

PSEUDO-LINEAR CONVERGENCE OF AN ADDITIVE SCHWARZ METHOD FOR DUAL TOTAL VARIATION MINIMIZATION*

JONGHO PARK[†]

Abstract. In this paper, we propose an overlapping additive Schwarz method for total variation minimization based on a dual formulation. The $O(1/n)$ -energy convergence of the proposed method is proven, where n is the number of iterations. In addition, we introduce an interesting convergence property of the proposed method called *pseudo-linear convergence*; the energy decreases as fast as for linearly convergent algorithms until it reaches a particular value. It is shown that this particular value depends on the overlapping width δ , and the proposed method becomes as efficient as linearly convergent algorithms if δ is large. As the latest domain decomposition methods for total variation minimization are sublinearly convergent, the proposed method outperforms them in the sense of the energy decay. Numerical experiments which support our theoretical results are provided.

Key words. domain decomposition method, additive Schwarz method, total variation minimization, Rudin–Osher–Fatemi model, convergence rate

AMS subject classifications. 65N55, 65Y05, 65K15, 68U10

1. Introduction. This paper is concerned with numerical solutions of total variation minimization by additive Schwarz methods as overlapping domain decomposition methods (DDMs). Total variation minimization was introduced first by Rudin, Osher, and Fatemi [25], and it has become one of the fundamental problems in mathematical imaging. Let $\Omega \subset \mathbb{R}^2$ be a bounded rectangular domain. The total variation minimization model problem on Ω is given by

$$(1.1) \quad \min_{u \in BV(\Omega)} \{F(u) + TV_{\Omega}(u)\},$$

where $F(u)$ is a convex function, $TV_{\Omega}(u)$ is the total variation of u on Ω defined by

$$TV_{\Omega}(u) = \sup \left\{ \int_{\Omega} u \operatorname{div} \mathbf{p} \, dx : \mathbf{p} \in (C_0^1(\Omega))^2 \text{ such that } |\mathbf{p}(x)| \leq 1 \text{ for almost all } x \in \Omega \right\},$$

with $|\mathbf{p}(x)| = \sqrt{p_1(x)^2 + p_2(x)^2}$, and $BV(\Omega)$ is the space of functions in $L^1(\Omega)$ with finite total variation. Equation (1.1) contains an extensive range of problems arising in mathematical imaging. For example, if we set $F(u) = \frac{\lambda}{2} \int_{\Omega} (u - f)^2 \, dx$ in (1.1) for $\lambda > 0$ and $f \in L^2(\Omega)$, then we get the celebrated Rudin–Osher–Fatemi (ROF) model [25]:

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

In the perspective of image processing, a solution u of (1.2) is a denoised image obtained from the noisy image f . A more complex example of (1.1) is the TV - H^{-1} model [6, 23, 26]:

$$(1.3) \quad \min_{u \in BV(\Omega)} \left\{ \frac{\lambda}{2} \int_{\Omega} |\nabla(-\Delta)^{-1}(u - f)|^2 \, dx + TV_{\Omega}(u) \right\},$$

where $\lambda > 0$ and $(-\Delta)^{-1}$ denotes the inverse of the negative Laplacian with homogeneous Dirichlet boundary conditions. Since (1.3) is known to incorporate several desirable properties

*Received November 25, 2019. Accepted December 5, 2020. Published online on February 3, 2021. Recommended by Martin Gander. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2019R1A6A3A01092549).

[†]Department of Mathematical Sciences, KAIST, Daejeon 34141, Korea (jongho.park@kaist.ac.kr).

TABLE 1.1
Difficulties in designing DDMs for (1.2) and (1.5).

problem	obstacle	grad-div
energy	$\frac{1}{2} \int_{\Omega} \nabla u ^2 dx - \int_{\Omega} f u dx$	$\frac{1}{2} \int_{\Omega} [(\operatorname{div} \mathbf{p})^2 + \mathbf{p} ^2] dx - \int_{\Omega} \mathbf{f} \cdot \mathbf{p} dx$
solution	scalar-valued, $H_0^1(\Omega)$	vector-valued , $H_0(\operatorname{div}; \Omega)$
constraint	yes	no
strong convexity	yes	yes
smoothness	yes	yes
separability	yes	yes
Schwarz methods	[2, 27]	[22, 29]
convergence to a global minimizer	linear	linear
problem	ROF (1.2)	dual ROF (1.5)
energy	$\frac{\lambda}{2} \int_{\Omega} (u - f)^2 dx + TV_{\Omega}(u)$	$\frac{1}{2\lambda} \int_{\Omega} (\operatorname{div} \mathbf{p} + \lambda f)^2 dx$
solution	scalar-valued, $BV(\Omega)$	vector-valued , $H_0(\operatorname{div}; \Omega)$
constraint	no	yes
strong convexity	yes	no
smoothness	no	yes
separability	no	yes
Schwarz methods	primal decomposition: [12, 13] dual decomposition: [16, 17]	[10, 15]
convergence to a global minimizer	primal decomposition: not guaranteed [17] dual decomposition: convergent [16, 17]	sublinear [10, 19]

of higher-order variational models for imaging, such as the smooth connection of shapes, it has various applications in advanced imaging problems including image decomposition [23] and image inpainting [6]. One may refer to [9] for various other examples of (1.1).

In view of designing DDMs, several difficulties lie in the total variation term in (1.1). The total variation functional is nonsmooth, i.e., it has no gradient, so that a careful consideration is required to solve (1.1). Furthermore, since it measures the jumps of a function across edges, it is nonseparable in the sense that

$$TV_{\Omega}(u) \neq \sum_{i=1}^N TV_{\Omega_i}(u)$$

for a nonoverlapping partition $\{\Omega_i\}_{i=1}^N$ of Ω in general. Due to those characteristics, it is challenging to design Schwarz methods for (1.1) that converge to a correct solution. Indeed, it was shown in [17] that Schwarz methods for (1.2), which is a special case of (1.1), introduced in [12, 13] may not converge to a correct minimizer. We also point out that the Schwarz framework for nonsmooth convex optimization proposed in [1] does not apply to (1.1) since $TV_{\Omega}(u)$ does not satisfy the condition in [1, equation (7)]; see [17, Claim 6.1].

Instead of (1.1), one may consider a Fenchel–Rockafellar dual formulation (see, e.g., [9]) of (1.1), which is given by

$$\min_{\mathbf{p} \in (C_0^1(\Omega))^2} F^*(\operatorname{div} \mathbf{p}) \quad \text{subject to } |\mathbf{p}(x)| \leq 1, \quad \forall x \in \Omega,$$

or an alternative formulation

$$(1.4) \quad \min_{\mathbf{p} \in H_0(\operatorname{div}; \Omega)} F^*(\operatorname{div} \mathbf{p}) \quad \text{subject to } |\mathbf{p}(x)| \leq 1, \quad \forall x \in \Omega,$$

in which the solution space is replaced by an appropriate Hilbert space

$$H_0(\operatorname{div}; \Omega) = \{\mathbf{p} \in (L^2(\Omega))^2 : \operatorname{div} \mathbf{p} \in L^2(\Omega) \text{ and } \mathbf{p} \cdot \mathbf{n} = 0 \text{ on } \partial\Omega\}$$

and where F^* is the Legendre–Fenchel conjugate of F and \mathbf{n} is the outer normal to $\partial\Omega$. In particular, a dual formulation of (1.2) is given by

$$(1.5) \quad \min_{\mathbf{p} \in H_0(\operatorname{div}; \Omega)} \frac{1}{2\lambda} \int_{\Omega} (\operatorname{div} \mathbf{p} + \lambda f)^2 dx \quad \text{subject to } |\mathbf{p}(x)| \leq 1, \quad \forall x \in \Omega.$$

The above dual formulation was first considered in [7]. If one has a solution of the dual problem (1.4), then a solution of the primal problem (1.1) can be easily obtained by the primal-dual relation; see [9]. One readily sees that (1.4) is a constrained minimization problem. We note that the energy functional of (1.4) is not strongly convex. Even for (1.2), where F is smooth, the energy functional of its dual problem (1.5) is not strongly convex due to the div-operator therein. Hence, Schwarz methods proposed in [2, 27] for constrained optimization (in particular, the obstacle problem) are not valid for either (1.4) or (1.5). Moreover, (1.4) is a vector-valued problem related to the div-operator; it is usually more difficult to design DDMs for vector-valued problems than for scalar-valued ones because of the huge null space of the div-operator [22, 29]. The above mentioned difficulties for (1.2) and (1.5), which are special cases of (1.1) and (1.4), respectively, are summarized in Table 1.1 with comparisons to some related problems in structural mechanics: the obstacle problem and the grad-div problem.

Despite these difficulties, several successful Schwarz methods for (1.5) have been developed [10, 15, 19]. In [15], subspace correction methods for (1.5) based on a nonoverlapping domain decomposition were proposed. Since then, the $O(1/n)$ -energy convergence of overlapping Schwarz methods for (1.5) in a continuous setting was derived in [10], where n is the number of iterations. In [19], it was shown that the methods proposed in [15] are also $O(1/n)$ -convergent. In addition, an $O(1/n^2)$ -convergent additive method was designed using an idea of pre-relaxation. Inspired by the dual problem (1.5), Schwarz methods for (1.2) based on dual decomposition were considered in [16, 17]. Recently, several iterative substructuring methods for more general problems of the form (1.4) were considered [18, 20].

In this paper, we propose an additive Schwarz method for (1.4) based on an overlapping domain decomposition. While the existing methods in [15, 19] for (1.5) are based on finite difference discretizations, the proposed method is based on finite element discretizations that were recently proposed in [14, 18, 20]. Compared to the methods in [10], the proposed method has the advantage that it does not depend on either a particular function decomposition or a constraint decomposition. We prove that the proposed method is $O(1/n)$ -convergent similarly to the existing methods in [10, 19]. In addition, we explicitly describe the dependency of the convergence rate on the condition number of F . We investigate another interesting convergence property of the proposed method, which we call *pseudo-linear* convergence. The precise definition of pseudo-linear convergence is given as follows:

DEFINITION 1.1. *A sequence $\{a_n\}_{n \geq 0}$ of positive real numbers is said to converge pseudo-linearly to 0 at rate γ with threshold ϵ if a_n converges to 0 as n tends to ∞ , and there exist constants $0 < \gamma < 1$, $c > 0$, and $\epsilon > 0$ such that*

$$a_n \leq \gamma^n c + \epsilon, \quad \forall n \geq 0.$$

Note that the above definition reduces to ordinary linear convergence if $\epsilon = 0$.

