MATHEMATICS OF COMPUTATION
https://doi.org/10.1090/mcom/3526
Article electronically published on April 7, 2020

CONVERGENCE ANALYSIS OF THE FAST SUBSPACE DESCENT METHOD FOR CONVEX OPTIMIZATION PROBLEMS

LONG CHEN, XIAOZHE HU, AND STEVEN M. WISE

Abstract. The full approximation storage (FAS) scheme is a widely used multigrid method for nonlinear problems. In this paper, a new framework to design and analyze FAS-like schemes for convex optimization problems is developed. The new method, the fast subspace descent (FASD) scheme, which generalizes classical FAS, can be recast as an inexact version of nonlinear multigrid methods based on space decomposition and subspace correction. The local problem in each subspace can be simplified to be linear and one gradient descent iteration (with an appropriate step size) is enough to ensure a global linear (geometric) convergence of FASD for convex optimization problems.

1. Introduction

Most real-world applications are inherently nonlinear. The design of fast algorithms for the solution or approximate solution of nonlinear equations is of fundamental interest to mathematicians, physicists, biologists, and others. In this paper, we consider solving nonlinear equations arising from the minimization of a convex functional in the abstract Hilbert space setting.

The well-known Newton-Raphson method is a traditional and popular approach for solving nonlinear equations. Basically, Newton's method iteratively finds the approximate solution by linearizing the problem near the current iterate. In the present case, a linear symmetric positive definite system (the Jacobian system) needs to be solved at each Newton's iteration, and fast linear multigrid (MG) methods are sometimes used as a solver. Practically, each linear problem can be approximately inverted by applying a few multigrid iterations. But, if this is done, the quadratic rate of convergence may be sacrificed.

One alternative to Newton's method for solving nonlinear PDE is the nonlinear multigrid method, better known as the full approximation storage (FAS) scheme. This method, developed by Brandt [3] in the late 1970s (see also [4]) often converges linearly and with optimal complexity in practice. Recall that the success of multigrid methods relies on two ingredients: 1) high frequency components of the error will be damped by smoothers; and 2) low frequency components of the error can be approximated well on a coarse grid. The smoother used in FAS is usually the nonlinear Gauss-Seidel smoother, which solves many small-sized nonlinear problems (typically with one degree of freedom) on small patches of the mesh. For

Received by the editor October 2, 2018, and, in revised form, July 1, 2019, October 19, 2019, and January 10, 2020.

2010 *Mathematics Subject Classification.* Primary 65N55, 65N22, 65K10, 65J15.

The second author was supported by NSF Grant DMS-1620063.

The third author was supported by the NSF Grant DMS-1719854.

the coarse grid problem, the FAS method uses the full approximation rather than the standard defect, which makes it essentially different from linear MG methods. Due to its high efficiency, the FAS method has been applied to many nonlinear PDE problems, such as in [14–16, 19, 24, 28, 30].

Although FAS is quite successful in practice, its theoretical analysis is limited. In [13], Hackbusch considered nonlinear MG methods for general nonlinear problems. By imposing conditions on the nonlinear operators and their derivatives, together with standard smoothing and approximation properties, he was able to show that the FAS converges in a sufficiently small neighborhood of the solution on a fine enough mesh. Moreover, the number of smoothing steps needs to be sufficiently large, and at least the W-cycle should be used. Later in [22, 23], Reusken considered FAS for a class of semi-linear second order elliptic boundary value problems with mild nonlinearity. Within this nice class of nonlinear problems, he was able to show the convergence of FAS under weaker assumptions on the nonlinear operators. We want to mention that the proofs in their work are based on the linearization of the FAS iterations, and the rate of convergence is in some sense local. For example, in [23], Reusken showed that the V-cycle FAS converges locally in a ball with radius shrinking from coarse to fine levels.

In this paper we consider a special class of nonlinear equations that can be viewed as Euler equations of certain convex objective functions. The convergence of MG methods for convex optimization problems has been studied in [26, 27] under the framework of subspace correction methods [29]. In [27], Tai and Xu considered some unconstrained convex optimization problems and developed global and uniform convergence estimates for a class of subspace correction iterative methods. Their approach is based on an abstract space decomposition which is assumed to satisfy the so-called stable decomposition property and strengthened Cauchy Schwarz inequality. We point out that in each subspace, the original objective function is used, which is, strictly speaking, naturally defined on the finest level. Furthermore, the local problem should be solved exactly, which is more expensive than what is required in the FAS scheme.

We shall borrow the theoretical framework established in [27] to analyze a hybrid of the FAS and subspace correction methods, what we will call the *fast subspace descent* (FASD) method. In contrast to the subspace correction method considered in [27], in which an exact subspace solver is used, we recast FASD as a subspace correction method with an inexact subspace solver, which reduces the computational cost significantly. In particular, we show that one step of preconditioned gradient descent iteration in each subspace is good enough to guarantee the global convergence for convex optimization problems.

Several other FAS-like algorithms for solving optimization problems have been considered in the literature [11, 12, 17, 19], including those that are line search-based recursive or trust region-based recursive algorithms. Only basic convergence is established in these works. Here we shall prove a global linear convergence for a class of strongly convex optimization problems.

