}.$$

It is noteworthy that the functional $D(\mathbf{p})$ is convex but not strictly convex, so the problem (2.5) admits at least one minimizer $\mathbf{p}^* \in K$. To guarantee the uniqueness, one may modify the energy functional (cf. [21, 10]) and consider

$$(2.6) \quad D_\epsilon(\mathbf{p}) = D(\mathbf{p}) + \epsilon \|\mathbf{p}\|_{L^2(\Omega)}^2, \quad \text{where } 0 < \epsilon \ll 1.$$

However, this requires a delicate convergence analysis of the limiting process $\epsilon \rightarrow 0$. In fact, it is desirable to work directly with the problem (2.5). Indeed, we need only one optimum $\mathbf{p}^*$ to resolve the primal variable via $u^* = g - \lambda \operatorname{div} \mathbf{p}^*$.

2.1. A general setup for DDM. The domain decomposition algorithm is built upon an appropriate decomposition of the domain Ω and the convex constraint set K . Without loss of generality, we assume that Ω is a rectangular domain. We partition Ω into M_c classes of overlapping subdomains and color each class with a different color, that is,

$$(2.7) \quad \Omega = \bigcup_{j=1}^{M_c} \Omega_j, \quad \Omega_i \cap \Omega_j \neq \emptyset \quad \text{if } i \neq j,$$

where Ω_j is a union of m_j disjoint subdomains with the same color. Hence, the total number of subdomains that cover Ω is $m_T = \sum_{j=1}^{M_c} m_j$. We illustrate in Figure 1 a typical decomposition of a rectangular domain with a total of $m_T = 25$ subdomains colored by four colors (i.e., $M_c = 4$).

In what follows, let δ be the overlapping size of subdomains with different colors, and let N_0 be the maximum number of subdomains, among all m_T subdomains of the partition, where a point $x \in \Omega$ can belong. It is evident that $N_0 \leq M_c$. For example, $N_0 = 4$ for the partition depicted in Figure 1.

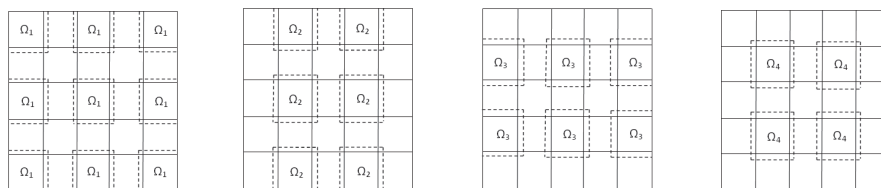


Figure 1. Domain decomposition with coloring technique. Here, $M_c = 4$, $m_1 = 9$, $m_2 = 6$, $m_3 = 6$, $m_4 = 4$, and $m_T = 25$.

An essential technique for our method is to decompose the convex constraint set K into a sum of convex constraint sets over the subdomains. For this purpose, we need to use the “partition of unity functions (PUFs)” [43], denoted by $\{\theta_j\}_{j=1}^{M_c}$, which one-to-one correspond to $\{\Omega_j\}_{j=1}^{M_c}$ and satisfy the following:

$$(2.8) \quad (i) \quad \sum_{j=1}^{M_c} \theta_j \equiv 1, \quad \theta_j \geq 0 \quad \text{on } \Omega;$$

$$(2.9) \quad (ii) \quad \theta_j \in H^1(\Omega), \quad \text{supp}(\theta_j) \subset \bar{\Omega}_j, \quad 1 \leq j \leq M_c;$$

$$(2.10) \quad (iii) \quad \|\nabla \theta_j\|_\infty \leq \frac{C_0}{\delta}, \quad 1 \leq j \leq M_c,$$

where C_0 is a positive constant independent of δ , and $\|\cdot\|_\infty$ is the L^∞ -norm. We refer to Figure 2 for the plots of the (discrete) PUFs associated with the partition of Ω in Figure 1.

With the PUFs at our disposal, we define the convex subsets as

$$(2.11) \quad K_j = \{\mathbf{p} \in H_0(\text{div}; \Omega) : |\mathbf{p}| \leq \theta_j\}, \quad 1 \leq j \leq M_c.$$

We infer from (2.9) that for any $\mathbf{p} \in K_j$, $\text{supp}(\mathbf{p}) \subset \bar{\Omega}_j$. The following identity provides the decomposition of the convex set K .

Proposition 2.1. Let $\{K_j\}_{j=1}^{M_c}$ be the sets defined in (2.11). Then we have

$$(2.12) \quad K = K_1 + K_2 + \cdots + K_{M_c}.$$

Proof. On the one hand, for any $\mathbf{p} \in K$, we can define $\mathbf{p}_j = \theta_j \mathbf{p}$ for $1 \leq j \leq M_c$. Thanks to (2.8)–(2.10), one verifies readily that $\mathbf{p}_j \in K_j$, and $\mathbf{p} = \sum_{j=1}^{M_c} \mathbf{p}_j$. On the other hand, given $\mathbf{p} = \sum_{j=1}^{M_c} \mathbf{p}_j$ with $\mathbf{p}_j \in K_j$, we have $\mathbf{p} \in H_0(\text{div}; \Omega)$, and

$$|\mathbf{p}| = \left| \sum_{j=1}^{M_c} \mathbf{p}_j \right| \leq \sum_{j=1}^{M_c} |\mathbf{p}_j| \leq \sum_{j=1}^{M_c} \theta_j = 1,$$

so $\mathbf{p} \in K$. ■

2.2. DDMs for the dual formulation (2.5). With the above setup, we are now ready to formulate the DDMs for (2.5), which include the PSC algorithm and the SSC algorithm, as described below. Hereafter, let $\alpha > 0$ and $\hat{\alpha} > 0$ be the relaxation parameters.

Algorithm I: Parallel Subspace Correction (PSC)

1. Initialization: choose $\mathbf{p}^0$ and select a relaxation parameter

$$(2.13) \quad \alpha \in (0, M_c^{-1}].$$

2. For $n = 0, 1, \dots$,
Compute $\mathbf{p}^{n+1}$ by

$$(2.14) \quad \mathbf{p}^{n+1} = (1 - \alpha)\mathbf{p}^n + \alpha \sum_{j=1}^{M_c} \hat{\mathbf{q}}_j^n,$$

where

$$(2.15) \quad \hat{\mathbf{q}}_j^n = \arg \min_{\mathbf{v} \in K_j} D\left(\mathbf{v} + \sum_{i \neq j} \theta_i \mathbf{p}^n\right), \quad 1 \leq j \leq M_c.$$

3. Endfor until some stopping criterion meets.
-

We remark that the selection of the relaxation parameter α in the range in (2.13) will be justified by the convergence analysis (cf. (3.13)). The stopping rule can be the error residual (cf. (4.5)) not exceeding a prescribed tolerance (see section 4), and likewise for SSC below. With the PSC algorithm, the minimization of (2.5) boils down to solving a sequence of subdomain problems (2.15), which has a much smaller scale and can be resolved in parallel.

It is noteworthy that when we update $\mathbf{p}^{n+1}$ by (2.14), we could use the available qualities $\{\hat{\mathbf{q}}_j^n\}$ to replace the corresponding values $\{\mathbf{p}^n|_{K_j}\}$. This leads to the SSC algorithm. Indeed, the situation is reminiscent of the Jacobi and Gauss–Seidel methods for solving linear systems.

Algorithm II: Successive Subspace Correction (SSC)

1. Initialization: choose $\mathbf{p}^0$ and select a relaxation parameter $\hat{\alpha} \in (0, 1]$.

2. For $n = 0, 1, \dots$,

- (i) For $j = 1, \dots, M_c$,

Compute

$$(2.16) \quad \hat{\mathbf{q}}_j^n = \arg \min_{\mathbf{v} \in K_j} D\left(\mathbf{v} + \sum_{1 \leq i < j} \mathbf{q}_i^n + \sum_{j < i \leq M_c} \theta_i \mathbf{p}^n\right),$$

and set

$$(2.17) \quad \mathbf{q}_j^n = (1 - \hat{\alpha})\theta_j \mathbf{p}^n + \hat{\alpha} \hat{\mathbf{q}}_j^n.$$

Endfor of j .

(ii) Update

$$(2.18) \quad \mathbf{p}^{n+1} = (1 - \hat{\alpha})\mathbf{p}^n + \hat{\alpha} \sum_{j=1}^{M_c} \hat{\mathbf{q}}_j^n.$$

3. Endfor of n until some stopping rule meets.

Remark 2.1. We can relax the range of the parameter $\hat{\alpha}$ to $(0, 1]$ in Algorithm II. Indeed, from (2.16), we have $\hat{\mathbf{q}}_j^n \in K_j$, so by (2.11), $|\hat{\mathbf{q}}_j^n| \leq \theta_j$. Then we find from (2.18) that if $\hat{\alpha} \in (0, 1]$ and $\mathbf{p}^n \in K$, i.e., $|\mathbf{p}^n| \leq 1$, we have

$$|\mathbf{p}^{n+1}| \leq (1 - \hat{\alpha})|\mathbf{p}^n| + \hat{\alpha} \sum_{j=1}^{M_c} |\hat{\mathbf{q}}_j^n| \leq (1 - \hat{\alpha}) + \hat{\alpha} \sum_{j=1}^{M_c} \theta_j = 1.$$

Note that the property $\mathbf{p}^n \in K$, i.e., $|\mathbf{p}^n| \leq 1$, can be preserved if we choose $\mathbf{p}^0 \in K$.

Remark 2.2. We highlight the significant difference between the above algorithms with the DDMs in [36]:

- (a) The underlying cost functional is neither strongly convex nor strictly convex, so the uniqueness of the minimizer is lacking. Moreover, it appears much more challenging to decompose the involved functional space than the setting in [36] (see the proof of Lemma 3.2 in this paper). In addition, the minimizers to the subproblems are also not unique.
- (b) The treatment of the constraint of the vector-valued dual variable $|\mathbf{p}| \leq 1$ is more difficult than the scalar case in [36].

3. Analysis of the convergence rate. This section is devoted to convergence analysis of the algorithms proposed in the previous section. We first state the main result, and then present some useful lemmas, followed by the proof of the main result. We analyze Algorithm I in detail and sketch just the proof for the convergence of Algorithm II. Indeed, the argument for the analysis of both algorithms is very similar.

3.1. Main results.

Theorem 3.1. Let $\mathbf{p}^*$ be a minimizer of (2.5), i.e.,

$$(3.1) \quad \mathbf{p}^* = \arg \min_{\mathbf{p} \in K} D(\mathbf{p}),$$

and let $\{\mathbf{p}^n\}$ be the sequence generated by Algorithm I with $\alpha \in (0, M_c^{-1}]$. Setting

$$(3.2) \quad u^n := g - \lambda \operatorname{div} \mathbf{p}^n, \quad u^* := g - \lambda \operatorname{div} \mathbf{p}^*,$$

we have

$$(3.3) \quad \|u^n - u^*\|_{L^2(\Omega)} \leq \frac{C}{\sqrt{n}},$$

where

$$(3.4) \quad C = \sqrt{\zeta^0} \left\{ \frac{2}{\alpha} (2M_c + 1)^2 + 8\sqrt{2}C_0\lambda|\Omega|^{\frac{1}{2}}(\zeta^0)^{-\frac{1}{2}} \frac{M_c\sqrt{N_0}}{\delta\sqrt{\alpha}} + \sqrt{2} - 1 \right\},$$

and $\zeta^0 = |D(\mathbf{p}^0) - D(\mathbf{p}^*)|$ is the initial error of the energy functional. Here, the constants λ, C_0, N_0, M_c , and δ are defined as before.

