SIAM J. NUMER. ANAL.
Vol. 58, No. 3, pp. 1495–1530

ADDITIVE SCHWARZ METHODS FOR CONVEX OPTIMIZATION AS GRADIENT METHODS*

JONGHO PARK†

Abstract. This paper gives a unified convergence analysis of additive Schwarz methods for general convex optimization problems. Resembling the fact that additive Schwarz methods for linear problems are preconditioned Richardson methods, we prove that additive Schwarz methods for general convex optimization are in fact gradient methods. Then an abstract framework for convergence analysis of additive Schwarz methods is proposed. The proposed framework applied to linear elliptic problems agrees with the classical theory. We present applications of the proposed framework to various interesting convex optimization problems such as nonlinear elliptic problems, nonsmooth problems, and nonsharp problems.

Key words. additive Schwarz method, gradient method, convex optimization, convergence analysis, domain decomposition method

AMS subject classifications. 65N55, 65J15, 65K15, 90C25

DOI. 10.1137/19M1300583

1. Introduction.

Many modern iterative methods such as block relaxation methods, multigrid methods, and domain decomposition methods for linear systems belong to Schwarz methods, also known as subspace correction methods. Because of this fact, constructing an abstract convergence theory for Schwarz methods has been considered an important task in the field of numerical analysis. There is a vast literature on the convergence theory of Schwarz methods for linear systems. The paper [35] by Xu contains an outstanding survey on some early results on Schwarz methods. Several variants of the convergence theory with various viewpoints were proposed in, e.g., [16, 17, 36]. For a modern representation of the abstract convergence theory with historical remarks, one may refer to the monograph [32] by Toselli and Widlund.

While the convergence theory of Schwarz methods for linear elliptic problems seems to be almost complete, there has still been much research on convergence analysis of Schwarz methods for nonlinear and nonsmooth problems. The papers [30, 31] are important early results on Schwarz methods for nonlinear problems. In [4, 5, 29], Schwarz methods for variational inequalities which arise in quadratic optimization with constraints were proposed. Convergence analysis for Schwarz methods was successfully extended to nonquadratic and nonsmooth variational inequalities in [1] and [3], respectively. Recently, overlapping Schwarz methods for convex optimization problems lacking strong convexity were proposed in [13, 28], especially for total variation minimization problems arising in mathematical imaging. On the other hand, it was shown in [21] that Schwarz methods may not converge to a correct solution in the case of nonsmooth convex optimization.

One of the most important observations done in the convergence theory of Schwarz methods for linear problems is that Schwarz methods can be viewed as preconditioned

*Received by the editors November 18, 2019; accepted for publication (in revised form) March 23, 2020; published electronically May 19, 2020.
https://doi.org/10.1137/19M1300583
Funding: The work of the author was supported by the Basic Science Research program through the National Research Foundation Korea (NRF) grant 2019R1A6A3A01092549 funded by the Ministry of Education.
†Department of Mathematical Sciences, KAIST, Daejeon 34141, Korea (jongho.park@kaist.ac.kr, https://sites.google.com/view/jonghopark).

Richardson methods; see, e.g., [32]. This observation makes the convergence analysis of a method fairly simple; convergence is obvious by the well-known convergence results on the Richardson method and one only need to estimate the condition number of the linear system to obtain an estimate for the convergence rate. However, such an observation does not exist for general nonlinear and nonsmooth problems. Due to this situation, all of the aforementioned works on nonlinear problems provided proofs on why their methods converge to a solution correctly with some complex computations. To the best of our knowledge, the only relevant result on nonlinear problems is [22]; it says that block Jacobi methods for a constrained quadratic optimization problem can be regarded as preconditioned forward-backward splitting algorithms [8].

In this paper, we show that additive Schwarz methods for general convex optimization can be represented as gradient methods. In the field of mathematical optimization, there has been much research on gradient methods for solving convex optimization problems; for example, see [8, 12, 25]. Therefore, by observing that additive Schwarz methods are interpreted as gradient methods, we can borrow many valuable tools on convergence analysis from the field of mathematical optimization in order to analyze Schwarz methods. Consequently, we propose a novel abstract convergence theory of additive Schwarz methods for convex optimization. The proposed framework directly generalizes the classical convergence theory presented in [32] for linear elliptic problems to general convex optimization problems. We also highlight that our framework gives a better convergence rate than existing works [1, 3, 31] for some applications.

Various applications of the proposed convergence theory are presented in this paper. A very broad range of convex optimization problems fits into our framework. In particular, we provide examples of nonlinear elliptic problems, nonsmooth problems, and problems without sharpness, where those classes of problems were considered in existing works [31], [4, 29], [3], and [13, 28], respectively.

The rest of this paper is organized as follows. In section 2, we provide some useful tools of convex analysis required in this paper. An abstract gradient method for solving general convex optimization is introduced in section 3 with the convergence analysis motivated by [25]. In section 4, we show that additive Schwarz methods for convex optimization are indeed gradient methods; a novel abstract convergence theory for additive Schwarz methods is proposed in this viewpoint. One- and two-level overlapping domain decomposition settings and some important stable decomposition estimates are summarized in section 5. Applications of the proposed convergence theory to various convex optimization problems are presented in section 6. We conclude this paper with remarks in section 7.

2. Preliminaries.

In this section, we introduce notation and basic notions of convex analysis that will be used throughout the paper.

Let V be a reflexive Banach space equipped with the norm $\|\cdot\|_V$. The topological dual space of V is denoted by V^*, and $\langle\cdot,\cdot\rangle_{V^*\times V}$ denotes the duality pairing of V, i.e.,

$$\langle p,u\rangle_{V^*\times V} = p(u), \quad u\in V,\ p\in V^*.$$

We may omit the subscripts if there is no ambiguity. We denote the collection of proper, convex, lower semicontinuous functionals from V to $\overline{\mathbb{R}}$ by $\Gamma_0(V)$.

The *effective domain* of a proper functional $F\colon V\to\overline{\mathbb{R}}$ is denoted by $\operatorname{dom} F$, i.e.,

$$\operatorname{dom} F = \{u\in V : F(u)<\infty\}.$$

For example, for a subset K of V, its *characteristic function* $\chi_K \colon V \to \overline{\mathbb{R}}$ defined by

$$\chi_K(u) = \begin{cases} 0 & \text{if } u \in K, \\ \infty & \text{if } u \notin K \end{cases} \tag{2.1}$$

has the effective domain $\operatorname{dom} \chi_K = K$.

A functional $F \colon V \to \overline{\mathbb{R}}$ is said to be *coercive* if

$$F(u) \to \infty \quad \text{as} \quad \|u\| \to \infty.$$

If $F \in \Gamma_0(V)$ is coercive, then the minimization problem

$$\min_{u \in V} F(u) \tag{2.2}$$

has a solution $u^* \in V$ with $F(u^*) > -\infty$ [6, Proposition 11.14]. If we further assume that F is strictly convex, then the solution of (2.2) is unique.

For a convex functional $F \colon V \to \overline{\mathbb{R}}$, the *subdifferential* of F at a point $u \in V$ is defined as

$$\partial F(u) = \left\{ p \in V^* : F(v) \geq F(u) + \langle p, v-u \rangle \quad \forall v \in V \right\}.$$

If F is Frechét differentiable at u, then the subdifferential $\partial F(u)$ agrees with the Frechét derivative $F'(u)$, i.e., $\partial F(u) = \{F'(u)\}$. It is clear from the definition of subdifferential that $u^* \in V$ is a global minimizer of F if and only if $0 \in \partial F(u^*)$.

If $F_k \colon V \to \overline{\mathbb{R}}$, $1 \leq k \leq N$ are proper convex functionals, one can obtain directly from the definition of subdifferential that

$$\partial \left(\sum_{k=1}^N F_k \right)(u) \supseteq \sum_{k=1}^N \partial F_k(u), \quad u \in V. \tag{2.3}$$

We have a similar result on the composition with a linear operator; let W be a reflexive Banach space. For a proper convex functional $F \colon V \to \overline{\mathbb{R}}$ and a bounded linear functional $A \colon W \to V$, one can show that

$$\partial (F \circ A)(w) \supseteq A^* \partial F(Aw), \quad w \in W. \tag{2.4}$$

The *Legendre–Fenchel conjugate* $F^* \colon V^* \to \overline{\mathbb{R}}$ of a functional $F \colon V \to \overline{\mathbb{R}}$ is defined by

$$F^*(p) = \sup_{u \in V} \left\{ \langle p, u \rangle - F(u) \right\}.$$

Clearly, F^* is convex lower semicontinuous regardless of whether F is. If we further assume that $F \in \Gamma_0(V)$, then ∂F and ∂F^* are inverses of each other [6, Theorem 16.23], i.e., we have

$$p \in \partial F(u) \; \Leftrightarrow \; u \in \partial F^*(p), \quad u \in V, \; p \in V^*. \tag{2.5}$$

For convex functionals F_k, $1 \leq k \leq N$, defined on V, the *infimal convolution* of F_k is given by

$$\left(\mathop{\square}_{k=1}^N F_k \right)(v) = \inf \left\{ \sum_{k=1}^N F_k(v_k) : v = \sum_{k=1}^N v_k, \; v_k \in V \right\}.$$

It is easy to check that $\square_{k=1}^N F_k$ is convex. If each F_k is in $\Gamma_0(V)$ and coercive, then we have $\square_{k=1}^N F_k \in \Gamma_0(V)$ [6, Proposition 12.14].

For another reflexive Banach space W and a bounded linear operator $A\colon V \to W$, the *infimal postcomposition* $A \triangleright F\colon W \to \overline{\mathbb{R}}$ of a convex functional $F\colon V \to \overline{\mathbb{R}}$ by A is given by

$$(A \triangleright F)(w) = \inf \{F(v) : Av = w,\ v \in V\}.$$

If there does not exist $v \in V$ such that $Av = w$, then we set $(A \triangleright F)(w) = \infty$. One can show that $A \triangleright F$ is also convex [6, Proposition 12.34]. If the adjoint $A^*\colon W^* \to V^*$ of A is surjective and $F \in \Gamma_0(V)$, then we get $A \triangleright F \in \Gamma_0(W)$ [10, Lemma 2.6]. We have the following formulas for the convex conjugates for infimal convolution and infimal postcomposition [6, Proposition 13.21]:

$$\left(\mathop{\square}_{k=1}^{N} F_k \right)^* = \sum_{k=1}^{N} F_k^* \quad \text{and} \quad (A \triangleright F)^* = F^* \circ A^*. \tag{2.6}$$

We state a useful identity on infimal convolution and infimal postcomposition in Lemma 2.1, whose proof will be given in Appendix A.1.

LEMMA 2.1. *For a positive integer N, let W_k, $1 \leq k \leq N$, and W be real vector spaces. For linear operators $A_k\colon W_k \to W$ and functionals $F_k\colon W_k \to \overline{\mathbb{R}}$, the following is satisfied:*

$$\left(\mathop{\square}_{k=1}^{N} (A_k \triangleright F_k) \right)(w) = \inf \left\{ \sum_{k=1}^{N} F_k(w_k) : w = \sum_{k=1}^{N} A_k w_k,\ w_k \in W_k \right\}, \quad w \in W.$$

For a convex and Frechét differentiable functional $F\colon V \to \mathbb{R}$, the *Bregman distance* of F is defined by

$$D_F(u,v) = F(u) - F(v) - \langle F'(v), u - v \rangle, \quad u, v \in V.$$

Note that D_F is convex and Frechét differentiable with respect to its first argument, i.e., for fixed $v \in V$ the map $u \mapsto D_F(u,v)$ is Frechét differentiable and convex.

Remark 2.2. Although the results in the references [6, 10] we cited in this section are stated in the Hilbert space setting, they are still valid for reflexive Banach spaces. Two main properties of Hilbert spaces used in [6, 10] are the weak compactness of closed bounded sets and the equivalence between the strong and weak lower semicontinuity of convex functions, and they are also true for reflexive Banach spaces.

3. Gradient methods. In this section, we propose an abstract gradient method that generalizes several existing first order methods for convex optimization. As we will see in section 4, conventional additive Schwarz methods for convex optimization are interpreted as abstract gradient methods. Therefore, the abstract gradient method and its convergence proof shall be very useful in the analysis of additive Schwarz methods.

Throughout this section, let V be a reflexive Banach space. We consider the following model problem:

$$\min_{u \in V} \{E(u) := F(u) + G(u)\}, \tag{3.1}$$

where $F\colon V \to \mathbb{R}$ is a Frechét differentiable convex function and $G \in \Gamma_0(V)$ is possibly nonsmooth. We further assume that E is coercive, so that a solution $u^* \in V$ of (3.1) exists. The optimality condition of u^* reads as

$$F'(u^*) + \partial G(u^*) \ni 0,$$

or equivalently,

$$\langle F'(u^*), u - u^* \rangle + G(u) - G(u^*) \geq 0, \quad u \in V. \tag{3.2}$$

Let B: $V \times V \to \overline{\mathbb{R}}$ be a functional which is proper, convex, and lower semicontinuous with respect to its first argument. In addition, we assume that B satisfies the following.

Assumption 3.1. There exists constants $q > 1$ and $\theta \in (0, 1]$ such that for any bounded and convex subset K of V, we have

$$\begin{aligned} &D_F(u, v) + G(u) \leq B(u, v) \\ &\quad\leq \frac{L_K}{q} \|u - v\|^q + \theta G\left(\frac{1}{\theta} u - \left(\frac{1}{\theta} - 1\right) v\right) + (1 - \theta) G(v), \quad u, v \in K \cap \operatorname{dom} G, \end{aligned}$$

where L_K is a positive constant depending on K.

With the functional B satisfying Assumption 3.1, the abstract gradient method for (3.1) is presented in Algorithm 3.1.

Algorithm 3.1 Abstract gradient method for (3.1).