We establish the convergence of the algorithm in the framework of subspace corrections [27]. We first show that, with a one dimensional line search approach, the FASD method converges globally and uniformly under the standard assumptions on the space decomposition. In addition, we borrow some techniques from the optimization literature [21] in order to properly handle the inexactness of the local

solver used in FASD. We introduce a fixed step size to guarantee that the objective function is decreasing globally. For the analysis of the original FAS method, which is obtained from the new FASD method via a simple modification, we impose an additional approximation property of the subspace problems and show that FASD converges globally and uniformly. We emphasize that our work represents not only a theoretical advance for the convergence analysis of FAS-type schemes, but also is algorithmically simpler, and even more flexible, than the original FAS. We show that, both theoretically and numerically, each local nonlinear problem can be approximated by a linear problem, and, consequently, the computational cost is reduced significantly.

The paper is organized as follows. In Section 2, we present the optimization problem, with its associated Euler equation, in a general Hilbert space framework. We conclude the section with the assumptions on the space decomposition. The successive subspace optimization (SSO) method is recalled in Section 3. The convergence analysis of SSO, based on slightly weaker assumptions compared with [27], is presented in the same section. The main global and uniform convergence analyses for FASD with the exact line search and approximate (quadratic) line search are derived in Sections 4 and 5, respectively. The original FAS method is analyzed in Section 6. In Section 7, an application problem is considered.

2. Problem and assumptions

Given an energy, or objective function, $E(v)$ defined on a Hilbert space $\mathcal{V}$, which is equipped with inner product $(\cdot,\cdot)_{\mathcal{V}}$ and norm $\|\cdot\|_{\mathcal{V}}$, we consider the following minimization problem:

$$u = \operatorname{argmin}_{v\in\mathcal{V}} E(v). \tag{1}$$

We now make some assumptions that guarantee that the minimizer exists and is unique.

2.1. Assumptions on the energy.

We assume that the energy functional $E(\,\cdot\,) : \mathcal{V} \to \mathbb{R}$ is Fréchet differentiable for all points $v \in \mathcal{V}$. For each fixed $v \in \mathcal{V}$, $E'(v) : \mathcal{V} \to \mathbb{R}$ is the continuous linear functional equal to the first Fréchet derivative at v. We further impose the following assumptions on the energy:

(E1) (Strong convexity): There is a constant $\mu > 0$ such that

$$\mu \|w - v\|_{\mathcal{V}}^2 \le \langle E'(w) - E'(v), w - v\rangle, \tag{2}$$

for all $v, w \in \mathcal{V}$, where $\langle\,\cdot\,,\,\cdot\,\rangle$ is the duality pairing between $\mathcal{V}'$ and $\mathcal{V}$.

(E2) (Lipschitz continuity of the first order derivative): For fixed $u_0 \in \mathcal{V}$, there exists a constant L such that, for all $v, w \in \mathcal{B} := \{v \in \mathcal{V} \mid E(v) \le E(u_0)\}$,

$$\|E'(w) - E'(v)\|_{\mathcal{V}'} \le L\|w - v\|_{\mathcal{V}}, \tag{3}$$

where

$$\|f\|_{\mathcal{V}'} := \sup_{\substack{v\in\mathcal{V}\\ \|v\|_{\mathcal{V}}=1}} \langle f, v\rangle = \sup_{v\in\mathcal{V}\setminus\{0\}} \frac{\langle f, v\rangle}{\|v\|_{\mathcal{V}}}.$$

Other authors, for example Ciarlet [8], use the term *elliptic* for the property in assumption (E1). We should also point out that assumption (E1) is equivalent to the property that the derivative is *strongly monotone* [1].

The following results are classical, and the proof, which is skipped for the sake of brevity, can be found in [9, p. 35], [8, Thm. 8.2-2], or [1, Thm. 3.3.13].

Theorem 2.1. *If E satisfies assumption* (E1)*, then, for all $w, v \in \mathcal{V}$,*

$$E(w) - E(v) \geq \langle E'(v), w - v\rangle + \frac{\mu}{2}\|w - v\|_{\mathcal{V}}^2. \tag{4}$$

Consequently, E is strongly convex and coercive. Furthermore, there is a unique element $u \in \mathcal{V}$ with the property that

$$E(u) \leq E(v) \quad \forall\, v \in \mathcal{V} \quad \textit{and} \quad E(u) < E(v) \quad \forall\, v \neq u,$$

and this global minimizer *satisfies Euler equation*

$$\langle E'(u), w\rangle = 0 \quad \forall\ w \in \mathcal{V}. \tag{5}$$

The strong convexity and the Lipschitz continuity imply the following estimates.

Lemma 2.2. *Suppose E satisfies assumptions* (E1) *and* (E2)*. For all $v, w \in \mathcal{B}$,*

$$\mu\|w - v\|_{\mathcal{V}}^2 \leq \langle E'(w) - E'(v), w - v\rangle \leq L\|w - v\|_{\mathcal{V}}^2.$$

Furthermore the lower bound holds for all $v, w \in \mathcal{V}$.