It is evident that the PSC algorithm converges at a rate of $O(n^{-1/2})$ for $\alpha \in (0, M_c^{-1}]$, where M_c is the number of colors (e.g., typically $M_c = 4$). Before we carry out the proof, we remark on the relationship between the convergence rate and the parameters.

Remark 3.1. Observe that the constant C depends on several parameters, including the tunable ones, that is, the overlapping size δ and the relaxation parameter α . It is seen that C increases as δ decreases, and so does α . Accordingly, one can choose $\alpha = M_c^{-1}$ and refer to section 4 for the selection of δ .

It is important to point out that if $\delta \rightarrow 0$ (corresponding to the nonoverlapping DDM), then $C \rightarrow \infty$. Notice that the aforementioned PUFs cannot be carried over to this nonoverlapping case. Therefore, the convergence result herein does not imply that of the nonoverlapping DDM. One can refer to [23] for the convergence analysis of the nonoverlapping DDM.

3.2. Useful lemmas. Hereafter, we simply denote the L^2 -inner product of scalar functions and vector-valued functions by $(\cdot, \cdot)$ and $\langle \cdot, \cdot \rangle$, respectively. Denote the induced norm by $\|\cdot\|$. We drop $d\mathbf{x}$ from the integral notation for simplicity. Define

$$(3.5) \quad |\mathbf{p}|_{1,*} = \|\lambda \operatorname{div} \mathbf{p}\|.$$

Let $D(\mathbf{p})$ be the cost functional defined in (1.4). We have

$$(3.6) \quad D(\mathbf{p}) - D(\mathbf{q}) - D'(\mathbf{q})(\mathbf{p} - \mathbf{q}) = |\mathbf{p} - \mathbf{q}|_{1,*}^2,$$

where $D'(\mathbf{q})$ is the Gâteaux derivative. One verifies readily that

$$(3.7) \quad \langle D'(\mathbf{p}) - D'(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle = 2|\mathbf{p} - \mathbf{q}|_{1,*}^2.$$

Lemma 3.2. *There hold the following two inequalities:*

$$(3.8) \quad \left(\sum_{i=1}^{M_c} |\theta_i \mathbf{p}|_{1,*}^2 \right)^{\frac{1}{2}} \leq \sqrt{2} |\mathbf{p}|_{1,*} + \frac{C_0 \sqrt{2N_0} \lambda}{\delta} \|\mathbf{p}\| \quad \forall \mathbf{p} \in K$$

and

$$(3.9) \quad \sum_{i,j=1}^{M_c} |\langle D'(\mathbf{q}_{ij} + \mathbf{p}_i) - D'(\mathbf{q}_{ij}), \hat{\mathbf{p}}_j \rangle| \leq 2M_c \left(\sum_{i=1}^{M_c} |\mathbf{p}_i|_{1,*}^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{M_c} |\hat{\mathbf{p}}_i|_{1,*}^2 \right)^{\frac{1}{2}}$$

for any $\mathbf{q}_{ij} \in K$, $\mathbf{p}_i \in K_i$, and $\hat{\mathbf{p}}_j \in K_j$ for $1 \leq i, j \leq M_c$.

Proof. We first prove (3.8). Observe that

$$|\theta_i \mathbf{p}|_{1,*}^2 = \lambda^2 \|\operatorname{div}(\theta_i \mathbf{p})\|^2 = \lambda^2 \|\nabla \theta_i \cdot \mathbf{p} + \theta_i \operatorname{div} \mathbf{p}\|^2 \leq 2\lambda^2 \|\nabla \theta_i \cdot \mathbf{p}\|^2 + 2\lambda^2 \|\theta_i \operatorname{div} \mathbf{p}\|^2.$$

Summing up the above equations and using (2.10) yields

$$\begin{aligned} \sum_{i=1}^{M_c} |\theta_i \mathbf{p}|_{1,*}^2 &\leq 2\lambda^2 \int_{\Omega} \left(\sum_{i=1}^{M_c} |\nabla \theta_i|^2 \right) \mathbf{p}^2 + 2\lambda^2 \int_{\Omega} \left(\sum_{i=1}^{M_c} \theta_i^2 \right) (\operatorname{div} \mathbf{p})^2 \\ &\leq 2N_0 \lambda^2 \left(\frac{C_0}{\delta} \right)^2 \|\mathbf{p}\|^2 + 2\lambda^2 \int_{\Omega} \left(\sum_{i=1}^{M_c} \theta_i \right) (\operatorname{div} \mathbf{p})^2 \leq 2N_0 \lambda^2 \left(\frac{C_0}{\delta} \right)^2 \|\mathbf{p}\|^2 + 2|\mathbf{p}|_{1,*}^2. \end{aligned}$$

Thus we have

$$\left(\sum_{i=1}^{M_c} |\theta_i \mathbf{p}|_{1,*}^2 \right)^{\frac{1}{2}} \leq \sqrt{2} |\mathbf{p}|_{1,*} + \frac{\sqrt{2N_0} C_0 \lambda}{\delta} \|\mathbf{p}\|.$$

Now, we turn to the second inequality (3.9). One readily verifies that

$$\langle D'(\mathbf{q}_{ij} + \mathbf{p}_i) - D'(\mathbf{q}_{ij}), \hat{\mathbf{p}}_j \rangle = 2\lambda^2 (\operatorname{div} \mathbf{p}_i, \operatorname{div} \hat{\mathbf{p}}_j).$$

Summing up the above over i, j from 1 to M_c leads to

$$\begin{aligned} \sum_{i,j=1}^{M_c} |\langle D'(\mathbf{q}_{ij} + \mathbf{p}_i) - D'(\mathbf{q}_{i,j}), \hat{\mathbf{p}}_j \rangle| &\leq 2 \sum_{i,j=1}^{M_c} \lambda^2 |(\operatorname{div} \mathbf{p}_i, \operatorname{div} \hat{\mathbf{p}}_j)| \\ &\leq 2 \sum_{i,j=1}^{M_c} |\mathbf{p}_i|_{1,*} |\hat{\mathbf{p}}_j|_{1,*} \leq 2M_c \left(\sum_{i=1}^{M_c} |\mathbf{p}_i|_{1,*}^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{M_c} |\hat{\mathbf{p}}_j|_{1,*}^2 \right)^{\frac{1}{2}}, \end{aligned}$$

where we used the fundamental inequality

$$\sum_{1 \leq i,j \leq M_c} |a_i| |b_j| \leq M_c \left(\sum_{i=1}^{M_c} a_i^2 \right)^{1/2} \left(\sum_{i=1}^{M_c} b_i^2 \right)^{1/2} \quad \forall a_i, b_j \in \mathbb{R}.$$

This ends the proof. ■

Define

$$\begin{aligned} \mathbf{e}_i^n &:= \hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^n, \quad \hat{\mathbf{q}}^n := \sum_{i=1}^{M_c} \hat{\mathbf{q}}_i^n = \mathbf{p}^n + \sum_{i=1}^{M_c} \mathbf{e}_i^n, \\ \mathbf{q}_{\frac{i}{M_c}}^n &:= \left(\sum_{j \neq i} \theta_j \mathbf{p}^n \right) + \hat{\mathbf{q}}_i^n = \mathbf{p}^n + \mathbf{e}_i^n. \end{aligned} \quad (3.10)$$

The first-order optimality condition of (2.15) reads as

$$\langle D'(\mathbf{q}_{\frac{i}{M_c}}^n), \tilde{\mathbf{q}} - \hat{\mathbf{q}}_i^n \rangle \geq 0 \quad \forall \tilde{\mathbf{q}} \in K_i, \quad 1 \leq i \leq M_c,$$

which is equivalent to

$$\int_{\Omega} (\lambda \operatorname{div} \mathbf{q}_{\frac{i}{M_c}}^n - g) \operatorname{div} (\tilde{\mathbf{q}} - \hat{\mathbf{q}}_i^n) \geq 0 \quad \forall \tilde{\mathbf{q}} \in K_i, \quad (3.11)$$

by computing the derivative of the functional $D(\mathbf{p})$ at $\mathbf{p} = \mathbf{q}_{\frac{i}{M_c}}^n$.

The following two lemmas provide error bounds of two consecutive iterations, which play an essential role in the convergence analysis.

Lemma 3.3. *If $\alpha \in (0, M_c^{-1}]$, we have*

$$D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}) \geq \alpha \sum_{i=1}^{M_c} |\mathbf{e}_i^n|_{1,*}^2. \quad (3.12)$$

Proof. It follows from (2.18) and (3.10) that

$$\begin{aligned}
 \mathbf{p}^{n+1} &= \mathbf{p}^n + \alpha \sum_{i=1}^{M_c} (\hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^n) \\
 &= \mathbf{p}^n + \alpha \sum_{i=1}^{M_c} (\mathbf{q}_{\frac{i}{M_c}}^n - \mathbf{p}^n) = (1 - M_c \alpha) \mathbf{p}^n + \alpha \sum_{i=1}^{M_c} \mathbf{q}_{\frac{i}{M_c}}^n.
 \end{aligned}
 \tag{3.13}$$

Thus by using (3.6), (3.13), and the assumption $1 - M_c \alpha \geq 0$, we obtain

$$\begin{aligned}
 D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}) &= D(\mathbf{p}^n) - D\left((1 - M_c \alpha) \mathbf{p}^n + \alpha \sum_{i=1}^{M_c} \mathbf{q}_{\frac{i}{M_c}}^n\right) \\
 &\geq D(\mathbf{p}^n) - \left((1 - M_c \alpha) D(\mathbf{p}^n) + \alpha \sum_{i=1}^{M_c} D(\mathbf{q}_{\frac{i}{M_c}}^n)\right) \\
 &= \alpha \sum_{i=1}^{M_c} \left(D(\mathbf{p}^n) - D(\mathbf{q}_{\frac{i}{M_c}}^n)\right) \\
 &= -\alpha \sum_{i=1}^{M_c} \langle D'(\mathbf{q}_{\frac{i}{M_c}}^n), \mathbf{e}_i^n \rangle + \alpha \sum_{i=1}^{M_c} |\mathbf{e}_i^n|_{1,*}^2,
 \end{aligned}
 \tag{3.14}$$

where the first inequality is deduced from the convexity of $D(\mathbf{p})$, i.e.,

$$D\left(\sum_{i=1}^M \lambda_i \mathbf{p}_i\right) \leq \sum_{i=1}^M \lambda_i D(\mathbf{p}_i)$$

for any integer $M \geq 0$ and for all $\lambda_i \geq 0$ such that $\sum_{i=1}^M \lambda_i = 1$.

