J. Korean Soc. Ind. Appl. Math. Vol.24, No.2, 161–197, 2020 http://doi.org/10.12941/jksiam.2020.24.161

RECENT ADVANCES IN DOMAIN DECOMPOSITION METHODS FOR TOTAL VARIATION MINIMIZATION

CHANG-OCK LEE[1†] AND JONGHO PARK[1]

[1]DEPARTMENT OF MATHEMATICAL SCIENCES, KAIST, DAEJEON, 34141, REPUBLIC OF KOREA
Email address: [†]colee@kaist.edu

ABSTRACT. Total variation minimization is standard in mathematical imaging and there have been numerous researches over the last decades. In order to process large-scale images in real-time, it is essential to design parallel algorithms that utilize distributed memory computers efficiently. The aim of this paper is to illustrate recent advances of domain decomposition methods for total variation minimization as parallel algorithms. Domain decomposition methods are suitable for parallel computation since they solve a large-scale problem by dividing it into smaller problems and treating them in parallel, and they already have been widely used in structural mechanics. Differently from problems arising in structural mechanics, energy functionals of total variation minimization problems are in general nonlinear, nonsmooth, and nonseparable. Hence, designing efficient domain decomposition methods for total variation minimization is a quite challenging issue. We describe various existing approaches on domain decomposition methods for total variation minimization in a unified view. We address how the direction of research on the subject has changed over the past few years, and suggest several interesting topics for further research.

1. INTRODUCTION

Along with the development of high-performance computing, there has been arisen a natural question for how to design numerical methods which make efficient use of distributed memory computers. Domain decomposition method (DDM) is a clever answer to that question, especially for numerical methods for solving large-scale algebraic systems arisen in discretization of boundary value problems. In DDM, we solve a boundary value problem by decomposing its domain into smaller subdomains to produce smaller problems on subdomains. Since such smaller problems in subdomains can be treated in parallel, DDM is suitable for making efficient use of distributed memory computers. In addition, DDMs can be regarded as preconditioned methods [1, Ch. 1]. Due to such advantages, they have been successfully developed over past decades. For more detailed explanation for DDMs, we refer readers to monographs [1, 2].

Received by the editors March 26 2020; Accepted June 1 2020; Published online June 25 2020.
2010 *Mathematics Subject Classification.* 65N55, 65K10, 65Y05, 68U10.
Key words and phrases. domain decomposition methods, total variation, mathematical imaging, parallel computation.
[†] Corresponding author.

There are two major approaches to design DDMs: *Schwarz methods* and *iterative substructuring methods*. In Schwarz methods, we decompose the solution space into a sum of its subspaces, where each subspace consists of functions whose supports are contained in the subdomain. Then, the next iterate is obtained by adding local corrections obtained from the residual of the previous iterate. Such correction procedure is done by either parallel (*additive Schwarz*) or successive (*multiplicative Schwarz*) manner. Well-known block relaxation schemes such as the block Jacobi method and the block Gauss–Seidel method belong to this class. There is a vast literature on the Schwarz methods; see [1, 3] and references therein.

On the other hand, in iterative substructuring methods, we solve a system which is composed of local problems and a global interface problem. The global interface problem comes from the interface condition such as the continuity of a solution on the subdomain interfaces. The interface condition is enforced by either *primal* or *dual* sense. In primal methods, the interface degrees of freedom (dofs) are shared by adjacent subdomains; for example, see [4, 5]. Alternatively, the interface condition is enforced by the method of Lagrange multipliers in dual methods [6, 7, 8]. Due to fast speed and scalability, iterative substructuring methods have been widely used for numerical solutions of linear elliptic PDEs.

The purpose of this paper is to review the recent notable advances in DDMs for total variation minimization. The general total variation minimization problem is given by

$$\min_{u\in BV(\Omega)} \{F(u) + TV(u)\}, \tag{1.1}$$

where $\Omega \subset \mathbb{R}^2$ is a bounded rectangular domain, F: $BV(\Omega) \to \overline{\mathbb{R}}$ is a convex functional, $TV(u)$ is the total variation of u, which is defined as

$$TV(u) = \sup\left\{ \int_\Omega u \operatorname{div} \mathbf{p}\, dx : \mathbf{p} \in (C_0^1(\Omega))^2,\ |\mathbf{p}(x)| \le 1,\ x \in \Omega \right\},$$

and $BV(\Omega)$ is the space of L^1 functions with finite total variation. Note that $TV(u)$ agrees with $\int_\Omega |\nabla u|\, dx$ for sufficiently smooth u. After a pioneering work of Rudin et al. [9], total variation minimization has been used as a standard regularizer in mathematical imaging. One of the most fundamental example of (1.1) is the Rudin–Osher–Fatemi (ROF) model [9] for image denoising:

$$\min_{u\in BV(\Omega)} \left\{ \frac{\alpha}{2} \int_\Omega (u-f)^2\, dx + TV(u) \right\}. \tag{1.2}$$

Here, $f \in L^2(\Omega)$ is an observed noisy image and α is a positive weight parameter. The fidelity term $\frac{\alpha}{2}\int_\Omega (u-f)^2\, dx$ in (1.2) measures a distance between the given data f and a solution u. On the other hand, the regularizer $TV(u)$ enforces some regularity of the solution. Thanks to the anisotropic diffusion property of the total variation term, the model (1.2) effectively removes Gaussian noise while preserving edges and discontinuities of the image [10]. There have been numerous approaches to solve (1.2): projected dual gradient method [11], alternating minimization algorithm [12], FISTA [13], split Bregman algorithm [14], and primal-dual algorithms [15, 16].

Even though DDMs for linear elliptic PDEs have been successfully developed over the past decades, there have been relatively modest achievements in total variation minimization problems due to their own difficulties. The total variation term in (1.1) is nonseparable, i.e., it cannot be expressed as the sum of local energy functionals in subdomains. Consequently, existing convergence results on DDMs for nonlinear problems such as [17, 18, 19] cannot be applied to (1.1), or even to the easier case (1.2). Even more, the solution space $BV(\Omega)$ allows discontinuities of a solution on the subdomain interfaces, so that it is difficult to impose appropriate boundary conditions to local problems in subdomains. This makes design of iterative substructuring methods hard.

Overcoming such difficulties, there have been several fruitful research on DDMs for total variation minimization. Schwarz methods for (1.2) were first considered in [20, 21]. Then they were generalized to (1.1) with L^2-L^1 mixed fidelity in [22]. Due to the nature of total variation minimization, it is difficult to impose the interface boundary condition to local problems of Schwarz methods. In [23, 24, 25], various methodologies for the efficient boundary process of local problems were proposed. While the convergence to a global minimizer is not guaranteed for Schwarz methods for *primal* total variation minimization [26], it was shown in [27, 28] that Schwarz methods for *dual* total variation minimization converge sublinearly to a global minimizer; we will state the precise definitions of primal and dual problems in Section 4. Incorporating with FISTA [13] acceleration, an accelerated block method for the dual ROF model was proposed in [29]. Iterative substructuring methods for the dual ROF model were considered in [30], based on finite element discretizations using the pixel grid. Then the primal-dual iterative substructuring method in [30] was generalized to general dual total variation minimization in [31].

This paper is intended to give readers a unified view on the above-mentioned works. In order to accomplish our goal, we present a concise summary on those works with comments. In Section 2, we review how the total variation minimization works in the field of mathematical imaging; we especially focus on the ROF model. Several early works on Schwarz methods for primal total variation minimization are summarized in Section 3. Fenchel–Rockafellar duality, one of the most important notions for designing DDMs for dual total variation minimization, is introduced in Section 4 with related Schwarz methods. Iterative substructuring methods for total variation minimization are presented in Section 5 with their basic finite element discretizations. Numerical comparison of DDMs for the ROF model is presented in Section 6. We conclude the paper in Section 7. In addition, we summarize several important algorithms for convex optimization in Appendix A.

We will not cover some interesting works which are not directly related to the above outline. There are several 'non-standard' DDMs for (1.1) in the sense that they do not generalize the classical methods developed for linear elliptic problems [1]. A DDM based on the method of Lagrange multipliers was proposed in [32] to solve the convex Chan–Vese model [33], a convex relaxation of the Chan–Vese model [34] for image segmentation. Even though it looks similar to the existing FETI method [7] at a first glance, it was proven in [35] that it corresponds to a particular overlapping domain decomposition in the continuous setting on the contrary to the claim that the FETI method corresponds to a nonoverlapping domain decomposition [32].

In addition, it was generalized to general total variation minimization (1.1) in [35]. In [36], DDMs using dual conversion for the TV-L^1 model [37] were proposed. Dual conversion is a special instance of Fenchel–Rockafellar duality that yields a saddle point problem with parallel structure which is equivalent to the original problem. We will not discuss in details on those methods in this paper.

We note that throughout this paper, the meaning of notations vary from section to section.

2. Total variation minimization in mathematical imaging

A digital image consists of a number of rows and columns of pixels. Each of them has an integer in $[0, 255]$, which is called the intensity. We denote black and white colors by 0 and 255, respectively. Usually we scale the intensity to the range $[0, 1]$ for the convenience of calculation. Then a grayscale digital image u of resolution $M \times N$ can be considered as a function

$$u\colon \{1, 2, \dots, M\} \times \{1, 2, \dots, N\} \to [0, 1].$$

In this perspective, it is convenient to treat a digital image as an L^∞ function satisfying $0 \le u \le 1$ a.e. on a bounded rectangular domain $\Omega \subset \mathbb{R}^2$.

Let X and Y be suitable Banach spaces for digital images. We consider the image restoration problem: for a given deteriorated image $f \in Y$, a solution $u \in X$ is a clean image without corruption obtained from f. We model the image restoration problem as the following linear inverse problem:

$$f = \mathcal{A}u + \eta, \tag{2.1}$$

where $\mathcal{A}\colon X \to Y$ is a linear operator (for instance, convolution by a blur kernel) and η is an unknown noise. Clearly, the problem (2.1) is ill-posed and we are only able to obtain an approximated solution of (2.1). An effective way to overcome the ill-posedness of (2.1) is to add a regularization term; we solve an optimization problem of the form

$$\min_{u \in X} \{F(\mathcal{A}u; f) + R(u)\}, \tag{2.2}$$

where $F(\mathcal{A}u; f)$ is a *fidelity* term that measures a distance between f and $\mathcal{A}u$ in a certain sense. The term $R(u)$ is a *regularizer* that resolves the ill-posedness of the problem (2.1) by imposing some desirable properties to the solution.

2.1. The Rudin–Osher–Fatemi model.

As an example, we consider additive Gaussian noise removal. The "Peppers" image displayed in Fig. 2.1 is corrupted by additive Gaussian noise with variance $\sigma^2 = 0.1$. We want to obtain a clean image which looks similar to the original image. The noise added to f makes the image "rough", so that the gradient of f becomes high. Thus, it is natural to consider the following energy functional with the Tikhonov regularizer [38]:

$$\min_{u \in H^1(\Omega)} \left\{ \frac{\alpha}{2} \int_\Omega (u - f)^2 \, dx + \frac{1}{2} \int_\Omega |\nabla u|^2 \, dx \right\}, \tag{2.3}$$

(a) Original image (b) With Gaussian noise (PSNR: 19.11) (c) Model (2.3), $\alpha = 0.1$ (PSNR: 22.38) (d) Model (1.2), $\alpha = 10$ (PSNR: 24.55)

FIGURE 2.1. Results of image denoising by the Tikhonov regularization (2.3) and the ROF model (1.2).

where $\alpha > 0$ is a weight parameter. However, (2.3) does not work well for image denoising. The Euler–Lagrange equation for (2.3) reads as

$$\begin{aligned} -\frac{1}{\alpha}\Delta u + u &= f \quad \text{in } \Omega, \\ \frac{\partial u}{\partial \mathbf{n}} &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{2.4}$$

where $\mathbf{n}$ is the unit outer normal to $\partial\Omega$. Due to the Laplacian term, a solution of (2.4) is highly diffusive. Consequently, edges or texture of the image disappear due to oversmoothing. Figure 2.1(c) shows the result of (2.3) applied to Fig. 2.1(b) with $\alpha = 0.1$. Therefore, we should use other regularizers which decrease ∇u but not penalize the discontinuity of u too much; see [39, Sect. 3.2] for a study on regularizers related to ∇u.

A suitable choice for the regularizer is the total variation, which results in the total variation-regularized problem (1.2). Note that the space $BV(\Omega)$ is a Banach space equipped with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + TV(u)$ and $BV(\Omega) \subset L^2(\Omega)$ [40]. Since $TV(u) = \int_\Omega |\nabla u|\, dx$ for smooth u, the total variation regularization penalizes large ∇u. In order to investigate the effect of the total variation, we consider the following formal Euler–Lagrange equation for (1.2):

$$\begin{aligned} -\frac{1}{\alpha}\,\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right) + u &= f \quad \text{in } \Omega, \\ \frac{\partial u}{\partial \mathbf{n}} &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.5}$$

One can observe that (2.5) has an anisotropic diffusion property so that it preserves edges and discontinuities in images. More precisely, by a suitable change of variables, the governing equation of (2.5) is transformed into

$$-\frac{1}{\alpha}\frac{u_{TT}}{|\nabla u|} + u = f \quad \text{in } \Omega,$$

where the subscript T denotes tangential differentiation to level lines of u [39]. It means that diffusion occurs along the tangential direction of level lines so that edges of the image are preserved. On the other hand, (2.5) can be viewed as the stationary equation for a mean curvature flow; see [41] for details. Figure 2.1(d) shows the result of (1.2) with $\alpha = 10$. Differently from Fig. 2.1(c), it does not smooth edges of the image while it successfully removes Gaussian noise.

Remark 2.1. In Fig. 2.1, we use the *peak-signal-to-noise ratio* (PSNR) as a measurement of the quality of denoising, which is defined by

$$\mathrm{PSNR}(u) = 10 \log_{10} \left(\frac{\mathrm{MAX}^2 \cdot |\Omega|}{\|u - u_{\mathrm{orig}}\|_2^2} \right),$$

where MAX is the maximum possible pixel value of the image ($\mathrm{MAX} = 1$ in our case), u_{orig} is the original clean image, and u is a denoised image. In general, a higher PSNR indicates that the image restoration is of higher quality.