With a suitable overlapping domain decomposition, it is shown that the proposed method is pseudo-linearly convergent with threshold $O(|\Omega|/\delta^2)$, where δ is the overlapping width parameter of the domain decomposition. Therefore, the proposed method is expected to converge to a minimizer much more rapidly than other sublinearly convergent methods if δ is large. We provide numerical experiments which illustrate this behaviour.

The rest of this paper is organized as follows. In Section 2, we present finite element discretizations for dual total variation minimization and an abstract space decomposition

setting. An abstract additive Schwarz method is introduced in Section 3 with the corresponding convergence results. In Section 4, overlapping domain decomposition settings for the proposed method are presented. Applications of the proposed method to various imaging problems of the form (1.1) are provided in Section 5. We conclude the paper with some remarks in Section 6.

2. General setting. First, we briefly review finite element discretizations proposed in [18, 20] for (1.4). All the results in this paper can naturally be generalized to polygonal domains with quasi-uniform meshes, but we restrict our discussion to rectangular domains with uniform meshes for simplicity; one may refer [14] for more general finite element discretizations.

Each pixel in an image is regarded as a square finite element whose side length equals h . Then the image domain composed of $m_1 \times m_2$ pixels becomes a rectangular region $(0, m_1 h) \times (0, m_2 h)$ in \mathbb{R}^2 . Let \mathcal{T}_h be the collection of all elements in Ω , and let \mathcal{E}_h be the collection of all interior element edges. The lowest-order Raviart–Thomas finite element space [24] on Ω with homogeneous essential boundary conditions is given by

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

where $\mathcal{RT}_0(T)$ is the collection of all vector fields $\mathbf{p}: T \rightarrow \mathbb{R}^2$ of the form

$$(2.1) \quad \mathbf{p}(x_1, x_2) = \begin{bmatrix} a_1 + b_1 x_1 \\ a_2 + b_2 x_2 \end{bmatrix}.$$

The degrees of freedom for Y_h are given by the average values of the normal components over the element edges. We denote the degree of freedom of $\mathbf{p} \in Y_h$ associated to an edge $e \in \mathcal{E}_h$ by $(\mathbf{p})_e$, i.e.,

$$(\mathbf{p})_e = \frac{1}{|e|} \int_e \mathbf{p} \cdot \mathbf{n}_e \, ds,$$

where \mathbf{n}_e is a unit normal to e . We define the space X_h by

$$X_h = \{u \in L^2(\Omega) : u|_T \text{ is constant}, \quad \forall T \in \mathcal{T}_h\}.$$

Then it is clear that $\operatorname{div} \mathbf{p} \in X_h$ for $\mathbf{p} \in Y_h$. Obviously, the degrees of freedom for X_h are the values on the elements; for $u \in X_h$, $T \in \mathcal{T}_h$, and $x_T \in T$, we write

$$(u)_T = u(x_T).$$

Let $\Pi_h: H_0(\operatorname{div}; \Omega) \rightarrow Y_h$ be the nodal interpolation operator, i.e., it satisfies

$$(\Pi_h \mathbf{p})_e = \frac{1}{|e|} \int_e \mathbf{p} \cdot \mathbf{n}_e \, ds, \quad \mathbf{p} \in H_0(\operatorname{div}; \Omega), e \in \mathcal{E}_h.$$

Then the following estimate holds [22, Lemma 5.1]:

LEMMA 2.1. *Let $\mathbf{p} \in Y_h$ and θ be any continuous, piecewise linear, scalar function supported in $S \subset \Omega$. Then, there exists a constant $c > 0$ independent of $|\Omega|$ and h such that*

$$\int_S [\operatorname{div}(\Pi_h(\theta \mathbf{p}))]^2 \, dx \leq c \int_S [\operatorname{div}(\theta \mathbf{p})]^2 \, dx.$$

The spaces X_h and Y_h are equipped with the inner products

$$\begin{aligned}\langle u, v \rangle_{X_h} &= h^2 \sum_{T \in \mathcal{T}_h} (u)_T (v)_T, & u, v \in X_h, \\ \langle \mathbf{p}, \mathbf{q} \rangle_{Y_h} &= h^2 \sum_{e \in \mathcal{E}_h} (\mathbf{p})_e (\mathbf{q})_e, & \mathbf{p}, \mathbf{q} \in Y_h,\end{aligned}$$

and their induced norms $\|\cdot\|_{X_h}$ and $\|\cdot\|_{Y_h}$, respectively. We have the following facts for the norms $\|\cdot\|_{X_h}$ and $\|\cdot\|_{Y_h}$.

LEMMA 2.2. *The norm $\|\cdot\|_{X_h}$ agrees with the $L^2(\Omega)$ -norm in X_h , i.e.,*

$$\|u\|_{X_h} = \|u\|_{L^2(\Omega)}, \quad \forall u \in X_h.$$

In addition, the norm $\|\cdot\|_{Y_h}$ is equivalent to the $(L^2(\Omega))^2$ -norm in Y_h in the sense that there exist constants $\underline{c}, \bar{c} > 0$ independent of $|\Omega|$ and h such that

$$\underline{c} \|\mathbf{p}\|_{(L^2(\Omega))^2} \leq \|\mathbf{p}\|_{Y_h} \leq \bar{c} \|\mathbf{p}\|_{(L^2(\Omega))^2}, \quad \forall \mathbf{p} \in Y_h.$$

Proof. See [18, Remark 2.2]. \square

In this setting, the following inverse inequality which is useful for the selection of parameters for solvers for total variation minimization (e.g. [4]) holds:

PROPOSITION 2.3. *For any $\mathbf{p} \in Y_h$, we have*

$$\|\operatorname{div} \mathbf{p}\|_{X_h}^2 \leq \frac{8}{h^2} \|\mathbf{p}\|_{Y_h}^2.$$

Proof. See [18, Proposition 2.5]. \square

For the rest of the paper, we may omit the subscripts X_h and Y_h from $\|\cdot\|_{X_h}$ and $\|\cdot\|_{Y_h}$, respectively, if there is no ambiguity. We define the subset C of Y_h as

$$C = \{\mathbf{p} \in Y_h : |(\mathbf{p})_e| \leq 1, \quad \forall e \in \mathcal{E}_h\}.$$

Then, we have the following discrete version of (1.4):

$$(2.2) \quad \min_{\mathbf{p} \in C} \{\mathcal{F}(\mathbf{p}) := F^*(\operatorname{div} \mathbf{p})\}.$$

One may refer to [14, 20] for further details on (2.2). In the sequel, we denote a solution of (2.2) by $\mathbf{p}^* \in Y_h$. In order to design a convergent additive Schwarz method for (2.2), we require the following assumption on F , which is common in the literature on Schwarz methods; see, e.g., [2, 27].

ASSUMPTION 2.4. The function F in (2.2) is α -strongly convex for some $\alpha > 0$, i.e., the map

$$u \mapsto F(u) - \frac{\alpha}{2} \|u\|^2$$

is convex. In addition, F is Frechét-differentiable, and its derivative F' is β -Lipschitz continuous for some $\beta > 0$, i.e.,

$$\|F'(u) - F'(v)\| \leq \beta \|u - v\|, \quad \forall u, v \in X_h.$$

We define the condition number κ of F by

$$(2.3) \quad \kappa = \frac{\beta}{\alpha},$$

where α and β are given in Assumption 2.4. In the case when $F(u) = \frac{1}{2} \langle u, Ku \rangle - \langle g, u \rangle$ for some symmetric positive definite operator $K: X_h \rightarrow X_h$ and $g \in X_h$, it is straightforward to see that κ agrees with the ratio of the extremal eigenvalues of K .

There are several interesting examples of F satisfying Assumption 2.4. First, we consider the ROF model (1.2), i.e., $F(u) = \frac{\lambda}{2} \|u - f\|^2$ for $\lambda > 0$ and $f \in X_h$. Recall that the discrete ROF model defined on X_h is given by

$$(2.4) \quad \min_{u \in X_h} \left\{ \frac{\lambda}{2} \|u - f\|^2 + TV_{\Omega}(u) \right\}.$$

It is clear that Assumption 2.4 is satisfied with $\alpha = \beta = \lambda$ and the condition number $\kappa = \beta/\alpha = 1$. The second example is the $TV-H^{-1}$ model: a natural discretization of (1.3) is given by

$$(2.5) \quad \min_{u \in X_h} \left\{ \frac{\lambda}{2} \|u - f\|_{K^{-1}}^2 + TV_{\Omega}(u) \right\},$$

where $K: X_h \rightarrow X_h$ is the standard 5-point-stencil approximation of $-\Delta$ with homogeneous essential boundary conditions [21] and $\|v\|_{K^{-1}} = \langle K^{-1}v, v \rangle^{1/2}$ for $v \in X_h$. Since K is nonsingular, (2.5) satisfies Assumption 2.4. It is well-known that the condition number of K becomes larger as the image size grows; a detailed estimate for the condition number can be found in [21]. We note that a domain decomposition method for the $TV-H^{-1}$ model was previously considered in [26].

If F satisfies Assumption 2.4, then we have the following properties for F^* [9]:

PROPOSITION 2.5. *Under Assumption 2.4, the function F^* in (2.2) is $(1/\beta)$ -strongly convex and Fréchet-differentiable with a $(1/\alpha)$ -Lipschitz continuous derivative. Equivalently, the following hold:*

$$\begin{aligned} F^*(u) - F^*(v) - \langle (F^*)'(v), u - v \rangle &\geq \frac{1}{2\beta} \|u - v\|^2, & \forall u, v \in X_h, \\ \|(F^*)'(u) - (F^*)'(v)\| &\leq \frac{1}{\alpha} \|u - v\|, & \forall u, v \in X_h. \end{aligned}$$

A solution of a discrete primal problem

$$\min_{u \in X_h} \{F(u) + TV_{\Omega}(u)\}$$

can be obtained from the solution \mathbf{p}^* of (2.2) by solving

$$(2.6) \quad \min_{u \in X_h} \{F(u) - \langle u, \operatorname{div} \mathbf{p}^* \rangle\}.$$

See [9] for details. Under Assumption 2.4, problem (2.6) is smooth and strongly convex. Therefore, one can solve (2.6) efficiently by linearly convergent first-order methods such as [9, Algorithm 5].