Setting $\tilde{\mathbf{q}} := \theta_i \mathbf{p}^n \in K_i$ in (3.11) yields

$$\langle D'(\mathbf{q}_{\frac{i}{M_c}}^n), \mathbf{e}_i^n \rangle \leq 0. \tag{3.15}$$

Therefore, we deduce (3.12) from (3.14) and (3.15). ■

Lemma 3.3 indicates the decay of energy, and the following lemma gives a more precise estimate of the decay rate.

Lemma 3.4. *Given $\mu \in (0, 1)$, we have*

$$D(\mathbf{p}^{n+1}) - D(\mathbf{p}^*) \leq \gamma(D(\mathbf{p}^n) - D(\mathbf{p}^*)) + C_3(D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}))^{\frac{1}{2}}, \tag{3.16}$$

where

$$C_3 := C_3(\mu) = \frac{2C_1\sqrt{\alpha}\mu|\Omega|^{\frac{1}{2}}}{\mu + C_2}, \quad \gamma := \gamma(\mu) = 1 - \frac{\alpha(1 - \mu)\mu}{\mu + C_2}, \tag{3.17}$$

with

$$C_1 = \frac{2\sqrt{2}C_0\lambda M_c\sqrt{N_0}}{\delta}, \quad C_2 = 2M_c(M_c + 1). \tag{3.18}$$

Proof. By the convexity of $D(\cdot)$, we have

$$(3.19) \quad \begin{aligned} D(\mathbf{p}^{n+1}) - D(\mathbf{p}^*) &\leq (1 - \alpha)D(\mathbf{p}^n) + \alpha D(\hat{\mathbf{q}}^n) - D(\mathbf{p}^*) \\ &= (1 - \alpha)(D(\mathbf{p}^n) - D(\mathbf{p}^*)) + \alpha(D(\hat{\mathbf{q}}^n) - D(\mathbf{p}^*)). \end{aligned}$$

Let us estimate $D(\hat{\mathbf{q}}^n) - D(\mathbf{p}^*)$. Introduce

$$(3.20a) \quad \phi_j^i = \mathbf{p}^n + \begin{cases} \sum_{k=i}^{j+i-1} \mathbf{e}_k^n, & j \in \{1, 2, \dots, M_c - i + 1\}, \\ \sum_{k=i}^{M_c} \mathbf{e}_k^n + \sum_{k=1}^{j-M_c+i-1} \mathbf{e}_k^n, & j \in \{M_c - i + 2, M_c - i + 3, \dots, M_c\}, \end{cases}$$

where $\{\mathbf{e}_k^n\}$ are defined in (3.10). Then we obtain

$$(3.21) \quad D'(\mathbf{p}^n + \sum_{j=1}^{M_c} \mathbf{e}_j^n) - D'(\mathbf{p}^n + \mathbf{e}_i^n) = D'(\phi_{M_c}^i) - D'(\phi_1^i) = \sum_{j=2}^{M_c} (D'(\phi_j^i) - D'(\phi_{j-1}^i)).$$

Using (3.9), (3.11), and (3.21), we derive

$$(3.22) \quad \begin{aligned} \langle D'(\hat{\mathbf{q}}^n), \hat{\mathbf{q}}^n - \mathbf{p}^* \rangle &= \sum_{i=1}^{M_c} \langle D'(\hat{\mathbf{q}}^n), \hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^* \rangle \leq \sum_{i=1}^{M_c} \langle D'(\hat{\mathbf{q}}^n) - D'(\mathbf{p}^n + \mathbf{e}_i^n), \hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^* \rangle \\ &= \sum_{i=1, j=2}^{M_c} \langle D'(\phi_j^i) - D'(\phi_{j-1}^i), \hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^* \rangle \leq 2M_c \left(\sum_{j=1}^{M_c} |\mathbf{e}_j^n|_{1,*}^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{M_c} |\hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^*|_{1,*}^2 \right)^{\frac{1}{2}} \\ &= 2M_c \left(\sum_{j=1}^{M_c} |\mathbf{e}_j^n|_{1,*}^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{M_c} |\mathbf{e}_i^n + \theta_i \mathbf{p}^n - \theta_i \mathbf{p}^*|_{1,*}^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Applying the fundamental inequality

$$\left(\sum_{i=1}^{M_c} |a_i + b_i|^2 \right)^{1/2} \leq \left(\sum_{i=1}^{M_c} |a_i|^2 \right)^{1/2} + \left(\sum_{i=1}^{M_c} |b_i|^2 \right)^{1/2},$$

we get

$$\begin{aligned}
 & \langle D'(\hat{\mathbf{q}}^n), \hat{\mathbf{q}}^n - \mathbf{p}^* \rangle \\
 & \leq 2M_c \left(\sum_{j=1}^{M_c} |e_j^n|_{1,*}^2 \right)^{\frac{1}{2}} \left(\left(\sum_{i=1}^{M_c} |e_i^n|_{1,*}^2 \right)^{\frac{1}{2}} + \left(\sum_{i=1}^{M_c} |\theta_i \mathbf{p}^n - \theta_i \mathbf{p}^*|_{1,*}^2 \right)^{\frac{1}{2}} \right) \\
 (3.23) \quad & \stackrel{(3.8)}{\leq} 2M_c \left(\sum_{j=1}^{M_c} |e_j^n|_{1,*}^2 \right)^{\frac{1}{2}} \left(\left(\sum_{i=1}^{M_c} |e_i^n|_{1,*}^2 \right)^{\frac{1}{2}} + \sqrt{2} \|\mathbf{p}^n - \mathbf{p}^*\|_{1,*} + \frac{C_0 \sqrt{2N_0} \lambda}{\delta} \|\mathbf{p}^n - \mathbf{p}^*\| \right) \\
 & = 2M_c \left(\sum_{j=1}^{M_c} |e_j^n|_{1,*}^2 \right) + 2\sqrt{2}M_c \|\mathbf{p}^n - \mathbf{p}^*\|_{1,*} \left(\sum_{j=1}^{M_c} |e_j^n|_{1,*}^2 \right)^{\frac{1}{2}} \\
 & \quad + \frac{2\sqrt{2}C_0 \lambda M_c \sqrt{N_0}}{\delta} \|\mathbf{p}^n - \mathbf{p}^*\| \left(\sum_{j=1}^{M_c} |e_j^n|_{1,*}^2 \right)^{\frac{1}{2}}.
 \end{aligned}$$

Using the first-order optimality condition of (2.5),

$$\langle D'(\mathbf{p}^*), \tilde{\mathbf{q}} - \mathbf{p}^* \rangle \geq 0 \quad \forall \tilde{\mathbf{q}} \in K,$$

and (3.6), we obtain

$$(3.24) \quad D(\mathbf{p}^n) - D(\mathbf{p}^*) \geq \|\mathbf{p}^n - \mathbf{p}^*\|_{1,*}^2.$$

By (3.12), (3.23), and (3.24),

$$\begin{aligned}
 & D(\hat{\mathbf{q}}^n) - D(\mathbf{p}^*) \leq \langle D'(\hat{\mathbf{q}}^n), \hat{\mathbf{q}}^n - \mathbf{p}^* \rangle \\
 & \leq \frac{2M_c}{\alpha} (D(\mathbf{p}^n) - D(\mathbf{p}^{n+1})) + \frac{2\sqrt{2}M_c}{\sqrt{\alpha}} (D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}))^{\frac{1}{2}} (D(\mathbf{p}^n) - D(\mathbf{p}^*))^{\frac{1}{2}} \\
 (3.25) \quad & \quad + \frac{2\sqrt{2}C_0 \lambda M_c \sqrt{N_0}}{\delta \sqrt{\alpha}} (D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}))^{\frac{1}{2}} \|\mathbf{p}^n - \mathbf{p}^*\| \\
 & \leq \frac{2}{\alpha \mu} (M_c^2 + M_c) (D(\mathbf{p}^n) - D(\mathbf{p}^{n+1})) + \mu (D(\mathbf{p}^n) - D(\mathbf{p}^*)) \\
 & \quad + \frac{2\sqrt{2}C_0 \lambda M_c \sqrt{N_0}}{\delta \sqrt{\alpha}} (D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}))^{\frac{1}{2}} \|\mathbf{p}^n - \mathbf{p}^*\|.
 \end{aligned}$$

In view of the inequality

$$ab \leq \frac{a^2}{4\mu} + \mu b^2 \quad \text{for } 0 < \mu < 1,$$

we introduce the constants in (3.18) and deduce that

$$\begin{aligned}
 (3.26) \quad & D(\hat{\mathbf{q}}^n) - D(\mathbf{p}^*) \leq \frac{C_2}{\alpha \mu} (D(\mathbf{p}^n) - D(\mathbf{p}^{n+1})) + \mu (D(\mathbf{p}^n) - D(\mathbf{p}^*)) \\
 & \quad + \frac{C_1}{\sqrt{\alpha}} \|\mathbf{p}^n - \mathbf{p}^*\| (D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}))^{\frac{1}{2}}.
 \end{aligned}$$

Applying (3.26) to (3.19) leads to

$$D(\mathbf{p}^{n+1}) - D(\mathbf{p}^*) \leq \left(1 - \alpha + \alpha\mu + \frac{C_2}{\mu}\right)(D(\mathbf{p}^n) - D(\mathbf{p}^*)) - \frac{C_2}{\mu}(D(\mathbf{p}^{n+1}) - D(\mathbf{p}^*)) \\ + C_1\sqrt{\alpha}\|\mathbf{p}^n - \mathbf{p}^*\|(D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}))^{\frac{1}{2}}.$$

Since $\mathbf{p}^n, \mathbf{p}^* \in K$ such that $\|\mathbf{p}^n - \mathbf{p}^*\| \leq 2|\Omega|^{\frac{1}{2}}$, we derive (3.16) from the above estimates. ■

In order to estimate the exact convergence rate from (3.16), we examine the following general “recursive inequality.”

Lemma 3.5. *If a positive monotone decreasing sequence $\{a^n\}_{n=0}^\infty$ satisfies*

$$(3.27) \quad a^{n+1} \leq \gamma a^n + C_3(a^n - a^{n+1})^{\frac{1}{2}},$$

with $0 \leq \gamma < 1$ and $0 \leq C_3 < \infty$, then

$$(3.28) \quad a^n \leq \frac{a^0}{\tilde{C}(a^0)a^0n + 1},$$

where $\tilde{C}$ is given by

$$\tilde{C}(a^0) = \frac{(1 - \gamma)^2}{2a^0(1 - \gamma)^2 + (\gamma\sqrt{a^0} + C_3)^2}.$$

Proof. We can essentially follow the argument used in [41] to derive the bound. By (3.27),

$$(1 - \gamma)a^{n+1} \leq \gamma(a^n - a^{n+1}) + C_3(a^n - a^{n+1})^{\frac{1}{2}} \\ = (\gamma(a^n - a^{n+1})^{\frac{1}{2}} + C_3)(a^n - a^{n+1})^{\frac{1}{2}} \leq (\gamma\sqrt{a^0} + C_3)(a^n - a^{n+1})^{\frac{1}{2}}.$$

Thus we obtain

$$(a^{n+1})^2 \leq \left(\frac{\gamma\sqrt{a^0} + C_3}{1 - \gamma}\right)^2 (a^n - a^{n+1}).$$

With this, we can obtain the desired estimate from [41, Lemma 3.2] and the proof on page 113 of [41]. ■

With the above preparations, we are now ready to prove the main results.