In order to solve (1.2) numerically, we need to discretize it. A natural discretization of (1.2) can be done by the first-order finite difference approximation. Since the image domain Ω of the resolution $M \times N$ is given by

$$\Omega = \{1, 2, \ldots, M\} \times \{1, 2, \ldots, N\},$$

we regard each pixel in an image as a discrete point. We define the function space V as the set of all functions from Ω into $\mathbb{R}$ and W as the set of all functions from Ω into $\mathbb{R}^2$ equipped with the usual Euclidean inner products

$$\begin{aligned} \langle u, v \rangle_V &= \sum_{(i,j) \in \Omega} u_{ij} v_{ij}, && u, v \in V, \\ \langle \mathbf{p}, \mathbf{q} \rangle_W &= \sum_{(i,j) \in \Omega} \left(p_{ij}^1 q_{ij}^1 + p_{ij}^2 q_{ij}^2 \right), && \mathbf{p} = (p^1, p^2), \mathbf{q} = (q^1, q^2) \in W, \end{aligned}$$

respectively. The discrete gradient operator $\nabla\colon V \to W$ is defined by using forward finite differences with the homogeneous Neumann boundary condition:

$$\begin{aligned} (\nabla u)_{ij}^1 &= \begin{cases} u_{i+1,j} - u_{ij} & \text{if} \quad i = 1, \ldots, M-1, \\ 0 & \text{if} \quad i = M, \end{cases} \\ (\nabla u)_{ij}^2 &= \begin{cases} u_{i,j+1} - u_{ij} & \text{if} \quad j = 1, \ldots, N-1, \\ 0 & \text{if} \quad j = N. \end{cases} \end{aligned}$$

It is natural to define the discrete total variation as the 1-norm of ∇u for $u \in V$:

$$\|\nabla u\|_1 = \sum_{(i,j) \in \Omega} |(\nabla u)_{ij}|_p,$$

where

$$|(\nabla u)_{ij}|_p = \left[\left| (\nabla u)_{ij}^1 \right|^p + \left| (\nabla u)_{ij}^2 \right|^p \right]^{\frac{1}{p}}$$

for $p = 1, 2$. The case $p = 1$ is called the *anisotropic* total variation and the case $p = 2$ is called the *isotropic* total variation. While some properties in the continuous setting such as the coarea formula are inherited to only the anisotropic case, the isotropic case has less dependency on the grid in terms of image restoration [42]. In this paper, we deal with the case $p = 2$ unless otherwise stated.

In conclusion, the following discrete ROF model is constructed:

$$\min_{u\in V}\left\{E_{\mathrm{p}}^{\mathrm{FD}}(u) := \frac{\alpha}{2}\|u-f\|_2^2 + \|\nabla u\|_1\right\}. \tag{2.6}$$

One important question is which algorithm should be used to solve the discretized problem (2.6). Several notable solvers for convex optimization, that can be adopted to solve (2.6) are summarized in Appendix A. For example, Algorithm A.4 with $O(1/n^2)$ energy convergence rate can be utilized to solve (2.6).

2.2. **General total variation minimization.** Now we consider more general cases; we revisit (2.2) with $R(u) = TV(u)$ in the discrete setting:

$$\min_{u\in V}\left\{F(Au; f) + \|\nabla u\|_1\right\}, \tag{2.7}$$

where A: $V \to V$ is a linear operator depending on the type of the problem and $f \in V$ is a given corrupted image. A number of problems in mathematical imaging can be represented in the form of (2.7).

Recall that the ROF model (2.6) solves the image denoising problem. In order to deal with other problems of the form (2.1), it is natural to consider the following generalization of (2.6):

$$\min_{u\in V}\left\{\frac{\alpha}{2}\|Au-f\|_2^2 + \|\nabla u\|_1\right\}. \tag{2.8}$$

We simply set $A = I$ for the denoising problem, so that (2.8) reduces to (2.6). For the inpainting problem, A is a block-diagonal matrix whose diagonal entries are 0 for the pixels in the inpainting region, and 1 elsewhere. For the deconvolution problem, A is defined by the matrix convolution $Au = k_A * u$ with a convolution kernel k_A. That is, each pixel value of Au is a linear combination of nearby pixel values of u. For instance, if k_A is the 3×3 Gaussian blur kernel

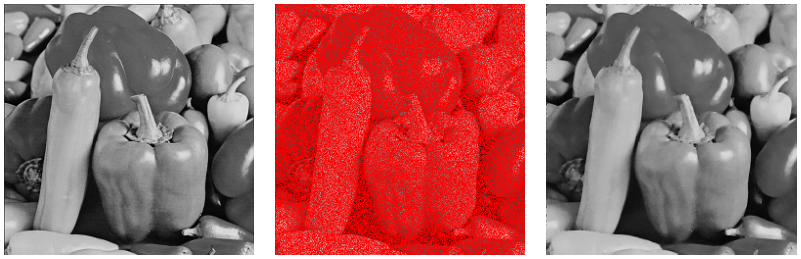
$$k_A = \frac{1}{16}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix},$$

then each entry of Au is given by

$$\begin{aligned}&(Au)_{ij}\\ &= \frac{1}{16}(u_{i-1,j-1}+2u_{i-1,j}+u_{i-1,j+1}+2u_{i,j-1}+4u_{ij}+2u_{i,j+1}+u_{i+1,j-1}+2u_{i+1,j}+u_{i+1,j+1}),\end{aligned}$$

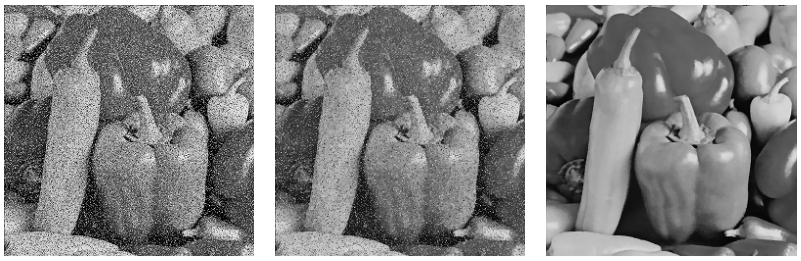
with the convention $u_{0j} = u_{M+1,j} = u_{i0} = u_{i,N+1} = 0$ for $1 \le i \le M$ and $1 \le j \le N$.

To implement a numerical algorithm for (2.8) efficiently, it is important to get a sharp bound for the operator norm of A. For the inpainting problem, it is trivial that $\|A\| = 1$. The following proposition gives a bound for the deblurring problem [36, Proposition 4].



(a) Original image (b) 80% missing data (c) $\alpha = 1000$ (PSNR: 28.22)

FIGURE 2.2. Results of image inpainting by the TV-L^2 model (2.8).



(a) Noisy image (PSNR: 12.29) (b) ROF, $\alpha = 10$ (PSNR: 17.44) (c) TV-L^1, $\alpha = 1$ (PSNR: 29.65)

FIGURE 2.3. Results of image denoising by the ROF model (2.6) and the TV-L^1 model (2.9).

Proposition 2.2. *Let A: $V \to V$ be the linear operator defined by the matrix convolution $Au = k_A * u$ with a deblurring kernel k_A. Then, the operator norm of A has a bound $\|A\| \leq 1$.*

Figure 2.2 shows the numerical result of image inpainting using the model (2.8). In the experiment, 80% of the image pixels are missing in the input f. As shown in Fig. 2.2(c), (2.8) gives a quite good restoration result; the missing part in Fig. 2.2(b) is recovered naturally while the edges of the original image are preserved.

A well-known drawback of the ROF model is loss of contrast [10, 43]. Using L^1 fidelity instead of L^2 fidelity can be a remedy to this problem [37, 44]:

$$\min_{u \in V} \{\alpha \|u - f\|_1 + \|\nabla u\|_1\}. \tag{2.9}$$

It is more difficult to deal with the TV-L^1 model (2.9) than the ROF model (2.6) by two reasons. First, the energy functional in (2.9) is convex but not strictly convex, so that solutions are not unique in general. In addition, both the fidelity term and total variation term lack smoothness. For more mathematical properties of the TV-L^1 model, readers may refer [37].

In the case of image restoration, (2.9) has remarkable characteristics. Unlike (2.6), the fidelity term in (2.9) is expressed in the L^1 norm, so that it penalizes less outliers than the L^2 norm. Thus, it performs better than the ROF model for removing noise with many outliers [44], for instance, salt-and-pepper noise. In Fig. 2.3(a), the input image is corrupted by 20% salt-and-pepper noise. Figures 2.3(b)-2.3(c) show the results of image denoising by (2.6) and (2.9), respectively. It is clear that the TV-L^1 model performs better than the ROF model in this case. Moreover, it was proven in [37] that the TV-L^1 model is contrast invariant.

Similarly to (2.8), one can generalize (2.9) to solve various imaging problems other than image denoising as follows:

$$\min_{u\in V} \{\alpha\|Au - f\|_1 + \|\nabla u\|_1\}. \tag{2.10}$$

Various numerical results for problems of the form (2.10) can be found in, e.g., [42].

3. SCHWARZ METHODS FOR THE PRIMAL PROBLEM

A natural attempt to design DDMs for total variation minimization is to apply existing Schwarz methods directly. In this section, we provide a survey on how the concept of Schwarz methods developed for problems in structural mechanics has been applied to the total variation minimization (2.7). The first target is (2.6), which is a typical case of (2.7). We call (2.6) the *primal* ROF model; a reason will be explained in Section 4. In [20, 21], overlapping and nonoverlapping Schwarz methods for (2.6) were proposed, respectively. Here, we present a unified framework for them.

For simplicity, we consider the two-subdomain case. Let Ω_1 and Ω_2 be subsets of Ω defined by

$$\Omega_1 = \{1,\dots,M\} \times \{1,\dots,N_1\},$$
$$\Omega_2 = \{1,\dots,M\} \times \{N_2,\dots,N\}.$$

The above decomposition is overlapping if $N_1 > N_2 - 1$ and is nonoverlapping if $N_1 = N_2 - 1$. The subdomain interface $\Gamma_{\rm p}$ is defined by

$$\Gamma_{\rm p} = \{1,\dots,M\} \times \{N_2 - 1\}.$$

For $k = 1, 2$, the local function space V_k is defined as

$$V_k = \{u \in V : \mathrm{supp}(u) \subset \Omega_k\}.$$

The restriction operator R_k: $V \to V_k$ is naturally defined. Note that the adjoint R_k^*: $V_k \to V$ becomes the natural prolongation operator. Then it is straightforward that

$$V = R_1^* V_1 + R_2^* V_2.$$

We construct a discrete partition of unity $\{D_1, D_2\}$ which satisfies

$$R_1^* D_1 R_1 + R_2^* D_2 R_2 = I.$$

The operator D_k is diagonal in its matrix representation so that its computational cost is almost negligible. We write

$$\widetilde{R}_k = D_k R_k.$$

Note that D_k is the identity operator on V_k if and only if the domain decomposition is nonoverlapping. Multiplicative and additive Schwarz methods for (2.6) are presented in Algorithms 3.1 and 3.2, respectively.

Algorithm 3.1 Multiplicative Schwarz method for the primal ROF model (2.6)

Let $u^{(0)} \in V$.
for $n = 0, 1, 2, \ldots$
$u_1^{(n+1)} = \arg\min_{u_1 \in V_1} E_{\mathrm{p}}^{\mathrm{FD}}(R_1^* u_1 + R_2^* \widetilde{R}_2 u^{(n)})$
$u_2^{(n+1)} = \arg\min_{u_2 \in V_2} E_{\mathrm{p}}^{\mathrm{FD}}(R_1^* u_1^{(n+1)} + R_2^* u_2)$
$u^{(n+1)} = R_1^* u_1^{(n+1)} + R_2^* u_2^{(n+1)}$
end

Algorithm 3.2 Additive Schwarz method for the primal ROF model (2.6)

Choose $\tau \in (0, 1/2]$. Let $u^{(0)} \in V$.
for $n = 0, 1, 2, \ldots$
$u_k^{(n+1)} = \arg\min_{u_k \in V_k} E_{\mathrm{p}}^{\mathrm{FD}}(R_k^* u_k + R_{3-k}^* \widetilde{R}_{3-k} u^{(n)}), \quad k = 1, 2$
$u^{(n+1)} = (1 - \tau) u^{(n)} + \tau(R_1^* u_1^{(n+1)} + R_2^* u_2^{(n+1)})$
end

One can show the energy decreasing property of Algorithms 3.1 and 3.2 without any difficulty; see, e.g., [25].

Theorem 3.1. *Let $\{u^{(n)}\}$ be the sequence generated by either Algorithm 3.1 or 3.2. Then we have*

$$E_{\mathrm{p}}^{\mathrm{FD}}(u^{(n+1)}) \le E_{\mathrm{p}}^{\mathrm{FD}}(u^{(n)}), \quad n = 0, 1, 2, \ldots.$$

3.1. **Boundary processing for local problems.** Differently from the case of linear elliptic problems, it is not quite straightforward to deal with the boundary condition on the subdomain interfaces of local problems. To be more precise, we observe that local problems of Algorithms 3.1 and 3.2 in Ω_1 have the following general form:

$$\min_{u_1 \in V_1} \left\{ \frac{\alpha}{2} \|u_1 - R_1 f\|_{2,V_1}^2 + \|\nabla_1 u_1 + \mathbf{g}_1\|_{1,W_1} \right\} \tag{3.1}$$

for some $\mathbf{g}_1 \in W_1$, where

$$W_1 = \{\mathbf{p} \in W : \mathrm{supp}(\mathbf{p}) \subset \Omega_1\}$$

and the local gradient operator $\nabla_1 \colon V_1 \to W_1$ in Ω_1 is given by

$$(\nabla_1 u)^1_{ij} = \begin{cases} u_{i+1,j} - u_{ij} & \text{if} \quad i = 1, \dots, M-1, \\ 0 & \text{if} \quad i = M, \end{cases}$$

$$(\nabla_1 u)^2_{ij} = \begin{cases} u_{i,j+1} - u_{ij} & \text{if} \quad j = 1, \dots, N_1 - 1, \\ -u_{ij} & \text{if} \quad j = N_1. \end{cases}$$

Note that ∇_1 has the Dirichlet boundary condition on the interface $\partial\Omega_1 \setminus \partial\Omega$, while ∇ has the Neumann boundary condition on the whole boundary of Ω. Due to such a difference, it is difficult to adopt existing solvers for the ROF model directly to (3.1). Moreover, one should clarify how to deal with $\mathbf{g}_1$ in (3.1).