Proof. The lower bound is just assumption (E1). To get the upper bound, observe that (E2) implies that, for all $w, v \in \mathcal{B}$, and for any $z \in \mathcal{V}$,

$$|\langle E'(w) - E'(v), z\rangle| \leq \|E'(w) - E'(v)\|_{\mathcal{V}'}\|z\|_{\mathcal{V}} \leq L\|w - v\|_{\mathcal{V}}\|z\|_{\mathcal{V}}.$$

Setting $z = w - v$ gives the desired inequality. □

Proposition 2.3. *If E satisfies* (E1)*, the sublevel set $\mathcal{B}$ is convex.*

Proof. Suppose that $v, w \in \mathcal{B}$. Then $E(w) \leq E(u_0)$ and $E(v) \leq E(u_0)$. Since E is strictly convex, for any $t \in [0, 1]$,

$$E(u_0) \geq (1 - t)E(w) + tE(v) \geq E((1 - t)w + tv).$$

Thus, $(1 - t)w + tv \in \mathcal{B}$ for any $t \in [0, 1]$. □

Now, we consider the relation between the energy and the norm centered at the minimizer. The following estimates can be easily proved using Taylor's theorem with integral remainder; see, e.g., [21].

Lemma 2.4 (Quadratic energy trap)**.** *Suppose E satisfies assumptions* (E1) *and* (E2)*. For all $v, w \in \mathcal{B}$,*

$$\frac{\mu}{2}\|w - v\|_{\mathcal{V}}^2 + \langle E'(v), w - v\rangle \leq E(w) - E(v) \leq \langle E'(v), w - v\rangle + \frac{L}{2}\|w - v\|_{\mathcal{V}}^2. \tag{6}$$

Furthermore the lower bound holds for all $v, w \in \mathcal{V}$. In addition, suppose $u \in \mathcal{B}$ is the minimizer of E; then for all $w \in \mathcal{B}$,

$$\frac{\mu}{2}\|w - u\|_{\mathcal{V}}^2 \leq E(w) - E(u) \leq \frac{L}{2}\|w - u\|_{\mathcal{V}}^2. \tag{7}$$

Again the lower bound holds for all $w \in \mathcal{V}$.

Based on assumption (E1), the upper bound can be replaced by a norm of the gradient. Since the proof is less standard, we include it here.

Lemma 2.5. *Suppose that E satisfies assumption* (E1) *and $u \in \mathcal{V}$ is the minimizer of E; then for all $v \in \mathcal{V}$,*

$$0 \leq E(v) - E(u) \leq \frac{1}{2\mu}\|E'(v)\|_{\mathcal{V}'}^2. \tag{8}$$

Proof. Fix the point $v \in \mathcal{V}$. Now, for any $w \in \mathcal{V}$, using the lower bound of (6), we have

$$E(w) \geq E(v) + \langle E'(v), w - v\rangle + \frac{\mu}{2}\|w - v\|_{\mathcal{V}}^2 =: g(w).$$

For fixed $v \in \mathcal{V}$, the minimizer of $g(w)$ is $w^* := v - \frac{1}{\mu}\mathfrak{R}E'(v)$, where $\mathfrak{R}E'(v)$ is the Riesz representation in $\mathcal{V}$ of $E'(v)$. Therefore,

$$E(w) \geq g(w) \geq g(w^*) = E(v) - \frac{1}{2\mu}\|\mathfrak{R}E'(v)\|_{\mathcal{V}}^2 = E(v) - \frac{1}{2\mu}\|E'(v)\|_{\mathcal{V}'}^2.$$

Then (8) is obtained by letting $w = u$ in the above inequality. □

We shall often use the following simple variant of Lemma 2.4.

Lemma 2.6 (Convexity of energy sections)**.** *Suppose that E satisfies* (E1) – (E2)*, $\xi \in \mathcal{B}$ is arbitrary, and $\mathcal{W} \subseteq \mathcal{V}$ is a closed subspace. Define the energy section*

$$J(w) := E(\xi + w) \quad \forall\ w \in \mathcal{W}.$$

Then $J : \mathcal{W} \to \mathbb{R}$ is differentiable, strongly convex, and there exists a unique element $\eta \in \mathcal{W}$ such that $\xi + \eta \in \mathcal{B}$, η is the unique global minimizer of J, and

$$\langle E'(\xi + \eta), w\rangle = \langle J'(\eta), w\rangle = 0 \quad \forall\ w \in \mathcal{W}.$$

Furthermore, for all $w \in \mathcal{W}$ with $w + \xi \in \mathcal{B}$,

$$\frac{\mu}{2}\|w - \eta\|_{\mathcal{V}}^2 \leq J(w) - J(\eta) = E(\xi + w) - E(\xi + \eta) \leq \frac{L}{2}\|w - \eta\|_{\mathcal{V}}^2.$$

The lower bound holds for any $w \in \mathcal{W}$, without restriction.

The ratio L/μ is called the condition number of the derivative E'; see [21, p. 63]. The rate of convergence of iterative methods for solving (1) usually depends on the condition number. Here we assume L/μ is uniformly bounded, as long as we remain in $\mathcal{B}$. Then the Riesz map $\mathfrak{R} : \mathcal{V}' \to \mathcal{V}$ can be used as a preconditioner and the corresponding preconditioned gradient descent method will converge [10].

Implementing preconditioned gradient descent methods in $\mathcal{V}$ requires the computation of the Riesz map $\mathfrak{R}$ which is equivalent to inverting a symmetric positive definite (SPD) operator (an SPD matrix of size $\dim \mathcal{V} \times \dim \mathcal{V}$ when $\dim \mathcal{V} < +\infty$). Of course we can also use multilevel methods to compute $\mathfrak{R}$ and use steepest descent, nonlinear conjugate gradient, or the Newton method as the outer iteration. In the following, we shall provide optimization methods that only require computing inverses with much smaller sizes.

