# Additive Schwarz Methods for Convex Optimization — Convergence Theory and Acceleration

Jongho Park

## 1 Introduction

This paper is concerned with additive Schwarz methods for convex optimization problems of the form

$$\min_{u \in V} \left\{ E(u) := F(u) + G(u) \right\}, \tag{1}$$

where $V$ is a reflexive Banach space, $F\colon V \to \mathbb{R}$ is a Frechét differentiable convex function, and $G\colon V \to \overline{\mathbb{R}}$ is a proper, convex, lower semicontinuous function which is possibly nonsmooth. We further assume that $E$ is coercive, so that (1) admits a solution $u^* \in V$. There are plenty of scientific problems of the form (1), e.g., nonlinear elliptic problems [13], variational inequalities [1, 12], and mathematical imaging problems [5, 10], and has been much research on Schwarz methods corresponding to them.

In this paper, we present a unified view to some notable recent results [8, 9] on additive Schwarz methods for convex optimization (1). The starting point is the generalized additive Schwarz lemma presented in [9]. Based on the relevancy between additive Schwarz methods and gradient methods for (1) investigated in the generalized additive Schwarz lemma, two main results are considered: the abstract convergence theory [9] that generalizes some important existing results [1, 13, 15] and the momentum acceleration scheme [8] that greatly improves the convergence rate for additive Schwarz methods. In addition, we propose a novel backtracking strategy for additive Schwarz methods that further improves the convergence rate. We present numerical results for additive Schwarz methods equipped with the proposed backtracking strategy in order to highlight numerical efficiency.

Jongho Park
Natural Science Research Institute, KAIST, Daejeon 34141, Korea
e-mail: jongho.park@kaist.ac.kr

## 2 Additive Schwarz methods

In this section, we present an abstract additive Schwarz method for (1). In what follows, 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 $V = \sum_{k=1}^N R_k^* V_k$ and its adjoint $R_k\colon V^* \to V_k^*$ is surjective. In order to describe local problems, we define $d_k\colon V_k \times V \to \overline{\mathbb{R}}$ and $G_k\colon V_k \times V \to \overline{\mathbb{R}}$ as functions which are proper, convex, and lower semicontinuous with respect to their first arguments. For positive constants $\tau$ and $\omega$, an *additive Schwarz operator* $\mathrm{ASM}_{\tau,\omega}\colon V \to V$ is defined by

$$\mathrm{ASM}_{\tau,\omega}(v) = v + \tau \sum_{k=1}^N R_k^* \tilde{w}_k,$$

where

$$\tilde{w}_k \in \operatorname*{arg\,min}_{w_k \in V_k} \left\{ F(v) + \langle F'(v), R_k^* w_k \rangle + \omega d_k(w_k, v) + G_k(w_k, v) \right\}. \tag{2}$$

We note that (2) may admits nonunique minimizers; we take $\tilde{w}_k$ as any one among them in this case. If we set

$$d_k(w_k, v) = D_F(v + R_k^* w_k, v), \quad G_k(w_k, v) = G(v + R_k^* w_k), \quad \omega = 1 \tag{3a}$$

in (2), then the minimization problem is reduced to

$$\min_{w_k \in V_k} E(v + R_k^* w_k), \tag{3b}$$

which is the case of exact local problems. Here $D_F$ denotes the Bregman distance

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

We note that other choices of $d_k$ and $G_k$, i.e., cases of inexact local problems, include various numerical methods such as block coordinate descent methods and constraint decomposition methods [5, 12]; see [9, Sect. 6.4] for details.

The abstract additive Schwarz method for (1) is presented in Algorithm 1. Constants $\tau_0$ and $\omega_0$ in Algorithm 1 will be given in Section 3. Note that $\operatorname{dom} G$ denotes the effective domain of $G$, i.e., $\operatorname{dom} G = \{ v \in V : G(v) < \infty \}$.

**Algorithm 1** Additive Schwarz method for (1)

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

$$u^{(n+1)} = \mathrm{ASM}_{\tau,\omega}(u^{(n)})$$

**end**

An important observation made in [9, Lemma 4.5] is that Algorithm 1 can be interpreted as a kind of a gradient method equipped with a nonlinear distance function [14]. A rigorous statement is presented in the following.