The Bregman distance [5] associated with \mathcal{F} is denoted by $D_{\mathcal{F}}$, i.e.,

$$(2.7) \quad D_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) = \mathcal{F}(\mathbf{p}) - \mathcal{F}(\mathbf{q}) - \langle \mathcal{F}'(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle, \quad \mathbf{p}, \mathbf{q} \in Y_h.$$

Note that

$$\mathcal{F}'(\mathbf{p}) = \operatorname{div}^* ((F^*)'(\operatorname{div} \mathbf{p})), \quad \mathbf{p} \in Y_h,$$

where $\operatorname{div}^*: X_h \rightarrow Y_h$ is the adjoint of $\operatorname{div}: Y_h \rightarrow X_h$. We have the following useful property of $D_{\mathcal{F}}$ [11, Lemma 3.1].

LEMMA 2.6. *For any $\mathbf{p}, \mathbf{q}, \mathbf{r} \in Y_h$, we have*

$$D_{\mathcal{F}}(\mathbf{r}, \mathbf{p}) - D_{\mathcal{F}}(\mathbf{r}, \mathbf{q}) = D_{\mathcal{F}}(\mathbf{q}, \mathbf{p}) - \langle \mathcal{F}'(\mathbf{p}) - \mathcal{F}'(\mathbf{q}), \mathbf{r} - \mathbf{q} \rangle.$$

For later use, we state the following trivial lemma for the set C .

LEMMA 2.7. *For any $\mathbf{p}, \mathbf{q} \in C$, we have*

$$\|\mathbf{p} - \mathbf{q}\|^2 \leq 8|\Omega|.$$

Proof. Note that $|\Omega| = m_1 m_2 h^2$ and $|\mathcal{E}_h| = (m_1 - 1)m_2 + m_1(m_2 - 1)$ for an image of size $m_1 \times m_2$. Using the inequality

$$\|\mathbf{p} - \mathbf{q}\|^2 = h^2 \sum_{e \in \mathcal{E}_h} [(\mathbf{p})_e - (\mathbf{q})_e]^2 \leq 4h^2 |\mathcal{E}_h|,$$

the conclusion is easily obtained. \square

Next, we present a space decomposition setting for $W = Y_h$. Let $W_k, k = 1, \dots, N$, be subspaces of W such that

$$(2.8) \quad W = \sum_{k=1}^N R_k^* W_k,$$

where $R_k: W \rightarrow W_k$ is the restriction operator, and its adjoint $R_k^*: W_k \rightarrow W$ is the natural extension operator. We state an additional assumption on (2.8) inspired by [2].

ASSUMPTION 2.8. There exist constants $c_1 > 0$ and $c_2 \geq 0$ which satisfy the following conditions: for any $\mathbf{p}, \mathbf{q} \in C$, there exists $\mathbf{r}_k \in W_k$ ($k = 1, \dots, N$) such that

$$(2.9a) \quad \mathbf{p} - \mathbf{q} = \sum_{k=1}^N R_k^* \mathbf{r}_k,$$

$$(2.9b) \quad \mathbf{q} + R_k^* \mathbf{r}_k \in C, \quad k = 1, \dots, N,$$

$$(2.9c) \quad \sum_{k=1}^N \|\operatorname{div} R_k^* \mathbf{r}_k\|^2 \leq c_1 \|\operatorname{div}(\mathbf{p} - \mathbf{q})\|^2 + c_2 \|\mathbf{p} - \mathbf{q}\|^2.$$

In Assumption 2.8, we call $\{\mathbf{r}_k\}$ a *stable decomposition* of $\mathbf{p} - \mathbf{q}$. A particular choice of spaces $\{W_k\}$ and functions $\{\mathbf{r}_k\}$ satisfying Assumption 2.8 based on an overlapping domain decomposition of Ω will be given in Section 4.

We conclude this section by presenting two useful estimates for sequences of positive real numbers.

LEMMA 2.9. *Let $\{a_n\}_{n \geq 0}$ be a sequence of positive real numbers which satisfy*

$$a_n - a_{n+1} \geq \frac{1}{c^2} (a_{n+1} - \gamma a_n)^2, \quad n \geq 0,$$

for some $c > 0$ and $0 \leq \gamma < 1$. Then we have

$$a_n \leq \frac{1}{\tilde{c}n + 1/a_0},$$

where

$$\tilde{c} = \frac{(1 - \gamma)^2}{2a_0(1 - \gamma)^2 + (\gamma\sqrt{a_0} + c)^2}.$$

Proof. See [10, Lemma 3.5]. \square

LEMMA 2.10. Let $\{a_n\}_{n \geq 0}$ be a sequence of positive real numbers which satisfy

$$a_{n+1} \leq \gamma a_n + c, \quad n \geq 0,$$

for some $c > 0$ and $0 \leq \gamma < 1$. Then we have

$$a_n \leq \gamma^n \left(a_0 - \frac{c}{1 - \gamma} \right) + \frac{c}{1 - \gamma}.$$

Proof. It is elementary. \square

REMARK 2.11. In [10, Lemma 3.5], it was proved that the sequence in Lemma 2.9 satisfies

$$(2.10) \quad a_n - a_{n+1} \geq \left(\frac{1 - \gamma}{\gamma\sqrt{a_0} + c} \right)^2 a_{n+1}^2.$$

Then (2.10) was combined with [28, Lemma 3.2] to yield the desired result. We note that several alternative estimates for the $O(1/n)$ -convergence of (2.10) with different constants are available; see [3, Lemmas 3.6 and 3.8] and [8, Proposition 3.1].

3. An additive Schwarz method. In this section, we propose an abstract additive Schwarz method for dual total variation minimization (2.2) in terms of the space decomposition (2.8). Then several remarkable convergence properties of the proposed method are investigated. Algorithm 1 shows the proposed additive Schwarz method for (2.2).

Algorithm 1 Additive Schwarz method for dual total variation minimization (2.2).

Choose $\mathbf{p}^{(0)} \in C$ and $\tau \in (0, 1/N]$.

for $n = 0, 1, 2, \dots$

$$\mathbf{r}_k^{(n+1)} \in \arg \min_{\mathbf{r}_k \in W_k, \mathbf{p}^{(n)} + R_k^* \mathbf{r}_k \in C} \mathcal{F} \left(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k \right), \quad k = 1, \dots, N$$

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

end

We note that an additive Schwarz method for the dual ROF model based on a constraint decomposition of C was proposed in [10], which is slightly different from our method. Differently from the method in [10], Algorithm 1 does not require an explicit decomposition

of the constraint set C as in [10, Proposition 2.1]. A similar situation was previously addressed in [2, 27] for the obstacle problem. In addition, the proposed method is applicable to not only the ROF model but also general total variation minimization problems that satisfy Assumption 2.4.

In order to analyze the convergence of Algorithm 1, we first present a descent rule for Algorithm 1 in Lemma 3.1, which is a main tool for the convergence analysis. We note that arguments using the descent rule are standard in the analysis of first-order methods for convex optimization (see [4, Lemma 2.3], [8, equation (3.6)], and [9, equations (4.37) and (C.1)], for instance). The proof of Lemma 3.1 closely follows that of [19, Lemma 3.2]. The main difference is that Lemma 3.1 allows any smooth \mathcal{F} , using the notion of Bregman distance, while [19] is restricted to the dual ROF case. In addition, the decomposition of $\mathbf{p} - \mathbf{q}$ is not unique due to the overlapping of the subdomains, while it is unique in the nonoverlapping case given in [19].

LEMMA 3.1. *Let $\mathbf{p}, \mathbf{q} \in C$ and $\tau \in (0, 1/N]$. We define $\hat{\mathbf{r}}_k \in W_k$, $\hat{\mathbf{q}} \in Y$, and $\bar{\mathbf{r}}_k \in W_k$, $k = 1, \dots, N$, as follows:*

- (i) $\hat{\mathbf{r}}_k \in \arg \min_{\mathbf{r}_k \in W_k, \mathbf{q} + R_k^* \mathbf{r}_k \in C} \mathcal{F}(\mathbf{q} + R_k^* \mathbf{r}_k)$, $k = 1, \dots, N$.
- (ii) $\hat{\mathbf{q}} = \mathbf{q} + \tau \sum_{k=1}^N R_k^* \hat{\mathbf{r}}_k$.
- (iii) $\{\bar{\mathbf{r}}_k\}$: a stable decomposition of $\mathbf{p} - \mathbf{q}$ as in Assumption 2.8.