On the other hand, local problems in Ω_2 have the general form

$$\min_{u_2 \in V_2} \left\{ \frac{\alpha}{2} \|u_2 - R_2 f\|^2_{2,V_2} + \|\nabla_2 u_2 + \mathbf{g}_2\|_{1,W_2} \right\}$$

for some $\mathbf{g}_2 \in W_2$, where

$$W_2 = \{\mathbf{p} \in W : \operatorname{supp}(\mathbf{p}) \subset \Omega_2 \cup \Gamma_{\mathrm{p}}\}$$

and the local gradient operator $\nabla_2 \colon V_2 \to W_2$ in Ω_2 is given by

$$(\nabla_2 u)^1_{ij} = \begin{cases} u_{i+1,j} - u_{ij} & \text{if} \quad i = 1, \dots, M-1, \\ 0 & \text{if} \quad i = M, \end{cases}$$

$$(\nabla_2 u)^2_{ij} = \begin{cases} u_{i,j+1} & \text{if} \quad j = N_2 - 1 \\ u_{i,j+1} - u_{ij} & \text{if} \quad j = N_2, \dots, N-1, \\ 0 & \text{if} \quad j = N. \end{cases}$$

One can readily observe that local problems in Ω_2 bear similar difficulties as ones in Ω_1.

Various methodologies to efficiently impose the interface boundary condition to local problems had been proposed: oblique thresholding [20, 21], graph cuts [23], Bregmanized operator splitting [24], augmented Lagrangian method [22], and primal-dual stitching [25]. We will not present these methods in detail in this paper because it was unfortunately shown in [26, Example 6.1] that Algorithms 3.1 and 3.2 may not converge to the global minimizer of (2.6). In the following, we present a counterexample for convergence.

Example 3.2 (counterexample). In (2.6), let $f = 1$ in Ω and $u^{(0)} = 0$. It is clear that the solution of (2.6) is given by $u^* = 1$. However, one can show that $u^{(n)} = 0$ for all $n \geq 0$ if $\alpha > 0$ is sufficiently small. Therefore, the sequence $\{u^{(n)}\}$ does not converge to u^* as n tends to infinity.

4. SCHWARZ METHODS FOR THE DUAL PROBLEM

As we observed in Section 3, the primal ROF model is not adequate for applying the Schwarz methods. Such inadequacy is due to the nonseparability of the total variation; it cannot be expressed as the sum of the local energy functionals in the subdomains. Since Schwarz methods work for separable nonsmooth terms (see [18, Eq. (7)]), it would be good if we were able to obtain an equivalent formulation to (2.6), whose nonsmooth term is separable. Fortunately, the

notion of *Fenchel–Rockafellar duality* gives us such an equivalent formulation, so called the dual ROF model. Recently, there have been a few notable works on the Schwarz methods for the dual ROF model [27, 28, 29]. In this section, we review those works in a unified view.

4.1. **Fenchel–Rockafellar duality.** For the sake of completeness, we present key features of the Fenchel–Rockafellar duality; see also [42, 45]. Let X and Y be Euclidean spaces. For a convex functional $F\colon X \to \overline{\mathbb{R}}$, the *effective domain* $\operatorname{dom} F$ of F is defined by

$$\operatorname{dom} F = \{x \in X : F(x) < \infty\}.$$

For a convex subset C of X, the *relative interior* of C, denoted by $\operatorname{ri} C$, is defined as the interior of C when C is regarded as a subset of its affine hull, i.e.,

$$\operatorname{ri} C = \{x \in C : \exists \epsilon > 0,\ B_\epsilon(x) \cap \operatorname{aff} C \subset C\},$$

where $B_\epsilon(x)$ is the open ball of radius ϵ centered at x and $\operatorname{aff} C$ is the affine hull of C. The *Legendre–Fenchel conjugate* $F^*\colon X \to \overline{\mathbb{R}}$ of $F\colon X \to \overline{\mathbb{R}}$ is defined by

$$F^*(x) = \sup_{\bar{x} \in X} \{\langle x, \bar{x}\rangle - F(\bar{x})\}, \quad x \in X.$$

We consider the minimization problem

$$\min_{x \in X} \{F(Kx) + G(x)\}, \tag{4.1}$$

where $K\colon X \to Y$ is a linear operator and $F\colon Y \to \overline{\mathbb{R}}$, $G\colon X \to \overline{\mathbb{R}}$ are proper, convex, lower semicontinuous functionals. We assume that a solution $x^* \in X$ of (4.1) exists. Under appropriate conditions, one can have several equivalent problems to (4.1) with different structures.

Proposition 4.1. *In* (4.1)*, assume that there exists* $x_0 \in X$ *such that* $Kx_0 \in \operatorname{ri}\operatorname{dom} F$ *and* $x_0 \in \operatorname{ri}\operatorname{dom} G$*. Then it satisfies that*

$$\begin{aligned}\min_{x \in X} \{F(Kx) + G(x)\} &= \min_{x \in X} \sup_{y \in Y} \{\langle Kx, y\rangle + G(x) - F^*(y)\} \\ &= \max_{y \in Y} \inf_{x \in X} \{\langle Kx, y\rangle + G(x) - F^*(y)\} \qquad (4.2a) \\ &= \max_{y \in Y} \{-F^*(y) - G(-K^*y)\}. \qquad (4.2b)\end{aligned}$$

In addition, (4.2b) *admits a solution* $y^* \in Y$*, and* (x^*, y^*) *is a solution of the saddle point problem* (4.2a)*.*

Proof. See [45, Corollary 31.2.1]. □

We call the problem (4.1) the *primal formulation*. The saddle point problem (4.2a) is called the *primal-dual formulation* of (4.1), and the maximization problem in (4.2b) is called the *dual formulation* of (4.1). It is obvious that (4.2b) is the same as the minimization problem

$$\min_{y \in Y} \{F^*(y) + G(-K^*y)\}.$$

We will solve either (4.2a) or (4.2b) instead of (4.1) because in many cases, either the primal-dual or dual formulation is easier to solve than the primal formulation.

Remark 4.2. Under suitable conditions on F and G, Proposition 4.1 holds even if either X or Y is infinite-dimensional; one may refer [46].

Now, we apply the Fenchel–Rockafellar duality to the general total variation minimization problem (1.1). By the definition of $TV(u)$, (1.1) is rewritten as

$$\min_{u\in BV(\Omega)} \sup_{\substack{\mathbf{p}\in(C_0^1(\Omega))^2,\\ |\mathbf{p}|\le 1}} \left\{ -\langle u, \operatorname{div}\mathbf{p}\rangle + F(u)\right\},$$

where $|\mathbf{p}|$ denotes the pointwise absolute value of $\mathbf{p}$, i.e., $|\mathbf{p}(x)| = \sqrt{(p^1(x))^2 + (p^2(x))^2}$ for $\mathbf{p} = (p^1, p^2)$ and $x \in \Omega$. Formally, we have a dual counterpart of the above formulation

$$\min_{\substack{\mathbf{p}\in H_0(\operatorname{div};\Omega)\\ |\mathbf{p}|\le 1 \text{ a.e.}}} F^*(\operatorname{div}\mathbf{p}), \tag{4.3}$$

where $H_0(\operatorname{div};\Omega)$ is a Hilbert space defined by

$$\begin{aligned} H(\operatorname{div};\Omega) &= \left\{\mathbf{p}\in (L^2(\Omega))^2 : \operatorname{div}\mathbf{p}\in L^2(\Omega)\right\},\\ H_0(\operatorname{div};\Omega) &= \left\{\mathbf{p}\in H(\operatorname{div};\Omega) : \mathbf{p}\cdot\mathbf{n} = 0 \text{ on } \partial\Omega\right\}. \end{aligned}$$

Note that, in (4.3), we use the appropriate Hilbert space $H_0(\operatorname{div};\Omega)$ instead of $(C_0^1(\Omega))^2$ which is not complete. Rigorous statements on the relation between (1.1) and (4.3) can be found in [47, 48].

In particular, the dual ROF model is written as

$$\min_{\substack{\mathbf{p}\in H_0(\operatorname{div};\Omega)\\ |\mathbf{p}|\le 1 \text{ a.e.}}} \frac{1}{2}\int_\Omega (\operatorname{div}\mathbf{p} + \alpha f)^2\, dx. \tag{4.4}$$

The dual problem (4.4) is a constrained optimization problem with the pointwise constraint $|\mathbf{p}| \le 1$ a.e. As Schwarz methods work well for variational inequalities with pointwise constraints [49, 50], we may expect that (4.4) is more suitable for Schwarz methods than the primal problem.

Similarly to (2.6), one can discretize (4.4) by the first-order finite difference approximation. The discrete divergence operator $\operatorname{div}\colon W \to V$ is defined as the minus adjoint of ∇, i.e.,

$$\begin{aligned} (\operatorname{div}\mathbf{p})_{ij} &= \begin{cases} p^1_{ij} & \text{if} \quad i = 1,\\ p^1_{ij} - p^1_{i-1,j} & \text{if} \quad i = 2,\dots,M-1,\\ -p^1_{i-1,j} & \text{if} \quad i = M \end{cases}\\ &\quad + \begin{cases} p^2_{ij} & \text{if} \quad j = 1,\\ p^2_{ij} - p^2_{i,j-1} & \text{if} \quad j = 2,\dots,N-1,\\ -p^2_{i,j-1} & \text{if} \quad j = N. \end{cases} \end{aligned}$$

Let C be the subset of W given by

$$C = \left\{\mathbf{p}\in W : |(\mathbf{p})_{ij}| \le 1,\ (i,j)\in\Omega\right\}.$$

Clearly, C is closed and convex. A discrete dual ROF model is given by

$$\min_{\mathbf{p}\in C}\left\{E_{\mathrm{d}}^{\mathrm{FD}}(\mathbf{p}) := \frac{1}{2}\|\operatorname{div}\mathbf{p} + \alpha f\|^2\right\}. \tag{4.5}$$

Since $E_{\mathrm{d}}^{\mathrm{FD}}(\mathbf{p})$ is not strictly convex, a solution of (4.5) is not unique in general. Using Proposition 4.1, it is straightforward to verify the equivalence between the primal problem (2.6) with the isotropic total variation and the dual one (4.5). From a solution $\mathbf{p}^* \in W$ of (4.5), one can recover a solution $u^* \in V$ of (2.6) by

$$u^* = f + \frac{1}{\alpha}\operatorname{div}\mathbf{p}^*.$$

The problem (4.5) can be efficiently solved by, e.g., Algorithm A.2.

4.2. **Schwarz methods for the dual ROF model.** Now, we present Schwarz methods for the dual ROF model and their convergence results [27, 28, 29]. We partition the image domain Ω into $\mathcal{N} = M_s \times N_s$ rectangular subdomains $\{\Omega_s\}_{s=1}^{\mathcal{N}}$. Let $\delta \geq 0$ be the overlapping width parameter, which is 0 in the nonoverlapping case. All subdomains can be classified into N_c colors by the usual coloring technique [1, Sect. 2.5.1]. We denote the union of all subdomains with color k, $1 \leq k \leq N_c$, by S_k. We define the local function space $\widetilde{W}_k$ in S_k by

$$\widetilde{W}_k = \{\mathbf{p} \in W : \operatorname{supp}(\mathbf{p}) \subset S_k\}.$$

Similarly to Section 3, we set the operators $R_k\colon W \to \widetilde{W}_k$, $D_k\colon \widetilde{W}_k \to \widetilde{W}_k$, and $\widetilde{R}_k\colon W \to \widetilde{W}_k$, so that

$$W = \sum_{k=1}^{N_c} R_k^* \widetilde{W}_k, \quad \widetilde{R}_k = D_k R_k, \quad \sum_{k=1}^{N_c} R_k^* \widetilde{R}_k = I.$$

In order to deal with the pointwise constraint C in each subdomain, we adopt an idea of constraint decomposition [27, 50]; with the discrete partition of unity $\{D_k\}_{k=1}^{N_c}$, one can construct a pointwise constraint $C_k \subset \widetilde{W}_k$ on S_k such that

$$C = \sum_{k=1}^{N_c} R_k^* C_k.$$

Indeed, one can show that a choice $C_k = \widetilde{R}_k C$ satisfies the above equation [27, Proposition 2.1]. In particular, if the decomposition $\{S_k\}_{k=1}^{N_c}$ is nonoverlapping, then we have

$$C = \sum_{k=1}^{N_c} R_k^* C_k = \bigoplus_{k=1}^{N_c} C_k.$$

In this setting, multiplicative and additive Schwarz methods for (4.5) are presented in Algorithms 4.1 and 4.2. In the case of nonoverlapping domain decomposition, Algorithms 4.1 and 4.2 reduce to the nonlinear block Gauss–Seidel and Jacobi methods, respectively.

Algorithm 4.1 Multiplicative Schwarz method for the dual ROF model (4.5)

Let $\mathbf{p}^{(0)} \in C$.
for $n = 0, 1, 2, \ldots$
  for $k = 1, 2, \ldots, N_c$

$$\mathbf{p}_k^{(n+1)} \in \operatorname*{arg\,min}_{\mathbf{p}_k \in C_k} E_{\mathrm{d}}^{\mathrm{FD}} \left(R_k^* \mathbf{p}_k + \sum_{j<k} R_j^* \mathbf{p}_j^{(n+1)} + \sum_{j>k} R_j^* \widetilde{R}_j \mathbf{p}^{(n)} \right)$$

  end

$$\mathbf{p}^{(n+1)} = \sum_{k=1}^{N_c} R_k^* \mathbf{p}_k^{(n+1)}$$

end

Algorithm 4.2 Additive Schwarz method for the dual ROF model (4.5)

Choose $\tau \in (0, 1/N_c]$. Let $\mathbf{p}^{(0)} \in C$.
for $n = 0, 1, 2, \ldots$

$$\mathbf{p}_k^{(n+1)} \in \operatorname*{arg\,min}_{\mathbf{p}_k \in C_k} E_{\mathrm{d}}^{\mathrm{FD}} \left(R_k^* \mathbf{p}_k + \sum_{j \neq k} R_j^* \widetilde{R}_j \mathbf{p}^{(n)} \right), \quad 1 \leq k \leq N_c$$

$$\mathbf{p}^{(n+1)} = (1 - \tau)\mathbf{p}^{(n)} + \tau \sum_{k=1}^{N_c} R_k^* \mathbf{p}_k^{(n+1)}$$

end

Note that local problems of Algorithms 4.1 and 4.2 may admit nonunique minimizers. In implementation, we just choose any one among them and it does not vary the convergence behavior of algorithms.

Differently from the primal ROF model, convergence of Schwarz methods to a minimizer is always guaranteed for the dual ROF model. Global convergence results for Algorithms 4.1 and 4.2 are summarized in Theorem 4.3.

Theorem 4.3. *Let $\{\mathbf{p}^{(n)}\}$ be the sequence generated by either Algorithm 4.1 or 4.2. Then $\{\mathbf{p}^{(n)}\} \subset C$ and we have*

$$E_{\mathrm{d}}^{\mathrm{FD}}(\mathbf{p}^{(n)}) - E_{\mathrm{d}}^{\mathrm{FD}}(\mathbf{p}^*) \leq \frac{c}{n}, \quad n \geq 1,$$

where c is a positive constant depending on τ (for Algorithm 4.2 only), $\mathbf{p}^{(0)}$, Ω, and its domain decomposition $\{\Omega_s\}_{s=1}^{\mathcal{N}}$.