3.3. Proof of Theorem 3.1. We take two steps to carry out the proof.

Step I. Let $\zeta^n = D(\mathbf{p}^n) - D(\mathbf{p}^*)$. We prove $\zeta^n \rightarrow 0$ as $n \rightarrow \infty$. From Lemma 3.3 and (3.24), we derive

$$(3.29) \quad D(\mathbf{p}^{n+1}) \leq D(\mathbf{p}^n), \quad 0 \leq \zeta^{n+1} \leq \zeta^n.$$

Thus, by Lemma 3.4,

$$(3.30) \quad D(\mathbf{p}^{n+1}) - D(\mathbf{p}^*) \leq \gamma(D(\mathbf{p}^n) - D(\mathbf{p}^*)) + C_3(D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}))^{\frac{1}{2}},$$

that is,

$$(3.31) \quad \zeta^{n+1} \leq \gamma\zeta^n + C_3(\zeta^n - \zeta^{n+1})^{\frac{1}{2}},$$

where $0 \leq \gamma < 1$. The sequence $\{\zeta^n\}_{n=1}^\infty$ is convergent due to the decreasing property (3.29). Thus, taking the limit in (3.31) and using the fact that $0 \leq \gamma < 1$ leads to $\zeta^n \rightarrow 0$ as $n \rightarrow \infty$.

Step II. It follows from (3.31) and Lemma 3.5 that

$$(3.32) \quad \zeta^n \leq \frac{\zeta^0}{\tilde{C}(\zeta^0)\zeta^0_{n+1}} \leq \frac{\zeta^0}{\tilde{C}(\zeta^0)\zeta^0_n} = \frac{C^*}{n},$$

where

$$C^* := C^*(\mu) = \frac{2\zeta^0(1-\gamma)^2 + (\gamma\sqrt{\zeta^0} + C_3)^2}{(1-\gamma)^2}.$$

Recall that $C_3 = C_3(\mu)$ is defined in (3.17). Then, by (3.2), (3.5), and (3.24),

$$(3.33) \quad \|u^n - u^*\|^2 = |\mathbf{p}^n - \mathbf{p}^*|_{1,*}^2 \leq \zeta^n \leq \frac{C^*(\mu)}{n},$$

with $\mu \in (0, 1)$. As the left-hand side of (3.33) is independent of μ , we then look for a tight lower bound of C^* in terms of μ and derive the optimal constant C in (3.4). For this purpose, we search for μ^* such that

$$(3.34) \quad \begin{aligned} \mu^* &= \arg \min_{0 < \mu < 1} C^*(\mu) \\ &= \arg \min_{0 < \mu < 1} \frac{2\zeta^0(1-\gamma)^2 + (\gamma\sqrt{\zeta^0} + C_3)^2}{(1-\gamma)^2} \\ &= \arg \min_{0 < \mu < 1} \left(2\zeta^0 + \left(\frac{\gamma\sqrt{\zeta^0} + C_3}{1-\gamma} \right)^2 \right) = \arg \min_{0 < \mu < 1} \frac{\gamma\sqrt{\zeta^0} + C_3}{1-\gamma} \\ &= \arg \min_{0 < \mu < 1} \frac{\left(1 - \frac{\alpha(1-\mu)\mu}{\mu + C_2} \right) \sqrt{\zeta^0} + \frac{2C_1\sqrt{\alpha}\mu}{\mu + C_2} |\Omega|^{\frac{1}{2}}}{\frac{\alpha(1-\mu)\mu}{\mu + C_2}} \\ &= \arg \min_{0 < \mu < 1} \frac{\alpha\mu + C_2/\mu + (1-\alpha + 2\sqrt{\alpha}C_1|\Omega|^{\frac{1}{2}}(\zeta^0)^{-\frac{1}{2}})}{1-\mu} \\ &= \arg \min_{0 < \mu < 1} \frac{\mu + C_4}{\mu(1-\mu)}, \end{aligned}$$

where we recall that C_1, C_2 are defined in (3.18), and denote

$$C_4 = \frac{C_2}{1 + 2\sqrt{\alpha}C_1|\Omega|^{\frac{1}{2}}(\zeta^0)^{-\frac{1}{2}}}.$$

A direct calculation leads to

$$(3.35) \quad \mu^* = \sqrt{C_4(C_4 + 1)} - C_4 \in (0, 1).$$

Substituting the constants into C^* , we derive

$$\begin{aligned}
 \frac{C^*(\mu^*)}{\zeta^0} &= 2 + \left(\left(\alpha^{-1} + 2 \frac{C_1}{\sqrt{\alpha}} |\Omega|^{\frac{1}{2}} (\zeta^0)^{-\frac{1}{2}} \right) \left(\sqrt{C_4} + \sqrt{C_4 + 1} \right)^2 - 1 \right)^2 \\
 &= 2 + \alpha^{-2} \left(\left(\sqrt{C_2} + \sqrt{1 + 2\sqrt{\alpha} C_1 |\Omega|^{\frac{1}{2}} (\zeta^0)^{-\frac{1}{2}} + C_2} \right)^2 - \alpha \right)^2 \\
 &\leq 2 + \left(\frac{2}{\alpha} \left(1 + 2C_2 + 2\sqrt{\alpha} C_1 |\Omega|^{\frac{1}{2}} (\zeta^0)^{-\frac{1}{2}} \right) - 1 \right)^2 \\
 &= 2 + \left(\frac{2}{\alpha} (1 + 2C_2) + \frac{4C_1}{\sqrt{\alpha}} |\Omega|^{\frac{1}{2}} (\zeta^0)^{-\frac{1}{2}} - 1 \right)^2 \\
 &\leq \left(\frac{2}{\alpha} (1 + 2C_2) + \frac{4C_1}{\sqrt{\alpha}} |\Omega|^{\frac{1}{2}} (\zeta^0)^{-\frac{1}{2}} + \sqrt{2} - 1 \right)^2.
 \end{aligned}$$

The proof of Theorem 3.1 is completed.

3.4. Convergence of Algorithm II. The main result on the convergence rate of Algorithm II is stated as follows.

Theorem 3.6. *Let $\mathbf{p}^*$ be a minimizer of (2.5), and let $\{\mathbf{p}^n\}$ be the sequence generated by Algorithm II with $\hat{\alpha} \in (0, 1]$. Then we have*

$$(3.36) \quad \|u^n - u^*\|_{L^2(\Omega)} \leq \frac{\hat{C}}{\sqrt{n}},$$

where

$$(3.37) \quad \hat{C} = \sqrt{\zeta^0} \left\{ \frac{2}{\hat{\alpha}} (2M_c + 1)^2 + 8\sqrt{2} C_0 \lambda |\Omega|^{\frac{1}{2}} (\zeta^0)^{-\frac{1}{2}} \frac{M_c \sqrt{N_0}}{\delta \sqrt{\hat{\alpha}}} + \sqrt{2} - 1 \right\},$$

and the involved constants are defined as in Theorem 3.1.

Proof. In view of the difference of two algorithms (cf. (2.14) and (2.16)), we redefine the corresponding notation in (3.10) as

$$\mathbf{q}_{\frac{i}{M_c}}^n := \sum_{j \leq i} \mathbf{q}_j^n + \sum_{j > i} \theta_j \mathbf{p}^n, \quad \hat{\mathbf{q}}_{\frac{i}{M_c}}^n := \sum_{j < i} \mathbf{q}_j^n + \hat{\mathbf{q}}_i^n + \sum_{j > i} \theta_j \mathbf{p}^n.$$

Thus, following along the same lines as in the proof of Lemma 3.3, we can derive

$$\begin{aligned}
 D(\mathbf{p}^n) - D(\mathbf{p}^{n+1}) &= \sum_{i=1}^{M_c} \left(D(\mathbf{q}_{\frac{i-1}{M_c}}^n) - D(\mathbf{q}_{\frac{i}{M_c}}^n) \right) \\
 &= \sum_{i=1}^{M_c} \left(D(\mathbf{q}_{\frac{i-1}{M_c}}^n) - D\left((1 - \hat{\alpha}) \mathbf{q}_{\frac{i-1}{M_c}}^n + \hat{\alpha} \hat{\mathbf{q}}_{\frac{i}{M_c}}^n \right) \right) \\
 &\geq \hat{\alpha} \sum_{i=1}^{M_c} \left(D(\mathbf{q}_{\frac{i-1}{M_c}}^n) - D(\hat{\mathbf{q}}_{\frac{i}{M_c}}^n) \right) \geq \hat{\alpha} \sum_{i=1}^{M_c} |\mathbf{e}_i^n|_{1,*}^2.
 \end{aligned}$$

Moreover, we need to modify ϕ_j^i in (3.20a)–(3.20b) as

$$(3.38a) \quad \phi_j^i = \mathbf{p}^n + \begin{cases} \hat{\alpha} \sum_{k=1}^{i-j} \mathbf{e}_k^n + \sum_{k=i-j+1}^i \mathbf{e}_k^n, & j \leq i, \\ \sum_{k=1}^j \mathbf{e}_k^n, & j > i. \end{cases}$$

One readily has

$$(3.39) \quad \phi_j^i - \phi_{j-1}^i = \begin{cases} (1 - \hat{\alpha}) \mathbf{e}_{i-j+1}, & j \leq i, \\ \mathbf{e}_j, & j > i, \end{cases}$$

where in contrast to $\phi_j^i - \phi_{j-1}^i$ defined in (3.20a)–(3.20b), we have the extra factor $1 - \hat{\alpha}$.

Therefore, similar to (3.22), we have

$$(3.40) \quad \begin{aligned} \langle D'(\hat{\mathbf{q}}^n), \hat{\mathbf{q}}^n - \mathbf{p}^* \rangle &= \sum_{i=1}^{M_c} \sum_{j=2}^{M_c} \langle D'(\phi_j^i) - D'(\phi_{j-1}^i), \hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^* \rangle \\ &\leq 2 \sum_{i=1}^{M_c} \sum_{j=2}^{M_c} |\phi_j^i - \phi_{j-1}^i|_{1,*} |\hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^*|_{1,*} \\ &= 2 \sum_{i=1}^{M_c} |\hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^*|_{1,*} \left(\sum_{j=2}^i (1 - \hat{\alpha}) |\mathbf{e}_{i-j+1}|_{1,*} + \sum_{j=i+1}^{M_c} |\mathbf{e}_j|_{1,*} \right), \end{aligned}$$

where the last inequality is deduced from (3.39). Thus, we have

$$(3.41) \quad \begin{aligned} \langle D'(\hat{\mathbf{q}}^n), \hat{\mathbf{q}}^n - \mathbf{p}^* \rangle &\leq 2 \sum_{i=1, j=2}^{M_c} |\mathbf{e}_j|_{1,*} |\hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^*|_{1,*} \\ &\leq 2M_c \left(\sum_{i=1}^{M_c} |\mathbf{e}_j|_{1,*}^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{M_c} |\hat{\mathbf{q}}_i^n - \theta_i \mathbf{p}^*|_{1,*}^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Following along the same lines as the rest of the proof of Lemma 3.4 and the proof of Theorem 3.1, we can obtain the desired estimate. ■

Remark 3.2. We see that $\hat{C}$ decreases as $\hat{\alpha}$ increases. In practice, one can choose $\hat{\alpha} = 1$. We inspect from (3.4) and (3.37) that $\hat{C} < C$ with the choice of the optimal $\hat{\alpha} = 1$ and $\alpha = M_c^{-1}$, respectively, and all other parameters being fixed. This predicts that SSC in Algorithm II is relatively faster than PSC in Algorithm I. Indeed, it is confirmed by Figure 7 in section 4, which also verifies that if $\alpha > M_c^{-1}$, PSC might diverge.