Then we have

$$\tau \mathcal{F}(\mathbf{p}) + (1 - \tau) \mathcal{F}(\mathbf{q}) - \mathcal{F}(\hat{\mathbf{q}}) \geq \tau \sum_{k=1}^N (D_{\mathcal{F}}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k, \mathbf{q} + R_k^* \hat{\mathbf{r}}_k) - D_{\mathcal{F}}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k, \mathbf{q})).$$

Proof. As $\mathbf{q} + R_k^* \bar{\mathbf{r}}_k \in C$ by (2.9b), from the optimality condition of $\hat{\mathbf{r}}_k$ (cf. [19, Lemma 2.2]), we have

$$\mathcal{F}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k) - \mathcal{F}(\mathbf{q} + R_k^* \hat{\mathbf{r}}_k) \geq D_{\mathcal{F}}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k, \mathbf{q} + R_k^* \hat{\mathbf{r}}_k).$$

Summation of the above equation over $k = 1, \dots, N$ yields

$$(3.1) \quad \tau \sum_{k=1}^N (\mathcal{F}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k) - \mathcal{F}(\mathbf{q} + R_k^* \hat{\mathbf{r}}_k)) \geq \tau \sum_{k=1}^N D_{\mathcal{F}}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k, \mathbf{q} + R_k^* \hat{\mathbf{r}}_k).$$

Note that

$$\hat{\mathbf{q}} = \mathbf{q} + \tau \sum_{k=1}^N R_k^* \hat{\mathbf{r}}_k = (1 - \tau N) \mathbf{q} + \tau \sum_{k=1}^N (\mathbf{q} + R_k^* \hat{\mathbf{r}}_k).$$

Since $1 - \tau N \geq 0$, we obtain by convexity of \mathcal{F} that

$$(3.2) \quad (1 - \tau N) \mathcal{F}(\mathbf{q}) + \tau \sum_{k=1}^N \mathcal{F}(\mathbf{q} + R_k^* \hat{\mathbf{r}}_k) \geq \mathcal{F}(\hat{\mathbf{q}}).$$

On the other hand, by the definition of Bregman distance (2.7), condition (iii), and the convexity of \mathcal{F} , we have

$$\begin{aligned}
 (3.3) \quad \tau \left(N\mathcal{F}(\mathbf{q}) - \sum_{k=1}^N \mathcal{F}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k) \right) &= -\tau \sum_{k=1}^N \left(\langle \mathcal{F}'(\mathbf{q}), R_k^* \bar{\mathbf{r}}_k \rangle + D_{\mathcal{F}}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k, \mathbf{q}) \right) \\
 &= -\tau \langle \mathcal{F}'(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle - \tau \sum_{k=1}^N D_{\mathcal{F}}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k, \mathbf{q}) \\
 &\geq -\tau(\mathcal{F}(\mathbf{p}) - \mathcal{F}(\mathbf{q})) - \tau \sum_{k=1}^N D_{\mathcal{F}}(\mathbf{q} + R_k^* \bar{\mathbf{r}}_k, \mathbf{q}).
 \end{aligned}$$

Summation of (3.1), (3.2), and (3.3) completes the proof. \square

With Lemma 3.1, the following property of sufficient decrease of \mathcal{F} for Algorithm 1 is straightforward (cf. [10, Lemma 3.3]).

LEMMA 3.2. *In Algorithm 1, we have*

$$\mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^{(n+1)}) \geq \frac{\tau}{2\beta} \sum_{k=1}^N \|\operatorname{div} R_k^* \mathbf{r}_k^{(n+1)}\|^2, \quad n \geq 0,$$

where β is given in Assumption 2.4.

Proof. Substitute \mathbf{p} by $\mathbf{p}^{(n)}$, \mathbf{q} by $\mathbf{p}^{(n)}$, and $\bar{\mathbf{r}}_k$ by 0 in Lemma 3.1. Then $\hat{\mathbf{r}}_k = \mathbf{r}_k^{(n+1)}$, $\hat{\mathbf{q}} = \mathbf{p}^{(n+1)}$, and we obtain

$$\mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^{(n+1)}) \geq \tau \sum_{k=1}^N D_{\mathcal{F}}(\mathbf{p}^{(n)}, \mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}).$$

On the other hand, for any k , it follows by Proposition 2.5 that

$$\begin{aligned}
 &D_{\mathcal{F}}(\mathbf{p}^{(n)}, \mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) \\
 &= \mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) - \left\langle \mathcal{F}'(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}), -R_k^* \mathbf{r}_k^{(n+1)} \right\rangle \\
 &= F^*(\operatorname{div} \mathbf{p}^{(n)}) - F^*(\operatorname{div}(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)})) \\
 &\quad - \left\langle (F^*)'(\operatorname{div}(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)})), -\operatorname{div} R_k^* \mathbf{r}_k^{(n+1)} \right\rangle \\
 &\geq \frac{1}{2\beta} \|\operatorname{div} R_k^* \mathbf{r}_k^{(n+1)}\|^2.
 \end{aligned}$$

Thus, we readily get the desired result. \square

Using Lemmas 3.1 and 3.2, one can show $O(1/n)$ -convergence of Algorithm 1 as follows:

THEOREM 3.3. *In Algorithm 1, let $\zeta_n = \alpha(\mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^*))$, for $n \geq 0$. Then there exists a constant $\tilde{c} > 0$ depending only on $|\Omega|$, κ , τ , ζ_0 , c_1 , and c_2 such that*

$$\zeta_n \leq \frac{1}{\tilde{c}n + 1/\zeta_0},$$

where κ is given in (2.3) and c_1, c_2 are given in Assumption 2.8.

Proof. In Lemma 3.1, replace \mathbf{p} by \mathbf{p}^* , \mathbf{q} by $\mathbf{p}^{(n)}$, and $\hat{\mathbf{r}}_s$ by $\mathbf{r}_s^{(n+1)}$. Then $\hat{\mathbf{q}} = \mathbf{p}^{(n+1)}$. We obtain that

$$\begin{aligned} & \tau \mathcal{F}(\mathbf{p}^*) + (1 - \tau) \mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^{(n+1)}) \\ & \geq \tau \sum_{k=1}^N \left(D_{\mathcal{F}}(\mathbf{p}^{(n)} + R_k^* \bar{\mathbf{r}}_k, \mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) - D_{\mathcal{F}}(\mathbf{p}^{(n)} + R_k^* \bar{\mathbf{r}}_k, \mathbf{p}^{(n)}) \right) \\ & = \tau \sum_{k=1}^N \left(D_{\mathcal{F}}(\mathbf{p}^{(n)}, \mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) - \left\langle \mathcal{F}'(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) - \mathcal{F}'(\mathbf{p}^{(n)}), R_k^* \bar{\mathbf{r}}_k \right\rangle \right), \end{aligned}$$

where the last equality is due to Lemma 2.6. It follows by Proposition 2.5 and the Cauchy–Schwarz inequality that

$$\begin{aligned} & \zeta_{n+1} - (1 - \tau) \zeta_n \\ & \leq -\tau \alpha \sum_{k=1}^N \left(D_{\mathcal{F}}(\mathbf{p}^{(n)}, \mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) \right. \\ & \quad \left. - \left\langle \mathcal{F}'(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) - \mathcal{F}'(\mathbf{p}^{(n)}), R_k^* \bar{\mathbf{r}}_k \right\rangle \right) \\ & \leq \tau \alpha \sum_{k=1}^N \left\langle \mathcal{F}'(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)}) - \mathcal{F}'(\mathbf{p}^{(n)}), R_k^* \bar{\mathbf{r}}_k \right\rangle \\ (3.4) \quad & = \tau \alpha \sum_{k=1}^N \left\langle (F^*)'(\operatorname{div}(\mathbf{p}^{(n)} + R_k^* \mathbf{r}_k^{(n+1)})) - (F^*)'(\operatorname{div} \mathbf{p}^{(n)}), \operatorname{div} R_k^* \bar{\mathbf{r}}_k \right\rangle \\ & \leq \tau \sum_{k=1}^N \|\operatorname{div} R_k^* \mathbf{r}_k^{(n+1)}\| \|\operatorname{div} R_k^* \bar{\mathbf{r}}_k\| \\ & \leq \tau \left(\sum_{k=1}^N \|\operatorname{div} R_k^* \mathbf{r}_k^{(n+1)}\|^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^N \|\operatorname{div} R_k^* \bar{\mathbf{r}}_k\|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

By (2.9c), Proposition 2.5, and Lemma 2.7, we have

$$\begin{aligned} (3.5) \quad & \sum_{k=1}^N \|\operatorname{div} R_k^* \bar{\mathbf{r}}_k\|^2 \leq c_1 \|\operatorname{div}(\mathbf{p}^{(n)} - \mathbf{p}^*)\|^2 + c_2 \|\mathbf{p}^{(n)} - \mathbf{p}^*\|^2 \\ & \leq 2\kappa c_1 \zeta_n + 8c_2 |\Omega| \\ & \leq 2\kappa c_1 \zeta_0 + 8c_2 |\Omega| =: c_3. \end{aligned}$$

Thus, it follows from (3.4) and (3.5) that

$$(3.6) \quad \zeta_{n+1} - (1 - \tau) \zeta_n \leq \tau c_3^{\frac{1}{2}} \left(\sum_{k=1}^N \|\operatorname{div} R_k^* \mathbf{r}_k^{(n+1)}\|^2 \right)^{\frac{1}{2}}.$$

Combining (3.6) with Lemma 3.2, we get

$$(3.7) \quad \zeta_n - \zeta_{n+1} \geq \frac{1}{2\tau\kappa c_3} [\zeta_{n+1} - (1 - \tau) \zeta_n]^2.$$

Finally, invoking Lemma 2.9 for (3.7) completes the proof. \square

To trace the dependency of the convergence of Algorithm 1 on parameters, we perform some additional calculations starting from (3.7). By Lemma 2.9, we have

$$\gamma = 1 - \tau \quad \text{and} \quad c = \sqrt{2\tau\kappa c_3} = 2\sqrt{\tau\kappa(\kappa c_1 \zeta_0 + 4c_2|\Omega|)},$$

so that

$$(3.8) \quad \begin{aligned} \frac{1}{\tilde{c}} &= 2\zeta_0 + \left(\frac{(1-\tau)\sqrt{\zeta_0} + 2\sqrt{\tau\kappa(\kappa c_1 \zeta_0 + 4c_2|\Omega|)}}{\tau} \right)^2 \\ &\leq 2 \left[1 + \left(\frac{1-\tau}{\tau} \right)^2 + \frac{4\kappa^2 c_1}{\tau} \right] \zeta_0 + \frac{32\kappa c_2 |\Omega|}{\tau}. \end{aligned}$$

Hence, the constant \tilde{c} in Theorem 3.3 depends on $|\Omega|$, κ , τ , ζ_0 , c_1 , and c_2 only. Moreover, we observe that (3.8) is decreasing with respect to $\tau \in (0, 1/N]$. Hence, we may choose $\tau = 1/N$; see also [10, Remark 3.1].

REMARK 3.4. In the ROF case (1.2), alternatively to Theorem 3.3, one can prove the $O(1/n)$ -convergence rate of Algorithm 1 by a similar argument as in [10]. Compared to [10], our proof is simpler due to the descent rule, Lemma 3.1. Moreover, our estimate is independent of N while that in [10] is not.