**Proposition 1 (generalized additive Schwarz lemma)**

*For $\tau, \omega > 0$, we have*

$$\mathrm{ASM}_{\tau,\omega}(v) = \operatorname*{arg\,min}_{u \in V} \left\{ F(v) + \langle F'(v), u - v \rangle + M_{\tau,\omega}(u,v) \right\}, \quad v \in V,$$

*where the functional $M_{\tau,\omega} \colon V \times V \to \overline{\mathbb{R}}$ is given by*

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

In the field of mathematical optimization, there has been numerous research on gradient methods for solving convex optimization problems [4, 6, 14]. Therefore, invoking Proposition 1, we can adopt many valuable tools from the field of mathematical optimization in order to analyze and improve Schwarz methods. In particular, we present two fruitful results in the remainder of the paper: novel convergence theory [9] and acceleration [8] for additive Schwarz methods.

## 3 Convergence theory

This section is devoted to an abstract convergence theory of additive Schwarz methods for convex optimization. The convergence theory introduced in this section directly generalizes the classical theory for linear problems [15, Chapter 2] to convex optimization problems. Similar to [15, Chapter 2], the following three conditions are considered: stable decomposition, strengthened convexity, and local stability.

*Assumption 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 $w_k \in V_k$, $1 \leq k \leq N$, with $u - v = \sum_{k=1}^{N} R_k^* w_k$, such that

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

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

*Assumption 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

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

*Assumption 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

$$D_F(v+R_k^*w_k, v) \leq \omega_0 d_k(w_k, v), \quad G(v+R_k^*w_k) \leq G_k(w_k, v).$$

Assumption 1 is compatible with various variants of stable decomposition presented in existing works [1, 13, 15]. Assumption 2 trivially holds with $\tau_0 = 1/N$ due to the convexity of $E$. However, a better value for $\tau_0$ independent of $N$ can be found by the usual coloring technique. In the same spirit as [15], Assumption 3 gives a one-sided measure of approximation properties of the local solvers. It was shown in [9, Sect. 4.1] that the above assumptions reduce to [15, Assumptions 2.2 to 2.4] if they are applied to linear elliptic problems. Under the above three assumptions, we have the following convergence theorem for Algorithm 1 [9, Theorem 4.7].

**Theorem 1** *Suppose that Assumptions 1, 2, and 3 hold. In Algorithm 1, we have*

$$E(u^{(n)})-E(u^*)=O\left(\frac{\kappa_{\mathrm{ASM}}}{n^{q-1}}\right),$$

*where $\kappa_{\mathrm{ASM}}$ is the additive Schwarz condition number defined by $\kappa_{\mathrm{ASM}} = \omega C_0^q/\tau^{q-1}$.*

Meanwhile, it is well-known that the Łojasiewicz inequality holds in many applications [11]; it says that the energy functional $E$ of (1) is sharp around the minimizer $u^*$. We summarize this property in Assumption 4.

*Assumption 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 $\mu_K > 0$.

We can obtain an improved convergence result for Algorithm 1 compared to Theorem 1 under an additional sharpness assumption on $E$ [9, Theorem 4.8].

**Theorem 2** *Suppose that Assumptions 1, 2, 3, and 4 hold. In Algorithm 1, we have*

$$E(u^{(n)})-E(u^*)=\begin{cases} O\left(\left(1-\left(1-\frac{1}{q}\right)\min\left\{\tau,\left(\frac{\mu}{q\kappa_{\mathrm{ASM}}}\right)^{\frac{1}{q-1}}\right\}\right)^n\right), & \text{if } p=q,\\ O\left(\frac{\left(\kappa_{\mathrm{ASM}}^p/\mu^q\right)^{\frac{1}{p-q}}}{n^{\frac{p(q-1)}{p-q}}}\right), & \text{if } p>q,\end{cases}$$

*where $\kappa_{\mathrm{ASM}}$ was defined in Theorem 1.*

Theorems 1 and 2 are direct consequences of Proposition 1 in the sense that they can be easily deduced by invoking theories of gradient methods for convex optimization [9, Sect. 2].