2.2. **Assumptions on the space decomposition.**

Suppose that

$$\mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2 + \cdots + \mathcal{V}_N, \qquad \mathcal{V}_i \subseteq \mathcal{V}, \quad i = 1, \ldots, N,$$

is a space decomposition of $\mathcal{V}$ using closed subspaces $\mathcal{V}_i$ for $i = 1, 2, \ldots, N$. We shall use the following assumptions on the space decomposition.

(SS1) (Stable decomposition): There is a constant $C_A > 0$, such that, for every $v \in \mathcal{V}$, there exists $v_i \in \mathcal{V}_i$, $i = 1, \cdots, N$, with the property that

$$v = \sum_{i=1}^{N} v_i \quad \text{and} \quad \sum_{i=1}^{N} \|v_i\|_{\mathcal{V}}^2 \leq C_A^2\|v\|_{\mathcal{V}}^2.$$

(SS2) (Strengthened Cauchy Schwarz inequality): There is a constant $C_S > 0$, such that, for any $w_{i,j} \in \mathcal{B}$, $u_i \in \mathcal{V}_i$, $v_i \in \mathcal{V}_i$, with $w_{i,j} + u_i \in \mathcal{B}$,

$$\sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle E'(w_{i,j}+u_j) - E'(w_{i,j}), v_i\rangle \leq \ C_S \left(\sum_{i=1}^{N} \|u_i\|_{\mathcal{V}}^2\right)^{1/2} \left(\sum_{i=1}^{N} \|v_i\|_{\mathcal{V}}^2\right)^{1/2}.$$

When E'' exists and is continuous, by the mean value theorem and standard Cauchy Schwarz inequality

$$\langle E'(w_{i,j}+u_j) - E'(w_{i,j}), v_i\rangle = \langle E''(\xi_i)u_j, v_i\rangle \leq \|E''(\xi_i)\|\|u_j\|_{\mathcal{V}}\|v_i\|_{\mathcal{V}}.$$

Thus a naive verification of (SS2) could use the constant $C_S = LN$, which would be large if N is large. When the inner product induced by the Hessian $E''(\xi_i)$ is spectrally equivalent to a $\mathcal{V}$-inner product, a better constant C_S, which is independent of N, can be obtained. This is the reason we call it the strengthened Cauchy Schwarz inequality.

We note that the constant $C_S > 0$ can be related to the Lipschitz constant in assumption (E2). Unless E was quadratic, we could not assume in general that the strengthened Cauchy Schwarz inequality would hold without restriction to the bounded sublevel set $\mathcal{B}$, as indicated in assumption (SS2).

3. Successive subspace optimization methods

For $k \geq 0$ and a given approximate solution $u^k \in \mathcal{V}$, one step of the successive subspace optimization (SSO) method [25] is given in Algorithm 1.

Algorithm: $u^{k+1} = \mathrm{SSO}(u^k)$
$v_0 = u^k$;
for $i = 1 : N$ **do**
  Define an energy section along $\mathcal{V}_i$:
  $$J_i(w) := E(v_{i-1} + w) \quad \forall w \in \mathcal{V}_i;$$
  Compute the subspace correction:
  $$e_i = \mathrm{argmin}_{w\in\mathcal{V}_i} J_i(w); \tag{9}$$
  Apply the subspace correction:
  $$v_i = v_{i-1} + e_i;$$
end
$u^{k+1} = v_N$

Algorithm 1: Successive subspace optimization method.

Remark 3.1. Note that e_i computed in (9) of Algorithm 1 is uniquely defined, owing to the strong convexity inherited by the energy section J_i. In fact, the correction satisfies

$$\langle E'(v_i), w\rangle = \langle E'(v_{i-1}+e_i), w\rangle = \langle J'(e_i), w\rangle = 0 \quad \forall\ w \in \mathcal{V}_i.$$

The orthogonality relation satisfied by the corrected approximation, v_i, specifically,

$$\langle E'(v_i), w\rangle = 0 \quad \forall\ w \in \mathcal{V}_i,$$

is sometimes referred to as the *fundamental orthogonality* (FO) of the solver.

Remark 3.2. We point out that, when $\mathcal{V}_i$ is one dimensional, then the computation of the subspace correction is identical to a nonlinear Gauss-Seidel method. In fact, the SSO method can be considered as a generalization of the nonlinear Gauss-Seidel methodology.

We aim to prove a linear reduction of the energy difference for one iteration of the SSO algorithm:

$$E(u^{k+1}) - E(u) \leq \rho(E(u^k) - E(u)), \tag{10}$$

where u is the minimizer of E and $u^{k+1} = \mathrm{SSO}(u^k)$, with a contraction factor $\rho \in (0,1)$. Ideally ρ is independent of the size of the problem. The algorithm and convergence theory has been developed in [25, 27] for a convex energy in Banach spaces. For completeness, we include a simplified version for Hilbert space here.

We will utilize the following simple result.