Choose $u^{(0)} \in \operatorname{dom} G$.
for $n = 0, 1, 2, \ldots$

$$u^{(n+1)} \in \operatorname*{arg\,min}_{u \in V} \left\{ Q(u, u^{(n)}) := F(u^{(n)}) + \langle F'(u^{(n)}), u - u^{(n)} \rangle + B(u, u^{(n)}) \right\}$$

end

Several fundamental first order methods for (3.1) can be represented as examples of Algorithm 3.1. Under the assumption that F' is Lipschitz continuous with modulus $M > 0$, setting

$$B(u, v) = \frac{1}{2\tau} \|u - v\|^2 + G(u)$$

for $\tau \in (0, 1/M]$ satisfies Assumption 3.1 with $q = 2$, $\theta = 1$, $L_K = 1/\tau$ and yields the forward-backward splitting algorithm [8] for (3.1). If we further assume that $G = 0$, then it reduces to the classical fixed-step gradient descent method.

First, we claim that the energy of Algorithm 3.1 always decreases under Assumption 3.1 in the following lemma; the proof will be given in Appendix A.2.

LEMMA 3.2. *Suppose that Assumption* 3.1 *holds. In Algorithm* 3.1*, the sequence* $\{E(u^{(n)})\}$ *is decreasing.*

We note that $E(u^{(0)}) < \infty$ because $u^{(0)} \in \operatorname{dom} G$. By Lemma 3.2 and the coercivity of E, the sequence $\{u^{(n)}\}$ generated by Algorithm 3.1 is contained in the bounded set

$$K_0 = \left\{ u \in V : E(u) \leq E(u^{(0)}) \right\}. \tag{3.3}$$

Clearly, K_0 is convex and $K_0 \subseteq \operatorname{dom} G$. Since K_0 is bounded, there exists a constant $R_0 > 0$ such that

$$K_0 \subseteq \{u \in V : \|u - u^*\| \leq R_0\}. \tag{3.4}$$

In what follows, we omit the subscript K_0 from L_{K_0} and write $L = L_{K_0}$.

We describe the convergence behavior of Algorithm 3.1. Although Algorithm 3.1 is written in a fairly general fashion, its convergence analysis can be done in a similar way to the vanilla gradient method described in [25]. The proof of the following convergence theorem for Algorithm 3.1 can be found in Appendix A.3.

THEOREM 3.3. *Suppose that Assumption* 3.1 *holds. In Algorithm* 3.1*, if* $E(u^{(0)}) - E(u^*) \geq \theta^{q-1} L R_0^q$*, then*

$$E(u^{(1)}) - E(u^*) \leq \left(1 - \theta\left(1 - \frac{1}{q}\right)\right)(E(u^{(0)}) - E(u^*)).$$

Otherwise, we have

$$E(u^{(n)}) - E(u^*) \leq \frac{C_{q,\theta} L R_0^q}{(n+1)^{q-1}}, \quad n \geq 0,$$

where $C_{q,\theta}$ *is a positive constant defined in* (A.10) *depending on* q *and* θ *only, and* R_0 *was defined in* (3.4)*.*

Theorem 3.3 means that the convergence rate of the energy error of Algorithm 3.1 is $O(1/n^{q-1})$. If the functional E in (3.1) is *sharp*, then a better convergence rate can be obtained. The sharpness condition of F is summarized in Assumption 3.4.

Assumption 3.4 (sharpness). There exists a constant $p > 1$ such that for any bounded and convex subset K of V satisfying $u^* \in K$, we have

$$\frac{\mu_K}{p}\|u - u^*\|^p \leq E(u) - E(u^*), \quad u \in K,$$

for some $\mu_K > 0$.

The inequality in Assumption 3.4 is also known as the Łojasiewicz inequality. It is known that quite many kinds of functions satisfy Assumption 3.4; see [11, 37]. Invoking (3.2), one can obtain the following simple criterion to check whether Assumption 3.4 holds.

PROPOSITION 3.5. *Consider the minimization problem* (3.1)*. For any bounded and convex subset* K *of* V*, assume that* F *is uniformly convex with parameters* p *and* μ_K *on* K*, i.e.,*

$$D_F(u, v) \geq \frac{\mu_K}{p}\|u - v\|^p, \quad u, v \in K. \tag{3.5}$$

Then Assumption 3.4 *holds.*

We write $\mu = \mu_{K_0}$, where K_0 was defined in (3.3). One can prove without major difficulty that p should be greater than or equal to q in order to satisfy Assumptions 3.1 and 3.4 simultaneously. Note that under Assumption 3.4, a solution of (3.1) is unique.

With Assumptions 3.1 and 3.4, the following improved convergence theorem for Algorithm 3.1 is available; see Appendix A.4 for the proof.

THEOREM 3.6. *Suppose that Assumptions* 3.1 *and* 3.4 *hold. In Algorithm* 3.1*, we have the following:*

1. *In the case $p = q$, we have*
$$E(u^{(n)}) - E(u^*) \leq \left(1 - \left(1 - \frac{1}{q}\right) \min\left\{\theta, \left(\frac{\mu}{qL}\right)^{\frac{1}{q-1}}\right\}\right)^n (E(u^{(0)}) - E(u^*)),\ n \geq 0.$$
2. *In the case $p > q$, if $E(u^{(0)}) - E(u^*) \geq \theta^{\frac{p(q-1)}{p-q}} p^{\frac{q}{p-q}} (L^p/\mu^q)^{\frac{1}{p-q}}$, then*
$$E(u^{(1)}) - E(u^*) \leq \left(1 - \theta\left(1 - \frac{1}{q}\right)\right)(E(u^{(0)}) - E(u^*)).$$
Otherwise, we have
$$E(u^{(n)}) - E(u^*) \leq \frac{C_{p,q,\theta}(L^p/\mu^q)^{\frac{1}{p-q}}}{(n+1)^{\frac{p(q-1)}{p-q}}}, \quad n \geq 0,$$
where $C_{p,q,\theta}$ is a positive constant defined in (A.16) *depending on p, q, and θ only.*

In Theorems 3.3 and 3.6, the decay rate of the energy error $E(u^{(n)}) - E(u^*)$ depends on only p, q, θ, L, and μ if the initial energy error $E(u^{(0)}) - E(u^*)$ is small enough. Therefore, in applications, it is enough to estimate those variables to get the convergence rate of the algorithm.

4. Additive Schwarz methods for convex optimization.

This section is devoted to an abstract convergence theory of additive Schwarz methods for convex optimization (3.1). We present an additive Schwarz method for (3.1) based on an abstract framework of space decomposition. Then we show that the proposed method is an instance of Algorithm 3.1. Such an observation makes the convergence analysis of the proposed method straightforward.

First, we present a space decomposition setting. Throughout this section, an index k runs from 1 to N. Let V_k be a reflexive Banach space and $R_k^*\colon V_k \to V$ be a bounded linear operator such that

$$V = \sum_{k=1}^{N} R_k^* V_k \tag{4.1}$$

and its adjoint $R_k\colon V^* \to V_k^*$ is surjective. For example, if V_k is a subspace of V, then one may choose R_k^* as the natural embedding.

In our framework, we allow inexact local solvers. Let $d_k\colon V_k \times V \to \overline{\mathbb{R}}$ and $G_k\colon V_k \times V \to \overline{\mathbb{R}}$ be functionals which are proper, convex, and lower semicontinuous with respect to their first arguments. Local problems of the proposed method shall have the following general form:

$$\min_{w_k \in V_k} \left\{F(v) + \langle F'(v), R_k^* w_k\rangle + \omega d_k(w_k, v) + G_k(w_k, v)\right\} \tag{4.2}$$

for some $v \in V$ and $\omega > 0$. In the case of exact local solvers, we set

$$\begin{aligned} d_k(w_k, v) &= D_F(v + R_k^* w_k, v), \\ G_k(w_k, v) &= G(v + R_k^* w_k) \end{aligned} \tag{4.3}$$

for $w_k \in V_k$, $v \in V$, and we set $\omega = 1$. Then (4.2) becomes

$$\min_{w_k \in V_k} E(v + R_k^* w_k).$$

We present a general additive Schwarz method for (3.1) with local problems (4.2) in Algorithm 4.1. The constants τ_0 and ω_0 in Algorithm 4.1 will be defined in Assumptions 4.2 and 4.3, respectively.

Algorithm 4.1 Additive Schwarz method for (3.1).

Choose $u^{(0)} \in \operatorname{dom} G$, $\tau \in (0, \tau_0]$, and $\omega \geq \omega_0$.
for $n = 0, 1, 2, \dots$

$$w_k^{(n+1)} \in \operatorname*{arg\,min}_{w_k \in V_k} \left\{ F(u^{(n)}) + \langle F'(u^{(n)}), R_k^* w_k \rangle + \omega d_k(w_k, u^{(n)}) + G_k(w_k, u^{(n)}) \right\}, \ 1 \leq k \leq N,$$

$$u^{(n+1)} = u^{(n)} + \tau \sum_{k=1}^{N} R_k^* w_k^{(n+1)}$$

end

In order to ensure convergence of Algorithm 4.1, the following three conditions should be considered: stable decomposition, strengthened convexity, and local stability.

Assumption 4.1 (stable decomposition). There exists a constant $q > 1$ such that for any bounded and convex subset K of V, the following holds: for any $u, v \in K \cap \operatorname{dom} G$, there exists $w_k \in V_k$, $1 \leq k \leq N$, such that

$$u - v = \sum_{k=1}^{N} R_k^* w_k,$$

$$\sum_{k=1}^{N} d_k(w_k, v) \leq \frac{C_{0,K}^q}{q} \|u - v\|^q,$$

and

$$\sum_{k=1}^{N} G_k(w_k, v) \leq G(u) + (N-1)G(v), \tag{4.4}$$

where $C_{0,K}$ is a positive constant depending on K.

Similar assumptions to Assumption 4.1 for Schwarz methods can be found in existing works, e.g., [3, Assumption 1 and equation (7)]. In those works, several function decompositions tailored for particular applications were proposed. We will see in section 6 that Assumption 4.1 is compatible with them. We also note that the assumption (4.4) for the nonsmooth part G of (3.1) is essential; a counterexample for the convergence of Schwarz methods for a problem not satisfying (4.4) was introduced in [21, Claim 6.1].

Assumption 4.2 (strengthened convexity). There exists a constant $\tau_0 \in (0, 1]$ which satisfies the following: for any $v \in V$, $w_k \in V_k$, $1 \leq k \leq N$, and $\tau \in (0, \tau_0]$, we have

$$(1 - \tau N) E(v) + \tau \sum_{k=1}^{N} E(v + R_k^* w_k) \geq E\left(v + \tau \sum_{k=1}^{N} R_k^* w_k\right).$$

By the convexity of E, Assumption 4.2 is valid $\tau_0 = 1/N$. However, a smaller value for τ_0 independent of N can be found by, for example, the coloring technique; details will be given in section 5.

Assumption 4.3 (local stability). There exists a constant $\omega_0 > 0$ which satisfies the following: for any $v \in \operatorname{dom} G$, and $w_k \in V_k$, $1 \leq k \leq N$, we have

$$\begin{aligned} D_F(v + R_k^* w_k, v) &\leq \omega_0 d_k(w_k, v), \\ G(v + R_k^* w_k) &\leq G_k(w_k, v). \end{aligned}$$

In the case of exact solvers, i.e., (4.3), Assumption 4.3 is trivial with $\omega_0 = 1$. In general, as explained in [32], Assumption 4.3 gives a one-sided measure of approximation properties of the local solvers. One can use any local solvers satisfying Assumption 4.3 for Algorithm 4.1.

For convergence analysis of Algorithm 4.1, we introduce a functional $M_{\tau,\omega} : V \times V \to \overline{\mathbb{R}}$: for two positive real numbers τ and ω, the functional $M_{\tau,\omega}$ is defined as

$$\begin{aligned} M_{\tau,\omega}(u, v) = \tau \inf &\left\{ \sum_{k=1}^N (\omega d_k + G_k)(w_k, v) : u - v = \tau \sum_{k=1}^N R_k^* w_k,\ w_k \in V_k \right\} \\ &+ (1 - \tau N)\, G(v),\ u, v \in V. \end{aligned} \tag{4.5}$$

The following lemma summarizes important properties of $M_{\tau,\omega}$.

LEMMA 4.4. *For τ, $\omega > 0$, the functional $M_{\tau,\omega}$: $V \times V \to \mathbb{R}$ defined in* (4.5) *is convex and lower semicontinuous with respect to its first argument.*

Proof. For convenience, we fix $v \in V$ and write

$$M_{\tau,\omega}(u) = M_{\tau,\omega}(u, v), \quad d_k(w_k) = d_k(w_k, v), \quad G_k(w_k) = G_k(w_k, v)$$

for $u \in V$ and $w_k \in V_k$. By Lemma 2.1 we have

(4.6)

$$\begin{aligned} M_{\tau,\omega}(u) &= \tau \inf \left\{ \sum_{k=1}^N (\omega d_k + G_k)(w_k) : u - v = \tau \sum_{k=1}^N R_k^* w_k,\ w_k \in V_k \right\} + (1 - \tau N) G(v) \\ &= \tau \left(\mathop{\square}_{k=1}^N \left((\tau R_k^*) \triangleright (\omega d_k + G_k) \right) \right) (u - v) + (1 - \tau N) G(v). \end{aligned}$$

Since R_k is surjective, by [10, Lemma 2.6] we get the desired result. ∎

The following lemma, named the *generalized additive Schwarz lemma*, shows that Algorithm 4.1 in fact belongs to a class of Algorithm 3.1 with $B(u, v) = M_{\tau,\omega}(u, v)$.

LEMMA 4.5 (generalized additive Schwarz lemma). *Let $\{u^{(n)}\}$ be the sequence generated by Algorithm* 4.1*. Then it satisfies*

$$u^{(n+1)} \in \operatorname*{arg\,min}_{u \in V} \left\{ F(u^{(n)}) + \langle F'(u^{(n)}), u - u^{(n)} \rangle + M_{\tau,\omega}(u, u^{(n)}) \right\}, \quad n \geq 0, \tag{4.7}$$

where $M_{\tau,\omega}(u, u^{(n)})$ was given in (4.5).