In addition to the $O(1/n)$ -convergence, we prove that Algorithm 1 converges pseudo-linearly, i.e., $\mathcal{F}(\mathbf{p}^{(n)})$ decreases as fast as linear convergence until it reaches a particular value. Theorem 3.5 provides a rigorous statement for the pseudo-linear convergence of Algorithm 1.

THEOREM 3.5. In Algorithm 1, let $\zeta_n = \alpha(\mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^*))$, for $n \geq 0$. Then we have

$$\zeta_n \leq \left(1 - \frac{\tau}{\kappa^2(\sqrt{c_1} + \sqrt{c_1 + \kappa^{-2}})^2} \right)^n \left(\zeta_0 - \frac{4c_2|\Omega|}{\kappa\sqrt{c_1}(c_1 + \kappa^{-2})} \right) + \frac{4c_2|\Omega|}{\kappa\sqrt{c_1}(c_1 + \kappa^{-2})},$$

where κ is given in (2.3) and c_1, c_2 are given in Assumption 2.8.

Proof. Take any $n \geq 0$. For the sake of convenience, we write

$$\Delta = \frac{1}{2} \sum_{k=1}^N \|\operatorname{div} R_k^* \mathbf{r}_k^{(n+1)}\|^2.$$

The starting points of the proof are (3.4) and (3.5):

$$\zeta_{n+1} \leq (1-\tau)\zeta_n + \tau(2\kappa c_1 \zeta_n + 8c_2|\Omega|)^{\frac{1}{2}} (2\Delta)^{\frac{1}{2}}.$$

Using the inequality

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

we readily get

$$(3.9) \quad \begin{aligned} \zeta_{n+1} &\leq (1-\tau)\zeta_n + \tau \left(\zeta_n + \frac{4c_2|\Omega|}{\kappa c_1} \right)^{\frac{1}{2}} (4\kappa c_1 \Delta)^{\frac{1}{2}} \\ &\leq (1-\tau)\zeta_n + \tau \left[\epsilon \left(\zeta_n + \frac{4c_2|\Omega|}{\kappa c_1} \right) + \frac{1}{4\epsilon} \cdot 4\kappa c_1 \Delta \right] \\ &= (1-\tau + \tau\epsilon)\zeta_n + \frac{\tau\kappa c_1}{\epsilon} \Delta + \frac{4\tau\epsilon c_2 |\Omega|}{\kappa c_1} \\ &\leq (1-\tau + \tau\epsilon)\zeta_n + \frac{\kappa^2 c_1}{\epsilon} (\zeta_n - \zeta_{n+1}) + \frac{4\tau\epsilon c_2 |\Omega|}{\kappa c_1}, \end{aligned}$$

where the last inequality is due to Lemma 3.2. Equation (3.9) can be rewritten as

$$\zeta_{n+1} \leq \left(1 - \frac{\tau\epsilon(1-\epsilon)}{\epsilon + \kappa^2 c_1}\right) \zeta_n + \frac{4\tau\epsilon^2 c_2 |\Omega|}{\kappa c_1 (\epsilon + \kappa^2 c_1)}.$$

We take

$$\epsilon = \kappa^2 \left(\sqrt{c_1(c_1 + \kappa^{-2})} - c_1\right) \in (0, 1),$$

which maximizes $\frac{\epsilon(1-\epsilon)}{\epsilon + \kappa^2 c_1}$ so that

$$\begin{aligned} \frac{\epsilon(1-\epsilon)}{\epsilon + \kappa^2 c_1} &= \frac{1}{\kappa^2(\sqrt{c_1} + \sqrt{c_1 + \kappa^{-2}})^2}, \\ \frac{\epsilon^2}{c_1(\epsilon + \kappa^2 c_1)} &= \frac{1}{\kappa^2 \sqrt{c_1(c_1 + \kappa^{-2})}(\sqrt{c_1} + \sqrt{c_1 + \kappa^{-2}})^2}. \end{aligned}$$

We note that similar computations were done in [2, 27]. Then it follows that

$$\zeta_{n+1} \leq \left(1 - \frac{\tau}{\kappa^2(\sqrt{c_1} + \sqrt{c_1 + \kappa^{-2}})^2}\right) \zeta_n + \frac{4\tau c_2 |\Omega|}{\kappa^3 \sqrt{c_1(c_1 + \kappa^{-2})}(\sqrt{c_1} + \sqrt{c_1 + \kappa^{-2}})^2}.$$

By Lemma 2.10, we have

$$\zeta_n \leq \left(1 - \frac{\tau}{\kappa^2(\sqrt{c_1} + \sqrt{c_1 + \kappa^{-2}})^2}\right)^n \left(\zeta_0 - \frac{4c_2 |\Omega|}{\kappa \sqrt{c_1(c_1 + \kappa^{-2})}}\right) + \frac{4c_2 |\Omega|}{\kappa \sqrt{c_1(c_1 + \kappa^{-2})}},$$

which is the desired result. \square

By Theorem 3.5, $\{\mathcal{F}(\mathbf{p}^{(n)})\}$ in Algorithm 1 converges pseudo-linearly with threshold $4c_2 |\Omega| / \kappa \sqrt{c_1(c_1 + \kappa^{-2})}$. It means that if one can make $c_2 |\Omega| / c_1$ sufficiently small, then the proposed method shows almost the same convergence pattern as a linearly convergent algorithm. We shall consider in Section 4 means of how to make $c_2 |\Omega| / c_1$ small, and observe the behavior of the proposed method in Section 5.

4. Domain decomposition. In this section, we present overlapping domain decomposition settings for the proposed DDM. We decompose the domain Ω into \mathcal{N} disjoint square subdomains $\{\Omega_s\}_{s=1}^{\mathcal{N}}$ in a checkerboard fashion. The side length of each subdomain Ω_s is denoted by H . For each $s = 1, \dots, \mathcal{N}$, let Ω'_s be an enlarged subdomain consisting of Ω_s and its surrounding layers of pixels with width δ for some $\delta > 0$. The overlapping subdomains $\{\Omega'_s\}_{s=1}^{\mathcal{N}}$ can be colored with $N_c \leq 4$ colors such that any two subdomains are of the same color if they are disjoint [10]. Let S_k be the union of all subdomains Ω'_s with color k for $k = 1, \dots, N_c$. We denote the collection of all elements of \mathcal{T}_h in S_k by $\mathcal{T}_{h,k}$. In what follows, for two positive real numbers A and B depending on the parameters $|\Omega|, H, h$, and δ , we use the notation $A \lesssim B$ to indicate that there exists a constant $c > 0$ such that $A \leq cB$, where c is independent of the parameters. In addition, we write $A \approx B$ if $A \lesssim B$ and $B \lesssim A$.

We consider a DDM based on the domain decomposition $\{\Omega'_s\}$. Let $N = N_c$, and for $k = 1, \dots, N$, we set $W_k = Y_h(S_k)$, where

$$(4.1) \quad Y_h(S_k) = \{\mathbf{p} \in H_0(\text{div}; S_k) : \mathbf{p}|_T \in \mathcal{RT}_0(T), \quad \forall T \in \mathcal{T}_{h,k}\},$$

where $\mathcal{RT}_0(T)$ is defined in (2.1).

By Lemma 3.4 in [29], there exists a continuous and piecewise linear partition of unity $\{\theta_k\}_{k=1}^N$ for Ω subordinate to the covering $\{S_k\}_{k=1}^N$ such that

$$(4.2a) \quad \text{supp } \theta_k \subset \bar{S}_k, \quad \theta_k \in W_0^{1,\infty}(S_k),$$

$$(4.2b) \quad 0 \leq \theta_k \leq 1, \quad \sum_{k=1}^N \theta_k = 1 \text{ in } \Omega,$$

$$(4.2c) \quad \|\nabla \theta_k\|_{L^\infty(S_k)} \lesssim \frac{1}{\delta},$$

where \bar{S}_k is the closure of S_k and $W_0^{1,\infty}(S_k)$ is defined as

$$W_0^{1,\infty}(S_k) = \{\theta \in L^\infty(S_k) : \nabla \theta \in L^\infty(S_k) \text{ and } \theta|_{\partial S_k} = 0\}.$$

One can show the following property of θ_k :

PROPOSITION 4.1. For $1 \leq k \leq N$ and $\mathbf{p} \in Y_h$, we have

$$\|\text{div}(\Pi_h(\theta_k \mathbf{p}))\|^2 \lesssim \|\text{div } \mathbf{p}\|^2 + \frac{1}{\delta^2} \|\mathbf{p}\|^2.$$

Proof. Invoking (4.2) and Lemmas 2.1 and 2.2 yields

$$\begin{aligned} \|\text{div}(\Pi_h(\theta_k \mathbf{p}))\|^2 &= \int_{\Omega} [\text{div}(\Pi_h(\theta_k \mathbf{p}))]^2 dx \\ &\lesssim \int_{\Omega} [\text{div}(\theta_k \mathbf{p})]^2 dx \\ &\lesssim \int_{\Omega} [\nabla \theta_k \cdot \mathbf{p}]^2 dx + \int_{\Omega} [\theta_k \text{div } \mathbf{p}]^2 dx \\ &\lesssim \frac{1}{\delta^2} \int_{\Omega} |\mathbf{p}|^2 dx + \int_{\Omega} (\text{div } \mathbf{p})^2 dx \\ &\lesssim \frac{1}{\delta^2} \|\mathbf{p}\|^2 + \|\text{div } \mathbf{p}\|^2. \end{aligned}$$

We note that a similar calculation was done in [10, Lemma 3.2]. \square

Using Proposition 4.1, the following stable decomposition estimate is obtained:

LEMMA 4.2. In the space decomposition setting (4.1), Assumption 2.8 holds with

$$c_1 \approx 1, \quad c_2 \lesssim \frac{1}{\delta^2}.$$

Proof. Clearly, we have $c_1 \geq 1$. Take any $\mathbf{p}, \mathbf{q} \in C$. For $k = 1, \dots, N$, we define $\mathbf{r}_k \in Y_k$ by

$$R_k^* \mathbf{r}_k = \Pi_h(\theta_k(\mathbf{p} - \mathbf{q})).$$

It is obvious that $\{\mathbf{r}_k\}$ satisfies (2.9a) and (2.9b). By Proposition 4.1, we get

$$\|\text{div } R_k^* \mathbf{r}_k\|^2 \lesssim \|\text{div}(\mathbf{p} - \mathbf{q})\|^2 + \frac{1}{\delta^2} \|\mathbf{p} - \mathbf{q}\|^2.$$

Summing the above equation over all k yields (2.9c) with $c_1 \lesssim 1$ and $c_2 \lesssim 1/\delta^2$. \square

REMARK 4.3. Lemma 4.2 cannot be applied to the nonoverlapping case ($\delta = 0$) since $1/\delta^2 \rightarrow \infty$ as $\delta \rightarrow 0$. On the other hand, in a finite difference discretization given in [7, 19], it

can be proved that the nonoverlapping decomposition satisfies Assumption 2.8 with a similar argument as in [19, Lemma 3.5].