4. Numerical results. In this section, we describe the algorithms for solving the subproblems in Algorithms I–II and provide ample illustrative numerical results to support the theoretical results in the previous section.

4.1. Algorithms for the subproblems (2.15)–(2.16). We follow [6] to formulate the schemes for these subproblems. Denote

$$\mathbf{q}_i^0 := \sum_{j \neq i} \theta_j \mathbf{p}, \quad g_i = \frac{g}{\lambda} - \operatorname{div} \mathbf{q}_i^0.$$

From the KKT condition, we deduce that there exists a Lagrange multiplier $\mu_i \geq 0$ such that

$$(4.1) \quad D'(\hat{\mathbf{q}}_i + \mathbf{q}_i^0) + 2\mu_i \hat{\mathbf{q}}_i = 0.$$

Thus, we have (cf. [6])

$$(4.2) \quad \theta_i \left(-\nabla(\operatorname{div} \hat{\mathbf{q}}_i - g_i) \right) + \left| -\nabla(\operatorname{div} \hat{\mathbf{q}}_i - g_i) \right| \hat{\mathbf{q}}_i = 0.$$

This leads to the iterative scheme

$$(4.3) \quad \hat{\mathbf{q}}_i^{n+1} = \frac{\theta_i \hat{\mathbf{q}}_i^n + \theta_i \tau \left(\nabla(\operatorname{div} \hat{\mathbf{q}}_i^n - g_i) \right)}{\theta_i + \tau \left| \nabla(\operatorname{div} \hat{\mathbf{q}}_i^n - g_i) \right|},$$

where $\tau > 0$ is a tunable parameter.

We discretize the differential operators by finite difference. Consider the reference domain $\Omega = (0, 1)^2$, and lay out the grids as

$$\Omega_h := \{(x_i, y_j) \mid x_i = ih_x, y_j = jh_y, 0 \leq i \leq m, 0 \leq j \leq n\},$$

with the grid size of h_x and h_y . The mesh size is defined by $h = \max\{h_x, h_y\}$. Then the discrete gradient at (x_i, y_j) is given by

$$(\nabla u)_{i,j} = ((\nabla u)_{i,j}^1, (\nabla u)_{i,j}^2),$$

where

$$(\nabla u)_{i,j}^1 = \begin{cases} \frac{1}{h_x}(u_{i+1,j} - u_{i,j}) & \text{if } i < m, \\ 0 & \text{if } i = m, \end{cases} \quad (\nabla u)_{i,j}^2 = \begin{cases} \frac{1}{h_y}(u_{i,j+1} - u_{i,j}) & \text{if } j < n, \\ 0 & \text{if } j = n. \end{cases}$$

Similarly, for a vector-valued function $\mathbf{p}$, we define

$$(\operatorname{div} \mathbf{p})_{i,j} = \begin{cases} \frac{1}{h_x}(p_{i,j}^1 - p_{i-1,j}^1) & \text{if } 0 < i < m, \\ \frac{1}{h_x}p_{i,j}^1 & \text{if } i = 0, \\ -\frac{1}{h_x}p_{i-1,j}^1 & \text{if } i = m, \end{cases} + \begin{cases} \frac{1}{h_y}(p_{i,j}^2 - p_{i,j-1}^2) & \text{if } 0 < j < n, \\ \frac{1}{h_y}p_{i,j}^2 & \text{if } j = 0, \\ -\frac{1}{h_y}p_{i,j-1}^2 & \text{if } j = n. \end{cases}$$

In real computation, we need to discretize the PUFs defined in (2.8)–(2.10). More precisely, we approximate θ_j by its piecewise linear interpolant: $\hat{\theta}_j := I^h(\theta_j)$. To fix the idea, we consider the one-dimensional case with $\Omega = (0, 1)$, which is partitioned into two overlapping intervals

$\Omega_1 = [0, a + \delta]$ and $\Omega_2 = [a - \delta, 1]$, with $0 < a < 1/2$ and $\delta > 0$ being the overlapping sizes. Then we have

$$\tilde{\theta}_1 = \begin{cases} 1, & x \in [0, a - \delta], \\ 1 - \frac{x - a + \delta}{2\delta}, & x \in [a - \delta, a + \delta], \\ 0, & x \in [a + \delta, 1], \end{cases} \quad \tilde{\theta}_2 = \begin{cases} 1, & x \in [a + \delta, 1], \\ \frac{x - a + \delta}{2\delta}, & x \in [a - \delta, a + \delta], \\ 0, & x \in [0, a - \delta]. \end{cases}$$

One readily verifies that $\tilde{\theta}_j$ satisfies (2.8)–(2.9) and infers from [43, p. 57] that it meets (2.10). The two-dimensional PUFs shown in Figure 2 are tensor products of the above one-dimensional functions, corresponding to the partition depicted in Figure 1.

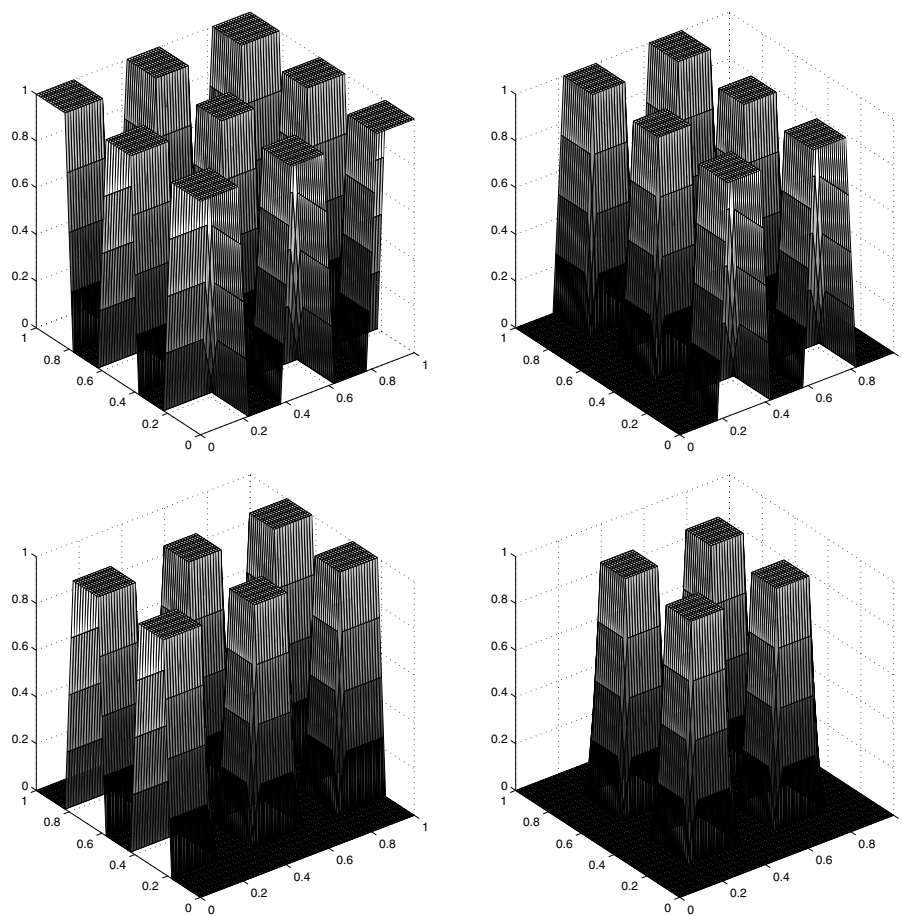


Figure 2. Unit decomposition functions $\{\theta_i\}_{i=1}^4$: the subdomains have four colors, with overlapping size $\delta = 2h$.

4.2. Verification of convergence. We next present numerical results obtained by the proposed algorithms. We consider two images in Figure 3 which are corrupted by the additive

normally distributed noise with mean zero and variance σ . Define the discrete energy as

$$(4.4) \quad E(\mathbf{p}) := \sum_{i=1}^m \sum_{j=1}^n (\lambda \operatorname{div} \mathbf{p} - g)_{i,j}^2,$$

where the image resolution is $m \times n$.

A measurement for the convergence is the normalized residual of the KKT condition of the dual problem similar to (4.2):

$$e_{res}(\mathbf{p}) = \frac{\tilde{e}_{res}(\mathbf{p})}{\tilde{e}_{res}(\mathbf{p}^0)},$$

where the residual is defined as

$$\tilde{e}_{res}(\mathbf{p}) := \left(\sum_{i=1}^m \sum_{j=1}^n |\mathbf{q} - |\mathbf{q}| \mathbf{p}|_{i,j}^2 \right)^{\frac{1}{2}},$$

with $\mathbf{q} = \nabla (\nabla \cdot \mathbf{p} - g/\lambda)$. It could allow the quantification of the convergence independent of the knowledge of the exact minimizer $\mathbf{p}^*$.

Meanwhile, the error is defined as

$$(4.5) \quad e := \left(\sum_{i=1}^m \sum_{j=1}^n (\lambda \operatorname{div}(\mathbf{p}_{DDM} - \mathbf{p}^*))_{i,j}^2 \right)^{\frac{1}{2}} / \left(\sum_{i=1}^m \sum_{j=1}^n (g - \lambda \operatorname{div} \mathbf{p}^*)_{i,j}^2 \right)^{\frac{1}{2}},$$

where $\mathbf{p}_{DDM}$ is computed by DDMs, and $\mathbf{p}^*$ is approximated by dual methods with the Nesterov acceleration technique [4] after 10^7 iterations without DDMs (the residual of the approximated solution is about 10^{-12}). All of the experiments in this paper were implemented in MATLAB on a desktop with Intel Core i3-2120 CPU @ 3.30GHz and 4G RAM.



Figure 3. Left: resolution 128×128 . Right: resolution 256×256 .

Hereafter, we use “*subsize*” to denote the size of subdomain, and we define $Energy := E(\mathbf{p}_{DDM}) - E(\mathbf{p}^*)$ and $Error := e$. We further denote N_{in} as the number of iterations for the solvers of the subproblems in subsection 4.1.

We first show the performance of Algorithm II. The restored images are depicted in Figure 4. The restored images and the differences between the solutions by the proposed DDMs and the exact minimizer are shown in Figure 5 within the first four iterations, and the convergence curve is shown in Figure 6.

The proposed DDMs perform well by observing Figure 4. Inferred from Figure 5, denoising results are satisfactory within the first four iterations of the proposed DDMs. Indeed, the proposed DDMs are convergent by observing Figure 6.