Proof. Choose any $n \geq 0$. We write

$$\begin{aligned} d_k^{(n)}(w_k) &= (\omega d_k + G_k)(w_k, u^{(n)}), \quad && w_k \in V_k, \\ M_{\tau,\omega}^{(n)}(u) &= M_{\tau,\omega}(u, u^{(n)}), && u \in V \end{aligned}$$

for convenience. The optimality condition of (4.7) is given by

$$F'(u^{(n)}) + \partial M_{\tau,\omega}^{(n)}(u) \ni 0,$$

or equivalently,

$$u \in \partial M_{\tau,\omega}^{(n)*}(-F'(u^{(n)}))$$

by (2.5) and Lemma 4.4. Therefore, it suffices to show that

$$u^{(n+1)} \in \partial M_{\tau,\omega}^{(n)*}(-F'(u^{(n)})). \tag{4.8}$$

The optimality condition for $w_k^{(n+1)}$ reads as

$$R_k F'(u^{(n)}) + \partial d_k^{(n)}(w_k^{(n+1)}) \ni 0.$$

Since $d_k^{(n)} \in \Gamma_0(V_k)$, one can obtain the following explicit formula for $w_k^{(n+1)}$:

$$w_k^{(n+1)} \in \partial d_k^{(n)*}(-R_k F'(u^{(n)})). \tag{4.9}$$

Summation of (4.9) over $1 \leq k \leq N$ yields

$$u^{(n+1)} \in u^{(n)} + \tau \sum_{k=1}^{N} R_k^* \partial d_k^{(n)*}(-R_k F'(u^{(n)})). \tag{4.10}$$

On the other hand, dualizing (4.6) with $v = u^{(n)}$ yields

$$\begin{aligned} M_{\tau,\omega}^{(n)*}(p) &= \tau \left(\mathop{\Box}_{k=1}^{N} ((\tau R_k^*) \triangleright d_k^{(n)}) \right)^* \left(\frac{1}{\tau} p \right) + \langle p, u^{(n)} \rangle \\ &\overset{(2.6)}{=} \tau \left(\sum_{k=1}^{N} d_k^{(n)*} \circ (\tau R_k) \right) \left(\frac{1}{\tau} p \right) + \langle p, u^{(n)} \rangle \\ &= \tau \sum_{k=1}^{N} d_k^{(n)*} \left(R_k p \right) + \langle p, u^{(n)} \rangle \end{aligned} \tag{4.11}$$

for $p \in V^*$. Consequently, by (2.3) and (2.4) we have

$$\partial M_{\tau,\omega}^{(n)*}(p) \supseteq \tau \sum_{k=1}^{N} R_k^* \partial d_k^{(n)*} \left(R_k p \right) + u^{(n)}. \tag{4.12}$$

If we substitute p by $-F'(u^{(n)})$ in (4.12), we get

$$\partial M_{\tau,\omega}^{(n)*}(-F'(u^{(n)})) \supseteq \tau \sum_{k=1}^{N} R_k^* \partial d_k^{(n)*}(-R_k F'(u^{(n)})) + u^{(n)}. \tag{4.13}$$

Combining (4.10) and (4.13), we have (4.8). ∎

Thanks to Lemma 4.5, it suffices to verify that Assumption 3.1 holds when $B(u, v) = M_{\tau,\omega}(u, v)$ in order to show the convergence of Algorithm 4.1. In the following, we prove that Assumptions 4.1 to 4.3 are sufficient to ensure Assumption 3.1.

LEMMA 4.6. *Suppose that Assumptions* 4.1 *to* 4.3 *hold. Let* $\tau \in (0, \tau_0]$ *and* $\omega \geq \omega_0$. *For any bounded and convex subset* K *of* V, *we have*

$$\begin{aligned} D_F(u,v) + G(u) &\leq M_{\tau,\omega}(u,v) \\ &\leq \frac{\omega C_{0,K'}^q}{q\tau^{q-1}} \|u-v\|^q + \tau G\left(\frac{1}{\tau}u - \left(\frac{1}{\tau}-1\right)v\right) \\ &\quad + (1-\tau)G(v), \quad u,v \in K \cap \operatorname{dom} G, \end{aligned}$$

where the functional $M_{\tau,\omega}$ *was given in* (4.5) *and*

$$K' = \left\{ \frac{1}{\tau}u - \left(\frac{1}{\tau}-1\right)v : u,v \in K \right\}. \tag{4.14}$$

Proof. Take any $w_k \in V_k$ such that

$$u - v = \tau \sum_{k=1}^N R_k^* w_k. \tag{4.15}$$

By Assumption 4.3 we get

$$\begin{aligned} \tau \sum_{k=1}^N \omega d_k(w_k, v) &\geq \tau \sum_{k=1}^N D_F(v + R_k^* w_k, v) \\ &= \tau \sum_{k=1}^N F(v + R_k^* w_k) - \tau N F(v) - \langle F'(v), u - v \rangle \end{aligned} \tag{4.16a}$$

and

$$\tau \sum_{k=1}^N G_k(w_k, v) \geq \tau \sum_{k=1}^N G(v + R_k^* w_k). \tag{4.16b}$$

Then using Assumption 4.2, we have

$$\begin{aligned} &\tau \sum_{k=1}^N (\omega d_k + G_k)(w_k, v) + (1 - \tau N)\, G(v) \\ &\quad \overset{(4.16)}{\geq} (1 - \tau N)\, E(v) + \tau \sum_{k=1}^N E(v + R_k^* w_k) - F(v) - \langle F'(v), u - v \rangle \\ &\quad \geq E(u) - F(v) - \langle F'(v), u - v \rangle \\ &\quad = D_F(u,v) + G(v). \end{aligned}$$

Taking the infimum on the left-hand side of the above equation over all w_k satisfying (4.15) yields

$$D_F(u,v) + G(u) \leq M_{\tau,\omega}(u,v).$$

On the other hand, let

$$\bar{u} = \frac{1}{\tau}u - \left(\frac{1}{\tau} - 1\right)v.$$

Since $\bar{u}, v \in K'$, by Assumption 4.1, there exist $\bar{w}_k \in V_k$, $1 \leq k \leq N$, such that

$$\bar{u} - v = \sum_{k=1}^{N} R_k^* \bar{w}_k, \tag{4.17}$$

$$\sum_{k=1}^{N} d_k(\bar{w}_k, v) \leq \frac{C_{0,K'}^q}{q} \|\bar{u} - v\|^q, \tag{4.18}$$

and

$$\sum_{k=1}^{N} G_k(\bar{w}_k, v) \leq G(\bar{u}) + (N-1)G(v).$$

Note that

$$u - v = \tau(\bar{u} - v) = \tau \sum_{k=1}^{N} R_k^* \bar{w}_k.$$

By (4.17) and (4.18), it follows that

$$\begin{aligned}
M_{\tau,\omega}(u, v) &\leq \tau \sum_{k=1}^{N} (\omega d_k + G_k)(\bar{w}_k, v) + (1 - \tau N)G(v) \\
&\leq \frac{\tau \omega C_{0,K'}^q}{q} \|\bar{u} - v\|^q + \tau G(\bar{u}) + (1 - \tau)G(v) \\
&= \frac{\omega C_{0,K'}^q}{q \tau^{q-1}} \|u - v\|^q + \tau G\left(\frac{1}{\tau} u - \left(\frac{1}{\tau} - 1\right) v\right) + (1 - \tau)G(v).
\end{aligned}$$

Now, the proof is complete. □

Lemma 4.6 means that Algorithm 4.1 satisfies Assumption 3.1 under Assumptions 4.1 to 4.3 with $\theta = \tau$, $L_K = \omega C_{0,K'}^q / \tau^{q-1}$. Then by Lemma 3.2, the energy sequence $\{E(u^{(n)})\}$ generated by Algorithm 4.1 always decreases. Thus, if we define the set $K_0 \subseteq V$ as (3.3), the sequence $\{u^{(n)}\}$ is contained in K_0. Recall that we can choose $R_0 > 0$ satisfying (3.4). In the following, we write $C_0 = C_{0,K_0'}$, where K_0' is defined in the same way as (4.14). If F additionally satisfies Assumption 3.4, we write $\mu = \mu_{K_0}$. We define the *additive Schwarz condition number* κ_{ASM} as follows:

$$\kappa_{\mathrm{ASM}} = \frac{\omega C_0^q}{\tau^{q-1}}. \tag{4.19}$$

Then the value of κ_{ASM} depends on τ, ω, and $u^{(0)}$. By Theorems 3.3 and 3.6, the following convergence theorems for Algorithm 4.1 are straightforward.

THEOREM 4.7. *Suppose that Assumptions* 4.1 *to* 4.3 *hold. In Algorithm* 4.1*, if* $E(u^{(0)}) - E(u^*) \geq \tau^{q-1} R_0^q \kappa_{ASM}$*, then*

$$E(u^{(1)}) - E(u^*) \leq \left(1 - \tau\left(1 - \frac{1}{q}\right)\right)(E(u^{(0)}) - E(u^*)).$$

Otherwise, we have

$$E(u^{(n)}) - E(u^*) \leq \frac{C_{q,\tau} R_0^q \kappa_{ASM}}{(n+1)^{q-1}}, \quad n \geq 0,$$

where $C_{q,\tau}$ *is a positive constant defined in* (A.10) *depending on* q *and* τ *only,* R_0 *was defined in* (3.4)*, and* κ_{ASM} *was defined in* (4.19)*.*

THEOREM 4.8. *Suppose that Assumptions* 3.4 *and* 4.1 *to* 4.3 *hold. In Algorithm* 4.1, *we have the following:*

1. *In the case $p = q$, we have*

$$E(u^{(n)}) - E(u^*) \\ \leq \left(1 - \left(1 - \frac{1}{q}\right)\min\left\{\tau, \left(\frac{\mu}{q\kappa_{ASM}}\right)^{\frac{1}{q-1}}\right\}\right)^n (E(u^{(0)}) - E(u^*)),\ n \geq 0.$$

2. *In the case $p > q$, if $E(u^{(0)}) - E(u^*) \geq p^{\frac{q}{p-q}}\tau^{\frac{p(q-1)}{p-q}}(\kappa_{ASM}^p/\mu^q)^{\frac{1}{p-q}}$, then*

$$E(u^{(1)}) - E(u^*) \leq \left(1 - \tau\left(1 - \frac{1}{q}\right)\right)(E(u^{(0)}) - E(u^*)).$$

Otherwise, we have

$$E(u^{(n)}) - E(u^*) \leq \frac{C_{p,q,\tau}\,(\kappa_{ASM}^p/\mu^q)^{\frac{1}{p-q}}}{(n+1)^{\frac{p(q-1)}{p-q}}}, \quad n \geq 0,$$

where $C_{p,q,\tau}$ is a positive constant defined in (A.16) *depending on p, q, and τ only, and κ_{ASM} was defined in* (4.19).

In Theorems 4.7 and 4.8, we observe that the asymptotic convergence rate of Algorithm 4.1 becomes faster as κ_{ASM} becomes smaller. Therefore, getting sharp estimates for C_0, τ_0, and ω_0 is important in the analysis of additive Schwarz methods. We will consider in section 6 how to estimate those constants.

4.1. Relation to the classical additive Schwarz theory. Here, we show that the additive Schwarz framework proposed in this section is a generalization of the classical theory for linear elliptic problems developed in [32]. Throughout this section, H denotes a Hilbert space.

Let $a(\cdot,\cdot)\colon H \times H \to \mathbb{R}$ be a continuous and symmetric positive definite (SPD) bilinear form on H, and $f \in H^*$. We consider the variational problem

$$a(u,v) = \langle f, v\rangle, \quad v \in H.$$

The above problem is standard in the field of elliptic partial differential equations. By the Lax–Milgram theorem, a unique solution $u^* \in H$ of the above problem is characterized as a solution of the minimization problem

$$\min_{u\in H}\left\{F(u) := \frac{1}{2}a(u,u) - \langle f, u\rangle\right\}. \tag{4.20}$$

If one has

$$a(u,v) = \langle Au, v\rangle, \quad u, v \in H$$

for some continuous and SPD linear operator $A\colon H \to H^*$, the energy functional $F(u)$ in (4.20) is Frechét differentiable with the derivative $F'(u) = Au - f \in H^*$ for $u \in H$. Hence, (4.20) is a particular instance of (3.1) and the theory developed in section 4 is applicable. In this case, the Bregman distance of F is given by

$$D_F(u,v) = \frac{1}{2}a(u-v, u-v), \quad u, v \in H.$$

We equip H with the energy norm $\|u\|_A = \sqrt{a(u,u)}$. Then Assumption 3.4 is true with $p = 2$ and $\mu_K = 1$ for all bounded and convex $K \subseteq H$.

In what follows, let an index k run from 1 to N. Similarly to (4.1), we assume that H admits a decomposition

$$H = \sum_{k=1}^{N} R_k^* H_k,$$

where H_k is a Hilbert space and R_k^*: $H_k \to H$ is a bounded linear operator with the surjective adjoint. We set d_k in (4.2) by

$$d_k(w_k, v) = \frac{1}{2}\tilde{a}_k(w_k, w_k), \quad w_k \in H_k,$$

for some continuous and SPD bilinear form $\tilde{a}_k(\cdot,\cdot)$ on H_k. Note that the above definition of $d_k(w_k, v)$ is independent of v, so that we may simply write $d_k(w_k) = d_k(w_k, v)$ for $w_k \in V_k$ and $v \in V$. In this setting, Assumption 4.1 with $q = 2$ is reduced to the following.

Assumption 4.9. There exists a constant $C_0 > 0$ which satisfies the following: for any $w \in H$, there exists $w_k \in H_k$, $1 \leq k \leq N$ such that

$$w = \sum_{k=1}^{N} R_k^* w_k$$

and

$$\sum_{k=1}^{N} \tilde{a}_k(w_k, w_k) \leq C_0^2 \|w\|_A^2. \tag{4.21}$$

Compared to Assumption 4.1, the dependency on the subset K is dropped in Assumption 4.9 since all terms in (4.21) are 2-homogeneous. Then Assumption 4.9 exactly agrees with [32, Assumption 2.2]. The following assumption is what Assumption 4.2 is reduced to.