Combining Theorem 3.5 and Lemma 4.2, we get the following result:

COROLLARY 4.4. *For fixed $\tau > 0$, Algorithm 1 with the domain decomposition (4.1) converges pseudo-linearly at a rate γ with threshold $\epsilon > 0$, where $\epsilon \lesssim |\Omega|/\delta^2$ and γ is independent of $|\Omega|$, H , h , and δ .*

From Corollary 4.4, we can deduce several notable facts about Algorithm 1. As $\epsilon \lesssim |\Omega|/\delta^2$, the proposed DDM converges as fast as a linear convergent algorithm until the energy error becomes very small if δ is chosen large. Indeed, we will see in Section 5 that the energy error decreases linearly to the machine error if δ is chosen such that $|\Omega|^{1/2}/\delta$ is less than about 2^7 . Moreover, since γ does not depend on $|\Omega|$, H , h or δ , the linear convergence rate of Algorithm 1, which dominates the convergence behavior, is the same regardless of $|\Omega|$, δ , and the number of subdomains. To the best of our knowledge, such an observation is new in the field of DDMs. Usually, the linear convergence rate of additive Schwarz methods depends on δ ; see [27], for example. However, in our case, the value of δ affects only the threshold ϵ but not the rate γ .

In the DDM described above, the local problems in Ω'_s ($s = 1, \dots, \mathcal{N}$) have the following general form:

$$(4.3) \quad \min_{\mathbf{r}_s \in Y_h(\Omega'_s), \mathbf{q} + (R^s)^* \mathbf{r}_s \in C} \mathcal{F}(\mathbf{q} + (R^s)^* \mathbf{r}_s),$$

where $\mathbf{q} \in Y_h$, $Y_h(\Omega'_s)$ is defined in the same manner as (4.1) and $(R^s)^*: Y_h(\Omega'_s) \rightarrow Y_h$ is the natural extension operator. Let $\mathbf{p}_s = \mathbf{r}_s + R^s \mathbf{q}$. Then (4.3) is equivalent to

$$(4.4) \quad \min_{\mathbf{p}_s \in C^s} \{\mathcal{F}_s(\mathbf{p}_s) := F^*(\operatorname{div}(R^s)^* \mathbf{p}_s + g_s)\},$$

where C^s is the subset of $Y_h(\Omega'_s)$ defined by

$$C^s = \{\mathbf{p}_s \in Y_h(\Omega'_s) : |(\mathbf{p}_s)_e| \leq 1, \quad \forall e \in \mathcal{E}_h \text{ such that } e \text{ is in the interior of } \Omega'_s\}$$

and

$$g_s = \operatorname{div}(I - (R^s)^* R^s) \mathbf{q}.$$

Existing state-of-the-art solvers for (2.2) (see [9]) can be utilized to solve (4.4); we have

$$\mathcal{F}'_s(\mathbf{p}_s) = R^s \operatorname{div}^* ((F^*)'(\operatorname{div}(R^s)^* \mathbf{p}_s + g_s)).$$

REMARK 4.5. For unconstrained, strongly convex, vector-valued problems such as the grad-div problem, one can obtain a stable decomposition such that c_1 is dependent on δ and $c_2 = 0$ by using the discrete Helmholtz decomposition (see, e.g., [22, Lemma 5.8]). In this case, a linear convergence rate depending on δ is obtained by the same argument as Theorem 3.5. However, it seems that such a stable decomposition is not available in our case, i.e., for constrained and non-strongly convex problem; see Table 1.1. Numerical experiments presented in Section 5 will show the following phenomena: Algorithm 1 converges not linearly but pseudo-linearly, i.e., the convergence rate deteriorates when $\mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^*)$ becomes sufficiently small, and the linearly convergent part of Algorithm 1 does not depend on δ .

5. Applications. In this section, we introduce several applications of the proposed method. We also provide numerical experiments which support our theoretical results presented above.

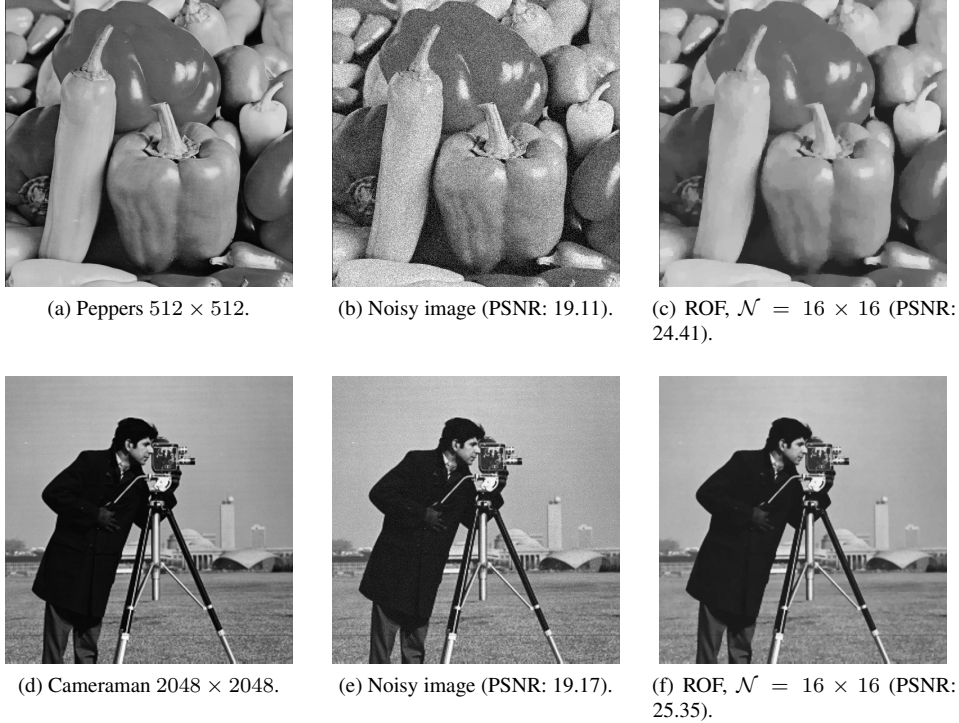


FIG. 5.1. Test images and their results of Algorithm 1 applied to (2.4) for $\mathcal{N} = 16 \times 16$ with $d/\delta = 2^6$.

All algorithms were implemented in C with MPI and performed on a computer cluster composed of seven machines, where each machine is equipped with two Intel Xeon SP-6148 CPUs (2.4GHz, 20C) and 192GB RAM. Two test images “Peppers 512×512 ” and “Cameraman 2048×2048 ” that we used in our experiments are displayed in Figures 5.1(a) and (d). As a measurement of the quality of image restoration, we provide the PSNR (peak signal-to-noise ratio); the PSNR of a corrupted image $u \in X_h$ with respect to the original clean image $u_{\text{orig}} \in X_h$ is defined by

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

where $\text{MAX} = 1$ is the maximum possible pixel value of the image. In the following, we take the side length of the elements $h = 1$ and denote the side length of Ω by d , i.e., $d = |\Omega|^{1/2}$ for square images such as in Figure 5.1. The scaled energy error $\alpha(\mathcal{F}(\mathbf{p}^{(n)}) - \mathcal{F}(\mathbf{p}^*))$ of the n th iterate $\mathbf{p}^{(n)}$ is denoted by ζ_n , where the minimum energy $\mathcal{F}(\mathbf{p}^*)$ is computed by 10^6 iterations of FISTA [4].

5.1. The Rudin–Osher–Fatemi model. The Fenchel–Rockafellar dual problem of the discrete ROF model (2.4) is stated as

$$(5.1) \quad \min_{\mathbf{p} \in C} \left\{ \mathcal{F}(\mathbf{p}) := \frac{1}{2\lambda} \|\text{div } \mathbf{p} + \lambda f\|^2 \right\},$$

and one can obtain the Frechét derivative of \mathcal{F} as

$$\mathcal{F}'(\mathbf{p}) = \frac{1}{\lambda} \text{div}^* (\text{div } \mathbf{p} + \lambda f).$$

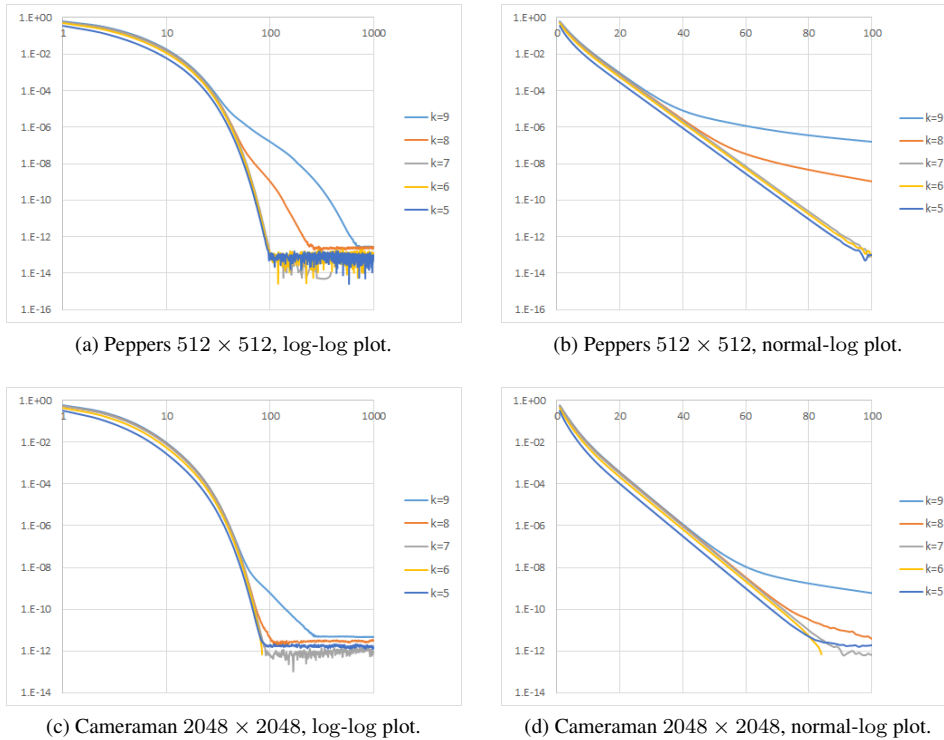


FIG. 5.2. Decay of the relative energy error ζ_n/ζ_0 of Algorithm 1 applied to (5.1) for $d/\delta = 2^k$ ($k = 5, 6, \dots, 9$) with $\mathcal{N} = 8 \times 8$.