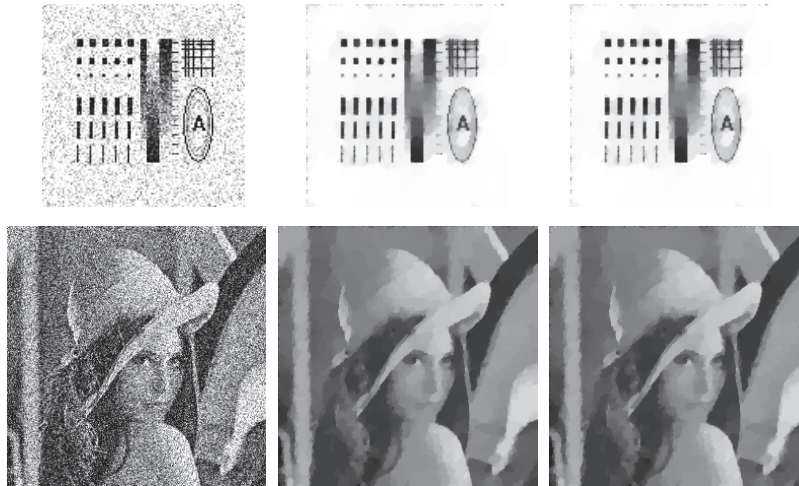


Figure 4. $\hat{\alpha} = 1$, $\tau = \frac{1}{4}$, $subsize = 64$, $\delta = 4$, $\sigma = 50$, $\lambda = 60$, $N_{in} = 500$. From left to right: noised image, denoised image without DDMs, and denoised by DDMs; from top to bottom: resolutions with 128×128 and 256×256 .

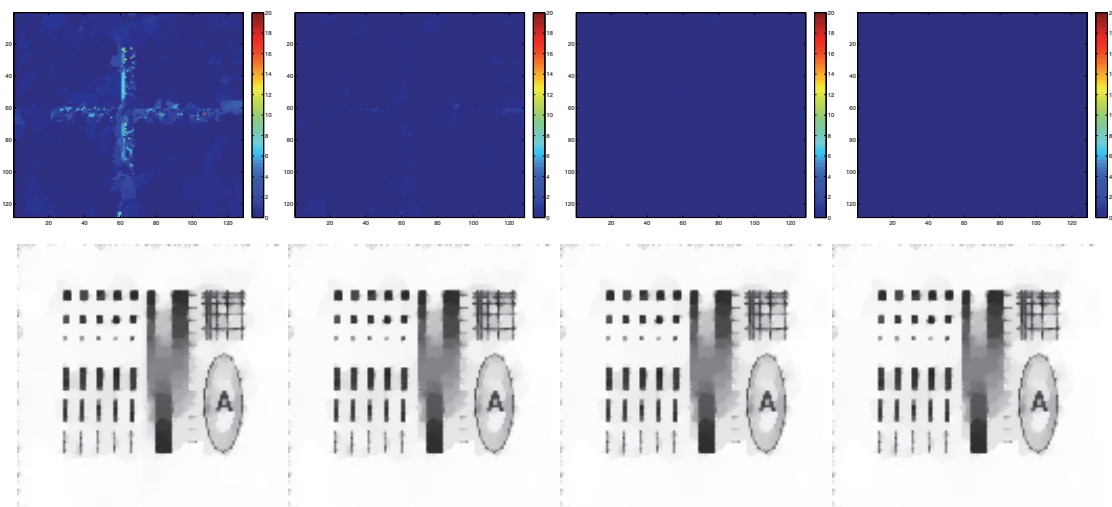


Figure 5. First row: difference between the DDMs ($\delta = 4$, $subsize = 64$, $\sigma = 50$, $\hat{\alpha} = 1$, $N_{in} = 500$, and $\lambda = 60$) and the exact minimizer within the first four iterations from the left to right; second row: restored images by proposed DDMs within the first four iterations.

We have deduced the same convergence rate for both Algorithm I (PSC) and Algorithm II (SSC) for the suitable relaxation parameters α and $\hat{\alpha}$ in the previous section. A numerical test for Figure 3(left) is done to verify the theoretical result; see Figure 7(a), where relaxation

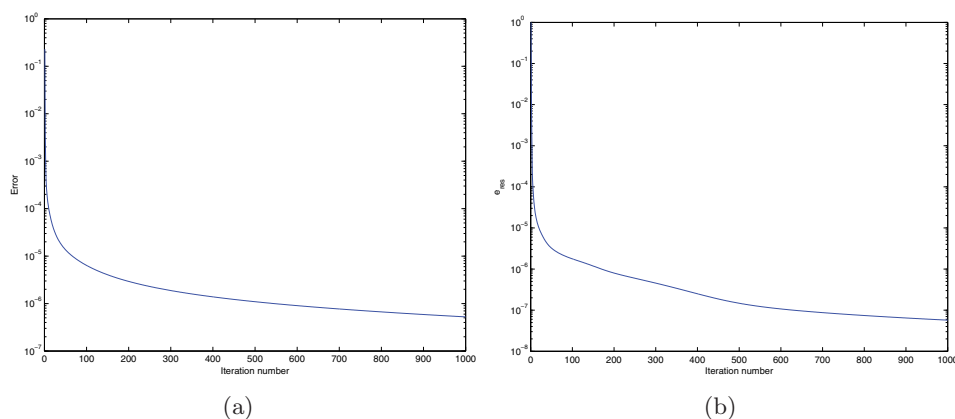


Figure 6. Convergence performance of the proposed DDMs after 1000 iterations. (a) The error is computed as $\|\text{div}(\mathbf{p}^k - \mathbf{p}^*)\| / \|\text{div}\mathbf{p}^*\|$, where $\mathbf{p}^k$ is the iterative solution by the proposed DDMs. (b) The error is computed by the residual e_{res} for the KKT condition for the dual problem.

parameters $\hat{\alpha} = \alpha = 1/4$. The convergence is proved in Theorem 3.1 for Algorithm I if $\alpha \in (0, 1/4]$. By observing Figure 7(a), these two algorithms are convergent, and Algorithm II is a bit faster than Algorithm I when using the same relaxation parameters. Inferred from Figure 7(b), if α is set to be bigger values, i.e., $1/2$ and 1 , Algorithm I might not converge, while Algorithm II is still convergent. In general, one can choose bigger relaxation parameters in Algorithm II leading to faster convergence. To have more insight into the performance of two algorithms, we choose the stopping criterion

$$(4.6) \quad E(\mathbf{p}_{DDM}) \leq E_{obj}$$

for the tests related to Figure 3(left), where E_{obj} is chosen to guarantee the quality of denoising.

We take the optimal relaxation parameters $\alpha = 1/4$ and $\hat{\alpha} = 1$ based on the analysis in Theorems 3.1 and 3.6 and choose $N_{in} = 10$, $subsize = 64$, $\delta = 4$. It takes 1.3 seconds after 53 iterations for Algorithm I, while it takes 0.3 seconds after 13 iterations for Algorithm II, where the energy decays to $E_{obj} = 8.8292 \times 10^8$.

We further compare the performance of both algorithms on the image “Lena” with resolution 2048×2048 with the setting similar to the above, but with $E_{obj} = 4.9301 \times 10^{10}$ and different overlapping sizes and numbers of subdomains. We tabulate the CPU times and iteration numbers of two algorithms in Tables 1 and 2, which show that with the choice of optimal relaxation parameters, Algorithm II is about 4 times faster than Algorithm I. This largely agrees with our analysis.

Hereafter, based on the above discussions, we present only numerical tests obtained by Algorithm II.

4.3. Convergence rate versus the parameters. The images in Figure 3 are tested to show the convergence rate with respect to the parameters. The iterative DDMs stop after a number of iterations (e.g., 1000 iterations) in order to verify the convergence rate. Actually, our numerical tests demonstrate that the noisy image can be restored well within a few iterations (cf. Figure 5).

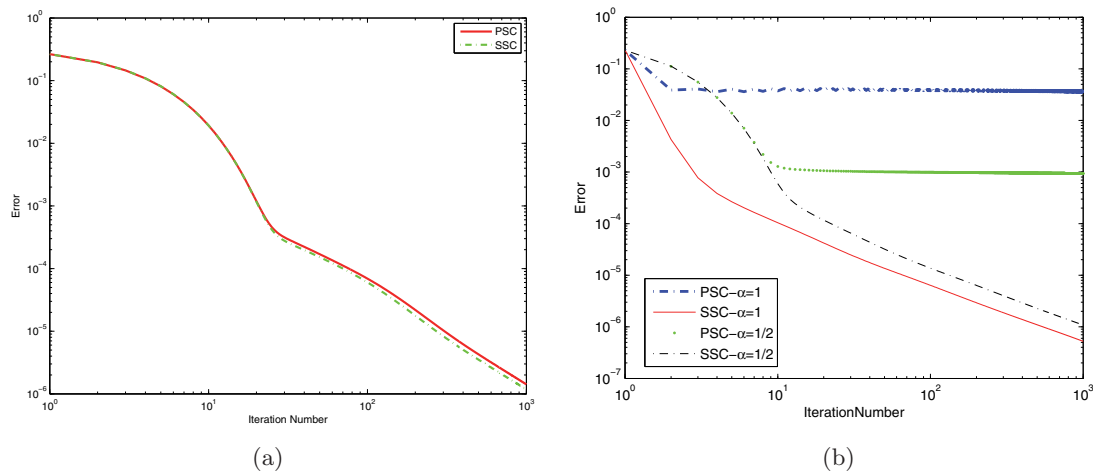


Figure 7. Comparison of PSC and SSC: (a) $\alpha = \hat{\alpha} = 1/4$; (b) different α and $\hat{\alpha}$, $\delta = 4$, $\text{subsize} = 64$, $\sigma = 50$, $\lambda = 60$, $N_{in} = 500$.

Table 1

Elapsed time in seconds and iteration numbers for Algorithms I and II with respect to different overlapping sizes with four subdomains ($\text{subsize} = 1024$, $N_{in} = 10$): “655(79)” in the table means that it takes 655 seconds after 79 iterations by using the given stopping condition.

δ	8	16	32	64	128	256
Algorithm I	655(79)	650(78)	658(76)	717(75)	752(71)	759(63)
Algorithm II	171(20)	164(19)	168(19)	174(18)	175(17)	180(15)

Table 2

Elapsed time in seconds and iteration numbers for Algorithms I and II with respect to different numbers of subdomains N_{sub} ($\text{subsize} = 2048/\sqrt{N_{sub}}$): “658(76)” in the table means that it takes 658 seconds after 76 iterations by using the given stopping condition, and the overlapping size is fixed to be $\delta = 32$, $N_{in} = 10$.

N_{sub}	2^2	4^2	8^2	16^2	24^2	32^2
Algorithm I	658(76)	693(75)	511(61)	503(50)	528(40)	560(32)
Algorithm II	168(19)	169(18)	129(15)	110(11)	120(9)	109(7)

Convergence rates are tested in Figure 8. From the tests we see that the convergence rate of energy decay is between $O(N^{-1})$ and $O(N^{-2})$, and the rate of the dual variable is a bit higher than $O(N^{-1})$, where N is the iterative number. The convergence rates seem a bit higher than our estimated results. More experiments are done to show the convergence with respect to the parameters in the following.