Assumption 4.10. There exists a constant $\tau_0 > 0$ which satisfies the following: for any $w_k \in H_k$, $1 \leq k \leq N$, and $\tau \in (0, \tau_0]$, we have

$$a\left(\sum_{k=1}^{N} R_k^* w_k, \sum_{k=1}^{N} R_k^* w_k\right) \leq \frac{1}{\tau} \sum_{k=1}^{N} a(R_k^* w_k, R_k^* w_k).$$

We recall that [32, Assumption 2.3], also known as strengthened Cauchy–Schwarz inequalities on spaces $\{H_k\}$, is written as follows: there exists constants $\epsilon_{ij} \in [0,1]$, $1 \leq i, j \leq N$, such that

$$|a(R_i^* w_i, R_j^* w_j)| \leq \epsilon_{ij} a(R_i^* w_i, R_i^* w_i)^{1/2} a(R_j^* w_j, R_j^* w_j)^{1/2}, \quad w_i \in H_i,\ w_j \in H_j. \tag{4.22}$$

Suppose that (4.22) holds. Then by the same argument as [32, Lemma 2.6], for

$w_k \in H_k$ we have

$$\begin{aligned} a\left(\sum_{k=1}^N R_k^* w_k, \sum_{k=1}^N R_k^* w_k\right) &= \sum_{i=1}^N \sum_{j=1}^N a(R_i^* w_i, R_j^* w_j) \\ &\leq \sum_{i=1}^N \sum_{j=1}^N \epsilon_{ij} a(R_i^* w_i, R_i^* w_i)^{1/2} a(R_j^* w_j, R_j^* w_j)^{1/2} \\ &\leq \rho(\mathcal{E}) \sum_{k=1}^N a(R_k^* w_k, R_k^* w_k), \end{aligned}$$

where $\rho(\mathcal{E})$ is the spectral radius of the matrix $\mathcal{E} = [\epsilon_{ij}]_{i,j=1}^N$. Therefore, $\tau_0 = 1/\rho(\mathcal{E})$ satisfies Assumption 4.10. In a trivial case of (4.22) when $\epsilon_{ij} = 1$ for all i and j, we have $\rho(\mathcal{E}) = N$ and it agrees with the trivial case $\tau_0 = 1/N$ of Assumption 4.10 noted in the previous section. In this sense, we can say that Assumption 4.2 is a generalization of [32, Assumption 2.3]. Finally, we consider a reduced version of Assumption 4.3.

Assumption 4.11. There exists a constant $\omega_0 > 0$ which satisfies the following: for any $w_k \in H_k$, $1 \leq k \leq N$, we have

$$a(R_k^* w_k, R_k^* w_k) \leq \omega_0 \tilde{a}_k(w_k, w_k).$$

Assumption 4.11 has the same form as [32, Assumption 2.4]. In summary, Assumptions 4.1 to 4.3 can be regarded as generalizations of the three assumptions in the abstract convergence theory of Schwarz methods for linear elliptic problems presented in [32].

Next, we claim that Lemma 4.5 is a direct generalization of the well-known *additive Schwarz lemma* (see [32, Lemma 2.5]), which plays a key role in the convergence analysis of Schwarz methods for linear problems. Let

$$\tilde{a}_k(w_k, w_k) = \langle \tilde{A}_k w_k, w_k \rangle, \quad w_k \in V_k,$$

for some continuous and SPD linear operator $\tilde{A}_k$: $H_k \to H_k^*$. We readily obtain

$$d_k^*(p_k) = \frac{1}{2} \langle p_k, \tilde{A}_k^{-1} p_k \rangle, \quad p_k \in V_k^*.$$

For fixed $v \in V$, we write $M_{\tau,\omega}(u) = M_{\tau,\omega}(u, v)$, where $M_{\tau,\omega}(u, v)$ was defined in (4.5). That is,

$$\begin{aligned} M_{\tau,\omega}(u) &= \tau \inf \left\{ \sum_{k=1}^N \frac{\omega}{2} \tilde{a}_k(w_k, w_k) : u - v = \tau \sum_{k=1}^N R_k^* w_k,\ w_k \in V_k \right\} \\ &= \frac{\omega}{2\tau} \inf \left\{ \sum_{k=1}^N \tilde{a}_k(w_k, w_k) : u - v = \sum_{k=1}^N R_k^* w_k,\ w_k \in V_k \right\}. \end{aligned}$$

By the same argument as (4.11), we have

$$M_{\tau,\omega}^*(p) = \frac{\tau}{2\omega} \left\langle p, \left(\sum_{k=1}^N R_k^* \tilde{A}_k^{-1} R_k \right) p \right\rangle + \langle p, v \rangle, \quad p \in V^*. \tag{4.23}$$

Dualizing (4.23) yields

$$M_{\tau,\omega}(u) = \frac{\omega}{2\tau}\left\langle \left(\sum_{k=1}^{N} R_k^* \tilde{A}_k^{-1} R_k\right)^{-1} (u-v), u-v \right\rangle, \quad u \in V. \tag{4.24}$$

That is, the functional $M_{\tau,\omega}$ is in fact a scaled quadratic form induced by the *additive Schwarz preconditioner* $M \colon V \to V^*$, which is defined by

$$M = \left(\sum_{k=1}^{N} R_k^* \tilde{A}_k^{-1} R_k\right)^{-1}.$$

Consequently, Lemma 4.5 and (4.24) imply that Algorithm 4.1 for (4.20) is the preconditioned Richardson method

$$u^{(n+1)} = u^{(n)} - \frac{\tau}{\omega} M^{-1}(Au^{(n)} - f), \quad n \geq 0.$$

Let $P_{\mathrm{ad}} = M^{-1}A$ be the additive operator introduced in [32, section 2.2]. Then we have

$$a(P_{\mathrm{ad}}^{-1}u, u) = \langle Mu, u\rangle = \inf\left\{ \sum_{k=1}^{N} \tilde{a}_k(u_k, u_k) : u = \sum_{k=1}^{N} R_k^* u_k,\ u_k \in V_k \right\},$$

which is the conclusion of the classical additive Schwarz lemma. In this sense, we call Lemma 4.5 the generalized additive Schwarz lemma.

Under Assumptions 4.9 to 4.11, one can easily prove using (4.24) that

$$\frac{\tau_0}{\omega_0}\|w\|_A^2 \leq \langle Mw, w\rangle \leq C_0^2 \|w\|_A^2.$$

Therefore, the condition number of the preconditioned operator P_{ad} is bounded by $\omega_0 C_0^2/\tau_0$. This bound agrees with [32, Theorem 2.7]. Moreover, it agrees with (4.19) in the case $\tau = \tau_0$ and $\omega = \omega_0$. Therefore, the additive Schwarz condition number κ_{ASM} introduced in section 4 generalizes the condition number of P_{ad}.

5. Overlapping domain decomposition.

In this section, we present overlapping domain decomposition settings for finite element spaces that will be used in this paper. In the remainder of the paper, let Ω be a bounded polygonal domain in $\mathbb{R}^d$, where d is a positive integer. The notation $A \lesssim B$ means that there exists a constant $c > 0$ such that $A \leq cB$, where c is independent of the parameters H, h, and δ which are related to the geometry of domain decomposition and will be defined later. We also write $A \approx B$ if $A \lesssim B$ and $B \lesssim A$.

As a coarse mesh, let $\mathcal{T}_H$ be a quasi-uniform triangulation of Ω with H the maximal element diameter. We refine the coarse mesh $\mathcal{T}_H$ to obtain a quasi-uniform triangulation $\mathcal{T}_h$ with $h < H$, which plays a role of a fine mesh. Let $S_H(\Omega) \subset W_0^{1,\infty}(\Omega)$ and $S_h(\Omega) \subset W_0^{1,\infty}(\Omega)$ be the continuous, piecewise linear finite element spaces on $\mathcal{T}_H$ and $\mathcal{T}_h$ with the homogeneous essential boundary condition, respectively. For sufficiently smooth functions, the nodal interpolation operators I_H and I_h onto $S_H(\Omega)$ and $S_h(\Omega)$, respectively, are well-defined.

We decompose Ω into N nonoverlapping subdomains $\{\Omega_k\}_{k=1}^N$ such that each Ω_k is the union of some coarse elements in $\mathcal{T}_H$, and the number of coarse elements consisting of Ω_k is uniformly bounded. For each Ω_k, we make a larger region Ω_k' by adding layers of fine elements with the width δ. We define $S_h(\Omega_k') \subset W_0^{1,\infty}(\Omega_k')$ as the continuous, piecewise linear finite element space on the $\mathcal{T}_h$-elements in Ω_k' with the homogeneous essential boundary condition.

In the additive Schwarz framework presented in section 4, we set

$$V = S_h(\Omega) \quad \text{and} \quad V_k = S_h(\Omega_k'), \quad 1 \le k \le N. \tag{5.1}$$

We also define

$$V_0 = S_H(\Omega).$$

We take $R_k^*\colon V_k \to V$ as the natural extension operator for $1 \le k \le N$, and $R_0^*\colon V_0 \to V$ as the natural interpolation operator. Then it is clear that

$$V = \sum_{k=1}^{N} R_k^* V_k \tag{5.2}$$

and

$$V = R_0^* V_0 + \sum_{k=1}^{N} R_k^* V_k. \tag{5.3}$$

We say that an additive Schwarz method is said to be *one-level* if it uses the space decomposition (5.2), while it is called *two-level* if it uses (5.3).

5.1. Coloring technique. In section 4, we noted that a constant τ_0 in Assumption 4.2 larger than $1/N$ can be found by the coloring technique. Now, we explain the details. In the proposed method, we say that two spaces V_i and V_j are *of the same color* if

$$\begin{aligned} &E(v + R_i^* w_i + R_j^* w_j) - E(v) \\ &\quad = (E(v + R_i^* w_i) - E(v)) + \left(E(v + R_j^* w_j) - E(v)\right), \quad v \in V,\ w_i \in V_i,\ w_j \in V_j. \end{aligned} \tag{5.4}$$

Inductively, one can prove the following: if $V_{k_1}, \dots, V_{k_m}$ are of the same color, then we have

$$E\left(v + \sum_{i=1}^{m} R_{k_i}^* w_{k_i}\right) - E(v) = \sum_{i=1}^{m} \left(E(v + R_{k_i}^* w_{k_i}) - E(v)\right), \quad v \in V,\ w_{k_i} \in V_{k_i}. \tag{5.5}$$

Assume that the local spaces $\{V_k\}_{k=1}^N$ are classified into N_c colors according to (5.4) for some $N_c \le N$. Let $\mathcal{I}_j$, $1 \le j \le N_c$ be the set of the indices k such that V_k is of the color j. Then for $\tau \in (0, 1/N_c]$, $v \in V$, and $w_k \in V_k$, we have

$$\begin{aligned} &(1 - \tau N)E(v) + \tau \sum_{k=1}^{N} E(v + R_k^* w_k) - E\left(v + \tau \sum_{k=1}^{N} R_k^* w_k\right) \\ &\overset{(5.5)}{=} \tau \sum_{j=1}^{N_c} \left(E\left(v + \sum_{k \in \mathcal{I}_j} w_k\right) - E(v)\right) - \left(E\left(v + \tau \sum_{j=1}^{N_c} \sum_{k \in \mathcal{I}_j} R_k^* w_k\right) - E(v)\right) \\ &= \tau \sum_{j=1}^{N_c} E\left(v + \sum_{k \in \mathcal{I}_j} w_k\right) + (1 - \tau N_c)E(v) - E\left(v + \tau \sum_{j=1}^{N_c} \sum_{k \in \mathcal{I}_j} R_k^* w_k\right) \\ &\ge 0, \end{aligned}$$

where the last inequality is due to the convexity of E and $1 - \tau N_c \geq 0$. Therefore, Assumption 4.2 is true with $\tau_0 = 1/N_c$.

In most applications, $E(u)$ has the integral structure and naturally satisfies (5.4). As a descriptive example, let V and V_k be given by (5.1) and

$$E(u) = \frac{1}{s} \int_\Omega |\nabla u|^s \, dx - \langle f, u \rangle$$

for $s > 1$ and $f \in V^*$. Then it is obvious that V_i and V_j are of the same color if $\bar{\Omega}_i' \cap \bar{\Omega}_j' = \emptyset$. Hence, for suitable overlap parameter δ, we have

$$N_c \leq \begin{cases} 2 & \text{if } d = 1, \\ 4 & \text{if } d = 2, \\ 8 & \text{if } d = 3. \end{cases}$$

For two-level methods, we have $\tau_0 = 1/(N_c + 1)$ because of the coarse space V_0. In summary, we have

$$\tau_0 = \begin{cases} \frac{1}{N_c} & \text{for one-level (5.2)}, \\ \frac{1}{N_c+1} & \text{for two-level (5.3)}. \end{cases} \tag{5.6}$$

We conclude the section by observing a special case when $V = H$ is a Hilbert space and

$$E(u) = \frac{1}{2} \langle Au, u \rangle - \langle f, u \rangle$$

for a continuous, symmetric, positive definite linear operator $A \colon H \to H^*$ and $f \in H^*$. Then (5.4) reduces to

$$\left\langle AR_i^* w_i, R_j^* w_j \right\rangle = 0, \quad w_i \in V_i, \ w_j \in V_j,$$

which agrees with [32, section 2.5.1]. In this sense, the proposed coloring technique generalizes the theory developed in [32].

5.2. One-level domain decomposition. First, we consider the one-level domain decomposition (5.2). By [32, Lemma 3.4], we can choose a continuous and piecewise linear partition of unity $\{\theta_k\}_{k=1}^N$ for Ω subordinate to the covering $\{\Omega_k'\}_{k=1}^N$ satisfying [32, equations (3.2) and (3.3)]. Invoking [32, Lemmas 3.4 and 3.9], the following lemma is straightforward under the space decomposition (5.2).