The projection onto C can be easily computed by the pointwise Euclidean projection [18]. Therefore, (5.1) can be solved efficiently by, e.g., FISTA [4]. Note that for the case of (5.1), the primal-dual relation (2.6) reduces to the following:

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

where $u^* \in X_h$ solves (2.4).

For our experiments, test images shown in Figures 5.1(a) and (d) were corrupted by additive Gaussian noise with mean 0 and variance 0.05; see Figures 5.1(b) and (e). The parameter λ in (2.4) is chosen as $\lambda = 10$. In Algorithm 1, we set $\tau = 1/4$. The local problems in Ω'_s , $s = 1, \dots, \mathcal{N}$, are solved by FISTA [4] with $L = 8/\lambda$ and the stopping criterion

$$(5.2) \quad \frac{\|\operatorname{div}(\mathbf{r}_s^{(n+1)} - \mathbf{r}_s^{(n)})\|^2}{|\Omega'_s|} \leq 10^{-18} \quad \text{or} \quad n = 1000.$$

We note that the parameter selection $L = 8/\lambda$ is due to Proposition 2.3. The resulting images for the case 16×16 are given in Figures 5.1(c) and (f), and they show no trace on the subdomain boundaries.

First, we observe how the convergence rate of Algorithm 1 is affected by d/δ . Figure 5.2 shows the decay of the relative energy error ζ_n/ζ_0 , for $d/\delta = 2^k$ ($k = 5, 6, \dots, 9$), when the number of subdomains is fixed by $\mathcal{N} = 8 \times 8$. As Figures 5.2(a) and (c) illustrate, the threshold of the pseudo-linear convergence decreases as d/δ decreases, which verifies

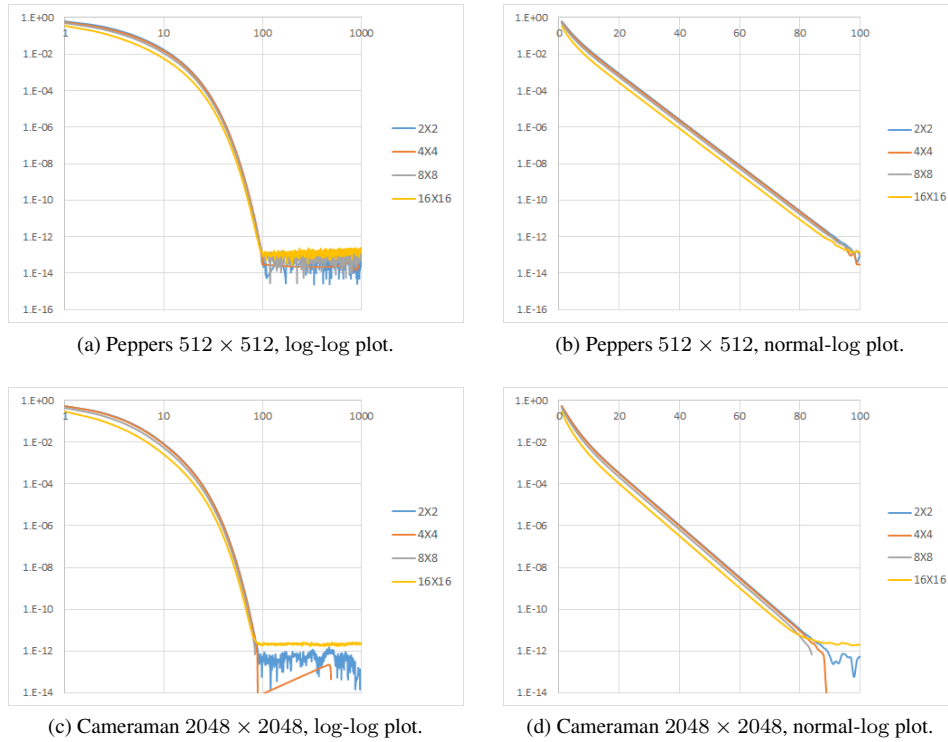


FIG. 5.3. Decay of the relative energy error ζ_n/ζ_0 of Algorithm 1 applied to (5.1) for $\mathcal{N} = 2 \times 2, \dots, 16 \times 16$ with $d/\delta = 2^6$.

Corollary 4.4. Furthermore, in the cases when $d/\delta \leq 2^7$, the threshold is so small that the behavior of Algorithm 1 is just as for linearly convergent algorithms. Thus, the proposed method is as efficient as linearly convergent algorithms in practice. We also observe from Figures 5.2(b) and (d) that the linear convergence rate of Algorithm 1 is independent of δ as noted in Corollary 4.4.

Next, we consider the performance of the proposed DDM with respect to the number of subdomains \mathcal{N} . Figure 5.3 shows the decay of ζ_n/ζ_0 when \mathcal{N} varies from 2×2 to 16×16 with $d/\delta = 2^6$. We readily see that the convergence behavior of Algorithm 1 is almost the same regardless of \mathcal{N} . Hence, we conclude that the convergence rate of Algorithm 1 does not depend on \mathcal{N} .

Finally, we compare the convergence behavior of the proposed method with two recently developed DDMs for the ROF model [18, 19]. The following algorithms were used in our experiments:

- ALG1: Algorithm 1, $\mathcal{N} = 8 \times 8$, $d/\delta = 2^6$.
- PDD: Primal DDM proposed in [18], $\mathcal{N} = 8 \times 8$, $L = 4$.
- FPJ: Fast pre-relaxed block Jacobi method proposed in [19], $\mathcal{N} = 8 \times 8$.

As shown in Figure 5.4, as the number of iterations increases, the convergence rate of ALG1 becomes eventually faster than those of PDD and FPJ, which were proven to be $O(1/n^2)$ -convergent. We note that numerical results that verify the superior convergence properties of PDD and FPJ compared to existing $O(1/n)$ -convergent DDMs were presented in [18, 19].

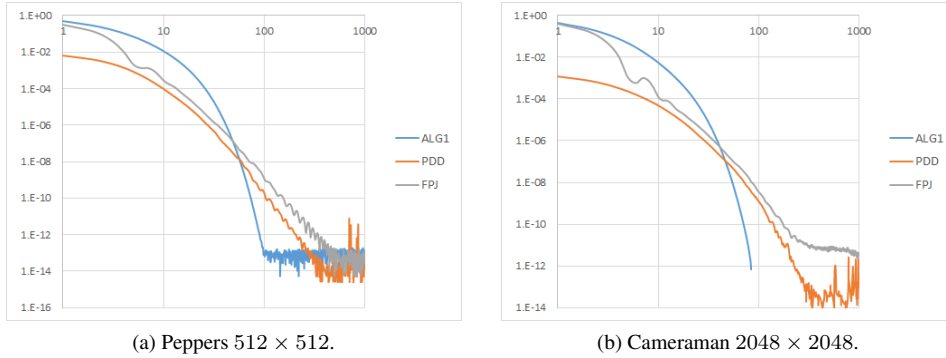


FIG. 5.4. Decay of the relative energy error ζ_n/ζ_0 of various DDMs applied to (5.1), $\mathcal{N} = 8 \times 8$.

REMARK 5.1. Even though the proposed method is based on finite element discretizations, a direct comparison with methods based on finite difference discretizations such as FPJ is possible by virtue of the equivalence relation presented in [18, Theorem 2.3].

5.2. The $TV-H^{-1}$ model. Now, we consider the discrete $TV-H^{-1}$ model (2.5). The dual problem of (2.5) is given by

$$(5.3) \quad \min_{\mathbf{p} \in C} \left\{ \mathcal{F}(\mathbf{p}) := \frac{1}{2\lambda} \|\operatorname{div} \mathbf{p}\|_K^2 + \langle f, \operatorname{div} \mathbf{p} \rangle \right\},$$

where $\|v\|_K = \langle Kv, v \rangle^{1/2}$, for $v \in X_h$. The Fréchet derivative $\mathcal{F}'(\mathbf{p})$ can be easily computed as

$$\mathcal{F}'(\mathbf{p}) = \frac{1}{\lambda} \operatorname{div}^*(K \operatorname{div} \mathbf{p} + \lambda f).$$

If we have a solution $\mathbf{p}^* \in Y_h$ of (5.3), then a solution $u^* \in X_h$ of (2.5) can be obtained by

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

Now, we present the numerical results of Algorithm 1 for (5.3). The corrupted test images Figures 5.1(b) and (e) are used as f in (5.3). We set $\lambda = 10$. In Algorithm 1, the parameter τ is chosen by $\tau = 1/4$, and the local problems in Ω'_s , $s = 1, \dots, \mathcal{N}$, are solved by FISTA [4] with $L = 64/\lambda$ and the stopping criterion (5.2). The parameter selection $L = 64/\lambda$ is derived from Proposition 2.3 and the Gershgorin circle theorem for K [21].