Theorem 3.3. *Suppose that* $\{d_k\}_{k=0}^{\infty}$, $\{\delta_k\}_{k=0}^{\infty}$, $\{\eta_k\}_{k=0}^{\infty}$ *are sequences of nonnegative real numbers, the first two having the relationship*

$$\delta_k = d_k - d_{k+1}, \quad k = 0, 1, 2, \cdots.$$

Assume that there are constants $C_L, C_U > 0$, *independent of* k, *such that*

$$C_L \eta_k \leq \delta_k \quad \textit{and} \quad d_{k+1} \leq C_U \eta_k.$$

Then

$$d_{k+1} \leq \frac{C_U}{C_L + C_U} d_k, \quad k = 0, 1, 2, \cdots. \tag{11}$$

Consequently $\{d_k\}$ *converges monotonically, and* (*at least*) *linearly to* 0.

Proof. Observe that

$$d_{k+1} \leq C_U \eta_k = \frac{C_U}{C_L} C_L \eta_k \leq \frac{C_U}{C_L} \delta_k = \frac{C_U}{C_L}(d_k - d_{k+1}),$$

which implies (11). Proving that $\{d_k\}$ is strictly decreasing to zero is straightforward, and the proof is omitted. □

We will apply the last result with the following definitions:

$$d_k := E(u^k) - E(u) \quad \text{and} \quad \delta_k := E(u^k) - E(u^{k+1}). \tag{12}$$

The quantity d_k is the difference between the current energy and the minimum energy, also known as the optimality gap, and δ_k is the energy decrease associated to the $k+1$th iteration. They are connected, as desired, by the trivial identity

$$\delta_k = d_k - d_{k+1}.$$

See Figure 1 for an illustration. We define η_k in terms of the subspace corrections via

$$\eta_k := \sum_{i=1}^{N} \|e_i\|_{\mathcal{V}}^2,$$

and we assume the following upper and lower bounds.

Lower bound on energy decay. There exists a positive constant C_L such that for any $k = 0, 1, 2, \cdots$

$$E(u^k) - E(u^{k+1}) = \delta_k \geq C_L \eta_k = C_L \sum_{i=1}^{N} \|e_i\|_{\mathcal{V}}^2. \tag{13}$$

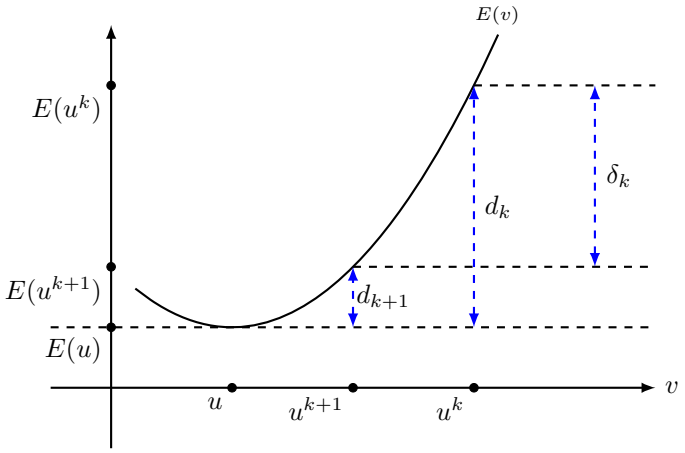

FIGURE 1. The sequences $\{d_k\}$ and $\{\delta_k\}$.

Upper bound on optimality gap. There exists a positive constant C_U such that for any $k = 0, 1, 2, \cdots$

$$E(u^{k+1}) - E(u) = d_{k+1} \leq C_U \eta_k = C_U \sum_{i=1}^{N} \|e_i\|_{\mathcal{V}}^2. \tag{14}$$

If these bounds hold, then, as a corollary to Theorem 3.3, we have the following.

Corollary 3.4. *Assume that the lower bound* (13) *and upper bound* (14) *hold with positive constants C_L and C_U, respectively. We then have*

$$E(u^{k+1}) - E(u) \leq \rho \left(E(u^k) - E(u) \right), \quad \rho := \frac{C_U}{C_L + C_U},$$

and $E(u^k)$ converges monotonically, and (at least) linearly to $E(u)$, at the linear rate ρ. Furthermore, u^k converges at least linearly to u.

Proof. The linear convergence of $E(u^k)$ to $E(u)$ at the rate ρ is guaranteed by Theorem 3.3. Using (7), with $w = u^k$, we have

$$\frac{\mu}{2} \left\| u^k - u \right\|_{\mathcal{V}}^2 \leq E(u^k) - E(u),$$

which guarantees the linear convergence of u^k to u. □

Verifying the lower bound is relatively easy since E is convex. Solving the convex optimization problem in each subspace will definitely decrease the energy, and this decrease can be quantified in terms of the norms of the corrections. We make essential use of the fundamental orthogonality property.