LEMMA 5.1. *Assume that the space V is decomposed according to* (5.2). *For $w \in V$, we choose $w_k \in V_k$, $1 \leq k \leq N$ such that*

$$R_k^* w_k = I_h(\theta_k w). \tag{5.7}$$

Then for $s \geq 1$, we have $w = \sum_{k=1}^N R_k^* w_k$ and

$$\sum_{k=1}^N \|R_k^* w_k\|_{W^{1,s}(\Omega)} \lesssim C_{N_c} \left(1 + \frac{1}{\delta} \right) \|w\|_{W^{1,s}(\Omega)},$$

where C_{N_c} is a positive constant depending on the number of colors N_c only.

5.3. Two-level domain decomposition. There are several results on stable decompositions for the two-level domain decomposition (5.3) which are counterparts to Lemma 5.1. If we choose a coarse component $w_0 \in V_0$ of $w \in V$ by the L^2-projection technique, we obtain the following estimate [31, Lemma 4.1].

LEMMA 5.2. *Assume that the space V is decomposed according to* (5.3). *For $w \in V$, let $w_0 \in V_0$ such that $R_0^* w_0$ is the L^2-projection of w, i.e.,*

$$\int_\Omega (R_0^* w_0 - w) R_0^* \phi_0 \, dx = 0, \quad \phi_0 \in V_0. \tag{5.8a}$$

Then we choose $w_k \in V_k$, $1 \leq k \leq N$ such that

$$R_k^* w_k = I_h(\theta_k(w - R_0^* w_0)). \tag{5.8b}$$

For $s \geq 1$, we have $w = R_0^* w_0 + \sum_{k=1}^N R_k^* w_k$ and

$$\|R_0^* w_0\|_{W^{1,s}(\Omega)} + \sum_{k=1}^N \|R_k^* w_k\|_{W^{1,s}(\Omega)} \lesssim C_{N_c} \left(1 + \left(\frac{H}{\delta} \right)^{\frac{s-1}{s}} \right) \|w\|_{W^{1,s}(\Omega)},$$

where C_{N_c} is a positive constant depending on the number of colors N_c only.

In applications to nonsmooth optimization problems, we may need a decomposition different from Lemma 5.2 in order to satisfy (4.4) [4, 29]. Let $I_H^\ominus \colon S_h(\Omega) \to S_H(\Omega)$ be the nonlinear interpolation operator defined in [29, section 4]. One may refer to [1, 29] for some useful estimates related to $I_H^\ominus$. Then we have the following estimate on a decomposition using $I_H^\ominus$ [1, Proposition 4.1].

LEMMA 5.3. *Assume that the space V is decomposed according to* (5.3). *For $w \in V$, we define $w_0 \in V_0$ by*

$$R_0^* w_0 = I_H^\ominus (\max(0, w)) - I_H^\ominus (\max(0, -w)). \tag{5.9a}$$

Then we choose $w_k \in V_k$, $1 \leq k \leq N$ such that

$$R_k^* w_k = I_h(\theta_k(w - R_0^* w_0)). \tag{5.9b}$$

For $s \geq 1$, we have $w = R_0^* w_0 + \sum_{k=1}^N R_k^* w_k$ and

$$\|R_0^* w_0\|_{W^{1,s}(\Omega)} + \sum_{k=1}^N \|R_k^* w_k\|_{W^{1,s}(\Omega)} \lesssim C_{N_c} C_{d,s}(H,h) \left(1 + \frac{H}{\delta} \right) \|w\|_{W^{1,s}(\Omega)},$$

where C_{N_c} is a positive constant depending on the number of colors N_c only and

$$C_{d,s}(H,h) = \begin{cases} 1 & \text{if } d = s = 1 \text{ or } 1 \leq d < s, \\ \left(1 + \log \frac{H}{h}\right)^{\frac{s-1}{s}} & \text{if } 1 < d = s, \\ \left(\frac{H}{h}\right)^{\frac{d-s}{s}} & \text{if } 1 \leq s < d. \end{cases}$$

6. Applications. In this section, we present various applications of the proposed abstract convergence theory for additive Schwarz methods. The proposed theory covers many interesting convex optimization problems: nonlinear elliptic problems, nonsmooth problems, and nonsharp problems. It also gives a unified analysis with some other decomposition methods such as block coordinate descent methods and constraint decomposition methods. The proposed theory can adopt stable decomposition estimates for Schwarz methods presented in existing works without modification. This makes the convergence analysis of additive Schwarz methods easy and gives an equivalent or even better estimate for the convergence rate compared to existing works.

6.1. Nonlinear elliptic problems. We present applications of the proposed additive Schwarz method to some nonlinear elliptic partial differential equations on Ω. We consider the minimization problem

$$\min_{u \in W_0^{1,s}(\Omega)} \left\{ E(u) := \frac{1}{s} \int_\Omega |\nabla u|^s \, dx - \langle f, u \rangle \right\} \tag{6.1}$$

for some $s > 1$ such that $s \neq 2$ and $f \in W^{-1,s^*}(\Omega)$, where s^* is the Hölder conjugate of s, i.e., $\frac{1}{s} + \frac{1}{s^*} = 1$. The unique solution of (6.1) is characterized by a solution of the well-known s-Laplacian equation

$$\begin{aligned} -\operatorname{div}\left(|\nabla u|^{s-2} \nabla u\right) &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

It is well-known that there exist two positive constants α_s and β_s such that for any $u, v \in W_0^{1,s}(\Omega)$, we have

$$\langle E'(u) - E'(v), u - v \rangle \geq \alpha_s \|u - v\|_{W^{1,s}(\Omega)}^s, \tag{6.2a}$$

$$\|E'(u) - E'(v)\|_{W^{-1,s^*}(\Omega)} \leq \beta_s \left(\|u\|_{W^{1,s}(\Omega)} + \|v\|_{W^{1,s}(\Omega)} \right)^{s-2} \|u - v\|_{W^{1,s}(\Omega)} \tag{6.2b}$$

if $s > 2$ and

$$\langle E'(u) - E'(v), u - v \rangle \geq \alpha_s \frac{\|u - v\|_{W^{1,s}(\Omega)}^2}{(\|u\|_{W^{1,s}(\Omega)} + \|v\|_{W^{1,s}(\Omega)})^{2-s}}, \tag{6.3a}$$

$$\|E'(u) - E'(v)\|_{W^{-1,s^*}(\Omega)} \leq \beta_s \|u - v\|_{W^{1,s}(\Omega)}^{s-1} \tag{6.3b}$$

if $1 < s < 2$, where $\|\cdot\|_{W^{-1,s^*}(\Omega)}$ is the dual norm of $\|\cdot\|_{W^{1,s}(\Omega)}$; see [14].

A conforming finite element approximation of (6.1) using $S_h(\Omega) \subset W_0^{1,s}(\Omega)$ is given by

$$\min_{u \in S_h(\Omega)} \left\{ E_h(u) := \frac{1}{s} \int_\Omega |\nabla u|^s \, dx - \langle f, u \rangle \right\}. \tag{6.4}$$

Clearly, (6.4) is an instance of (3.1) with

$$V = S_h(\Omega), \quad F(u) = \frac{1}{s} \int_\Omega |\nabla u|^s \, dx - \langle f, u \rangle, \quad G(u) = 0.$$

One can show that E_h is coercive without difficulty [31].

We take any bounded and convex subset K of V and define $M_K > 0$ by

$$M_K = \sup_{u \in K} \|u\|_{W^{1,s}(\Omega)} < \infty.$$

We choose $u, v \in K$ arbitrarily. Note that $v + t(u - v) \in K$ for any $t \in [0, 1]$. It follows by the fundamental theorem of calculus that

$$\begin{aligned} D_F(u, v) &= \int_0^1 \langle F'(v + t(u - v)), u - v \rangle \, dt - \langle F'(v), u - v \rangle \\ &= \int_0^1 \frac{1}{t} \langle F'(v + t(u - v)) - F'(v), t(u - v) \rangle \, dt. \end{aligned} \tag{6.5}$$

If $s > 2$, combining (6.2) and (6.5), we have

$$D_F(u,v) \geq \frac{\alpha_s}{s}\|u-v\|_{W^{1,s}(\Omega)}^s, \tag{6.6a}$$

$$D_F(u,v) \leq \frac{\beta_s(2M_K)^{s-2}}{2}\|u-v\|_{W^{1,s}(\Omega)}^2. \tag{6.6b}$$

Similarly, if $1 < s < 2$, then we obtain

$$D_F(u,v) \geq \frac{\alpha_s}{(2M_K)^{2-s}}\|u-v\|_{W^{1,s}(\Omega)}^2, \tag{6.7a}$$

$$D_F(u,v) \leq \frac{\beta_s}{s}\|u-v\|_{W^{1,s}(\Omega)}^s \tag{6.7b}$$

by using (6.3) and (6.5).

Suppose that we want to solve (6.4) by Algorithm 4.1. We should verify Assumptions 4.1 to 4.3 to ensure the convergence, and Assumption 3.4 if possible. If we use the domain decompositions given by either (5.2) or (5.3), then Assumption 4.2 is straightforward with (5.6). Assumption 4.3 holds with $\omega_0 = 1$ in the case of the exact local solvers. Moreover, invoking Proposition 3.5 to either (6.6) or (6.7) implies Assumption 3.4.

Next, we show that Assumption 4.1 is valid for the two domain decompositions (5.2) and (5.3). Take any bounded and convex subset K of V and let $u, v \in V$ with $w = u - v$. For the one-level domain decomposition (5.2), we set $w_k \in V_k$, $1 \leq k \leq N$ as (5.7). For the two-level case domain decomposition (5.3), we set $w_k \in V_k$, $0 \leq k \leq N$ as (5.8). In the case of $s > 2$ and the one-level domain decomposition, by (6.6) and Lemma 5.1 we have

$$\begin{aligned}\sum_{k=1}^N d_k(w_k, v) &\lesssim \sum_{k=1}^N \|R_k^* w_k\|_{W^{1,s}(\Omega)}^2 \\ &\lesssim \left(1 + \frac{1}{\delta^2}\right)\|w\|_{W^{1,s}(\Omega)}^2.\end{aligned}$$

Similarly, the following estimate can be obtained using (6.6) and Lemma 5.2 in the two-level domain decomposition:

$$d_0(w_0, v) + \sum_{k=1}^N d_k(w_k, v) \lesssim \left(1 + \left(\frac{H}{\delta}\right)^{\frac{2(s-1)}{s}}\right)\|w\|_{W^{1,s}(\Omega)}^2.$$

Therefore, Assumption 4.1 is satisfied if $s > 2$. Results corresponding to the case $1 < s < 2$ can be obtained by the same argument.

In summary, by Theorem 4.8 we have the following convergence results of Algorithm 4.1 for (6.4).

THEOREM 6.1. *In Algorithm* 4.1 *for* (6.4), *suppose that we set* τ_0 *as* (5.6) *and use the exact local solvers. If* $E(u^{(0)}) - E(u^*)$ *is small enough, then we have*

$$E_h(u^{(n)}) - E_h(u^*) \lesssim \frac{1 + 1/\delta^q}{(n+1)^{\frac{p(q-1)}{p-q}}} \qquad \textit{for one-level,}$$

$$E_h(u^{(n)}) - E_h(u^*) \lesssim \frac{1 + (H/\delta)^{\frac{q(s-1)}{s}}}{(n+1)^{\frac{p(q-1)}{p-q}}} \qquad \textit{for two-level,}$$

where

$$\begin{aligned} &p = s,\ q = 2 \quad &&\textit{if}\ \ s > 2, \\ &p = 2,\ q = s \quad &&\textit{if}\ \ 1 < s < 2. \end{aligned}$$

A remarkable property of Algorithm 4.1 for (6.4) is that the convergence of the method is not significantly affected by the initial energy error $E_h(u^{(0)}) - E_h(u^*)$, even though the asymptotic convergence rate is only sublinear. In Theorem 4.8, we showed that the energy error decays linearly if it is sufficiently large. Therefore, the number of iterations required to meet a prescribed stop condition does not become too large even if $E_h(u^{(0)}) - E_h(u^*)$ is very big. We note that a similar discussion was done in [7].

Finally, we compare the above estimates with existing ones. In [31], it was proven that Algorithm 4.1 applied to (6.4) has the $O(1/n^{\frac{q(q-1)}{(p-q)(p+q-1)}})$ convergence rate of the energy error. More recently, the $O(1/n^{\frac{q-1}{p-q}})$ convergence of the energy error was shown in [1, 3]. Since

$$\frac{q(q-1)}{(p-q)(p+q-1)} < \frac{q-1}{p-q} < \frac{p(q-1)}{p-q}$$

for $1 < q < p$, we conclude that Theorem 6.1 is sharper than the existing results mentioned above.

6.2. Nonsmooth problems. We deal with the problems of the form (3.1) with the nonzero nonsmooth parts, i.e., $G \neq 0$. Suppose that G satisfies the following assumption, which was previously stated in [2, 3].

Assumption 6.2. Let $\mathcal{N}_h$ be the set of vertices in the triangulation $\mathcal{T}_h$. Then G can be expressed as

$$G(u) = \sum_{x \in \mathcal{N}_h} s_x(h)\phi(u(x))$$

for some convex functions $\phi\colon \mathbb{R} \to \overline{\mathbb{R}}$ and $s_x(h) \geq 0$.

Assumption 6.2 means that G is the sum of pointwisely defined convex functions. Various examples satisfying Assumption 6.2 can be found in [3]. Here, we consider the following L^1-regularized obstacle problem [33]:

$$\min_{u \in H_0^1(\Omega)} \left\{ E(u) := \frac{1}{2}\int_\Omega |\nabla u|^2\,dx - \langle f, u\rangle + \lambda \int_\Omega |u|\,dx \right\}, \tag{6.8}$$

where $f \in H^{-1}(\Omega)$ and $\lambda > 0$. Note that (6.8) has an equivalent variational inequality of the second kind of the form (3.2). We show that the additive Schwarz method for variational inequalities of the second kind proposed in [2] can be represented in our framework.