Proof. Proofs for both Algorithms 4.1 and 4.2 in the overlapping case can be found in [27]. Meanwhile, in the nonoverlapping case, one can observe that Algorithm 4.1 is the proximal alternating minimization method [51] applied to the dual ROF model (4.5). Therefore, its

$O(1/n)$ energy convergence is guaranteed; see, e.g., [52, 53]. Algorithm 4.2 in the nonoverlapping case was shown to converge at $O(1/n)$ rate in [29]. □

We have additional remarks on the relaxed block Jacobi method, a special case of Algorithm 4.2 when the domain decomposition is nonoverlapping. In [29], it was observed that the relaxed block Jacobi method has a similar but not the exactly same structure as the forward-backward splitting algorithm (see Algorithm A.1; one may compare [29, Lemma 3.2] with [13, Lemma 2.3]. Motivated by this fact, the authors modified the relaxed block Jacobi method to have the same structure as forward-backward splitting. Then FISTA acceleration (see Algorithm A.2) was successfully applied and an accelerated method was obtained. We summarize the accelerated method called the *Fast prerelaxed block Jacobi method* in Algorithm 4.3.

Algorithm 4.3 Fast prerelaxed block Jacobi method for the dual ROF model (4.5)

Choose $\tau \in (0, 1/N_c]$. Let $\mathbf{p}^{(0)} = \mathbf{q}^{(0)} \in C$ and $t_0 = 1$.
for $n = 0, 1, 2, \dots$

$$\mathbf{p}_k^{(n+1)} \in \operatorname*{arg\,min}_{\mathbf{p}_k \in C_k} E_{\mathrm{d}}^{\mathrm{FD}} \left(R_k^* \left[\frac{1}{\tau} \mathbf{p}_k - \left(\frac{1}{\tau} - 1 \right) \widetilde{R}_k \mathbf{q}^{(n)} \right] + \sum_{j \neq k} R_j^* \widetilde{R}_j \mathbf{q}^{(n)} \right), \quad 1 \leq k \leq N_c$$

$$\mathbf{p}^{(n+1)} = \sum_{k=1}^{N_c} R_k^* \mathbf{p}_k^{(n+1)}$$

$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}$$

$$\mathbf{q}^{(n+1)} = \mathbf{p}^{(n+1)} + \frac{t_n - 1}{t_{n+1}} (\mathbf{p}^{(n+1)} - \mathbf{p}^{(n)})$$

end

Local problems of Algorithm 4.3 have a slightly different form from ones of Algorithm 4.2; the term $\mathbf{p}_k$ in Algorithm 4.2 is replaced by the relaxed term $\tau^{-1}\mathbf{p}_k - (\tau^{-1} - 1) R_k \mathbf{q}^{(n)}$ in Algorithm 4.3. Note that their computational costs are the same. On the other hand, Algorithm 4.3 does not have the relaxation step in computation of $\mathbf{p}^{(n+1)}$ from $\mathbf{p}_k^{(n+1)}$ on the contrary to Algorithm 4.2. In this sense, we say that Algorithm 4.3 is prerelaxed. In Algorithm 4.3, computational costs for t_{n+1} and $\mathbf{q}^{(n+1)}$ are negligible compared to that of $\mathbf{p}^{(n+1)}$.

One can prove that a map $\mathbf{q}^{(n)} \mapsto \mathbf{p}^{(n+1)}$ in Algorithm 4.3 is in fact a proximal descent step with respect to a certain pseudometric; see [29, Lemma 3.10]. Consequently, we obtain $O(1/n^2)$ energy convergence of Algorithm 4.3 as a simple corollary of Theorem A.2.

Theorem 4.4. *Let $\{\mathbf{p}^{(n)}\}$ be the sequence generated by Algorithm 4.3. Then $\{\mathbf{p}^{(n)}\} \subset C$ and we have*

$$E_{\mathrm{d}}^{\mathrm{FD}}(\mathbf{p}^{(n)}) - E_{\mathrm{d}}^{\mathrm{FD}}(\mathbf{p}^*) \leq \frac{c}{(n+1)^2}, \quad n \geq 1,$$

where c is a positive constant depending on τ, $\mathbf{p}^{(0)}$, Ω, and its domain decomposition $\{\Omega_s\}_{s=1}^{\mathcal{N}}$.

Proof. See [29, Theorem 3.14]. □

We note that a similar acceleration technique for the block Gauss–Seidel method (Algorithm 4.1 with a nonoverlapping domain decomposition) with $N_c = 2$ was presented in [52, 53], but we omit details.

Remark 4.5. In Algorithm 4.2, it was shown in [27] that $\tau = 1/N_c$ is optimal in both theoretical and practical senses. One can verify without major difficulty that the same applies to Algorithm 4.3.

4.3. **Schwarz methods for the primal problem based on dual decompositions.** As shown in Section 3, Schwarz methods for the primal ROF model have troubles in the convergence to the minimizer. Lee and Nam [26] proposed nonoverlapping Schwarz methods for the primal ROF model based on the dual formulation which are guaranteed to converge to the minimizer. Recently, Langer and Gaspoz [54] generalized the work [26] to overlapping domain decomposition. Here, we present the Schwarz methods for the primal ROF model based on dual decompositions proposed in [26, 54]. For simplicity, we only consider the nonoverlapping case with two subdomains.

The discrete primal ROF model (2.6) is revisited:

$$\min_{u\in V}\left\{E_{\mathrm{p}}^{\mathrm{FD}}(u;f) := \frac{\alpha}{2}\|u-f\|_2^2 + \|\nabla u\|_1\right\}.$$

Let $\{\Omega_1, \Omega_2\}$ be a nonoverlapping domain decomposition of Ω, i.e.,

$$\begin{aligned}\Omega_1 &= \{1,\dots,M\}\times\{1,\dots,N_1\},\\ \Omega_2 &= \{1,\dots,M\}\times\{N_1+1,\dots,N\},\end{aligned}$$

for some N_1. Let Γ_{d} be the subdomain interface

$$\Gamma_{\mathrm{d}} = \{1,\dots,M\}\times\{N_1+1\}.$$

Note that Γ_{d} is different from Γ_{p} given in Section 3. The local primal spaces $\widetilde{V}_1$, $\widetilde{V}_2$ and local dual spaces W_1, W_2 are defined by

$$\begin{aligned}\widetilde{V}_1 &= \{u\in V : \operatorname{supp}(u)\subset\Omega_1\cup\Gamma_{\mathrm{d}}\},\\ \widetilde{V}_2 &= \{u\in V : \operatorname{supp}(u)\subset\Omega_2\},\\ W_1 &= \{\mathbf{p}\in W : \operatorname{supp}(\mathbf{p})\subset\Omega_1\},\\ W_2 &= \{\mathbf{p}\in W : \operatorname{supp}(\mathbf{p})\subset\Omega_2\}.\end{aligned}$$

The restriction operator onto $\widetilde{V}_k$, $k = 1, 2$, is denoted by $R_k\colon V \to \widetilde{V}_k$. While the dual decomposition $W = W_1 \oplus W_2$ is nonoverlapping, one line of pixels is overlapped in the primal decomposition $V = R_1^*\widetilde{V}_1 + R_2^*\widetilde{V}_2$ [26, Remark 3.1].

One can observe that $\operatorname{div} \mathbf{p}_k \in \widetilde{V}_k$ for all $\mathbf{p}_k \in W_k$. Consequently, the local gradient operator $\tilde{\nabla}_k \colon \widetilde{V}_k \to W_k$ can be defined as the minus adjoint of $\operatorname{div} \colon W_k \to \tilde{V}_k$, i.e.,

$$
\begin{aligned}
(\tilde{\nabla}_1 \tilde{u})^1_{ij} &= \begin{cases} \tilde{u}_{i+1,j} - \tilde{u}_{ij} & \text{if} \quad i = 1, \dots, M-1, \\ 0 & \text{if} \quad i = M, \end{cases} \\
(\tilde{\nabla}_1 \tilde{u})^2_{ij} &= \begin{cases} \tilde{u}_{i,j+1} - \tilde{u}_{ij} & \text{if} \quad j = 1, \dots, N_1, \\ 0 & \text{if} \quad j = N_1 + 1, \end{cases} \\
(\tilde{\nabla}_2 \tilde{u})^1_{ij} &= \begin{cases} \tilde{u}_{i+1,j} - \tilde{u}_{ij} & \text{if} \quad i = 1, \dots, M-1, \\ 0 & \text{if} \quad i = M, \end{cases} \\
(\tilde{\nabla}_2 \tilde{u})^2_{ij} &= \begin{cases} \tilde{u}_{i,j+1} - \tilde{u}_{ij} & \text{if} \quad j = N_1 + 1, \dots, N-1, \\ 0 & \text{if} \quad j = N. \end{cases}
\end{aligned}
$$

In this case, $\tilde{\nabla}_k$ has the same form as the global gradient operator $\nabla \colon V \to W$ with the homogeneous Neumann boundary condition. Note that it is different from the local gradient operator ∇_k in Section 3.

Now, we present the relaxed block Jacobi method proposed in [26] based on the decomposition $V = R_1^* \widetilde{V}_1 + R_2^* \widetilde{V}_2$ in Algorithm 4.4. For the sake of simplicity, we omit the corresponding block Gauss–Seidel method; see [26] for the method. Global convergence theorems for Algorithm 4.4 can be found in [26, 54].

Algorithm 4.4 Relaxed block Jacobi method for the primal ROF model (2.6) based on dual decomposition

Choose $\tau \in (0, 1/2]$. Let $\tilde{v}_1^{(0)} \in \widetilde{V}_1$ and $\tilde{v}_2^{(0)} \in \widetilde{V}_2$.
for $n = 0, 1, 2, \dots$
$\tilde{f}_k^{(n+1)} = R_k(-R_{3-k}^* \tilde{v}_{3-k}^{(n)} + f), \quad k = 1, 2$
$\tilde{u}_k^{(n+1)} = \underset{\tilde{u}_k \in \widetilde{V}_k}{\arg\min} \left\{ \frac{\alpha}{2} \|\tilde{u}_k - \tilde{f}_k^{(n+1)}\|^2_{2,\widetilde{V}_k} + \|\tilde{\nabla}_k \tilde{u}_k\|_{1,W_k} \right\}, \quad k = 1, 2$
$\tilde{v}_k^{(n+1)} = (1 - \tau)\tilde{v}_k^{(n)} + \tau(-\tilde{u}_k^{(n+1)} + \tilde{f}_k^{(n+1)}), \quad k = 1, 2$
$u^{(n+1)} = -R_1^* \tilde{v}_1^{(n+1)} - R_2^* \tilde{v}_2^{(n+1)} + f$
end

One of the advantages of Algorithm 4.4 is that it can be easily generalized to the general TV-L^2 model

$$
\min_{u \in V} \left\{ \frac{\alpha}{2} \|Au - f\|_2^2 + \|\nabla u\|_1 \right\}. \tag{4.6}
$$

Here $A \colon V \to V$ is a linear operator which satisfies $\|A\| \leq 1$. This assumption is not too restrictive for imaging problems; see Proposition 2.2. In addition, we assume that $\ker A$ does not contain constant functions, so that the energy functional of (4.6) is coercive.

The main idea is to combine Algorithm 4.4 and a forward-backward splitting algorithm (see Algorithm A.1) for (4.6) with inexact proximity operators. Since the fidelity term $\frac{\alpha}{2}\|Au - f\|_2^2$

in (4.6) is smooth and has the Lipschitz continuous gradient $A^*(Au-f)$, the following forward-backward splitting iteration for (4.6) is available:

$$u^{(n+1)} = \arg\min_{u\in V} E_{\mathrm{p}}^{\mathrm{FD}}(u; f - A^*(Au^{(n)} - f)).$$

The above minimization problem can be solved by Algorithm 4.4, but in order to reduce the computation time, the minimization problem is solved approximately; see Algorithm 4.5.

Algorithm 4.5 Relaxed block Jacobi method for the TV-L^2 model (4.6) based on dual decomposition

Choose $\tau \in (0, 1/2]$. Let $u^{(0)} \in V$, $\tilde{v}_1^{(0)} \in \widetilde{V}_1$, and $\tilde{v}_2^{(0)} \in \widetilde{V}_2$.
for $n = 0, 1, 2, \ldots$
  $f^{(n+1)} = u^{(n)} - A^*(Au^{(n)} - f)$
  $\tilde{v}_k^{(n,0)} = \tilde{v}_k^{(n)}, \qquad k = 1, 2$
  for $j = 0, 1, 2, \ldots$
    $\tilde{f}_k^{(n,j+1)} = R_k(-R_{3-k}^* \tilde{v}_{3-k}^{(n,j)} + f^{(n+1)}), \quad k = 1, 2$
    $\tilde{u}_k^{(n,j+1)} = \arg\min_{\tilde{u}_k \in \widetilde{V}_k} \left\{ \frac{\alpha}{2} \|\tilde{u}_k - \tilde{f}_k^{(n,j+1)}\|_{2,\widetilde{V}_k}^2 + \|\tilde{\nabla}_k \tilde{u}_k\|_{1,W_k} \right\}, \quad k = 1, 2$
    $\tilde{v}_k^{(n,j+1)} = (1-\tau)\tilde{v}_k^{(n,j)} + \tau(-\tilde{u}_k^{(n,j+1)} + \tilde{f}_k^{(n,j+1)}), \quad k = 1, 2$
  end until $E_{\mathrm{den}}(-R_1^* \tilde{v}_1^{(n,j+1)} - R_2^* \tilde{v}_2^{(n,j+1)} + f^{(n+1)}; f^{(n+1)}) \le E_{\mathrm{den}}\left(u^{(n)}; f^{(n+1)}\right)$
  $\tilde{v}_k^{(n+1)} = \tilde{v}_k^{(n,j+1)}, \qquad k = 1, 2$
  $u^{(n+1)} = f^{(n+1)} - R_1^* \tilde{v}_1^{(n+1)} - R_2^* \tilde{v}_2^{(n+1)}$
end

Since inner denoising problems are solved inexactly, Theorem A.1 cannot be directly applied to Algorithm 4.5. That is, the convergence of the algorithm is nontrivial. Rigorous convergence analysis for Algorithm 4.5 can be found in [26].