Figure 5.5 shows the decay of the relative energy error ζ_n/ζ_0 for various values of d/δ when $\mathcal{N} = 8 \times 8$. We observe the same dependency of the convergence rate on d/δ as the ROF case: Algorithm 1 behaves as a linearly convergent algorithm if $d/\delta \leq 2^7$, and the rate of linear convergence is independent of δ . As Figure 5.6 indicates, the dependency of the convergence rate of Algorithm 1 is independent of \mathcal{N} ; the convergence rates when $\mathcal{N} = 2 \times 2, \dots, 16 \times 16$ are almost the same. In conclusion, Corollary 4.4 is verified for the $TV-H^{-1}$ model, as well as the ROF model.

It is interesting to observe that the pseudo-linear convergence of Algorithm 1 is not contaminated even in the case of a large condition number κ . While the condition number of (2.5) is much larger than the one of (2.4) in general, the pseudo-linear convergence is evident for both problems. The threshold of the pseudo-linear convergence presented in Theorem 3.5

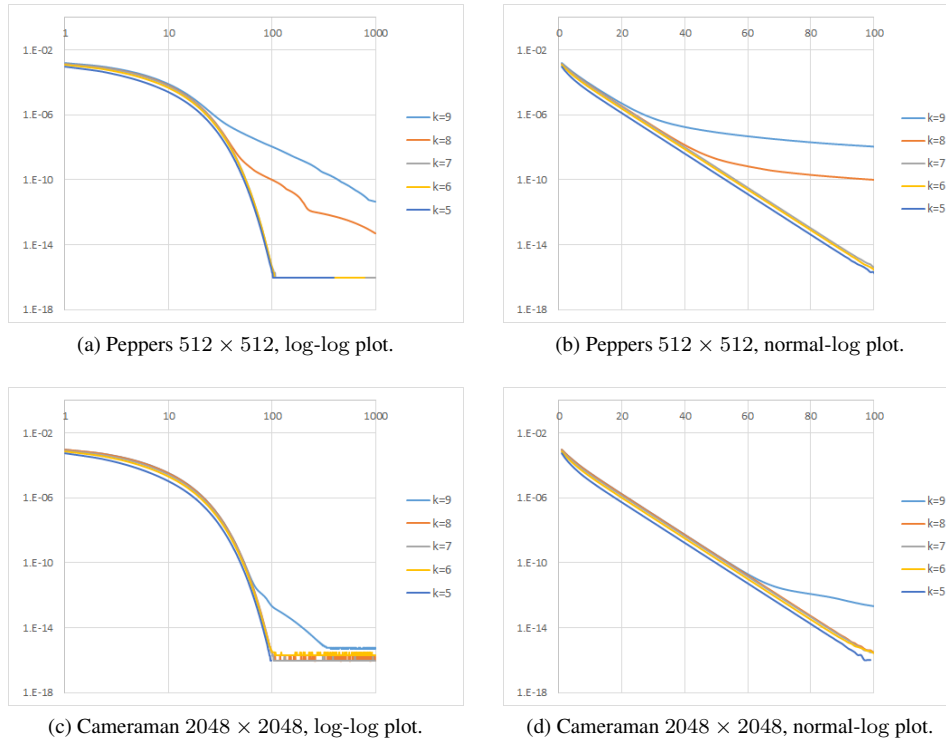


FIG. 5.5. Decay of the relative energy error ζ_n/ζ_0 of Algorithm 1 applied to (5.3) for $d/\delta = 2^k$ ($k = 5, 6, \dots, 9$) with $\mathcal{N} = 8 \times 8$.

has an upper bound independent of κ as follows:

$$\frac{4c_2|\Omega|}{\kappa\sqrt{c_1(c_1 + \kappa^{-2})}} \leq \frac{4c_2|\Omega|}{c_1}.$$

Therefore, one can conclude that this observation is indeed reflected in Theorem 3.5.

6. Conclusion. We proposed an additive Schwarz method based on an overlapping domain decomposition for total variation minimization. Contrary to the existing work [10], we showed that our method is applicable to not only the ROF model but also to more general total variation minimization problems. A novel technique using a descent rule for the convergence analysis of the additive Schwarz method was presented. With this technique, we obtained the convergence rate of the proposed method as well as the dependency of the rate on the condition number of the model problem. In addition, we showed the pseudo-linear convergence property of the proposed method, in which the convergence behavior of the proposed method is as for linearly convergent algorithms if the overlapping width δ is large. Numerical experiments verified our theoretical results.

Recently, the acceleration technique proposed in [4] was successfully applied to nonoverlapping DDMs for the ROF model, and accelerated methods were developed [18, 19]. However, it is still open how to adopt the acceleration technique to overlapping DDMs for general total variation minimization.

As a final remark, we note that the convergence analysis in this paper can be easily applied to either a continuous setting or a finite difference discretization with slight modification.

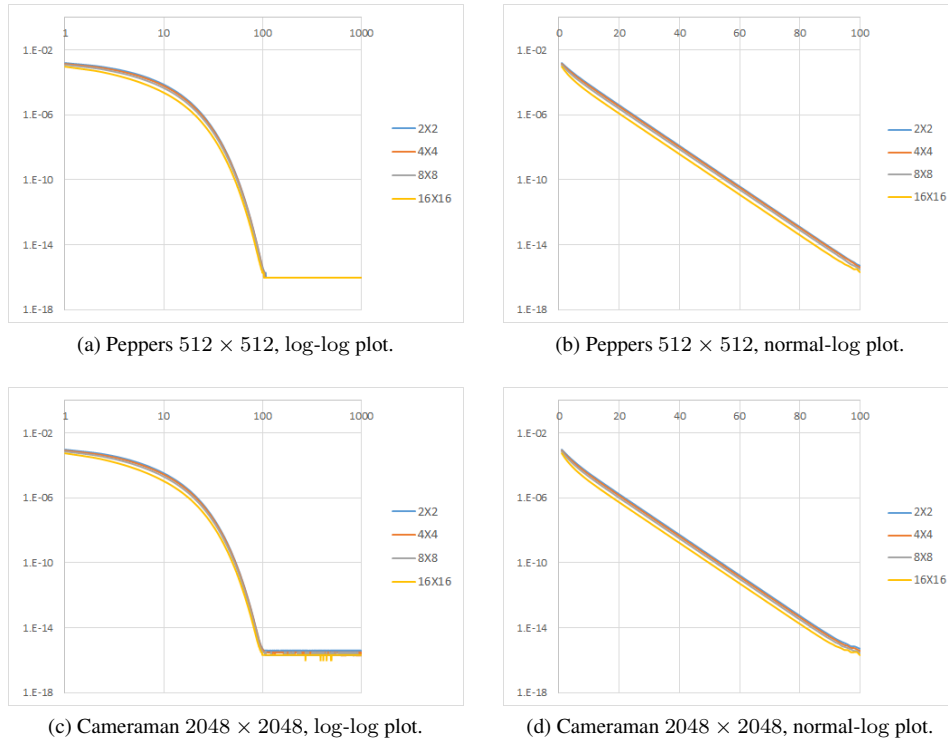


FIG. 5.6. Decay of the relative energy error ζ_n/ζ_0 of Algorithm 1 applied to (5.3) for $\mathcal{N} = 2 \times 2, \dots, 16 \times 16$ with $d/\delta = 2^6$.

Acknowledgments. The author would like to thank Professor Chang-Ock Lee for insightful discussions and comments.

REFERENCES

- [1] 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.
- [2] 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.
- [3] A. BECK, *On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes*, SIAM J. Optim., 25 (2015), pp. 185–209.
- [4] A. BECK AND M. TEBOLLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
- [5] L. M. BRÈGMAN, *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming*, Ž. Vyčisl. Mat i Mat. Fiz., 7 (1967), pp. 620–631.
- [6] M. BURGER, L. HE, AND C.-B. SCHÖNLIEB, *Cahn-Hilliard inpainting and a generalization for grayvalue images*, SIAM J. Imaging Sci., 2 (2009), pp. 1129–1167.
- [7] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.
- [8] 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. Comput. Math., 1 (2015), pp. 29–54.
- [9] ———, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319.
- [10] 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.

- [11] G. CHEN AND M. TEBoulLE, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., 3 (1993), pp. 538–543.
- [12] 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.
- [13] M. FORNASIER AND C.-B. SCHÖNLIEB, *Subspace correction methods for total variation and l_1 -minimization*, SIAM J. Numer. Anal., 47 (2009), pp. 3397–3428.
- [14] 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.
- [15] 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.
- [16] A. LANGER AND F. GASPOZ, *Overlapping domain decomposition methods for total variation denoising*, SIAM J. Numer. Anal., 57 (2019), pp. 1411–1444.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] ———, *A finite element nonoverlapping domain decomposition method with Lagrange multipliers for the dual total variation minimizations*, J. Sci. Comput., 81 (2019), pp. 2331–2355.
- [21] R. J. LEVEQUE, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, SIAM, Philadelphia, 2007.
- [22] D.-S. OH, *An overlapping Schwarz algorithm for Raviart-Thomas vector fields with discontinuous coefficients*, SIAM J. Numer. Anal., 51 (2013), pp. 297–321.
- [23] S. OSHER, A. SOLÉ, AND L. VESE, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Model. Simul., 1 (2003), pp. 349–370.
- [24] P.-A. RAVIART AND J.-M. THOMAS, *A mixed finite element method for 2nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, I. Galligani and E. Magenes, eds., Lecture Notes in Math., vol. 606, Springer, Berlin, 1977, pp. 292–315.
- [25] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D., 60 (1992), pp. 259–268.
- [26] C.-B. SCHÖNLIEB, *Total variation minimization with an H^{-1} constraint*, in Singularities in Nonlinear Evolution Phenomena and Applications, M. Novaga and G. Orlandi, eds., vol. 9 of CRM Series, Ed. Norm., Pisa, 2009, pp. 201–234.
- [27] X.-C. TAI, *Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities*, Numer. Math., 93 (2003), pp. 755–786.
- [28] X.-C. TAI AND J. XU, *Global and uniform convergence of subspace correction methods for some convex optimization problems*, Math. Comp., 71 (2002), pp. 105–124.
- [29] A. TOSELLI AND O. WIDLUND, *Domain Decomposition Methods—Algorithms and Theory*, Springer, Berlin, 2005.