A finite element approximation of (6.8) using $S_h(\Omega) \subset H_0^1(\Omega)$ is written as

$$\min_{u \in S_h(\Omega)} \left\{ E_h(u) := \frac{1}{2}\int_\Omega |\nabla u|^2\,dx - \langle f, u\rangle + G_h(u) \right\}, \tag{6.9}$$

where $G_h(u)$ is the numerical integration of $\lambda|u|$ using the piecewise linear approximation, i.e.,

$$G_h(u) = \lambda \int_\Omega I_h|u|\,dx.$$

It is clear that G_h satisfies Assumption 6.2. We focus on the fact that (6.9) is an instance of (3.1) with

$$V = S_h(\Omega), \quad F(u) = \frac{1}{2}\int_\Omega |\nabla u|^2 \, dx - \langle f, u \rangle, \quad G(u) = G_h(u).$$

We analyze the convergence behavior of Algorithm 4.1 applied to (6.9) with the space decompositions (5.2) and (5.3). Assume that we use the exact local solvers. Then Assumptions 4.2 and 4.3 are trivially satisfied with (5.6) and $\omega_0 = 1$, respectively. Assumption 3.4 with $p = 2$ and $\mu \approx 1$ is verified by

$$\frac{1}{2}\|u-v\|_{H^1(\Omega)}^2 \approx \frac{1}{2}|u-v|_{H^1(\Omega)}^2 = D_F(u,v), \quad u, v \in V,$$

followed by an application of Proposition 3.5, where we used the Poincaré–Friedrichs inequality for $H_0^1(\Omega)$ in $\approx$.

Now, we prove Assumption 4.1 for two domain decompositions (5.2) and (5.3). Take any $u, v \in V$ and let $w = u - v$. For the one-level domain decomposition (5.2), we set $w_k \in V_k$, $1 \leq k \leq N$ as (5.7). For the two-level case, we set $w_k \in V_k$, $0 \leq k \leq N$ as (5.9). Then using Lemma 5.1 and Assumption 6.2, it is straightforward to show that Assumption 4.1 holds in the one-level case with $q = 2$ and

$$C_{0,K} \lesssim 1 + \frac{1}{\delta}$$

for all K; see [2, Proposition 5.1]. In addition, by using Lemma 5.3 and closely following [2, Proposition 5.2], Assumption 4.1 is satisfied for the two-level domain decomposition with $q = 2$ and

$$C_{0,K} \lesssim C_{d,2}(H,h)\left(1 + \frac{H}{\delta}\right)$$

for all K.

In conclusion, invoking Theorem 4.8 yields the following convergence theorem for Algorithm 4.1 applied to (6.9).

THEOREM 6.3. *In Algorithm* 4.1 *for* (6.9)*, suppose that we set* τ_0 *as* (5.6) *and use the exact local solvers. Then we have*

$$\begin{aligned}
&E_h(u^{(n)}) - E_h(u^*) \\
&\quad \leq \left(1 - \frac{1}{2}\min\left\{\tau, \frac{C}{1 + 1/\delta^2}\right\}\right)^n (E_h(u^{(0)}) - E_h(u^*)) \quad \textit{for one-level}, \\
&E_h(u^{(n)}) - E_h(u^*) \\
&\quad \leq \left(1 - \frac{1}{2}\min\left\{\tau, \frac{C}{C_{d,2}(H,h)^2(1 + (H/\delta)^2)}\right\}\right)^n (E_h(u^{(0)}) - E_h(u^*)) \ \textit{for two-level},
\end{aligned}$$

where $C > 0$ *is a generic constant independent of* H*,* h*, and* δ*.*

Theorem 6.3 agrees with the existing results in [2] in the sense that the linear convergence rate is dependent on the bounds for $C_{0,K}$.

We remark that constrained problems belong to the class of nonsmooth problems. Indeed, for a nonempty, convex, and closed subset K_h of $V = S_h(\Omega)$, the constrained minimization problem

$$\min_{u \in K_h} F(u)$$

can be represented as an unconstrained and nonsmooth minimization problem

$$\min_{u\in V} \left\{ F(u) + \chi_{K_h}(u) \right\}, \tag{6.10}$$

where the functional χ_{K_h} was defined in (2.1). Therefore, additive Schwarz methods for constrained problems can be analyzed in the same way as above. If we let $G = \chi_{K_h}$ in (3.1), then Assumption 6.2 reduces to the following.

Assumption 6.4. Let $\theta \in S_h(\Omega)$ with $0 \leq \theta \leq 1$. Then for $u, v \in K_h$, we have $I_h(\theta u + (1-\theta)v) \in K_h$.

It is clear that Assumption 6.4 holds when K_h is defined in terms of pointwise constraints. The same assumptions as Assumption 6.4 were used in existing works [1, 4] on obstacle problems.

6.3. Absence of the sharpness. The examples we had presented above satisfied Assumption 3.4. Now, we provide an application of the proposed framework to a problem lacking the sharpness, i.e., not satisfying Assumption 3.4. As a model problem, we consider the following:

$$\min_{\mathbf{u}\in H_0(\mathrm{div};\Omega)} \left\{ E(\mathbf{u}) := \tilde{F}(\mathrm{div}\,\mathbf{u}) + \chi_K(\mathbf{u}) \right\}, \tag{6.11}$$

where $\tilde{F}\colon L^2(\Omega) \to \mathbb{R}$ is a convex, Frechét differentiable functional and K is the subset of $H_0(\mathrm{div};\Omega)$ defined by

$$K = \left\{ \mathbf{u} \in H_0(\mathrm{div};\Omega) : |\mathbf{u}| \leq 1 \ \text{ a.e. in } \Omega \right\}.$$

We further assume that $\tilde{F}'$ is Hölder continuous with parameters $q-1 \in (0,1]$ and $\tilde{L} > 0$, so that

$$D_{\tilde{F}}(u,v) \leq \frac{\tilde{L}}{q} \|u-v\|_{L^2(\Omega)}^q, \quad u, v \in L^2(\Omega); \tag{6.12}$$

see [31, Lemma 2.1]. Problems of the form (6.11) are typical in mathematical imaging. More precisely, (6.11) appears in Fenchel–Rockafellar dual problems of total variation regularized problems which are standard in mathematical imaging [19, 23]. Schwarz methods for (6.11) have been studied recently in [13, 28].

A discrete counterpart of (6.11) can be obtained by replacing $H_0(\mathrm{div};\Omega)$ by the lowest order Raviart–Thomas finite element space $\mathbf{S}_h(\Omega)$ [18, 23]:

$$\min_{\mathbf{u}\in \mathbf{S}_h(\Omega)} \left\{ E_h(\mathbf{u}) := \tilde{F}(\mathrm{div}\,\mathbf{u}) + \chi_{K_h}(\mathbf{u}) \right\}. \tag{6.13}$$

In (6.13), K_h is the convex subset of $\mathbf{S}_h(\Omega)$ given by

$$K_h = \left\{ \mathbf{u} \in \mathbf{S}_h(\Omega) : \frac{1}{e}\int_e |\mathbf{u}\cdot \mathbf{n}_e|\, dS \leq 1,\ e\text{: interior faces of } \mathcal{T}_h \right\},$$

where $\mathbf{n}_e$ is the unit outer normal to e. See, e.g., [27] for further properties of the space $\mathbf{S}_h(\Omega)$. We denote a solution of (6.13) by $\mathbf{u}^*$. We observe that (6.13) is of the form (3.1) with

$$V = \mathbf{S}_h, \quad F(\mathbf{u}) = \tilde{F}(\mathrm{div}\,\mathbf{u}), \quad G(\mathbf{u}) = \chi_{K_h}(\mathbf{u}).$$

The energy functional E is coercive due to the χ_{K_h}-term. Because of the large null space of div operator, (6.11) does not satisfy Assumption 3.4.

Based on the overlapping domain decomposition $\{\Omega_k\}$ introduced in section 5, we define

$$V_k = \mathbf{S}_h(\Omega_k'), \quad 1 \leq k \leq N,$$

where $\mathbf{S}_h(\Omega_k')$ is the Raviart–Thomas finite element space on Ω_k' with the homogeneous essential boundary condition. Then it satisfies that

$$V = \sum_{k=1}^{N} R_k^* V_k, \tag{6.14}$$

where R_k^*: $V_k \to V$ is the natural embedding.

We investigate the convergence property of Algorithm 4.1 applied to (6.13) based on the space decomposition (6.14). If we use the exact local solvers, then Assumptions 4.2 and 4.3 are trivial with $\tau_0 = 1/N_c$ and $\omega_0 = 1$. In order to verify Assumption 4.1, we choose any $\mathbf{u}, \mathbf{v} \in K_h$. We define $\mathbf{w}_k \in V_k$, $1 \leq k \leq N$ such that

$$R_k^* \mathbf{w}_k = \Pi_h(\theta_k(\mathbf{u} - \mathbf{v})),$$

where Π_h is the nodal interpolation operator onto $\mathbf{S}_h(\Omega)$. Then we clearly have $\mathbf{v} + R_k^* \mathbf{w}_k \in K_h$ and (4.4) holds. Moreover, we get

$$\begin{aligned}
\sum_{k=1}^{N} d_k(\mathbf{w}_k, \mathbf{v}) &= \sum_{k=1}^{N} D_{\tilde{F}\circ \mathrm{div}}(\mathbf{v} + R_k^* \mathbf{w}_k, \mathbf{v}) \\
&\overset{(6.12)}{\leq} \frac{\tilde{L}}{q} \| \operatorname{div} R_k^* \mathbf{w}_k \|_{L^2(\Omega)}^q \\
&\overset{(a)}{\lesssim} C_{N_c} \tilde{L} \left(1 + \frac{1}{\delta^q} \right) \| \mathbf{u} - \mathbf{v} \|_{H(\mathrm{div};\Omega)}^q,
\end{aligned}$$

where C_{N_c} is a positive constant depending on N_c, and (a) is due to [28, Proposition 4.1]. Hence, Assumption 4.1 also holds.

By Theorem 4.7, we have the following convergence theorem for Algorithm 4.1 applied to (6.13).

Theorem 6.5. *In Algorithm* 4.1 *for* (6.13) *with the space decomposition* (6.14), *suppose that we set $\tau_0 = 1/N_c$ and use the exact local solvers. We also assume that* (6.12) *holds. If $E(u^{(0)}) - E(u^*)$ is small enough, then we have*

$$E_h(u^{(n)}) - E_h(u^*) \lesssim \frac{1 + 1/\delta^q}{(n+1)^{q-1}}.$$

Similarly to the case of (6.4), one can observe from Theorem 4.8 that the initial energy error $E_h(u^{(0)}) - E_h(u^*)$ does not affect the number of required iterations much.

Compared to the analysis in [28], the result presented in Theorem 6.5 only requires the Hölder continuity of $\tilde{F}'$, while [28] requires much stronger conditions: the Lipschitz continuity of $\tilde{F}'$ and the strong convexity of $\tilde{F}$.

6.4. Inexact local solvers. We present two notable instances of the proposed method with inexact local solvers: block coordinate descent methods and constraint decomposition methods.

Block coordinate descent methods are popular in convex optimization and there is a vast literature about them; for example, see [9, 15, 34]. Here, we show that parallel block coordinate descent methods are instances of Algorithm 4.1 with inexact local solvers. In Algorithm 4.1, we set

$$V = \prod_{k=1}^{N} V_k \quad \text{with} \quad V_k = \mathbb{R}^{m_k},\ 1 \leq k \leq N.$$

Let $\tilde{R}_k$: $V \to V_k$ be the natural restriction operator, i.e.,

$$u = [\tilde{R}_k u]_{k=1}^N := (\tilde{R}_1 u, \dots, \tilde{R}_N u), \quad u \in V.$$

We set R_k^*: $V_k \to V$ to be the extension-by-zero operator. Then we clearly have

$$V = \sum_{k=1}^{N} R_k^* V_k$$

and

$$[u_k]_{k=1}^N = \sum_{k=1}^{N} R_k^* u_k, \quad u_k \in V_k,\ 1 \leq k \leq N.$$

In addition, it is satisfied that

$$\sum_{k=1}^{N} R_k^* \tilde{R}_k = I. \tag{6.15}$$

The following assumptions are imposed on F and G.

Assumption 6.6. The functional F: $V \to \mathbb{R}$ has the Lipschitz continuous derivative. That is, there exists a constant $L > 0$ such that

$$\|F'(u) - F'(v)\| \leq L\|u - v\|, \quad u, v \in V.$$

Assumption 6.7. The functional G: $V \to \mathbb{R}$ is block-separable. That is, there exist functionals G^k: $V_k \to \mathbb{R}$, $1 \leq k \leq N$, such that

$$G\left([u_k]_{k=1}^N\right) = \sum_{k=1}^{N} G^k(u_k).$$

In this setting, a simple parallel block coordinate descent method to solve (3.1) is presented in Algorithm 6.1.

In Algorithm 6.1, let $u^{(n)} = [u_k^{(n)}]_{k=1}^N$. Then it is straightforward to observe that the sequence $\{u^{(n)}\}$ generated by Algorithm 6.1 is the same as the one generated by Algorithm 4.1 with $\tau_0 = 1/N$, $\omega = L$, and

$$\begin{aligned} d_k(w_k, v) &= \frac{1}{2}\|R_k^* w_k\|^2, \\ G_k(w_k, v) &= G^k(\tilde{R}_k v + w_k) + \sum_{j \neq k} G^j(\tilde{R}_j v) \end{aligned}$$

for $w_k \in V_k$, $v \in V$.

Algorithm 6.1 Parallel block coordinate descent method for (3.1).