5. Iterative substructuring methods for the dual problem

In this section, we cover iterative substructuring methods, another pillar of DDMs, for total variation minimization. While iterative substructuring methods are popular in DDMs for problems arising in structural mechanics (see e.g., [4, 6, 8]), relatively little achievement has been made on mathematical imaging problems. Due to the grid structure of images, it seems natural to employ finite difference discretizations which use image pixels as discrete points; see Section 2. Such discretizations do not have interface dofs, so that designing iterative substructuring methods is not straightforward. However, iterative substructuring methods have an advantage as DDMs for dual total variation minimization compared to overlapping Schwarz methods; since the constraint $|\mathbf{p}| \le 1$ in (4.3) is fully separable into subdomains, treating the constraint $|\mathbf{p}| \le 1$ in iterative substructuring methods is as easy as doing in full-dimension problems.

Earlier in this section, we present finite element discretizations of dual total variation minimization (4.3) proposed in [30, 31]. Based on these discretizations, we describe primal and primal-dual iterative substructuring methods for the dual ROF model (4.4). Each method has its own merits; the primal method is easily accelerated by FISTA acceleration (see Algorithm A.2), and the primal-dual method can be generalized to more general total variation minimization problems of the form (4.3).

We mention that there are other approaches for total variation minimization using finite elements [55, 56]. However, we will not give details on them in this paper.

5.1. **Finite element discretizations.** In finite element discretizations of (4.3) proposed in [30, 31], each pixel in the image domain Ω is regarded as a square finite element with side length 1. Let $\mathcal{T}$ be the collection of all elements in Ω and let $\mathcal{E}$ be the collection of all interior element edges. We define the space X by

$$X = \left\{ u \in L^2(\Omega) : u|_T \text{ is constant, } T \in \mathcal{T} \right\}.$$

Dofs for X are values on the elements; for $u \in X_h$, $T \in \mathcal{T}$, and $x_T \in T$, we write $(u)_T = u(x_T)$. We also define the space Y as the lowest order Raviart–Thomas finite element space [57] on Ω with the pure essential boundary condition, i.e.,

$$Y = \{ \mathbf{p} \in H_0(\operatorname{div}; \Omega) : \mathbf{p}|_T \in \mathcal{RT}_0(T),\ T \in \mathcal{T} \},$$

where $\mathcal{RT}_0(T)$ is the collection of vector functions $\mathbf{q}$: $T \to \mathbb{R}^2$ of the form

$$\mathbf{q}(x_1, x_2) = \begin{bmatrix} a_1 + b_1 x_1 \\ a_2 + b_2 x_2 \end{bmatrix}$$

for $a_1, a_2, b_1, b_2 \in \mathbb{R}$. Dofs for Y are values of the normal components over the interior element edges. We denote the dof of $\mathbf{p} \in Y$ associated to an edge $e \in \mathcal{E}$ by $(\mathbf{p})_e$. It is straightforward to observe that $\operatorname{div} Y \subset X$.

We equip spaces X and Y by the Euclidean inner products of dofs, i.e.,

$$\begin{aligned} \langle u, v \rangle_X &= \sum_{T \in \mathcal{T}} (u)_T (v)_T, \quad u, v \in X, \\ \langle \mathbf{p}, \mathbf{q} \rangle_Y &= \sum_{e \in \mathcal{E}} (\mathbf{p})_e (\mathbf{q})_e, \quad \mathbf{p}, \mathbf{q} \in Y, \end{aligned}$$

and their induced norms $\|\cdot\|_X$ and $\|\cdot\|_Y$. One can verify that $\langle \cdot, \cdot \rangle_X$ agrees with the $L^2(\Omega)$-inner product and $\langle \cdot, \cdot \rangle_Y$ is spectrally equivalent to the $(L^2(\Omega))^2$-inner product [30, Remark 2.2]. In the sequel, we may drop subscripts X and Y if there is no ambiguity.

In order to treat the constraint $|\mathbf{p}| \le 1$ in (4.3), we define the convex subset C of Y as

$$C = \{ \mathbf{p} \in Y : |(\mathbf{p})_e| \le 1,\ e \in \mathcal{E} \}.$$

Then the finite element discretization of (4.3) with $Y \subset H_0(\operatorname{div}; \Omega)$ is given by

$$\min_{\mathbf{p} \in C} \left\{ E^{\mathrm{FE}}(\mathbf{p}) := F^*(\operatorname{div} \mathbf{p}) \right\}. \tag{5.1}$$

Application of Proposition 4.1 yields that a solution of the primal problem

$$\min_{u\in X}\{\alpha F(u)+TV(u)\} \tag{5.2}$$

can be recovered from a solution of the dual problem (5.1) by the following equation:

$$0\in -\operatorname{div}\mathbf{p}+\alpha\partial F(u). \tag{5.3}$$

In a special case of the ROF model, i.e., when (5.1) becomes

$$\min_{\mathbf{p}\in C}\left\{E_{\mathrm{d}}^{\mathrm{FE}}(\mathbf{p}):=\frac{1}{2}\|\operatorname{div}\mathbf{p}+\alpha f\|^2\right\}, \tag{5.4}$$

the equation (5.3) reduces to the following:

$$u=f+\frac{1}{\alpha}\operatorname{div}\mathbf{p}.$$

It was shown in [31, Proposition 3.4] that a finite element solution of (5.2) is indeed a finite difference solution (2.6) with the anisotropic total variation.

5.2. **Primal iterative substructuring methods.** We present the primal iterative substructuring method for the dual ROF model (5.4) proposed in [30]. Primal iterative substructuring methods are based on nonoverlapping domain decomposition. The interior dofs of the subdomains are eliminated so that only the interface dofs remain. Then the resulting equivalent problem with respect to the interface dofs is solved by a suitable iterative solver.

The image domain Ω is decomposed into $\mathcal{N}$ disjoint rectangular subdomains. Assume that each subdomain is a union of elements in $\mathcal{T}$. In the sequel, let the indices s and t ($s<t$) run from 1 to $\mathcal{N}$. We denote the subdomain interface between two adjacent subdomains Ω_s and Ω_t by Γ_{st}. We also define the union of all subdomain interfaces by Γ, i.e., $\Gamma=\bigcup_{s<t}\Gamma_{st}$.

Let $\mathcal{T}_s$ be the collection of all elements of $\mathcal{T}$ in Ω_s. The local function space Y_s is given by

$$Y_s=\{\mathbf{p}_s\in H_0(\operatorname{div};\Omega_s):\mathbf{p}_s|_T\in\mathcal{RT}_0(T),\ T\in\mathcal{T}_s\}.$$

In addition, we set Y_I by

$$Y_I=\bigoplus_{s=1}^{\mathcal{N}}Y_s.$$

The interface function space Y_Γ is given by $Y_\Gamma=Y/Y_I$. That is, for $\mathbf{p}\in Y$, there exists a unique decomposition

$$\mathbf{p}=\mathbf{p}_I\oplus\mathbf{p}_\Gamma=\left(\bigoplus_{s=1}^{\mathcal{N}}\mathbf{p}_s\right)\oplus\mathbf{p}_\Gamma$$

for some $\mathbf{p}\in Y_s$ and $\mathbf{p}_\Gamma\in Y_\Gamma$. In order to deal with inequality constraints in (4.3), let C_s be the subset of Y_s such that

$$C_s=\{\mathbf{p}_s\in Y_s:|(\mathbf{p}_s)_e|\le 1,\ e\in\mathcal{E}_s\},$$

where $\mathcal{E}_s$ is the collection of interior element edges of Ω_s. Similarly, we set

$$C_\Gamma = \left\{ \mathbf{p}_\Gamma \in Y_\Gamma : |(\mathbf{p}_\Gamma)_e| \le 1,\ e \in \mathcal{E} \setminus \bigcup_{s=1}^{\mathcal{N}} \mathcal{E}_s \right\}.$$

For fixed $\mathbf{p}_\Gamma \in C_\Gamma$, we define $\mathcal{H}_I \mathbf{p}_\Gamma \in C_I$ by a solution of the minimization problem

$$\min_{\mathbf{p}_I \in C_I} E_{\mathrm{d}}^{\mathrm{FE}}(\mathbf{p}_I \oplus \mathbf{p}_\Gamma).$$

Note that $\mathcal{H}_I \mathbf{p}_\Gamma$ is the direct sum of $\{\mathbf{p}_s\}_{s=1}^{\mathcal{N}}$, where $\mathbf{p}_s \in Y_s$ is a solution of the minimization problem

$$\min_{\mathbf{p}_s \in C_s} \frac{1}{2} \int_{\Omega_s} (\operatorname{div}(\mathbf{p}_s + \mathbf{p}_\Gamma|_{\Omega_s}) + \alpha f)^2 \, dx.$$

Since each $\mathbf{p}_s$ can be solved independently in each subdomain, evaluating $\mathcal{H}_I \mathbf{p}_\Gamma$ is suitable for parallel computation. Elimination of $\mathbf{p}_I$ in (5.4) by using $\mathcal{H}_I$ produces the following:

$$\min_{\mathbf{p}_\Gamma \in C_\Gamma} \left\{ E_\Gamma(\mathbf{p}_\Gamma) := E_{\mathrm{d}}^{\mathrm{FE}}(\mathcal{H}_I \mathbf{p}_\Gamma \oplus \mathbf{p}_\Gamma) \right\}. \tag{5.5}$$

It is interesting to note that the functional E_Γ in (5.5) is well-defined even though the value of $\mathcal{H}_I \mathbf{p}_\Gamma$ is nonunique in general; see [30] for details. The following proposition summarizes a relation between (5.4) and (5.5).

Proposition 5.1. *If $\mathbf{p}^* \in Y$ is a solution of* (5.4)*, then $\mathbf{p}_\Gamma^* = \mathbf{p}^*|_{Y_\Gamma}$ is a solution of* (5.5)*. Conversely, if $\mathbf{p}_\Gamma^* \in Y_\Gamma$ is a solution of* (5.5)*, then $\mathbf{p}^* = \mathcal{H}_I \mathbf{p}_\Gamma^* \oplus \mathbf{p}_\Gamma^*$ is a solution of* (5.4)*.*

Proof. See [30, Proposition 3.1]. □

Thanks to Proposition 5.1, it suffices to solve (5.5) in order to obtain a solution of (5.4). A surprising fact on (5.5) is that, nevertheless the nonlinearity and nonuniqueness of $\mathcal{H}_I$, the functional E_Γ is differentiable with the Lipschitz continuous derivative [30, Corollary 3.4]. Therefore, one may adopt Algorithm A.2 to solve (5.5) and it automatically yields an $O(1/n^2)$ energy convergent iterative substructuring method for the dual ROF model. We present the resulting algorithm in Algorithm 5.1.

Algorithm 5.1 Primal iterative substructuring method for the dual ROF model (5.4)

Choose $\tau \in (0, 1/4]$. Let $\mathbf{q}_\Gamma^{(0)} = \mathbf{p}_\Gamma^{(0)} \in C_\Gamma$ and $t_0 = 1$.
for $n = 0, 1, 2, \ldots$

$$\mathbf{p}_\Gamma^{(n+1)} = \operatorname{proj}_{C_\Gamma} \left(\mathbf{q}_\Gamma^{(n)} - \tau \operatorname{div}^* \left(\operatorname{div}(\mathcal{H}_I \mathbf{q}_\Gamma^{(n)} \oplus \mathbf{q}_\Gamma^{(n)}) + \alpha f \right) \Big|_{Y_\Gamma} \right)$$

$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}$$

$$\mathbf{q}_\Gamma^{(n+1)} = \mathbf{p}_\Gamma^{(n+1)} + \frac{t_n - 1}{t_{n+1}} (\mathbf{p}_\Gamma^{(n+1)} - \mathbf{p}_\Gamma^{(n)})$$

end

We have the following convergence theorem of Algorithm 5.1 as a direct consequence of Theorem A.2.

Theorem 5.2. *Let $\{\mathbf{p}_\Gamma^{(n)}\}$ be the sequence generated by Algorithm 5.1, and let $\mathbf{p}_\Gamma^* \in Y_\Gamma$ be a solution of* (5.5)*. Then we have*

$$E_\Gamma(\mathbf{p}_\Gamma^{(n)}) - E_\Gamma(\mathbf{p}_\Gamma^*) \le \frac{c}{(n+1)^2}, \quad n \ge 1,$$

where c is a positive constant depending on τ, $\mathbf{p}_\Gamma^{(0)}$, Ω, and its domain decomposition $\{\Omega_s\}_{s=1}^{\mathcal{N}}$.

As a historical remark, we mention that Algorithm 5.1 is the first DDM that yields the $O(1/n^2)$ convergence rate for total variation minimization. Later, the idea of FISTA acceleration was successfully implanted to block Jacobi methods in order to obtain an $O(1/n^2)$ convergent block method, Algorithm 4.3.

5.3. **Primal-dual iterative substructuring methods.** Unlike the primal iterative substructuring method, the continuity of a solution on the subdomain can be imposed by the method of Lagrange multipliers; see [6, 7]. This results a saddle point problem of $\mathbf{p}$ and Lagrange multiplier λ.