## 4 Acceleration

An important observation on Schwarz methods for linear problems is that they can be interpreted as preconditioned Richardson iterations with appropriate preconditioners. Replacing Richardson iterations by conjugate gradient iterations with the same preconditioners, we can obtain improved algorithms that converge faster. Since Proposition 1 says that additive Schwarz methods for (1) are in fact gradient methods, in the same spirit, we may adopt some acceleration schemes for gradient methods (see, e.g., [4, 7]) in order to improve additive Schwarz methods. Motivated by the FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) momentum [2] and the gradient adaptive restarting scheme [7], the following accelerated variant of Algorithm 1 was considered in [8].

**Algorithm 2** Accelerated additive Schwarz method for (1)

Let $u^{(0)} = v^{(0)} \in \operatorname{dom} G$, $\tau > 0$, and $t_0 = 1$.
**for** $n = 0, 1, 2, \ldots$

$$u^{(n+1)} = \mathrm{ASM}_{\tau,\omega}(v^{(n)})$$

$$\begin{cases} t_{n+1} = 1, \ \beta_n = 0, & \text{if } \langle v^{(n)} - u^{(n+1)}, u^{(n+1)} - u^{(n)} \rangle > 0, \\ t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}, \ \beta_n = \frac{t_n - 1}{t_{n+1}}, & \text{otherwise.} \end{cases}$$

$$v^{(n+1)} = u^{(n+1)} + \beta_n (u^{(n+1)} - u^{(n)})$$

**end**

The major part of each iteration of Algorithm 2 is to compute the additive Schwarz operator $\mathrm{ASM}_{\tau,\omega}$; the computational cost for momentum parameters $t_n$ and $\beta_n$ is marginal. Therefore, the main computational cost of Algorithm 2 is the same as the one of Algorithm 1. Nevertheless, it was shown numerically in [8] that Algorithm 2 achieves much faster convergence to the energy minimum compared to Algorithm 1.

In the remainder of this section, we consider how to further improve Algorithm 2. More precisely, we present a backtracking strategy for additive Schwarz methods that allows for local optimization of the parameter $\tau$. Mimicking [3, 6], at each iteration of additive Schwarz methods, we choose $\tau$ as large as possible satisfying

$$E(u^{(n+1)}) \leq F(u^{(n)}) + \langle F'(u^{(n)}), u^{(n+1)} - u^{(n)} \rangle + M_{\tau,\omega}(u^{(n+1)}, u^{(n)}).$$

An optimal $\tau$ can be found by a logarithmic grid search. Algorithm 2 accompanied with the backtracking strategy is presented in Algorithm 3. Note that the parameter $\rho \in (0,1)$ in Algorithm 3 plays a role of an adjustment parameter for the grid search.

**Algorithm 3** Accelerated additive Schwarz method for (1) with backtracking

Let $u^{(0)} = v^{(0)} \in \operatorname{dom} G$, $\tau > 0$, $t_0 = 1$, and $\rho \in (0,1)$.
**for** $n = 0, 1, 2, \dots$

$$\tau \leftarrow \tau/\rho$$

**repeat**

$$u^{(n+1)} = \mathrm{ASM}_{\tau,\omega}(v^{(n)})$$

**if** $E(u^{(n+1)}) > F(u^{(n)}) + \langle F'(u^{(n)}), u^{(n+1)} - u^{(n)} \rangle + M_{\tau,\omega}(u^{(n+1)}, u^{(n)})$

$$\tau \leftarrow \rho\tau$$

**end if**
**until** $E(u^{(n+1)}) \le F(u^{(n)}) + \langle F'(u^{(n)}), u^{(n+1)} - u^{(n)} \rangle + M_{\tau,\omega}(u^{(n+1)}, u^{(n)})$

$$\begin{cases} t_{n+1} = 1,\ \beta_n = 0, & \text{if } \langle v^{(n)} - u^{(n+1)}, u^{(n+1)} - u^{(n)} \rangle > 0, \\ t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2},\ \beta_n = \frac{t_n - 1}{t_{n+1}}, & \text{otherwise.} \end{cases}$$

$$v^{(n+1)} = u^{(n+1)} + \beta_n (u^{(n+1)} - u^{(n)})$$

**end**

Different from the existing works [3, 6], adopting the backtracking strategy for additive Schwarz methods has an own difficulty that evaluation of $M_{\tau,\omega}(u^{(n+1)}, u^{(n)})$ is not straightforward due to its complicated definition. The following proposition provides a way to evaluate $M_{\tau,\omega}(u^{(n+1)}, u^{(n)})$ without major computational cost.