Choose $u_k^{(0)} \in \operatorname{dom} G^k$, $1 \leq k \leq N$, and $\tau \in (0, 1/N]$.
for $n = 0, 1, 2, \ldots$

$$v_k^{(n+1)} = \operatorname*{arg\,min}_{u_k \in V_k} \left\{ F(u^{(n)}) + \langle F'(u^{(n)}), R_k^*(u_k - u_k^{(n)}) \rangle + \frac{L}{2} \|u_k - u_k^{(n)}\|^2 + G^k(u_k) \right\}, \ 1 \leq k \leq N$$
$$u_k^{(n+1)} = (1-\tau) u_k^{(n)} + \tau v_k^{(n+1)}, \quad 1 \leq k \leq N$$

end

By using the Cauchy–Schwarz inequality and Assumption 6.7, it is easy to check that Assumption 4.1 holds with $q = 2$ and $C_{0,K} = \sqrt{N}$ for all K. Indeed, for any $u, v \in V$ with $u - v = [w_k]_{k=1}^N$ for $w_k \in V_k$, $1 \leq k \leq N$, it follows that

$$\sum_{k=1}^N d_k(w_k, v) = \frac{1}{2} \sum_{k=1}^N \|R_k^* w_k\|^2 \leq \frac{N}{2} \left\| \sum_{k=1}^N R_k^* w_k \right\|^2 = \frac{N}{2} \|u - v\|^2$$

and

$$\begin{aligned} \sum_{k=1}^N G_k(w_k, v) &= \sum_{k=1}^N \left(G^k(\tilde{R}_k v + w_k) + \sum_{j \neq k} G^j(\tilde{R}_j v) \right) \\ &= G\left(R_k^*(\tilde{R}_k v + w_k) \right) + (N-1) G\left(\sum_{k=1}^N R_k^* \tilde{R}_k v_k \right) \\ &= G(u) + (N-1) G(v). \end{aligned}$$

Assumption 4.2 is obvious. Assumption 4.3 with $\omega_0 = L$ is a direct consequence of Assumption 6.6. In conclusion, Assumptions 4.1 to 4.3 are verified and the $O(1/n)$ convergence of Algorithm 6.1 is obtained by Theorem 4.7.

Now, we turn our attention to constraint decomposition methods. In [13, 29], constraint decomposition methods were proposed as domain decomposition methods for nonlinear variational inequalities. We show that those methods can be regarded as instances of Algorithm 4.1 with inexact local solvers. In particular, we consider the one-level constraint decomposition method proposed in [29] for (6.10) with

$$F(u) = \frac{1}{2} \int_\Omega |\nabla u|^2 \, dx - \langle f, u \rangle;$$

the two-level method can be treated in a similar way.

We assume that the constraint K_h in (6.10) is one-obstacle, i.e.,

$$K_h = \left\{ u \in S_h(\Omega) : u \geq \underline{g} \right\}$$

for some $\underline{g} \in S_h(\Omega)$. We also assume that the space $V = S_h(\Omega)$ is decomposed as (5.2). We define operators $\tilde{R}_k$: $V \to V_k$, $1 \leq k \leq N$ as

$$\tilde{R}_k u = \left(I_h(\theta_k u) \right) |_{\Omega'_k}, \quad u \in V,$$

so that (6.15) holds. If we set

$$K_h^k = \left\{ u_k \in V_k : u_k \geq \tilde{R}_k \underline{g} \right\}, \quad 1 \leq k \leq N,$$

then it is clear that

$$K_h = \sum_{k=1}^{N} R_k^* K_h^k.$$

The constraint decomposition method proposed in [29] in the above setting is summarized in Algorithm 6.2.

Algorithm 6.2 Constraint decomposition method for (6.10).

Choose $u^{(0)} \in K_h$ and $\tau \in (0, 1/N]$.
for $n = 0, 1, 2, \dots$

$$v_k^{(n+1)} \in \operatorname*{arg\,min}_{u_k \in V_k} \left\{ F\left(R_k^* u_k + \sum_{j \neq k} R_k^* \tilde{R}_k u^{(n)} \right) + \chi_{K_h^k}(u_k) \right\}, \quad 1 \leq k \leq N,$$

$$u^{(n+1)} = (1-\tau) u^{(n)} + \tau \sum_{k=1}^{N} R_k^* v_k^{(n+1)}$$

end

One can check without major difficulty that Algorithm 6.2 is an instance of Algorithm 4.1 with $\tau_0 = 1/N$, $\omega = 1$, and

$$\begin{aligned} d_k(w_k, v) &= D_F(v + R_k^* w_k, v), \\ G_k(w_k, v) &= \chi_{K_h^k}(\tilde{R}_k v + w_k) \end{aligned}$$

for $w_k \in V_k$, $v \in V$. In this sense, in order to prove the convergence of Algorithm 6.2, it suffices to verify Assumptions 4.1 and 4.3.

To verify Assumption 4.1, for any $u, v \in \operatorname{dom} G$, we set $w_k = \tilde{R}_k u - \tilde{R}_k v$, $1 \leq k \leq N$. Then we have

$$u - v = \sum_{k=1}^{N} R_k^* w_k.$$

Also, we get

$$\sum_{k=1}^{N} d_k(w_k, v) \lesssim \left(1 + \frac{1}{\delta^2}\right) \|u - v\|^2$$

by Lemma 5.1, and

$$G_k(w_k, v) = \chi_{K_h^k}(\tilde{R}_k u) = 0, \quad 1 \leq k \leq N.$$

That is, Assumption 4.1 holds with $q = 2$ and

$$C_{0,K} \lesssim 1 + \frac{1}{\delta}$$

for all K.

In Assumption 4.3, clearly we have $\omega_0 = 1$. Moreover, we can prove

$$\chi_{K_h}(v + R_k^* w_k) \leq \chi_{K_h^k}(\tilde{R}_k v + w_k), \quad v \in \operatorname{dom} \chi_{K_h},\ w_k \in V_k,$$

as follows: for any interior node x of $S_h(\Omega'_k)$, we have

$$\begin{aligned}(\tilde{R}_k v + w_k)(x) \geq (\tilde{R}_k \underline{g})(x) \;&\Leftrightarrow\; \theta_k(x)v(x) + w_k(x) \geq \theta_k(x)\underline{g}(x)\\ &\Rightarrow\; v(x) + w_k(x) \geq \underline{g}(x) \quad (\because v \in K_h)\\ &\Leftrightarrow\; (v + R_k^* w_k)(x) \geq \underline{g}(x).\end{aligned}$$

Therefore, Assumption 4.3 is proven.

Since the energy functional of (6.10) satisfies Assumption 3.4, we conclude that Algorithm 6.2 converges linearly. This result agrees with [29].

7. Conclusion. Motivated by the fact that additive Schwarz methods for linear elliptic problems can be represented as preconditioned Richardson methods, we showed that additive Schwarz methods for general convex optimization belong to a class of gradient methods. From this observation, we presented a novel abstract convergence theory for additive Schwarz methods for convex optimization. We also noted that the proposed theory directly generalizes the one presented in [32], a standard framework for analyzing Schwarz methods for linear elliptic problems. The proposed theory covers a fairly large range of convex optimization problems including constrained ones, nonsmooth ones, and nonsharp ones. Moreover, the proposed theory is compatible with many existing works in the sense that it can adopt stable decomposition estimates from existing works with little modification.

There are several interesting topics for future research. Due to the nonsymmetry of multiplicative Schwarz methods, they have no minimization structure like Lemma 4.5. Since the proposed theory relies on the minimization structure of additive Schwarz methods, it is not applicable to multiplicative Schwarz methods. Indeed, in the field of mathematical optimization, analyzing multiplicative or alternating methods are considered to be much harder work than analyzing additive or parallel ones. Recently, the minimization structure of the symmetric block Gauss–Seidel method for quadratic programming was revealed in [24]. We expect that a convergence theory for symmetric multiplicative Schwarz methods for general convex optimization can be designed by adopting the idea of [24].

In the perspective of gradient methods, it is worth considering acceleration of additive Schwarz methods. After a pioneering work of Nesterov [26], acceleration of gradient methods becomes a central topic in convex optimization. In particular, an accelerated gradient method for the problem (3.1) was presented in [8]. Recently, an accelerated block Jacobi method for a constrained quadratic optimization problem was proposed in [22]. Obtaining accelerated additive Schwarz methods for (3.1) by generalizing [22] should be considered as a future work.

Appendix A. Technical proofs. In this appendix, we provide missing proofs of lemmas and theorems that appeared in sections 2 and 3.

A.1. Proof of Lemma 2.1. We note that the following proof only requires the vector space structure of spaces W and W_k, $1 \leq k \leq N$; even the convexity of functionals F_k is not assumed.

Proof of Lemma 2.1. In this proof, an index k runs from 1 to N. Take any $w \in W$. For $w_k \in W_k$ satisfying $w = \sum_{k=1}^N A_k w_k$, by the definitions of infimal postcomposition and infimal convolution, we have

$$\sum_{k=1}^N F_k(w_k) \geq \sum_{k=1}^N (A_k \triangleright F_k)(A_k w_k) \geq \left(\mathop{\square}_{k=1}^N (A_k \triangleright F_k) \right)(w).$$

Taking the infimum in the left-hand side of the above equation yields

$$\left(\mathop{\square}_{k=1}^{N}(A_k \triangleright F_k)\right)(w) \leq \inf\left\{\sum_{k=1}^{N} F_k(w_k) : w = \sum_{k=1}^{N} A_k w_k,\ w_k \in W_k\right\}.$$

Now, we show the reverse direction. For convenience, we write

$$\Delta = \left(\mathop{\square}_{k=1}^{N}(A_k \triangleright F_k)\right)(w).$$

If $\Delta = \infty$, there is nothing to show. For the case $\Delta \in \mathbb{R}$, choose any $\epsilon > 0$. Then there exists $w^k \in V$ with $w = \sum_{k=1}^{N} w^k$ such that

$$\Delta \leq \sum_{k=1}^{N}(A_k \triangleright F_k)(w^k) \leq \Delta + \frac{\epsilon}{2}. \tag{A.1}$$

Thus $(A_k \triangleright F_k)(w^k)$ is finite for every k, and there exists $\bar{w}_k \in V_k$ with $w^k = A_k \bar{w}_k$ such that

$$F_k(\bar{w}_k) \leq (A_k \triangleright F_k)(w^k) + \frac{\epsilon}{2N}. \tag{A.2}$$

Summation of (A.1) and (A.2) over all k yields

$$\sum_{k=1}^{N} F_k(\bar{w}_k) \leq \sum_{k=1}^{N}(A_k \triangleright F_k)(w^k) + \frac{\epsilon}{2} \leq \Delta + \epsilon.$$

Since $w = \sum_{k=1}^{N} w^k = \sum_{k=1}^{N} A_k \bar{w}_k$ and ϵ was chosen arbitrary, we get

$$\inf\left\{\sum_{k=1}^{N} F_k(w_k) : w = \sum_{k=1}^{N} A_k w_k,\ w_k \in W_k\right\} \leq \Delta.$$

Finally, we consider the case $\Delta = -\infty$. Take any $M > 0$. One can choose $w^k \in V$ with $w = \sum_{k=1}^{N} w^k$ such that

$$\sum_{k=1}^{N}(A_k \triangleright F_k)(w^k) \leq -2M. \tag{A.3}$$

We define the following two index sets $\mathcal{I}_1$ and $\mathcal{I}_2$ as follows:

$$\mathcal{I}_1 = \left\{k : (A_k \triangleright F_k)(w^k) \in \mathbb{R}\right\},$$
$$\mathcal{I}_2 = \left\{k : (A_k \triangleright F_k)(w^k) = -\infty\right\}.$$

Clearly, $\mathcal{I}_1 \cup \mathcal{I}_2 = \{1, \ldots, N\}$. If $\mathcal{I}_2 = \emptyset$, there exist $\bar{w}_k \in V_k$ for all k satisfying $w^k = A_k \bar{w}_k$ such that

$$F_k(\bar{w}_k) \leq (A_k \triangleright F_k)(w^k) + \frac{M}{N}. \tag{A.4}$$

Combining with (A.3) followed by summing (A.4) over all k yields

$$\sum_{k=1}^{N} F_k(\bar{w}_k) \leq \sum_{k=1}^{N} (A_k \triangleright F_k)(w^k) + M \leq -M. \tag{A.5}$$

If $\mathcal{I}_2 \neq \emptyset$, one may choose $\bar{w}_k$ with $w^k = A_k \bar{w}_k$ such that

$$\begin{aligned} F_j(\bar{w}_j) &\leq (A_j \triangleright F_j)(w^j) + \frac{M}{N}, && j \in \mathcal{I}_1, \\ F_j(\bar{w}_j) &\leq -\frac{1}{|\mathcal{I}_2|}\left(\sum_{i \in \mathcal{I}_1} (A_i \triangleright F_i)(w^i) + 2M \right), && j \in \mathcal{I}_2. \end{aligned}$$

Summing $F_k(\bar{w}_k)$ over all k yields

$$\begin{aligned} \sum_{k=1}^{N} F_k(\bar{w}_k) &= \sum_{j \in \mathcal{I}_1} F_j(\bar{w}_j) + \sum_{j \in \mathcal{I}_2} F_j(\bar{w}_j) \\ &\leq \left(\sum_{j \in \mathcal{I}_1} (A_j \triangleright F_j)(w^j) + \frac{|\mathcal{I}_1| M}{N} \right) - \left(\sum_{j \in \mathcal{I}_1} (A_j \triangleright F_j)(w^j) + 2M \right) \\ &= \left(-2 + \frac{|\mathcal{I}_1|}{N} \right) M \\ &< -M. \end{aligned} \tag{A.6}$$

In both cases (A.5) and (A.6), we conclude that

$$\inf \left\{ \sum_{k=1}^{N} F_k(w_k) : w = \sum_{k=1}^{N} A_k w_k, \ w_k \in W_k \right\} = -\infty,$$

as M can be arbitrarily large. ◻

A.2. Proof of Lemma 3.2. Take any $n \geq 0$. It is obvious that there exists a bounded and convex subset K of V such that $u^{(n)}, u^{(n+1)} \in K$. By Assumption 3.1 and the minimization property of $u^{(n+1)}$, we have

$$\begin{aligned} E(u^{(n+1)}) &= F(u^{(n)}) + \langle F'(u^{(n)}), u^{(n+1)} - u^{(n)} \rangle + D_F(u^{(n+1)}, u^{(n)}) + G(u^{(n+1)}) \\ &\leq Q(u^{(n+1)}, u^{(n)}) \\ &\leq Q(u^{(n)}, u^{(n)}) \\ &= F(u^{(n)}) + B(u^{(n)}, u^{(n)}) \\ &\leq F(u^{(n)}) + G(u^{(n)}). \end{aligned}$$

Therefore, we conclude that $E(u^{(n+1)}) \leq E(u^{(n)})$.