4.3.1. Overlapping size δ and number of subdomains. The first two tests are based on the image with resolution 128×128 , and the last one is based on the image with resolution 256×256 .

- (1) Fixing $\text{subsize} = 64$, we test the convergence rate with respect to the overlapping sizes $\delta = 2, 4, 8$, and 16 (see Figure 9). From the tests, we see that the convergence becomes fast when the overlapping size increases. It implies that the overlapping size δ affects

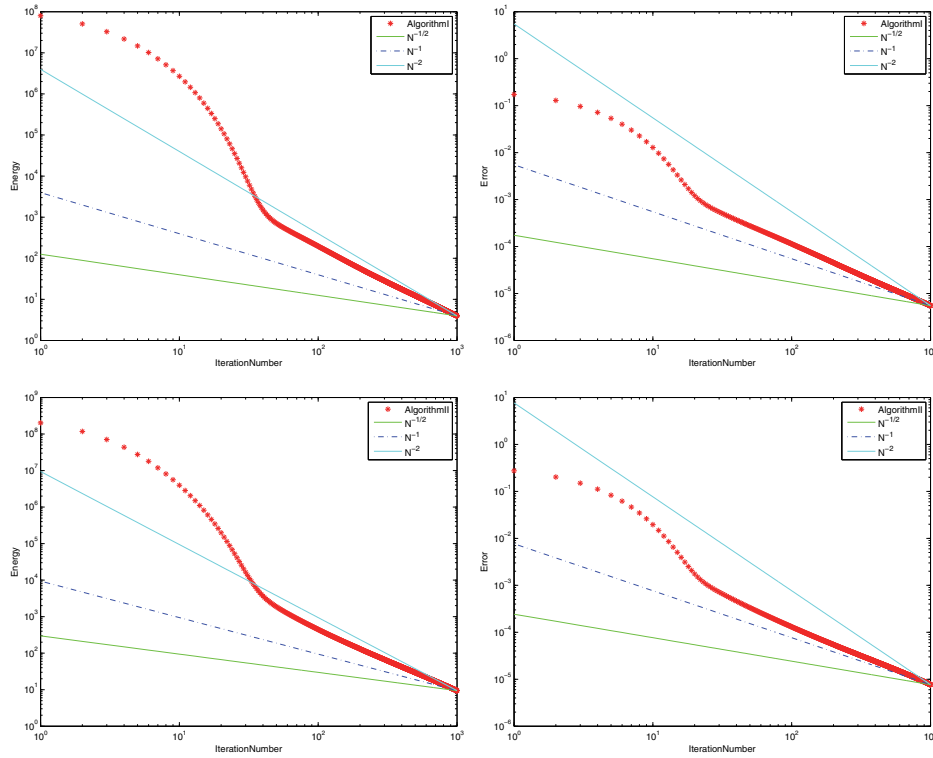


Figure 8. $\hat{\alpha} = \frac{1}{4}$, $\tau = \frac{1}{4}$, $subsize = 64$, $\delta = 4$, $\sigma = 50$, $\lambda = 60$, $N_{in} = 500$. From left to right: energy and error; resolutions from top to bottom: 128×128 and 256×256 .

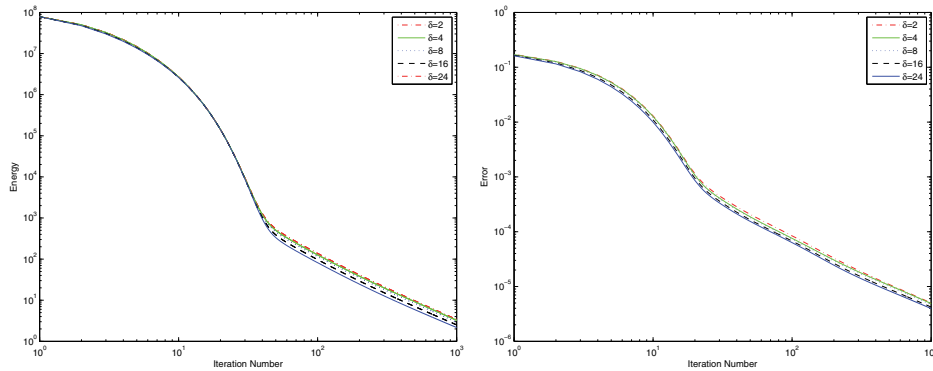


Figure 9. $\hat{\alpha} = \frac{1}{4}$, $\tau = \frac{1}{4}$, $subsize = 64$, $\sigma = 50$, $\lambda = 60$, $N_{in} = 500$. From left to right: energy and error.

only the coefficient of the convergence order.

- (2) Fixing $\delta = 4$, we test the convergence rate with respect to the number of subdomains: $subsize = 8, 16, 32$, and 64 (see Figure 10). The convergence becomes fast when the size of the subdomain decreases. We use color techniques that fix $M_c = 4$. Then the smaller the size of the subdomain, the relatively larger the overlapping size.
- (3) Fixing $subsize = 128$ when the testing image is Lena (256×256), we test the conver-

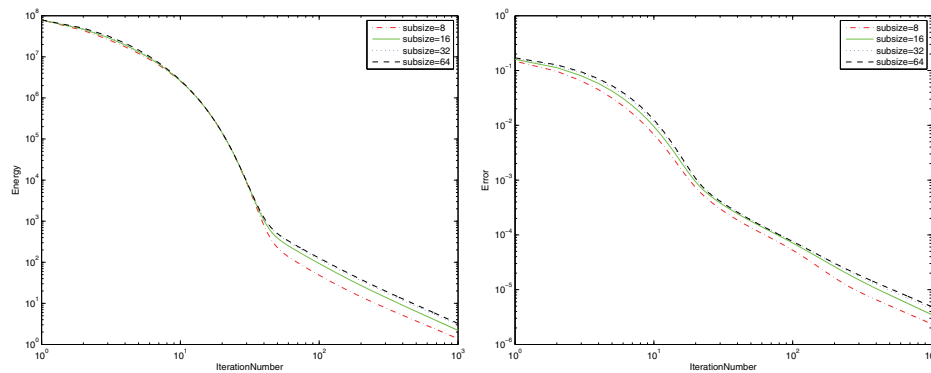


Figure 10. $\hat{\alpha} = \frac{1}{4}$, $\tau = \frac{1}{4}$, $\sigma = 50$, $\lambda = 60$, $N_{in} = 500$. From left to right: energy and error.

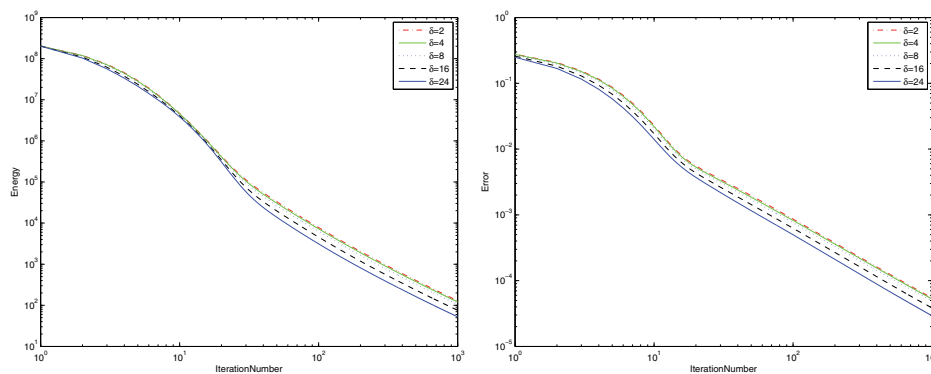


Figure 11. $\hat{\alpha} = \frac{1}{4}$, $\tau = \frac{1}{4}$, $subsize = 128$, $\sigma = 50$, $\lambda = 60$, $N_{in} = 500$. From left to right: energy and error.

gence rate with respect to the overlapping sizes $\delta = 2, 4, 8, 16$, and 24 . See Figure 11.

The convergence varies more obviously as the δ varies than that of Figure 9.

We also give the iteration numbers to reach the same error for SSC with respect to different overlapping sizes δ ; see Table 3. One can see that to reach the same error, fewer iterations are needed if using bigger overlapping size δ . It seems that the coarse grid correction might enhance the convergence behavior.

Table 3

Iteration numbers for SSC with different overlapping sizes.

δ	2	4	8	16	24	Error
Resolution 128×128	1000	981	937	857	806	5.0111×10^{-6}
Resolution 256×256	1000	968	909	733	587	9.7559×10^{-6}

4.3.2. Relaxation parameter $\hat{\alpha}$. We test four different values by setting $\hat{\alpha} = 1/8, 1/4, 1/2$, and 1 . See the results in Figure 12. The closer to 1 the parameter $\hat{\alpha}$ is, the faster Algorithm II is. This is consistent with Theorem 3.6, where the constant $\hat{C}$ is proved to be a decreasing function with respect to the parameter $\hat{\alpha}$. In real computation, $\hat{\alpha}$ shall be set to be 1 in order to gain better performance.

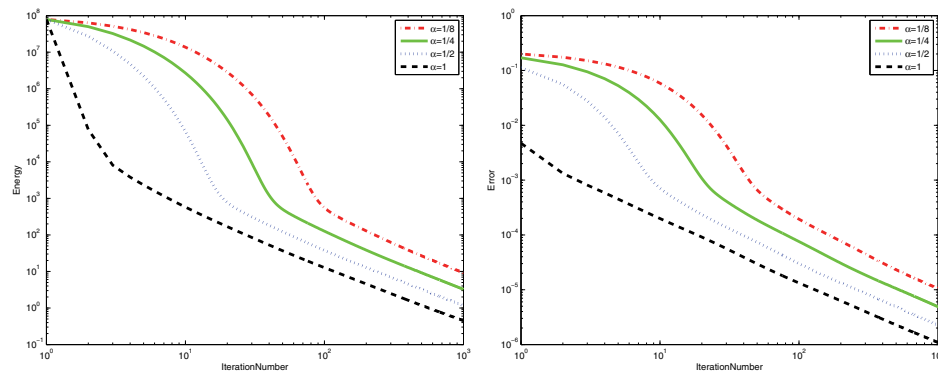


Figure 12. $\tau = \frac{1}{4}$, $\sigma = 50$, $subsize = 64$, $\delta = 4$, $\lambda = 60$, $N_{in} = 500$. From left to right: energy and error.

Remark 4.1. From the tests we see that this algorithm converges fast in the beginning but stagnates after some iterations with the order $O(N^{-\sigma})$. This may be deduced by (3.31). For the initial iterations the errors are governed by

$$\zeta^{n+1} \leq \gamma \zeta^n,$$

which implies the convergence rate is linear. Then after many iterations the errors are governed by

$$\zeta^{n+1} \leq C_3 \left(\zeta^n - \zeta^{n+1} \right)^{\frac{1}{2}},$$

which implies the convergence rate is $O(N^{-1})$.