Theorem 3.5. *Let u^k be the kth iteration and $u^{k+1} = \mathrm{SSO}(u^k)$. If E is strongly convex in the sense of satisfying* (E1)*, then*

$$\delta_k = E(u^k) - E(u^{k+1}) \geq C_L \sum_{i=1}^{N} \|e_i\|_{\mathcal{V}}^2, \quad C_L := \frac{\mu}{2}.$$

Proof. Recalling Lemma 2.6, we observe that J_i (defined in Algorithm 1) is strictly convex over $\mathcal{V}_i$ and is Fréchet differentiable, as it inherits the structure of E. It follows that

$$\langle J_i'(e_i), w\rangle = 0 \quad \forall\, w \in \mathcal{V}_i.$$

But the object on the left-hand side is simply a directional derivative of the full energy, and it is easy to see that

$$\langle J_i'(e_i), w\rangle = \langle E'(v_{i-1} + e_i), w\rangle = \langle E'(v_i), w\rangle \quad \forall\, w \in \mathcal{V}_i.$$

Therefore, the fundamental orthogonality, $E'(v_i) = 0$ in $\mathcal{V}_i'$, holds. As $e_i = v_i - v_{i-1} \in \mathcal{V}_i$, in view of Lemma 2.6, we have

$$E(v_{i-1}) - E(v_i) = J_i(0) - J_i(e_i) \geq \frac{\mu}{2} \|e_i\|_{\mathcal{V}}^2. \tag{15}$$

The sum of the left-hand side telescopes, and we have

$$E(u^k) - E(u^{k+1}) = \sum_{i=1}^{N} (E(v_{i-1}) - E(v_i)) \geq \frac{\mu}{2} \sum_{i=1}^{N} \|e_i\|_{\mathcal{V}}^2.$$

□

Remark 3.6. In view of (15), the convexity of E can be relaxed to the local convexity of the energy sections J_i in each subspace $\mathcal{V}_i$. Namely we may have a nonconvex energy E which, restricted to each subspace, is convex and the lower bound still holds. For example, the energy used in the optimal delaunay triangulation (ODT) [7] is nonconvex globally. But restricted to one vertex, it is convex, and the corresponding 1-D optimization problem has a closed form solution, which is known as ODT mesh smoothing [5]. Theorem 3.5 guarantees the energy decreasing property of ODT mesh smoothing.

The upper bound is more delicate and relies on the assumptions about the decomposition of spaces. The result is given in the following theorem.

Theorem 3.7. *Let u^{k+1} be the $k+1$st iteration in the* SSO *algorithm. Suppose that the space decomposition satisfies assumptions* (SS1) *and* (SS2) *and the energy E satisfies assumption* (E1)*; then we have*

$$d_{k+1} = E(u^{k+1}) - E(u) \leq C_U \sum_{i=1}^{N} \|e_i\|_{\mathcal{V}}^2, \quad C_U := \frac{C_S^2 C_A^2}{2\mu}.$$

Proof. Using Lemma 2.5, with the choice $v = u^{k+1}$ in (8), we have

$$d_{k+1} = E(u^{k+1}) - E(u) \leq \frac{1}{2\mu} \|E'(u^{k+1})\|_{\mathcal{V}'}^2.$$

For any $w \in \mathcal{V}$, we choose a stable decomposition $w = \sum_{i=1}^N w_i$; then

$$\begin{aligned}
\langle E'(u^{k+1}), w\rangle &= \sum_{i=1}^N \langle E'(u^{k+1}), w_i\rangle \\
&= \sum_{i=1}^N \langle E'(u^{k+1}) - E'(v_i), w_i\rangle \\
&= \sum_{i=1}^N \sum_{j=i+1}^N \langle E'(v_j) - E'(v_{j-1}), w_i\rangle \\
&\leq C_S \left(\sum_{i=1}^N \|e_i\|_{\mathcal{V}}^2\right)^{1/2} \left(\sum_{i=1}^N \|w_j\|_{\mathcal{V}}^2\right)^{1/2} \\
&\leq C_S C_A \left(\sum_{i=1}^N \|e_i\|_{\mathcal{V}}^2\right)^{1/2} \|w\|_{\mathcal{V}}.
\end{aligned}$$

Here we use the fact that we solve the minimization problem on each subspace exactly and the energy decreases, therefore, $v_j \in \mathcal{B}$ for all j and $E'(v_i) = 0$ in $\mathcal{V}_i'$. Then we have

$$\begin{aligned}
E(u^{k+1}) - E(u) &\leq \frac{1}{2\mu}\|E'(u^{k+1})\|_{\mathcal{V}'}^2 \\
&= \frac{1}{2\mu}\left(\sup_{w\in\mathcal{V}\setminus\{0\}} \frac{\langle E'(u^{k+1}), w\rangle}{\|w\|_{\mathcal{V}}}\right)^2 \\
&\leq \frac{1}{2\mu} C_S^2 C_A^2 \sum_{i=1}^N \|e_i\|_{\mathcal{V}}^2,
\end{aligned}$$

which finishes the proof. □

Based on the lower bound given in Theorem 3.5 and the upper bound given in Theorem 3.7, we can conclude the convergence of SSO. Comparing with the results in [27], we use slightly weaker assumptions, and the constant C_U seems to be slightly better.

Corollary 3.8. *Let u^k be the kth iteration and $u^{k+1} = \mathrm{SSO}(u^k)$. Suppose that the space decomposition satisfies assumptions* (SS1) *and* (SS2) *and the energy E satisfies assumption* (E1)*; then we have*

$$E(u^{k+1}) - E(u) \leq \rho(E(u^k) - E(u)) \quad \textit{with} \quad \rho = \frac{C_S^2 C_A^2}{C_S^2 C_A^2 + \mu^2}.$$

As we have pointed out previously, the Lipschitz continuity and constant L are implicitly contained in assumption (SS2) (the strengthened Cauchy Schwarz inequality) and the constant C_S. We will show how this can be so with an application at the end of the paper.