A.3. Proof of Theorem 3.3. In order to estimate the convergence rate of Algorithm 3.1, we need the following useful lemmas.

LEMMA A.1. *Let $\{a_n\}$ be a sequence of positive real numbers which satisfies*

$$a_n - a_{n+1} \geq C a_n^{\gamma}, \quad n \geq 0,$$

for some $C > 0$ *and* $\gamma > 1$. *Then with* $\beta = \frac{1}{\gamma - 1}$, *we have*

$$a_n \leq \frac{1}{(n+1)^\beta} \max \left\{ a_0, \left(\frac{2^\beta - 1}{C} \right)^\beta \right\}, \quad n \geq 0.$$

Proof. See [20, Lemma 1]. ∎

LEMMA A.2. *Let* $a, b > 0$, $q > 1$, *and* $\theta \in (0, 1]$. *The minimum of the function* $g(t) = \frac{a}{q} t^q - bt$, $t \in [0, \theta]$, *is given as follows:*

$$\min_{t \in [0,\theta]} g(t) = \begin{cases} \frac{a}{q}\theta^q - b\theta \leq -b\theta \left(1 - \frac{1}{q}\right) & \text{if } a\theta^{q-1} - b \leq 0, \\ -b \left(\frac{b}{a}\right)^{\frac{1}{q-1}} \left(1 - \frac{1}{q}\right) & \text{if } a\theta^{q-1} - b > 0. \end{cases}$$

Proof. It is elementary. ∎

We notice that the proof of Theorem 3.3 is motivated by [25, Theorem 4], where a special case, forward-backward splitting with $q = 2$ and $\theta = 1$, was analyzed.

Proof of Theorem 3.3. Take any $n \geq 0$. For $u \in K_0$, we write

$$u_\theta = \frac{1}{\theta} u - \left(\frac{1}{\theta} - 1 \right) u^{(n)}, \tag{A.7}$$

so that $u - u^{(n)} = \theta(u_\theta - u^{(n)})$. Note that if we set $u = tu^* + (1-t)u^{(n)}$ for $t \in [0, \theta]$, then $u \in K_0$ and

$$u_\theta = \frac{t}{\theta} u^* + \left(1 - \frac{t}{\theta} \right) u^{(n)} \in K_0. \tag{A.8}$$

We denote $E(u^{(n)}) - E(u^*)$ by ζ_n. It follows that

(A.9)

$$\begin{aligned} E(u^{(n+1)}) &= F(u^{(n)}) + \langle F'(u^{(n)}), u^{(n+1)} - u^{(n)} \rangle + D_F(u^{(n+1)}, u^{(n)}) + G(u^{(n+1)}) \\ &\overset{\text{(a)}}{\leq} Q_n(u^{(n+1)}) \\ &= \min_{u \in K_0} \left\{ F(u^{(n)}) + \langle F'(u^{(n)}), u - u^{(n)} \rangle + B(u, u^{(n)}) \right\} \\ &\overset{\text{(b),(A.7)}}{\leq} \min_{u \in K_0} \Bigg\{ F(u^{(n)}) + \theta \langle F'(u^{(n)}), u_\theta - u^{(n)} \rangle \\ &\qquad + \frac{L\theta^q}{q} \| u_\theta - u^{(n)} \|^q + \theta G(u_\theta) + (1-\theta) G(u^{(n)}) \Bigg\} \\ &\overset{\text{(c)}}{\leq} \min_{u \in K_0} \left\{ (1-\theta) E(u^{(n)}) + \theta E(u_\theta) + \frac{L\theta^q}{q} \| u_\theta - u^{(n)} \|^q \right\} \\ &\overset{\text{(A.8)}}{\leq} \min_{t \in [0,\theta]} \Bigg\{ (1-\theta) E(u^{(n)}) + \theta E\left(\frac{t}{\theta} u^* + \left(1 - \frac{t}{\theta} \right) u^{(n)} \right) \\ &\qquad + \frac{Lt^q}{q} \| u^* - u^{(n)} \|^q \Bigg\} \\ &\overset{\text{(d)}}{\leq} \min_{t \in [0,\theta]} \left\{ E(u^{(n)}) - t\zeta_n + \frac{Lt^q}{q} R_0^q \right\}, \end{aligned}$$

where (a), (b) are because of Assumption 3.1, (c) is due to the convexity of F, and (d) is due to the convexity of E. Invoking Lemma A.2, we have

$$E(u^{(n+1)}) \leq E(u^{(n)}) - \theta\left(1 - \frac{1}{q}\right)\zeta_n$$

if $\zeta_n \geq \theta^{q-1} L R_0^q$, which is equivalent to

$$\zeta_{n+1} \leq \left(1 - \theta\left(1 - \frac{1}{q}\right)\right)\zeta_n.$$

Otherwise, if $\zeta_n < \theta^{q-1} L R_0^q$, we get

$$E(u^{(n+1)}) \leq E(u^{(n)}) - \frac{1}{(LR_0^q)^{\frac{1}{q-1}}}\left(1 - \frac{1}{q}\right)\zeta_n^{\frac{q}{q-1}}.$$

Hence, we have

$$\zeta_n - \zeta_{n+1} \geq \frac{1}{(LR_0^q)^{\frac{1}{q-1}}}\left(1 - \frac{1}{q}\right)\zeta_n^{\frac{q}{q-1}}.$$

Invoking Lemma A.1 yields

$$\zeta_n \leq \frac{1}{(n+1)^{q-1}} \max\left\{\zeta_0, \left(\frac{q(2^{q-1}-1)}{q-1}\right)^{q-1} L R_0^q\right\}, \quad n \geq 0.$$

Since $\zeta_0 < \theta^{q-1} L R_0^q$, setting

$$C_{q,\theta} = \left(\max\left\{\theta, \left(\frac{q(2^{q-1}-1)}{q-1}\right)\right\}\right)^{q-1} \tag{A.10}$$

completes the proof. □

A.4. Proof of Theorem 3.6. The proof of Theorem 3.6 is done with a similar argument to [25, Theorem 5], where the convergence analysis for $p = q = 2$ was given.

Proof of Theorem 3.6. Again, we denote $E(u^{(n)}) - E(u^*)$ by ζ_n. By the same way as (A.9), one can obtain

$$E(u^{(n+1)}) \leq \min_{t \in [0,\theta]} \left\{E(u^{(n)}) - t\zeta_n + \frac{Lt^q}{q}\|u^* - u^{(n)}\|^q\right\}. \tag{A.11}$$

By Assumption 3.4, we have

$$\|u^* - u^{(n)}\|^q \leq \left(\frac{p}{\mu}\zeta_n\right)^{\frac{q}{p}}. \tag{A.12}$$

Combining (A.11) and (A.12) yields

$$E(u^{(n+1)}) \leq \min_{t \in [0,\theta]} \left\{E(u^{(n)}) - t\zeta_n + \frac{p^{\frac{q}{p}} L t^q}{q\mu^{\frac{q}{p}}}\zeta_n^{\frac{q}{p}}\right\}. \tag{A.13}$$

We consider the two cases $p = q$ and $p > q$ separately. First, we assume that $p = q$. Then (A.13) is simplified to

$$\zeta_{n+1} \leq \min_{t \in [0,\theta]} \left(1 - t + \frac{Lt^q}{\mu}\right)\zeta_n.$$

By Lemma A.2, if $\mu \geq \theta^{q-1}qL$, then we get

$$\zeta_{n+1} \leq \left(1 - \theta\left(1 - \frac{1}{q}\right)\right)\zeta_n. \tag{A.14}$$

Otherwise, if $\mu < \theta^{q-1}qL$, we have

$$\zeta_{n+1} \leq \left(1 - \left(1 - \frac{1}{q}\right)\left(\frac{\mu}{qL}\right)^{\frac{1}{q-1}}\right)\zeta_n. \tag{A.15}$$

Recursive applications of (A.14) and (A.15) yield the desired results for the case $p = q$.

Next, we consider the case $p > q$. Let $r = \frac{q}{p} < 1$. By Lemma A.2, if

$$\zeta_n^{1-r} \geq \frac{\theta^{q-1}p^r L}{\mu^r},$$

one can obtain from (A.13) that

$$\zeta_{n+1} \leq \left(1 - \theta\left(1 - \frac{1}{q}\right)\right)\zeta_n.$$

Otherwise, it follows that

$$\zeta_{n+1} \leq \zeta_n - \left(1 - \frac{1}{q}\right)\left(\frac{\mu^r}{p^r L}\right)^{\frac{1}{q-1}} \zeta_n^{\frac{q(p-1)}{p(q-1)}}.$$

Application of Lemma A.1 yields

$$\zeta_n \leq \frac{1}{(n+1)^\beta}\max\left\{\zeta_0, \left(\frac{q(2^\beta - 1)}{q-1}\right)^\beta \left(\frac{p^r L}{\mu^r}\right)^{\frac{1}{1-r}}\right\},$$

where $\beta = \frac{p(q-1)}{p-q}$. Since $\zeta_0 \leq \theta^{\frac{q-1}{1-r}}(\frac{p^r L}{\mu^r})^{\frac{1}{1-r}}$, we conclude that

$$\zeta_n \leq \frac{C_{p,q,\theta}(L/\mu^r)^{\frac{1}{1-r}}}{(n+1)^\beta}$$

with

$$C_{p,q,\theta} = p^{\frac{q}{p-q}}\left(\max\left\{\theta, \left(\frac{q(2^{\frac{p(q-1)}{p-q}} - 1)}{q-1}\right)\right\}\right)^{\frac{p(q-1)}{p-q}}. \tag{A.16}$$

This completes the proof. ∎

REFERENCES

[1] 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.

[2] L. Badea, *One- and Two-Level Additive Methods for Variational and Quasi-Variational Inequalities of the Second Kind*, preprint, Ser. Inst. Math. Rom. Acad., 2010; also available online from http://imar.ro/~dbeltita/IMAR_preprints/2010/2010_5.pdf.

[3] 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.

[4] 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.
[5] L. Badea and J. Wang, *An additive Schwarz method for variational inequalities*, Math. Comp., 69 (2000), pp. 1341–1354.
[6] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.
[7] 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.
[8] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
[9] A. Beck and L. Tetruashvili, *On the convergence of block coordinate descent type methods*, SIAM J. Optim., 23 (2013), pp. 2037–2060.
[10] S. Becker and P. L. Combettes, *An algorithm for splitting parallel sums of linearly composed monotone operators, with applications to signal recovery*, J. Nonlinear Convex Anal., 15 (2014), pp. 137–159.
[11] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2007), pp. 1205–1223.
[12] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319.
[13] 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.
[14] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, SIAM, Philadelphia, 2002.
[15] O. Fercoq and P. Richtárik, *Accelerated, parallel, and proximal coordinate descent*, SIAM J. Optim., 25 (2015), pp. 1997–2023.
[16] A. Frommer and D. B. Szyld, *An algebraic convergence theory for restricted additive Schwarz methods using weighted max norms*, SIAM J. Numer. Anal., 39 (2001), pp. 463–479.
[17] M. Griebel and P. Oswald, *On the abstract theory of additive and multiplicative Schwarz algorithms*, Numer. Math., 70 (1995), pp. 163–180.
[18] 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.
[19] M. Hintermüller and K. Kunisch, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM J. Appl. Math., 64 (2004), pp. 1311–1333.
[20] Y. Huang, R. Li, and W. Liu, *Preconditioned descent algorithms for p-Laplacian*, J. Sci. Comput., 32 (2007), pp. 343–371.
[21] 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.
[22] 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.
[23] 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.
[24] X. Li, D. Sun, and K.-C. Toh, *A block symmetric Gauss–Seidel decomposition theorem for convex composite quadratic programming and its applications*, Math. Program., 175 (2019), pp. 395–418.
[25] Y. Nesterov, *Gradient methods for minimizing composite functions*, Math. Program., 140 (2013), pp. 125–161.
[26] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, Soviet Math. Dokl., 269 (1983), pp. 543–547.
[27] D.-S. Oh, *An overlapping Schwarz algorithm for Raviart–Thomas vector fields with discontinuous coefficients*, SIAM J. Numer. Anal., 51 (2013), pp. 297–321.
[28] J. Park, *Pseudo-Linear Convergence of an Additive Schwarz Method for Dual Total Variation Minimization*, arXiv:1911.06639 [math.NA], 2019.
[29] X.-C. Tai, *Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities*, Numer. Math., 93 (2003), pp. 755–786.
[30] X.-C. Tai and M. Espedal, *Rate of convergence of some space decomposition methods for linear and nonlinear problems*, SIAM J. Numer. Anal., 35 (1998), pp. 1558–1570.
[31] 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.

[32] A. Toselli and O. Widlund, *Domain Decomposition Methods-Algorithms and Theory*, Springer, New York, 2005.

[33] G. Tran, H. Schaeffer, W. M. Feldman, and S. J. Osher, *An L^1 penalty method for general obstacle problems*, SIAM J. Appl. Math., 75 (2015), pp. 1424–1444.

[34] P. Tseng, *Convergence of a block coordinate descent method for nondifferentiable minimization*, J. Optim. Theory Appl., 109 (2001), pp. 475–494.

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

[36] J. Xu and L. Zikatanov, *The method of alternating projections and the method of subspace corrections in Hilbert space*, J. Amer. Math. Soc., 15 (2002), pp. 573–597.

[37] Y. Xu and W. Yin, *A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion*, SIAM J. Imaging Sci., 6 (2013), pp. 1758–1789.