**Proposition 2** *If $u = \mathrm{ASM}_{\tau,\omega}(v)$, then it satisfies that*

$$M_{\tau,\omega}(u, v) = \tau \sum_{k=1}^{N} (\omega d_k + G_k)(\tilde{w}_k, v) + (1 - \tau N) G(v),$$

*where $\tilde{w}_k$, $1 \le k \le N$, were defined in* (2). *In particular, if the exact local problems* (3) *are used, then we have*

$$F(v) + \langle F'(v), u - v \rangle + M_{\tau,\omega}(u, v) = (1 - \tau N) E(v) + \tau \sum_{k=1}^{N} E(v + R_k^* \tilde{w}_k).$$

***Proof*** See the proof of [9, Lemma 4.5]. □

Thanks to Proposition 2, one can compute $M_{\tau,\omega}(u^{(n+1)}, u^{(n)})$ in Algorithm 3 without solving the infimum in the definition of $M_{\tau,\omega}$. As discussed in [3], the backtracking strategy improves the convergence rate because it allows for adaptive adjustment of $\tau$ depending on the local flatness of the energy functional.

In order to show the computational efficiency of Algorithm 3, we present numerical results applied to a finite element $s$-Laplacian problem ($s \geq 1$). We set $\Omega = [0,1]^2 \subset \mathbb{R}^2$. We decompose the domain $\Omega$ into $\mathcal{N} = N \times N$ square subdomains $\{\Omega_k\}_{k=1}^{\mathcal{N}}$ in which each subdomain has the sidelength $H = 1/N$. Each subdomain $\Omega_k$, $1 \leq k \leq \mathcal{N}$, is partitioned into $2 \times H/h \times H/h$ uniform triangles to form a global triangulation $\mathcal{T}_h$ of $\Omega$. Similarly, we partition each $\Omega_k$ into two uniform triangles and let $\mathcal{T}_H$ be a coarse triangulation of $\Omega$ consisting of such triangles. Overlapping subdomains $\{\Omega_k'\}_{k=1}^{\mathcal{N}}$ are constructed in a way that $\Omega_k'$ is a union of $\Omega_k$ and its surrounding layers of fine elements in $\mathcal{T}_h$ with the width $\delta$ such that $0 < \delta < H/2$. The model finite element $s$-Laplacian problem is written as

$$\min_{u \in S_h(\Omega)} \left\{ \frac{1}{s} \int_\Omega |\nabla u|^s \, dx - \int_\Omega f u \, dx \right\}, \tag{4}$$

where $f \in (L^s(\Omega))^*$ and $V = S_h(\Omega)$ is the continuous piecewise linear finite element space on $\mathcal{T}_h$ with the homogeneous Dirichlet boundary condition. We set $V_k = S_h(\Omega_k')$, $1 \leq k \leq \mathcal{N}$, and take $R_k^*\colon V_k \to V$ as the natural extension operator, where $S_h(\Omega_k')$ is the continuous piecewise linear finite element space on the $\mathcal{T}_h$-elements in $\Omega_k'$ with the homogeneous Dirichlet boundary condition. As a coarse space, we set $V_0$ by the continuous piecewise linear space $S_H(\Omega)$ on $\mathcal{T}_H$ and take $R_0^*\colon V_0 \to V$ as the natural interpolation operator.

For numerical experiments, we set $s = 4$, $f = 1$, and $u^{(0)} = 0$. Exact local and coarse solvers (3) were used; they were solved numerically by FISTA with gradient adaptive restarts [7]. The initial step size $\tau$ was chosen as $1/5$ (cf. [9, Sect. 5.1]).