4. Fast subspace descent method with exact line search

In this section, we present the theory for the convergence of the fast subspace descent (FASD) method listed in Algorithm 2. To recap, in the SSO method,

Algorithm 1, we need to solve the optimization problem

$$\min_{w\in\mathcal{V}_i} E(v_{i-1}+w)$$

in each subspace exactly, which requires evaluation of the global energy E and its derivative E' in the space $\mathcal{V}$. Although the size of the optimization problem is reduced to $\dim\mathcal{V}_i$, such evaluations are still in the original space of size $\dim\mathcal{V}$, which may be expensive.

4.1. **Algorithm definition.** Denote by $I_i : \mathcal{V}_i \hookrightarrow \mathcal{V}$ the natural inclusion and $R_i = I_i^\top : \mathcal{V}' \to \mathcal{V}_i'$ the natural restriction of functionals. Thus, for all $w \in \mathcal{V}_i$,

$$\langle R_i E'(v_{i-1}), w\rangle = \langle E'(v_{i-1}), R_i^\top w\rangle = \langle E'(v_{i-1}), I_i w\rangle.$$

Often times we just drop R_i and I_i, as their actions can be assumed implicitly. We need to evaluate the gradient $R_iE'(v_{i-1}+I_iw)$, as well as the Hessian $R_iE''(v_{i-1}+I_iw)I_i$ and its inverse, if Newton's method is used, several times. This is practical only if the natural inclusion I_i is efficient to realize, e.g., a one dimensional subspace generated by one basis function of $\mathcal{V}$ and the resulting method is the so-called non-linear Gauss-Seidel iteration.

Instead of solving the minimization problem using the original energy E, in our FASD algorithm (Algorithm 2) we utilize a locally-defined energy E_i in each subspace $\mathcal{V}_i$ and solve a perturbed optimization problem. For the moment, let us assume that $E_i : \mathcal{V}_i \to \mathbb{R}$ is Fréchet differentiable in $\mathcal{V}_i$. We will give further assumptions shortly. In addition to prolongation and restriction operators, we also need a projection operator $Q_i : \mathcal{V} \to \mathcal{V}_i$. Ideally, Q_iv yields a good approximation of v in the subspace $\mathcal{V}_i$. Recall that as a projection operator $Q_iv_i = v_i$ for $v_i \in \mathcal{V}_i$.

Algorithm: $u^{k+1} = \mathrm{FASD}(u^k)$
$v_0 = u^k$;
for $i = 1 : N$ **do**
Compute the so-called subspace τ-*perturbation*: let $\xi_i = Q_iv_{i-1}$ and

$$\tau_i := E_i'(\xi_i) - R_iE'(v_{i-1}) \in \mathcal{V}_i'; \tag{16}$$

Solve the subspace residual problem: Find $\eta_i \in \mathcal{V}_i$, such that

$$\langle E_i'(\eta_i), w\rangle = \langle \tau_i, w\rangle \quad \forall\, w \in \mathcal{V}_i; \tag{17}$$

Compute the search direction:

$$s_i := \eta_i - \xi_i \in \mathcal{V}_i; \tag{18}$$

Orthogonalize the subspace correction via the exact line search:

$$\varepsilon_i := \alpha_i^* s_i, \tag{19}$$

where

$$\alpha_i^* = \mathrm{argmin}_{\alpha\in\mathbb{R}} E(v_{i-1}+\alpha s_i); \tag{20}$$

Apply the subspace correction:

$$v_i := v_{i-1} + \varepsilon_i; \tag{21}$$

end
$u^{k+1} := v_N$;

Algorithm 2: Fast subspace descent (FASD) method.

We shall view our fast subspace descent (FASD) method as a hybrid of the SSO and FAS methods. For the proof of convergence, it helps to treat FASD as an SSO iteration with an inexact local solver. We could also say that FASD (Algorithm 2) is essentially FAS with an additional line search step.

Our FASD algorithm is listed in Algorithm 2. In the orthogonalization step, cf. (19), we perform a line search to find the optimal step size which still requires the evaluation of some of the "fine level" functions $E(v_{i-1}+\alpha s_i)$, $E'(v_{i-1}+\alpha s_i)$, and $E''(v_{i-1}+\alpha s_i)$ in $\mathcal{V}$. The computational cost is reduced compared with evaluation of $v_{i-1}+w$ for multiple $w \in \mathcal{V}_i$. Algorithm 2 is an intermediate step towards the convergences proof of original FAS. In Section 5, we shall analyze an algorithm that uses a simpler choice of step size, one that is closer to the original FAS method. In Section 6, we shall consider the original FAS, which corresponds to FASD with the step size $\alpha_i = 1$.

4.2. **Strong convexity of local energy and well-posedness.** To show the well-posedness of the local problem (17), and therefore Algorithm 2, we need some assumptions on the energies E_i. As mentioned, we assume $E_i : \mathcal{V}_i \to \mathbb{R}$ is Fréchet differentiable for all points $v \in \mathcal{V}_i$. In addition, we introduce the following assumptions on the local energy, E_i, which is just the local version of (E1):

(E3) (Strong convexity/ellipticity): There exists a constant μ_i such that for all $v, w \in \mathcal{V}_i$
$$\langle E_i'(w) - E_i'(v), w - v\rangle \geq \mu_i \|w - v\|_{\mathcal{V}}^2.$$

For the local optimization problem, expressed in equation (17), we are not minimizing an approximated energy E_i, i.e., not solving $E_i'(Q_i v_{i-1} + s_i) = 0$. Instead a so-called τ-perturbation is added to the right-hand side. Still, this optimization problem is uniquely solvable.