First, we consider the primal-dual iterative substructuring method for the dual ROF model (5.4) proposed in [30]. We begin with the same nonoverlapping domain decomposition setting $\{\Omega_s\}_{s=1}^{\mathcal{N}}$ as the primal method. Let $\mathbf{n}_s$ denotes the unit outer normal to $\partial\Omega_s$. In the primal-dual method, we use a different local function space $\widetilde{Y}_s$ defined by

$$\widetilde{Y}_s = \{\tilde{\mathbf{p}}_s \in H(\mathrm{div};\Omega_s) : \tilde{\mathbf{p}}_s \cdot \mathbf{n}_s = 0 \text{ on } \partial\Omega_s \setminus \Gamma,\ \tilde{\mathbf{p}}_s|_T \in \mathcal{RT}_0(T),\ T \in \mathcal{T}_s\}.$$

Differently from the local space Y_s in Section 5.2, the essential boundary condition is not imposed on $\Gamma \subset \partial\Omega_s$ to the functions in $\widetilde{Y}_s$. In other words, $\widetilde{Y}_s$ has dofs on $\partial\Omega_s \cap \Gamma$. Let $\tilde{\mathcal{E}}_s$ be the collection of element edges inside Ω_s and on $\partial\Omega_s \cap \Gamma$. The inequality-constrained subset $\widetilde{C}_s$ of $\widetilde{Y}_s$ is defined by

$$\widetilde{C}_s = \left\{\tilde{\mathbf{p}}_s \in \widetilde{Y}_s : |(\tilde{\mathbf{p}}_s)_e| \le 1,\ e \in \tilde{\mathcal{E}}_s\right\}.$$

We write

$$\widetilde{Y} = \bigoplus_{s=1}^{\mathcal{N}} \widetilde{Y}_s, \quad \widetilde{C} = \bigoplus_{s=1}^{\mathcal{N}} \widetilde{C}_s.$$

By definition, functions in $\widetilde{Y}$ have discontinuities on the subdomain interfaces Γ in general. If we define the jump operator B: $\widetilde{Y} \to \mathbb{R}^{|\mathcal{E}_\Gamma|}$ by

$$B\tilde{\mathbf{p}}|_{\Gamma_{st}} = \tilde{\mathbf{p}}_s \cdot \mathbf{n}_s + \tilde{\mathbf{p}}_t \cdot \mathbf{n}_t, \quad s < t,$$

where $\tilde{\mathbf{p}}_s = \tilde{\mathbf{p}}|_{\Omega_s}$, it is clear that $\tilde{\mathbf{p}} \in \widetilde{Y}$ is continuous along Γ if and only if $\tilde{\mathbf{p}} \in \ker B$. Therefore, the continuity on the subdomain interfaces can be imposed by the constraint $B\tilde{\mathbf{p}} = 0$ using the method of Lagrange multipliers. Then we have the following saddle point problem:

$$\min_{\mathbf{p}\in\widetilde{C}} \max_{\lambda\in\mathbb{R}^{|\mathcal{E}_\Gamma|}} \left\{L(\tilde{\mathbf{p}},\lambda) := \widetilde{E}_{\mathrm{d}}^{\mathrm{FE}}(\tilde{\mathbf{p}}) + \langle B\tilde{\mathbf{p}},\lambda\rangle\right\}, \tag{5.6}$$

where

$$\widetilde{E}_{\mathrm{d}}^{\mathrm{FE}}(\tilde{\mathbf{p}}) = \sum_{s=1}^{\mathcal{N}} \frac{1}{2} \int_{\Omega_s} (\operatorname{div} \tilde{\mathbf{p}}_s + \alpha f)^2 \, dx, \quad \tilde{\mathbf{p}} = \bigoplus_{s=1}^{\mathcal{N}} \tilde{\mathbf{p}}_s \in \widetilde{Y}.$$

An equivalence relation between (5.4) and (5.6) can be found in [30, Proposition 4.2]. We observe that (5.6) is suitable for the primal-dual algorithm (see Algorithm A.3). Application of Algorithm A.3 to (5.6) yields the proposed primal-dual iterative substructuring method; see Algorithm 5.2.

Algorithm 5.2 Primal-dual iterative substructuring method for the dual ROF model (5.4)

Choose $\tau, \sigma > 0$ with $\tau\sigma \leq 1/2$. Let $\tilde{\mathbf{p}}^{(0)} = \tilde{\mathbf{p}}^{(-1)} \in \widetilde{Y}$ and $\lambda^{(0)} \in \mathbb{R}^{|\mathcal{E}_\Gamma|}$.
for $n = 0, 1, 2, \ldots$
$\lambda^{(n+1)} = \lambda^{(n)} + \sigma B(2\tilde{\mathbf{p}}^{(n)} - \tilde{\mathbf{p}}^{(n-1)})$
$\hat{\mathbf{p}}^{(n+1)} = \tilde{\mathbf{p}}^{(n)} - \tau B^* \lambda^{(n+1)}$
$\tilde{\mathbf{p}}^{(n+1)} \in \operatorname*{arg\,min}_{\tilde{\mathbf{p}} \in \widetilde{C}} \left\{ \widetilde{E}_{\mathrm{d}}^{\mathrm{FE}}(\tilde{\mathbf{p}}) + \frac{1}{2\tau} \|\tilde{\mathbf{p}} - \hat{\mathbf{p}}^{(n+1)}\|^2 \right\}$
end

As a direct consequence of Theorem A.3, we get the following $O(1/n)$ ergodic convergence theorem for Algorithm 5.2.

Theorem 5.3. *Let* $\{(\tilde{\mathbf{p}}^{(n)}, \lambda^{(n)})\}$ *be the sequence generated by Algorithm 5.2. Then, it converges to a saddle point of* (5.6) *and satisfies that*

$$L\left(\frac{1}{n}\sum_{k=1}^{n} \tilde{\mathbf{p}}^{(k)}, \lambda\right) - L\left(\tilde{\mathbf{p}}, \frac{1}{n}\sum_{k=1}^{n} \lambda^{(k)}\right) \leq \frac{1}{n}\left(\frac{1}{\tau}\|\tilde{\mathbf{p}} - \tilde{\mathbf{p}}^{(0)}\|_{2,\widetilde{Y}}^2 + \frac{1}{\sigma}\|\lambda - \lambda^{(0)}\|_{2,\mathbb{R}^{|\mathcal{E}_\Gamma|}}^2\right)$$

for any $\tilde{\mathbf{p}} \in Y$ *and* $\lambda \in \mathbb{R}^{|\mathcal{E}_\Gamma|}$.

Even though the convergence rate of Algorithm 5.2 is slower than that of Algorithm 5.1, it has several advantages. First, we observe that local problems of Algorithm 5.2 has the general form

$$\min_{\tilde{\mathbf{p}}_s \in \widetilde{C}_s} \left\{ \frac{1}{2}\|\operatorname{div} \tilde{\mathbf{p}}_s + \alpha f\|^2 + \frac{1}{2\tau}\|\tilde{\mathbf{p}}_s - \hat{\mathbf{p}}_s\|^2 \right\} \tag{5.7}$$

for $\hat{\mathbf{p}}_s \in \widetilde{Y}_s$. Thanks to the strongly convex term $\frac{1}{2\tau}\|\tilde{\mathbf{p}}_s - \hat{\mathbf{p}}_s\|^2$ in (5.7), one can adopt linearly convergent algorithms such as Algorithm A.5 for (5.7) while existing standard solvers (see Algorithms A.2 and A.4) are only $O(1/n^2)$ convergent. Improvement of performance due to this fact can be found in [30, Sect. 5].

Next, we note that Algorithm 5.2 does not utilize the differentiability of the dual ROF functional at all. It means that Algorithm 5.2 can be generalized to general total variation minimization (5.1) without major modification [31]. We assume that the functional F^* in (5.1) is

separable, i.e, there exists convex functionals F_s^*: $\operatorname{div} \widetilde{Y}_s \to \overline{\mathbb{R}}$ such that

$$F^*(\operatorname{div} \tilde{\mathbf{p}}) = \sum_{s=1}^{\mathcal{N}} F_s^*(\operatorname{div} \tilde{\mathbf{p}}|_{\Omega_s}), \quad \tilde{\mathbf{p}} \in \widetilde{Y}.$$

Many problems arising from mathematical imaging satisfy the above assumption; see [31]. We write

$$E^{\mathrm{FE}}(\tilde{\mathbf{p}}) = \sum_{s=1}^{\mathcal{N}} F_s^*(\operatorname{div} \tilde{\mathbf{p}}|_{\Omega_s}), \quad \tilde{\mathbf{p}} \in \widetilde{Y}.$$

Then for $\tau > 0$ and $\hat{\mathbf{p}} \in \widetilde{Y}$, a solution of the minimization problem

$$\min_{\tilde{\mathbf{p}} \in \widetilde{C}} \left\{ E^{\mathrm{FE}}(\tilde{\mathbf{p}}) + \frac{1}{2\tau} \|\tilde{\mathbf{p}} - \hat{\mathbf{p}}\|^2 \right\}$$

can be assembled from solutions of local problems

$$\min_{\tilde{\mathbf{p}}_s \in \widetilde{C}_s} \left\{ F_s^*(\operatorname{div} \tilde{\mathbf{p}}_s) + \frac{1}{2\tau} \|\tilde{\mathbf{p}}_s - \hat{\mathbf{p}}|_{\Omega_s}\|^2 \right\}.$$

In summary, we state a generalization of Algorithm 5.2 for general total variation minimization (5.1) in Algorithm 5.3. Algorithm 5.3 shares the same convergence theorem as Algorithm 5.2 [31, Theorem 4.3].

Algorithm 5.3 Primal-dual iterative substructuring method for general total variation minimization (5.1)

Choose $\tau, \sigma > 0$ with $\tau\sigma \le 1/2$. Let $\tilde{\mathbf{p}}^{(0)} = \tilde{\mathbf{p}}^{(-1)} \in \widetilde{Y}$ and $\lambda^{(0)} \in \mathbb{R}^{|\mathcal{E}_\Gamma|}$.
for $n = 0, 1, 2, \ldots$
  $\lambda^{(n+1)} = \lambda^{(n)} + \sigma B(2\tilde{\mathbf{p}}^{(n)} - \tilde{\mathbf{p}}^{(n-1)})$
  $\hat{\mathbf{p}}^{(n+1)} = \tilde{\mathbf{p}}^{(n)} - \tau B^* \lambda^{(n+1)}$
  for $s = 1, \ldots, \mathcal{N}$ **in parallel**
    $\tilde{\mathbf{p}}_s^{(n+1)} \in \arg\min_{\tilde{\mathbf{p}}_s \in \widetilde{C}_s} \left\{ F_s^*(\tilde{\mathbf{p}}_s) + \frac{1}{2\tau} \|\tilde{\mathbf{p}}_s - \hat{\mathbf{p}}^{(n+1)}|_{\Omega_s}\|^2 \right\}$
  end
  $\tilde{\mathbf{p}}^{(n+1)} = \bigoplus_{s=1}^{\mathcal{N}} \tilde{\mathbf{p}}_s^{(n+1)}$
end

6. NUMERICAL COMPARISON

Until now, we have discussed on various approaches on designing DDMs for total variation minimization: Schwarz methods for the dual problem, Schwarz methods for the primal problem based on dual decompositions, primal iterative substructuring methods for the dual problem, and primal-dual iterative substructuring methods for the dual problem. In order to verify that

all of those approaches produce DDMs that are efficient parallel solvers, we provide numerical comparison of Algorithms 4.3, 4.4, 5.1, and 5.2 on distributed memory computers. Note that Algorithms 4.3, 4.4, 5.1, and 5.2 belong to dual Schwarz methods, primal Schwarz methods with dual decompositions, primal iterative substructuring methods, and primal-dual iterative substructuring methods, respectively. The characteristics of these algorithms are summarized in Table 6.1.

We performed our computations on a computer cluster composed of seven nodes; each node possesses two Intel Xeon SP-6148 CPUs (2.4GHz, 20 cores) and 192GB RAM. The operating system for the cluster is CentOS 7.4 64bit. All algorithms were programmed in ANSI C with OpenMPI, and then compiled by Intel C Compiler.

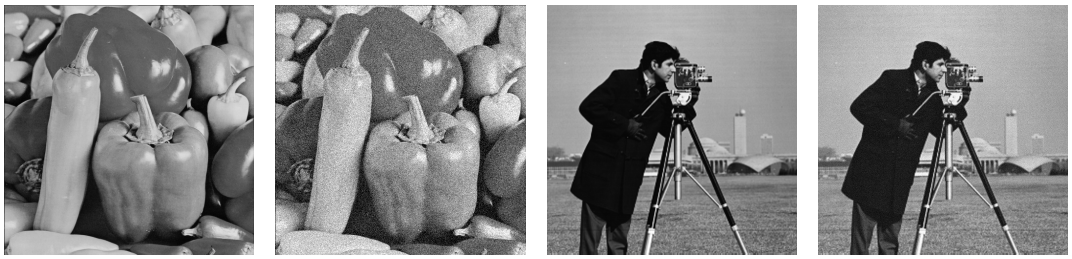
Two test images "Peppers 512×512", "Cameraman 2048×2048" were used in our experiments. Figure 6.1 shows original clean images and noisy images corrupted by additive Gaussian noise with mean 0 and variance 0.05. The weight parameter α in (1.2) was chosen as 10 heuristically.

At each iteration of DDMs, we have to solve local problems in subdomains. In our experiments, we used the same number of cores as subdomains, and each subdomain is assigned to a core one by one. Consequently, all local problems were solved simultaneously by the optimally accelerated primal-dual algorithms, i.e., Algorithm A.4 was adopted as the local solver for Algorithms 4.3, 4.4, and 5.1, while Algorithm A.5 was used for local problems of Algorithm 5.2. The stop condition for local problems is

$$\frac{\|u_s^{(n)} - u_s^{(n-1)}\|_2}{\|u_s^{(n)}\|_2} < 10^{-7} \quad \text{or} \quad n = 100,$$

TABLE 6.1. Characteristics of Algorithms 4.3, 4.4, 5.1, and 5.2.

	Algorithm 4.3	Algorithm 4.4	Algorithm 5.1	Algorithm 5.2
Problem	dual	primal	dual	dual
Discretization	finite difference	finite difference	finite element	finite element
Convergence rate	$O(1/n^2)$	not shown	$O(1/n^2)$	$O(1/n)$
Local solver	sublinear	sublinear	sublinear	linear



(a) Peppers 512×512 (b) Noisy image (PSNR: 19.11) (c) Cameraman 2048×2048 (d) Noisy image (PSNR: 19.17)

FIGURE 6.1. Test images for numerical experiments.

where $u_s^{(n)}$ is the nth iterate for the local primal variable in the subdomain Ω_s, $1 \leq s \leq \mathcal{N}$. In order to reduce the time elapsed in solving local problems, the local solutions from the previous iteration were chosen as initial guesses for the local problems at each iteration.

TABLE 6.2. Performance of Algorithms 4.3 and 4.4 with respect to various numbers of subdomains.