4.4. Convergence with respect to iteration number for subproblems. We show the convergence with respect to N_{in} . The image in the left of Figure 3 is tested, and $subsize = 64$, $\delta = 4$. We give the iteration number ($NumIter$) and elapsed time to reach the same error for SSC with respect to different N_{in} in Table 4. By observing the table, one can infer that the SSC with fewer inner iterations N_{in} needs more iterations $NumIter$ in order to reduce to the same energy value, and the algorithm becomes slower with respect to the iteration number. However, the elapsed time decreases per iteration, since fewer inner iteration numbers N_{in} are adopted. Thus, SSC takes less time to decrease to the same energy value by setting $N_{in} = 10$. Using smaller N_{in} will greatly increase the iteration number $NumIter$, which leads to increase of the elapsed time.

Table 4

Iteration number ($NumIter$) and elapsed time in seconds for SSC with different N_{in} , $subsize = 64$, and $\delta = 4$.

N_{in}	1	5	10	100	200	500	Energy value
$NumIter$	1000	206	107	26	24	24	8.8282×10^8
Time (s)	12.99	5.09	3.88	6.91	12.07	30.62	

4.5. Comparison with gradient projection algorithm [6]. We show the convergence behavior comparing with the gradient projection (GP) algorithm by Chambolle [6] with respect

to the iteration count and the elapsed time. One can readily infer from the error in Figure 13(left) that our proposed DDMs converge much faster per iteration. Inferred from Figure 13(right), the DDMs induce the smaller error per second. Moreover, it is not necessary to run the DDMs for so many iterations, and several iterations are usually enough (see Figure 5).

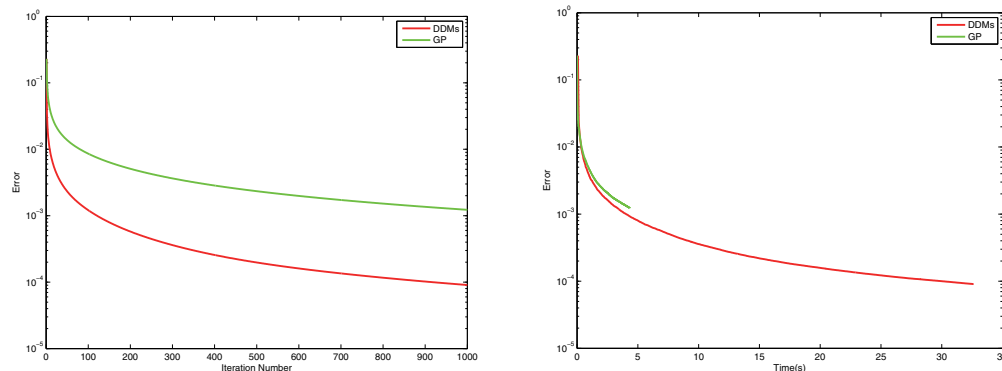


Figure 13. Comparison of error and elapsed time for the GP method and the proposed DDMs (Algorithm II, $N_{in} = 10$, $subsize = 64$, $\delta = 4$).

5. Conclusion. We analyzed and implemented overlapping DDMs for the ROF model based on a dual formulation. We provided an efficient technique to decompose the global constraint of the vector-valued dual variable. We obtained the convergence rate for both the SSC and the PSC algorithms and presented ample numerical results in support of the theoretical results. The proposed DDMs are viable for large scale image processing. To further enhance the performance of the methods, we intend to integrate them with the coarse mesh correction, which will be reported in a future work.

Acknowledgment. We thank the reviewers for the valuable comments, which greatly helped us to improve the paper.

REFERENCES

- [1] R. ACAR AND C. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1230.
- [2] B. APPLETON AND H. TALBOT, *Globally optimal geodesic active contours*, J. Math. Imaging Vis., 23 (2005), pp. 67–86.
- [3] K. ARROW, L. HURWICZ, AND H. UZAWA, *Studies in Linear and Non-linear Programming*, with contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, and R. M. Solow, Stanford Mathematical Studies in the Social Sciences II, Stanford University Press, Stanford, CA, 1958.
- [4] A. BECK AND M. TEBoulLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
- [5] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [6] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., 20 (2004), pp. 89–97.
- [7] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145.

- [8] T. CHAN AND T. MATHEW, *Domain decomposition algorithms*, in Acta Numerica 1994, Vol. 3, Cambridge University Press, Cambridge, UK, 1994, pp. 61–143.
- [9] H. CHANG, X. ZHANG, X. C. TAI, AND D. YANG, *Domain decomposition methods for nonlocal total variation image restoration*, J. Sci. Comput., 60 (2014), pp. 79–100.
- [10] Y. DONG, M. HINTERMÜLLER, AND M. NERI, *An efficient primal-dual method for L^1 TV image restoration*, SIAM J. Imaging Sci., 2 (2009), pp. 1168–1189.
- [11] M. DRYJA AND O. B. WIDLUND, *Towards a unified theory of domain decomposition algorithms for elliptic problems*, in Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, SIAM, Philadelphia, 1990, pp. 3–21.
- [12] Y. DUAN AND X. C. TAI, *Domain decomposition methods with graph cuts algorithms for total variation minimization*, Adv. Comput. Math., 36 (2012), pp. 175–199.
- [13] E. ESSER, X. ZHANG, AND T. F. CHAN, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci., 3 (2010), pp. 1015–1046.
- [14] D. FIRSOV AND S. H. LUI, *Domain decomposition methods in image denoising using Gaussian curvature*, J. Comput. Appl. Math., 193 (2006), pp. 460–473.
- [15] M. FORNASIER, A. LANGER, AND C.-B. SCHÖNLIEB, *Domain decomposition methods for compressed sensing*, in Proceedings of the 8th International Conference on Sampling Theory and Applications (SampTA'09), Marseilles, France, 2009.
- [16] M. FORNASIER, A. LANGER, AND C. SCHÖNLIEB, *A convergent overlapping domain decomposition method for total variation minimization*, Numer. Math., 116 (2010), pp. 645–685.
- [17] 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.
- [18] R. GLOWINSKI AND P. LE TALLEC, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM, Philadelphia, 1989.
- [19] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for L_1 -regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.
- [20] M. GRIEBEL AND P. OSWALD, *On the abstract theory of additive and multiplicative Schwarz algorithms*, Numer. Math., 70 (1995), pp. 163–180.
- [21] M. HINTERMÜLLER AND K. KUNISCH, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM J. Appl. Math., 64 (2004), pp. 1311–1333.
- [22] M. HINTERMÜLLER AND A. LANGER, *Subspace correction methods for a class of nonsmooth and nonadditive convex variational problems with mixed L^1/L^2 data-fidelity in image processing*, SIAM J. Imaging Sci., 6 (2013), pp. 2134–2173.
- [23] M. HINTERMÜLLER AND A. LANGER, *Non-overlapping domain decomposition methods for dual total variation based image denoising*, J. Sci. Comput., 62 (2015), pp. 456–481.
- [24] 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.
- [25] A. LANGER, S. OSHER, AND C. SCHÖNLIEB, *Bregmanized domain decomposition for image restoration*, J. Sci. Comput., 54 (2013), pp. 549–576.
- [26] P.-L. LIONS, *On the Schwarz alternating method I*, in First International Symposium on Domain Decomposition Methods for Partial Differential Equations, T. Chan, R. Glowinski, J. Périaux, and O. B. Widlund, eds., SIAM, Philadelphia, 1988, pp. 1–42.
- [27] P.-L. LIONS, *On the Schwarz alternating method II*, in Domain Decomposition Methods, SIAM, Philadelphia, 1989, pp. 47–70.
- [28] P.-L. LIONS, *On the Schwarz alternating method III*, in Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, SIAM, Philadelphia, 1990, pp. 202–223.
- [29] A. MARQUINA AND S. OSHER, *Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal*, SIAM J. Sci. Comput., 22 (2000), pp. 387–405.
- [30] P. MONK, *Finite Element Methods for Maxwell's Equations*, Oxford University Press, New York, 2003.
- [31] J. MÜLLER, *Parallel Total Variation Minimization*, Diploma Thesis, University of Münster, Münster, Germany, 2008.
- [32] 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.

- [33] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [34] O. SCHERZER, ED., *Handbook of Mathematical Methods in Imaging*, Springer, New York, 2011.
- [35] B. SMITH, P. BJØRSTAD, AND W. GROPP, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, Cambridge, UK, 1996.
- [36] X. C. TAI, *Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities*, Numer. Math., 93 (2003), pp. 755–786.
- [37] X. C. TAI AND Y. DUAN, *Domain decomposition methods with graph cuts algorithms for image segmentation*, Int. J. Numer. Anal. Model., 8 (2011), pp. 137–155.
- [38] X. C. TAI AND M. ESPEDAL, *Applications of a space decomposition method to linear and nonlinear elliptic problems*, Numer. Methods Partial Differential Equations, 14 (1998), pp. 717–737.
- [39] X.-C. TAI AND M. ESPEDAL, *Rate of convergence of some space decomposition methods for linear and nonlinear problems*, SIAM J. Numer. Anal., 35 (1998), pp. 1558–1570.
- [40] X. C. TAI AND P. TSENG, *Convergence rate analysis of an asynchronous space decomposition method for convex minimization*, Math. Comp., 71 (2002), pp. 1105–1136.
- [41] X. C. TAI AND J. XU, *Global and uniform convergence of subspace correction methods for some convex optimization problems*, Math. Comp., 71 (2002), pp. 105–124.
- [42] X. C. TAI AND C. WU, *Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model*, in Scale Space and Variational Methods in Computer Vision, Springer, Berlin, Heidelberg, 2009, pp. 502–513.
- [43] A. TOSELLI AND O. WIDLUND, *Domain Decomposition Methods—Algorithm and Theory*, Springer-Verlag, Berlin, Heidelberg, 2005.
- [44] C. VOGEL, *A multigrid method for total variation-based image denoising*, in Computation and Control IV, Progr. Systems Control Theory 20, Birkhäuser, Boston, 1995, pp. 323–331.
- [45] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
- [46] C. R. VOGEL AND M. E. OMAN, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.
- [47] C. VOGEL AND M. OMAN, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Trans. Image Process., 7 (1998), pp. 813–824.
- [48] Y. WANG, J. YANG, W. YIN, AND Y. ZHANG, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1 (2008), pp. 248–272.
- [49] Y. WANG, W. YIN, AND Y. ZHANG, *A Fast Algorithm for Image Deblurring with Total Variation Regularization*, Rice University CAAM Technical Report TR07-10, Rice University, Houston, TX, 2007.
- [50] C. WU AND X.-C. TAI, *Augmented Lagrangian method, dual methods, and split Bregman iterations for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3 (2010), pp. 300–339.
- [51] J. XU, *Iterative methods by space decomposition and subspace correction*, SIAM Rev., 34 (1992), pp. 581–613.
- [52] J. XU, H. CHANG, AND J. QIN, *Domain decomposition method for image deblurring*, J. Comput. Appl. Math., 271 (2014), pp. 401–414.
- [53] J. XU, X. C. TAI, AND L. L. WANG, *A two-level domain decomposition method for image restoration*, Inverse Probl. Imaging, 4 (2010), pp. 523–545.
- [54] M. ZHU AND T. CHAN, *An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration*, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.