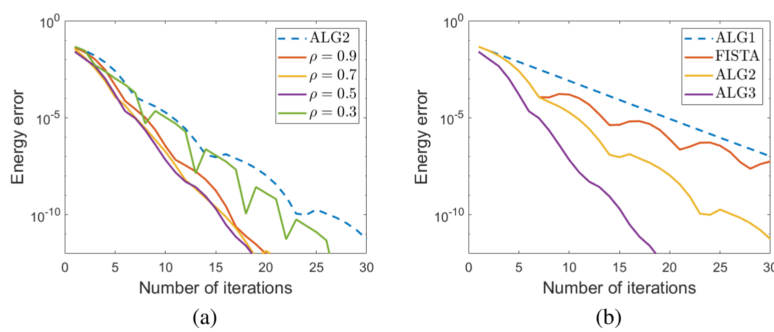


**Fig. 1:** Decay of the energy error $E(u^{(n)}) - E(u^*)$ in additive Schwarz methods ($\tau = 1/5$, $\omega = 1$) for the $s$-Laplacian problem (4) ($h = 1/2^6$, $H = 1/2^3$, $\delta = 4h$). (a) Algorithm 3 with various values of $\rho$. (b) Comparison of various additive Schwarz methods. FISTA denotes the FISTA momentum without restarts and ALG3 denotes Algorithm 3 with $\rho = 0.5$.

Figure 1 plots the energy error $E(u^{(n)}) - E(u^*)$ of various additive Schwarz methods when $h = 1/2^6$, $H = 1/2^3$, and $\delta = 4h$. As shown in Figure 1(a), Algorithm 3 shows faster convergence to the energy minimum compared to Algorithm 2 for various values of $\rho$. Hence, we can say that the backtracking strategy proposed in this paper is effective for acceleration of convergence. Although Algorithm 3 shows better performance than Algorithm 2 for all values of $\rho$, it remains as a future work to discover how to find an optimal $\rho$. Figure 1(b) presents a numerical comparison of Algorithm 1, Algorithm 1 equipped with the FISTA momentum, Algorithms 2 and 3. We can observe that all of the FISTA momentum, adaptive restarting technique, and backtracking strategy provide positive effects on the convergence rate of additive Schwarz methods. Consequently, Algorithm 3, which assembles all of the aforementioned acceleration schemes, show the best convergence rate among all methods. Since the main computational costs of all algorithms are essentially the same, we conclude that Algorithm 3 numerically outperforms all the others.

## References

1. Badea, L.: One- and two-level additive methods for variational and quasi-variational inequalities of the second kind. preprint, Ser. Inst. Math. Rom. Acad. (5) (2010)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
3. Calatroni, L., Chambolle, A.: Backtracking strategies for accelerated descent methods with smooth composite objectives. SIAM J. Optim. **29**(3), 1772–1798 (2019)
4. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numer. **25**, 161–319 (2016)
5. Chang, H., Tai, X.C., Wang, L.L., Yang, D.: Convergence rate of overlapping domain decomposition methods for the Rudin–Osher–Fatemi model based on a dual formulation. SIAM J. Imaging Sci. **8**(1), 564–591 (2015)
6. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1), 125–161 (2013)
7. O'Donoghue, B., Candes, E.: Adaptive restart for accelerated gradient schemes. Found. Comput. Math. **15**(3), 715–732 (2015)
8. Park, J.: Accelerated additive Schwarz methods for convex optimization with adaptive restart (2020). arXiv:2011.02695
9. Park, J.: Additive Schwarz methods for convex optimization as gradient methods. SIAM J. Numer. Anal. **58**(3), 1495–1530 (2020)
10. Park, J.: Pseudo-linear convergence of an additive Schwarz method for dual total variation minimization. Electron. Trans. Numer. Anal. **54**, 176–197 (2021)
11. Roulet, V., d'Aspremont, A.: Sharpness, restart, and acceleration. SIAM J. Optim. **30**(1), 262–289 (2020)
12. Tai, X.C.: Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. Numer. Math. **93**(4), 755–786 (2003)
13. Tai, X.C., Xu, J.: Global and uniform convergence of subspace correction methods for some convex optimization problems. Math. Comp. **71**(237), 105–124 (2002)
14. Teboulle, M.: A simplified view of first order methods for optimization. Math. Program. **170**(1), 67–96 (2018)
15. Toselli, A., Widlund, O.: Domain Decomposition Methods—Algorithms and Theory. Springer, Berlin (2005)