Test image	$\mathcal{N}$	Algorithm 4.3			Algorithm 4.4		
		PSNR	iter	wall-clock time (sec)	PSNR	iter	wall-clock time (sec)
Peppers 512×512	1	24.55	-	6.91	24.55	-	6.91
	2×2	24.55	39	4.94	24.55	45	5.08
	4×4	24.55	48	1.41	24.55	58	1.72
	8×8	24.55	62	0.68	24.55	67	0.71
	16×16	24.55	69	0.46	24.55	80	0.47
Cameraman 2048×2048	1	25.38	-	254.83	25.38	-	254.83
	2×2	25.38	34	160.62	25.38	30	123.49
	4×4	25.38	41	46.97	25.38	42	44.97
	8×8	25.38	50	14.47	25.38	53	14.77
	16×16	25.38	58	6.40	25.38	64	6.07

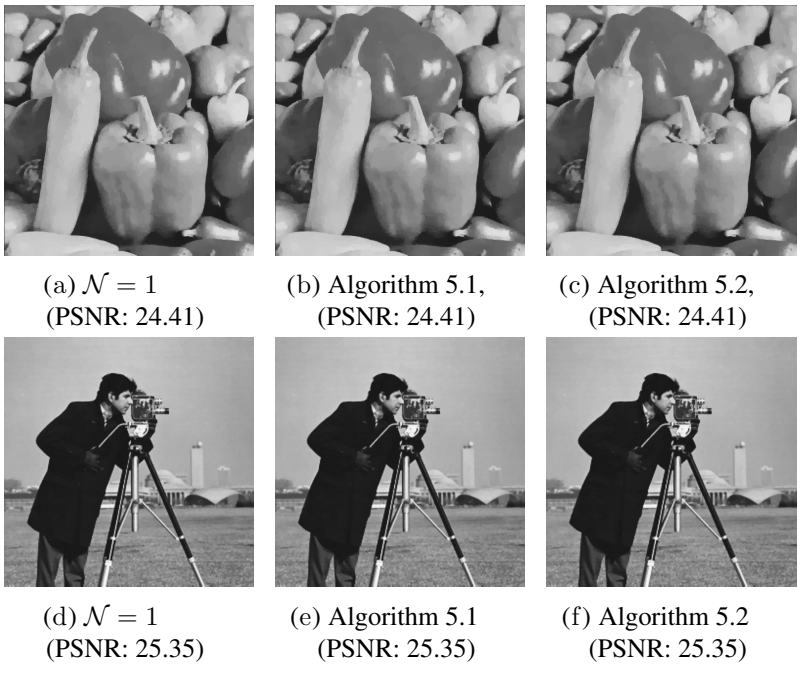
(a) $\mathcal{N} = 1$ (PSNR: 24.55) (b) Algorithm 4.3, (PSNR: 24.55) (c) Algorithm 4.4, (PSNR: 24.55)

(d) $\mathcal{N} = 1$ (PSNR: 25.38) (e) Algorithm 4.3, (PSNR: 25.38) (f) Algorithm 4.4, (PSNR: 25.38)

FIGURE 6.2. Results of Algorithms 4.3 and 4.4 with $\mathcal{N} = 16 \times 16$ for each finite difference model (4.5) and (2.6).

TABLE 6.3. Performance of Algorithms 5.1 and 5.2 with respect to various numbers of subdomains.

Test image	$\mathcal{N}$	Algorithm 5.1			Algorithm 5.2		
		PSNR	iter	wall-clock time (sec)	PSNR	iter	wall-clock time (sec)
Peppers 512×512	1	24.41	-	4.31	24.41	-	4.31
	2×2	24.41	46	3.14	24.41	105	3.01
	4×4	24.41	57	0.91	24.41	113	0.73
	8×8	24.41	69	0.40	24.41	123	0.32
	16×16	24.41	82	0.28	24.41	139	0.26
Cameraman 2048×2048	1	25.35	-	182.57	25.35	-	182.57
	2×2	25.35	34	73.48	25.35	103	107.80
	4×4	25.35	46	34.69	25.35	107	34.56
	8×8	25.35	58	11.40	25.35	113	9.83
	16×16	25.35	71	2.18	25.35	119	1.42



(a) $\mathcal{N} = 1$ (PSNR: 24.41) (b) Algorithm 5.1, (PSNR: 24.41) (c) Algorithm 5.2, (PSNR: 24.41)

(d) $\mathcal{N} = 1$ (PSNR: 25.35) (e) Algorithm 5.1 (PSNR: 25.35) (f) Algorithm 5.2 (PSNR: 25.35)

FIGURE 6.3. Results of Algorithms 5.1 and 5.2 with $\mathcal{N} = 16 \times 16$ for the finite element model (5.4).

Table 6.2 and Fig. 6.2 show the numerical results of Algorithms 4.3 and 4.4 for the finite difference ROF model (4.5) and (2.6), respectively, with respect to various numbers of subdomains, while Table 6.3 and Fig. 6.3 are for the results of the finite element ROF model (5.4). We used the following stop condition for all algorithms:

$$\frac{\|u^{(n)} - u^{(n-1)}\|_2}{\|u^{(n)}\|_2} < 10^{-5}, \tag{6.1}$$

where

$$u^{(n)} = f + \frac{1}{\alpha} \operatorname{div} \mathbf{p}^{(n)} \tag{6.2}$$

for Algorithms 4.3 and 5.2 and

$$u^{(n)} = f + \frac{1}{\alpha} \operatorname{div}(\mathcal{H}_I \mathbf{p}_\Gamma^{(n)} \oplus \mathbf{p}_\Gamma^{(n)})$$

for Algorithm 5.1. The case $\mathcal{N} = 1$ denotes Algorithm A.4 for the full-dimension problems. Since Algorithms 4.3 and 4.4 are based on the finite difference discretizations (4.5) and (2.6), respectively, while Algorithms 5.1 and 5.2 are based on the finite element discretization (5.4), they produce different results.

As shown in Figures 6.2 and 6.3, the results of the full-dimension problems and DDMs are not visually distinguishable. The resulting images of DDMs show no trace of the subdomain interfaces. In addition, in Tables 6.2 and 6.3, since PSNRs are constant regardless of the number of subdomains, we can say that the results of DDMs agree with those of the full-dimension problems.

In Table 6.2, the wall-clock time of Algorithm 4.3 is comparable to that of Algorithm 4.4 despite the less number of iterations. Since Algorithm 4.3 produces a seqence of dual variables $\{\mathbf{p}^{(n)}\}$, one have to recover its primal counterpart by (6.2) in each iteration in order to check whether the stop condition (6.1) is satisfied. On the other hand, Algorithm 4.4 produces a primal sequence $\{u^{(n)}\}$ so that no additional computation is required to check (6.1). Therefore, even though the computational cost for each iteration of Algorithm 4.3 is slightly lower than Algorithm 4.4, both algorithms show the similar performance.

In Table 6.3, we observe that the number of iterations of Algorithm 5.1 is less than Algorithm 5.2. This verifies the superior convergence property of Algorithm 5.1 shown in Theorem 5.2. However, the wall-clock time of Algorithm 5.2 is smaller than Algorithm 5.1. This is because Algorithm 5.2 utilized the linearly convergence local solver Algorithm A.5.

For all algorithms, we observe that the number of iterations increases as the number of subdomain increases. Indeed, it was proven in [29, Corollary 3.8] that the convergence rate of Algorithm 4.3 deteriorates as the total length of the subdomain interfaces increases. Similar discussions can be made for other DDMs.

Finally, for all DDMs presented in this section, we can see that the wall-clock time is decreasing as the number of subdomains $\mathcal{N}$ increases. This verifies the worth of DDMs as parallel solvers for total variation minimization.

7. Concluding remarks

In this paper, we presented a brief review of DDM's various approaches to total variation minimization developed over the past decade. The earliest attempts were direct applications of Schwarz methods to the primal total variation minimization (1.1). However, we noted that they were shown to have a counterexample to their global convergence. Almost all of recent works on DDM dealt with the dual total variation minimization (4.3) instead of the primal one. We introduced three approaches to design DDMs for (4.3): Schwarz methods for the dual problem, Schwarz methods for the primal problem based on dual decompositions, and iterative substructuring methods. We also provided numerical results that verify the efficiency of those DDMs as parallel solvers on distributed memory computers.

There is a great deal of future research topics related to DDMs for the total variation minimization. Even though two-level method design is one of the most important topics in DDMs (see, e.g., [1, Chapter 3]), to the best of our knowledge, there have been no successful two-level methods for total variation minimization and even for the ROF model. In order to get the scalability of DDMs, designing two-level methods is necessary and should be considered in future works.

Acceleration of DDMs for total variation minimization is also an interesting topic. We presented several accelerated nonoverlapping DDMs for the ROF model in this paper; see Algorithms 4.3 and 5.1. However, those acceleration schemes are not directly applicable to the overlapping domain decomposition. Future tasks will need to design acceleration schemes suitable for overlapping DDMs.

Other interesting topics may arise in the generalization of DDMs to nonsmooth optimization problems. For example, supervised machine learning with a neural network is modelled as a nonsmooth optimization problem of the following form [58]:

$$\min_{\theta} \left\{ J(\theta; \mathbf{x}, \mathbf{y}) + R(\theta) \right\}, \tag{7.1}$$

where $J(\theta; \mathbf{x}, \mathbf{y})$ is a loss function depending on the structure of the neural network, θ is a vector of parameters, $(\mathbf{x}, \mathbf{y})$ is a supervised training dataset, and $R(\theta)$ is a regularizer for parameters. Due to the huge size of the datasets, machine learning problems are large in general. Consequently, construction of fast and efficient parallel algorithms for problems of the form (7.1) is one of the active research topics in the field of machine learning. Since (7.1) has a similar structure to (2.2), we expect that the ideas and theories developed for DDMs on total variation minimization will be useful in designing DDMs for (7.1). Indeed, block coordinate descent algorithms [59] which are broadly used in machine learning can be regarded as particular forms of DDMs (see [3]), so that there are extensive possibilities of improvement using the theories of DDMs.

Appendix A. Convex optimization algorithms

This appendix covers representative algorithms for solving convex optimization problems. Throughout this chapter, X and Y denote Euclidean spaces equipped with the inner product

$\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$. The two most popular algorithms are *forward-backward splitting algorithms* [13, 60] and *primal-dual algorithms* [15]. For other state-of-the-art algorithms to solve convex optimization, one may refer a monograph [42] and references therein.

For a proper, convex, lower semicontinuous functional $F\colon X \to \overline{\mathbb{R}}$, the *proximal operator* $\mathrm{prox}_F\colon X \to X$ is defined as

$$\mathrm{prox}_F(\bar{x}) = \arg\min_{x\in X}\left\{F(x) + \frac{1}{2}\|x-\bar{x}\|^2\right\}, \quad \bar{x}\in X.$$

It is easy to check that the proximal operation generalizes the implicit gradient descent with the unit step size [42].

A.1. **Forward-backward splitting algorithms.** We consider the following general convex optimization problem:

$$\min_{x\in X}\{E(x) := F(x) + G(x)\}, \tag{A.1}$$

where $F\colon X \to \overline{\mathbb{R}}$ and $G\colon X \to \overline{\mathbb{R}}$ are proper, convex, and lower semicontinuous. We assume that F has the Lipschitz continuous gradient ∇F with a Lipschitz constant $M > 0$, i.e., it satisfies

$$\|\nabla F(x) - \nabla F(y)\| \le M\|x-y\|, \quad x, y \in X,$$

while G is possibly nonsmooth. In addition, we assume that a solution $x^* \in X$ of (A.1) exists. The forward-backward splitting scheme for (A.1) is a combination of an explicit gradient descent for the smooth part $F(x)$ and an (formal) implicit gradient descent for the nonsmooth part $G(x)$. The forward-backward splitting algorithm with fixed step size is given in Algorithm A.1.

Algorithm A.1 Forward-backward splitting algorithm

Choose $\tau > 0$ with $\tau \le 1/M$. Let $x^{(0)} \in X$.
for $n = 0, 1, 2, \ldots$
  $x^{(n+1)} = \mathrm{prox}_{\tau G}(x^{(n)} - \tau\nabla F(x^{(n)}))$
end

If $G = 0$, then Algorithm A.1 reduces to the gradient descent method with fixed step. Thus, it can be regarded as a generalization of the gradient descent method to nonsmooth convex optimization. The following sublinear convergence of Algorithm A.1 can be proven.

Theorem A.1. *Let $\{x^{(n)}\}$ be the sequence generated by Algorithm A.1. Then for any $n \ge 1$, it satisfies*

$$E(x^{(n)}) - E(x^*) \le \frac{1}{2\tau n}\|x^{(0)} - x^*\|^2.$$

Proof. See [13, Theorem 3.1]. □

Algorithm A.1 can be accelerated by an appropriate over-relaxation technique called the *Nesterov's momentum* [61]. The accelerated method is called *FISTA* (Fast Iterative Shrinkage-Thresholding Algorithm) and was first proposed in [13]; see Algorithm A.2. Note that Algorithm A.2 requires only a single gradient evaluation at each iteration, so that its computational cost per iteration is almost equal to Algorithm A.1.

Algorithm A.2 FISTA

Choose $\tau > 0$ with $\tau \leq 1/M$. Let $y^{(0)} = x^{(0)} \in X$ and $t_0 = 1$.
for $n = 0, 1, 2, \ldots$
$x^{(n+1)} = \operatorname{prox}_{\tau G}(y^{(n)} - \tau \nabla F(y^{(n)}))$
$t_{n+1} = \dfrac{1 + \sqrt{1 + 4t_n^2}}{2}$
$y^{(n+1)} = x^{(n+1)} + \dfrac{t_n - 1}{t_{n+1}}(x^{(n+1)} - x^{(n)})$
end

The following theorem states the $O(1/n^2)$ convergence rate of Algorithm A.2.

Theorem A.2. *Let $\{x^{(n)}\}$ be the sequence generated by Algorithm A.2. Then for any $n \geq 1$, it satisfies*

$$E(x^{(n)}) - E(x^*) \leq \frac{2}{\tau(n+1)^2}\|x^{(0)} - x^*\|^2.$$

Proof. See [13, Theorem 4.4]. □

We note that if $E(x)$ is strongly convex, it can be shown that the forward-backward splitting algorithm is linearly convergent [42]. However, we omit details here.

A.2. **Primal-dual algorithms.** Primal-dual algorithms are intended to solve problems of the following form:

$$\min_{x \in X} \{F(Kx) + G(x)\}, \tag{A.2}$$

where $K\colon X \to Y$ is a continuous linear operator and $F\colon Y \to \overline{\mathbb{R}}$, $G\colon X \to \overline{\mathbb{R}}$ are proper, convex, lower semicontinuous functionals. By Proposition 4.1, we may solve the following primal-dual formulation of (A.2) to obtain a solution of (A.2):

$$\min_{x \in X} \max_{y \in Y} \{L(x, y) := \langle Kx, y\rangle + G(x) - F^*(y)\}. \tag{A.3}$$

The primal-dual scheme for (A.3) is a combination of partial forward-backward steps for x and y with suitable step sizes. That is, primal-dual algorithms consist of the following building blocks:

$$\begin{aligned} \hat{x} &= \operatorname{prox}_{\tau G}(\bar{x} - \tau K^* \tilde{y}), \\ \hat{y} &= \operatorname{prox}_{\sigma F^*}(\bar{y} + \sigma K \tilde{x}), \end{aligned}$$

where $\hat{x}, \bar{x}, \tilde{x} \in X$, $\hat{y}, \bar{y}, \tilde{y} \in Y$, and $\tau, \sigma > 0$. Thus, primal-dual algorithms are useful when the proximal operators for both F and G can be computed efficiently. In addition, suitable over-relaxation steps should be accompanied in the algorithm to ensure convergence [62, Sect. 3]. In summary, the primal-dual algorithm to solve (A.3) is presented in Algorithm A.3 [15].

Algorithm A.3 Primal-dual algorithm

Choose $\tau, \sigma > 0$ with $\tau\sigma \leq 1/\|K\|^2$. Let $x^{(0)} = x^{(-1)} \in X$ and $y^{(0)} \in Y$.
for $n = 0, 1, 2, \ldots$
$y^{(n+1)} = \mathrm{prox}_{\sigma F^*}(y^{(n)} + \sigma K(2x^{(n)} - x^{(n-1)}))$
$x^{(n+1)} = \mathrm{prox}_{\tau G}(x^{(n)} - \tau K^* y^{(n+1)})$
end

In [63], it was shown that Algorithm A.3 is in fact equivalent to a preconditioned proximal point algorithm [64] applied to a monotone inclusion problem. The following ergodic convergence rate of Algorithm A.3 was proven in [65, Theorem 1].

Theorem A.3. *Let $\{x^{(n)}\}$ and $\{y^{(n)}\}$ be the sequences generated by Algorithm A.3. Then for any $(x, y) \in X \times Y$ and $n \geq 1$, it satisfies*

$$L\left(\frac{1}{n}\sum_{i=1}^{n} x^{(i)}, y\right) - L\left(x, \frac{1}{n}\sum_{i=1}^{n} y^{(i)}\right) \leq \frac{1}{n}\left(\frac{1}{\tau}\|x - x^{(0)}\|^2 + \frac{1}{\sigma}\|y - y^{(0)}\|^2\right).$$

Similarly to FISTA, Algorithm A.3 can be accelerated under additional assumptions. First, we assume that either G or F^* is strongly convex. By symmetry, it suffices to deal with the case when G is μ_G-strongly convex for some $\mu_G > 0$. Algorithm A.4 shows an accelerated primal-dual algorithm adapted in this case [65]. There are several alternative choices on the selection of step sizes; see, e.g., [15, Algorithm 2].

Algorithm A.4 $O(1/n^2)$-convergent accelerated primal-dual algorithm

Choose $\tau_0, \sigma_0 > 0$ with $\tau_0\sigma_0 \leq 1/\|K\|^2$ and $\theta_0 = 0$. Let $x^{(0)} = x^{(-1)} \in X$ and $y^{(0)} \in Y$.
for $n = 0, 1, 2, \ldots$
$y^{(n+1)} = \mathrm{prox}_{\sigma F^*}(y^{(n)} + \sigma K[(1 + \theta_n)x^{(n)} - \theta_n x^{(n-1)}])$
$x^{(n+1)} = \mathrm{prox}_{\tau G}(x^{(n)} - \tau K^* y^{(n+1)})$
$\theta_{n+1} = 1/\sqrt{1 + \mu_G\tau_n}$, $\tau_{n+1} = \theta_{n+1}\tau_n$, $\sigma_{n+1} = \sigma_n/\theta_{n+1}$
end

The convergence result for Algorithm A.4 is given in the following.

Theorem A.4. *Let $\{x^{(n)}\}$ and $\{y^{(n)}\}$ be the sequences generated by Algorithm A.4. Also, we define*

$$t_n = \frac{\sigma_{n-1}}{\sigma_0}, \quad T_n = \sum_{i=1}^{n} t_i, \quad n \geq 1.$$

Then for any $(x, y) \in X \times Y$ *and* $n \geq 1$*, it satisfies*

$$L\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i x^{(i)}, y\right) - L\left(x, \frac{1}{T_n}\sum_{i=1}^{n} t_i y^{(i)}\right) \leq \frac{1}{2T_n}\left(\frac{1}{\tau}\|x - x^{(0)}\|^2 + \frac{1}{\sigma}\|y - y^{(0)}\|^2\right).$$

Proof. See [65, Theorem 4]. In this case, we have $1/T_n = O(1/n^2)$ [15]. □

Next, we consider the case when both G and F^* are strongly convex; say G is μ_G-strongly convex and F^* is μ_{F^*}-strongly convex. Note that F^* is μ_{F^*}-strongly convex if and only if F is continuously differentiable and ∇F is Lipschitz continuous with a Lipschitz constant $1/\mu_{F^*}$. We present a particular choice of step sizes which was given in [42]. For the sake of convenience, let

$$\gamma = \frac{\mu_G \mu_{F^*}}{2\|K\|^2}\left(\sqrt{1 + \frac{4\|K\|^2}{\mu_G \mu_{F^*}}} - 1\right).$$

It is clear that $0 < \gamma < 1$. Algorithm A.5 shows a linearly convergent primal-dual algorithm for this case.

Algorithm A.5 Linearly convergent accelerated primal-dual algorithm

Choose $\tau = \frac{\gamma}{\mu_G(1-\gamma)}$, $\sigma = \frac{\gamma}{\mu_{F^*}(1-\gamma)}$, and $\theta = 1 - \gamma$. Let $x^{(0)} = x^{(-1)} \in X$ and $y^{(0)} \in Y$.
for $n = 0, 1, 2, \ldots$
$y^{(n+1)} = \operatorname{prox}_{\sigma F^*}(y^{(n)} + \sigma K[(1+\theta)x^{(n)} - \theta x^{(n-1)}])$
$x^{(n+1)} = \operatorname{prox}_{\tau G}(x^{(n)} - \tau K^* y^{(n+1)})$
end

The following ergodic convergence result for Algorithm A.5 is available.

Theorem A.5. *Let* $\{x^{(n)}\}$ *and* $\{y^{(n)}\}$ *be the sequences generated by Algorithm A.5. Also, we define*

$$t_n = \theta^{1-n}, \quad T_n = \sum_{i=1}^{n} t_i, \quad n \geq 1.$$

Then for any $(x, y) \in X \times Y$ *and* $n \geq 1$*, it satisfies*

$$L\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i x^{(i)}, y\right) - L\left(x, \frac{1}{T_n}\sum_{i=1}^{n} t_i y^{(i)}\right) \leq \frac{1}{2T_n}\left(\frac{1}{\tau}\|x - x^{(0)}\|^2 + \frac{1}{\sigma}\|y - y^{(0)}\|^2\right).$$

Proof. See [65, Theorem 5]. In this case, we have $1/T_n = O(\theta^n)$. □

ACKNOWLEDGMENT

The first author's work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2017R1A2B4011627). The second author's work was supported by NRF funded by the Ministry of Education (No. 2019R1A6A3A 01092549).

REFERENCES

[1] A. TOSELLI AND O. WIDLUND, *Domain Decomposition Methods-Algorithms and Theory*, vol. 34, Springer, Berlin, 2005.

[2] A. QUARTERONI AND A. VALLI, *Domain Decomposition Methods for Partial Differential Equations*, Oxford University Press, New York, 1999.

[3] J. XU, *Iterative methods by space decomposition and subspace correction*, SIAM Rev., 34 (1992), pp. 581–613.

[4] C. R. DOHRMANN, *A preconditioner for substructuring based on constrained energy minimization*, SIAM J. Sci. Comput., 25 (2003), pp. 246–258.

[5] J. MANDEL, *Balancing domain decomposition*, Commun. Numer. Methods Engrg., 9 (1993), pp. 233–241.

[6] C. FARHAT, M. LESOINNE, AND K. PIERSON, *A scalable dual-primal domain decomposition method*, Numer. Linear Algebra Appl., 7 (2000), pp. 687–714.

[7] C. FARHAT AND F.-X. ROUX, *A method of finite element tearing and interconnecting and its parallel solution algorithm*, Internat. J. Numer. Methods Engrg., 32 (1991), pp. 1205–1227.

[8] C.-O. LEE AND E.-H. PARK, *A dual iterative substructuring method with a penalty term*, Numer. Math., 112 (2009), pp. 89–113.

[9] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[10] D. STRONG AND T. CHAN, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Problems, 19 (2003), pp. S165–S187.

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

[12] 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.

[13] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[14] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for $L1$-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.

[15] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[16] 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.

[17] L. BADEA, *Convergence rate of a Schwarz multilevel method for the constrained minimization of nonquadratic functionals*, SIAM J. Numer. Anal., 44 (2006), pp. 449–477.

[18] L. BADEA AND R. KRAUSE, *One-and two-level Schwarz methods for variational inequalities of the second kind and their application to frictional contact*, Numer. Math., 120 (2012), pp. 573–599.

[19] X.-C. TAI AND J. XU, *Global and uniform convergence of subspace correction methods for some convex optimization problems*, Math. Comp., 71 (2001), pp. 105–124.

[20] M. FORNASIER, A. LANGER, AND C.-B. SCHÖNLIEB, *A convergent overlapping domain decomposition method for total variation minimization*, Numer. Math., 116 (2010), pp. 645–685.

[21] M. FORNASIER AND C.-B. SCHÖNLIEB, *Subspace correction methods for total variation and $l1$-minimization*, SIAM J. Numer. Anal., 47 (2009), pp. 3397–3428.

[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] 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.

[24] A. LANGER, S. OSHER, AND C.-B. SCHÖNLIEB, *Bregmanized domain decomposition for image restoration*, J. Sci. Comput., 54 (2013), pp. 549–576.
[25] C.-O. LEE, J. H. LEE, H. WOO, AND S. YUN, *Block decomposition methods for total variation by primal–dual stitching*, J. Sci. Comput., 68 (2016), pp. 273–302.
[26] C.-O. LEE AND C. NAM, *Primal domain decomposition methods for the total variation minimization, based on dual decomposition*, SIAM J. Sci. Comput., 39 (2017), pp. B403–B423.
[27] H. CHANG, X.-C. TAI, L.-L. WANG, AND D. YANG, *Convergence rate of overlapping domain decomposition methods for the Rudin–Osher–Fatemi model based on a dual formulation*, SIAM J. Imaging Sci., 8 (2015), pp. 564–591.
[28] 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.
[29] C.-O. LEE AND J. PARK, *Fast nonoverlapping block Jacobi method for the dual Rudin–Osher–Fatemi model*, SIAM J. Imaging Sci., 12 (2019), pp. 2009–2034.
[30] C.-O. LEE, E.-H. PARK, AND J. PARK, *A finite element approach for the dual Rudin–Osher–Fatemi model and its nonoverlapping domain decomposition methods*, SIAM J. Sci. Comput., 41 (2019), pp. B205–B228.
[31] C.-O. LEE AND J. PARK *A finite element nonoverlapping domain decomposition method with Lagrange multipliers for the dual total variation minimizations*, J. Sci. Comput., 81 (2019), pp. 2331–2355.
[32] Y. DUAN, H. CHANG, AND X.-C. TAI, *Convergent non-overlapping domain decomposition methods for variational image segmentation*, J. Sci. Comput., 69 (2016), pp. 532–555.
[33] T. F. CHAN, S. ESEDOGLU, AND M. NIKOLOVA, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math., 66 (2006), pp. 1632–1648.
[34] T. F. CHAN AND L. A. VESE, *Active contours without edges*, IEEE Trans. Image Process., 10 (2001), pp. 266–277.
[35] J. PARK, *An overlapping domain decomposition framework without dual formulation for variational imaging problems*. arXiv:2002.10070 [math.NA], 2019. To appear in Adv. Comput. Math.
[36] C.-O. LEE, C. NAM, AND J. PARK, *Domain decomposition methods using dual conversion for the total variation minimization with L^1 fidelity term*, J. Sci. Comput., 78 (2019), pp. 951–970.
[37] T. F. CHAN AND S. ESEDOḠLU, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.
[38] A. TIKHONOV, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Dokl., 4 (1963), pp. 1035–1038.
[39] G. AUBERT AND P. KORNPROBST, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Springer, New York, 2006.
[40] E. GIUSTI, *Minimal Surfaces and Functions of Bounded Variation*, Birkhäuser, Boston, 1984.
[41] X. FENG AND A. PROHL, *Analysis of total variation flow and its finite element approximations*, ESAIM Math. Model. Numer. Anal., 37 (2003), pp. 533–556.
[42] A. CHAMBOLLE AND T. POCK, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319.
[43] Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: the Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, vol. 22, American Mathematical Society, Providence, 2001.
[44] M. NIKOLOVA, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vision, 20 (2004), pp. 99–120.
[45] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, New Jersey, 2015.
[46] I. EKELAND AND R. TEMAM, *Convex Analysis and Variational Problems*, vol. 28, SIAM, Philadelphia, 1999.
[47] Y. DONG, M. HINTERMÜLLER, AND M. NERI, *An efficient primal-dual method for L^1TV image restoration*, SIAM J. Imaging Sci., 2 (2009), pp. 1168–1189.
[48] K. KUNISCH AND M. HINTERMÜLLER, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM J. Appl. Math., 64 (2004), pp. 1311–1333.

[49] L. BADEA, X.-C. TAI, AND J. WANG, *Convergence rate analysis of a multiplicative Schwarz method for variational inequalities*, SIAM J. Numer. Anal., 41 (2003), pp. 1052–1073.
[50] X.-C. TAI, *Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities*, Numer. Math., 93 (2003), pp. 755–786.
[51] J. BOLTE, S. SABACH, AND M. TEBOULLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., Ser. A, 146 (2014), pp. 459–494.
[52] A. CHAMBOLLE, AND T POCK, *A remark on accelerated block coordinate descent for computing the proximity operators of a sum of convex functions*, SMAI J. Comp. Math., 1 (2015), pp. 29–54.
[53] R. SHEFI AND M. TEBOULLE, *On the rate of convergence of the proximal alternating linearized minimization algorithm for convex problems*, EURO J. Comput. Optim., 4 (2016), pp. 27–46.
[54] A. LANGER AND F. GASPOZ, *Overlapping domain decomposition methods for total variation denoising*, SIAM J. Numer. Anal., 57 (2019), pp. 1411–1444.
[55] S. BARTELS, *Total variation minimization with finite elements: convergence and iterative solution*, SIAM J. Numer. Anal., 50 (2012), pp. 1162–1180.
[56] M. HERRMANN, R. HERZOG, S. SCHMIDT, J. VIDAL-NÚÑEZ, AND G. WACHSMUTH, *Discrete total variation with finite elements and applications to imaging*, J. Math. Imaging Vision, 61 (2019), pp. 411–431.
[57] P.-A. RAVIART AND J.-M. THOMAS, *A mixed finite element method for 2-nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, Springer, 1977, pp. 292–315.
[58] I. GOODFELLOW, Y. BENGIO, AND A. COURVILLE, *Deep Learning*, MIT Press, Cambridge, 2016.
[59] S. J. WRIGHT, *Coordinate descent algorithms*, Math. Program., 151 (2015), pp. 3–34.
[60] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
[61] Y. E. NESTEROV, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.
[62] B. HE, Y. YOU, AND X. YUAN, *On the convergence of primal-dual hybrid gradient algorithm*, SIAM J. Imaging Sci., 7 (2014), pp. 2526–2537.
[63] B. HE AND X. YUAN, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci., 5 (2012), pp. 119–149.
[64] R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
[65] A. CHAMBOLLE, AND T POCK, *On the ergodic convergence rates of a first-order primal–dual algorithm*, Math. Program., Ser. A, 159 (2016), pp. 253–287.