Lemma 4.1. *Assume E_i satisfies the strong convexity assumption* (E3)*. Then there exists a unique solution to the residual equation* (17)*.*

Proof. The residual equation (17) is the Euler equation for the minimization problem
$$\min_{v \in \mathcal{V}_i} \left(E_i(v) - \langle \tau_i, v\rangle\right). \tag{22}$$
As E_i is strictly convex, and the linear shift $\langle \tau_i, v\rangle$ will not affect the convexity, the global minimizer of (22) exists, is unique, and satisfies the Euler equation (17). Detailed proofs can be found in [9, p. 35], [8, Thm. 8.2-2], or [1, Thm. 3.3.13]. □

Remark 4.2. We note that Algorithm 2 (FASD) generalizes Algorithm 1 (SSO). They yield the same approximations in the case that
$$E_i(\eta) := E(v_{i-1} - Q_i v_{i-1} + \eta) \quad \forall\, \eta \in \mathcal{V}_i.$$
The projection Q_i just needs to satisfy the usual property $Q_i\eta = \eta$ for all $\eta \in \mathcal{V}_i$. As a consequence of this choice, $\tau_i \equiv 0$ and, for all $w \in \mathcal{V}_i$,
$$\langle E'(v_{i-1}+s_i), w\rangle = \langle E'(v_{i-1} - Q_i v_{i-1} + \eta_i), w\rangle = \langle E_i'(\eta_i), w\rangle = 0.$$
With these choices in FASD, the last step (orthogonalization) is redundant because
$$\langle E'(v_{i-1}+s_i), s_i\rangle = 0$$
upon taking $w = s_i$. In other words, the orthogonality is valid with $\alpha_i^* = 1$ for SSO.

4.3. **Lower bound.** The first correction that we obtain in Algorithm 2, namely, $s_i = \eta_i - \xi_i$, where $\xi_i = Q_i v_{i-1}$ is the full approximation, is used as the search direction for a line optimization. The line optimization confers an orthogonalization property to the corrected approximation v_i. Due to this orthogonalization and the convexity of E, the proof of the lower bound for FASD is almost exactly the same as that for the SSO method.

Theorem 4.3. *Suppose that E satisfies* (E1) *and E_i satisfies* (E3)*, and let u^k be the kth iteration in the FASD algorithm* (*Algorithm* 2)*. Then*

$$E(u^k) - E(u^{k+1}) \geq \frac{\mu}{2} \sum_{i=1}^{N} \|\varepsilon_i\|_{\mathcal{V}}^2.$$

Proof. We apply a similar technique as in the proof of Theorem 3.5. Due to the line search, we still have an orthogonality property that can be utilized, namely,

$$\langle E'(v_i), w \rangle = 0, \quad w \in \operatorname{span}\{s_i\}.$$

Then, applying Lemma 2.6, with the subspace $\mathcal{W} = \operatorname{span}\{s_i\}$, and noting that

$$v_i - v_{i-1} = \varepsilon_i = \alpha_i^* s_i \in \operatorname{span}\{s_i\},$$

we have

$$E(v_{i-1}) - E(v_i) \geq \frac{\mu}{2} \|v_{i-1} - v_i\|_{\mathcal{V}}^2 = \frac{\mu}{2} \|\varepsilon_i\|_{\mathcal{V}}^2,$$

and consequently

$$E(u^k) - E(u^{k+1}) = \sum_{i=1}^{N} \left(E(v_{i-1}) - E(v_i) \right) \geq \frac{\mu}{2} \sum_{i=1}^{N} \|\varepsilon_i\|_{\mathcal{V}}^2.$$

□

We will later need the following simple result, which follows because of the strong convexity assumption (E3).

Lemma 4.4. *Let s_i be computed as in Algorithm* 2 *and suppose that E_i satisfies assumption* (E3)*. Then s_i is a descent direction in the sense that*

$$\langle -E'(v_{i-1}), s_i \rangle \geq \mu_i \|s_i\|_{\mathcal{V}}^2 > 0.$$

Proof. The local problem (17) can be rewritten as follows: find $\eta_i \in \mathcal{V}_i$ s.t.

$$\langle E_i'(\eta_i) - E_i'(\xi_i), w \rangle = -\langle R_i E'(v_{i-1}), w \rangle \quad \forall\ w \in \mathcal{V}_i. \tag{23}$$

Here recall that $\xi_i = Q_i v_{i-1} \in \mathcal{V}_i$ and $\eta_i = \xi_i + s_i$. Choosing $w = s_i$ and using the strong convexity of E_i, we obtain the inequality

$$\langle -R_i E'(v_{i-1}), s_i \rangle = \langle E_i'(\eta_i) - E_i'(\xi_i), s_i \rangle \geq \mu_i \|s_i\|_{\mathcal{V}}^2 > 0.$$

□

4.4. **Lipschitz continuity of E_i' and estimates of α_i^*.** As Theorem 4.3 implies, the energy is always decreasing and iterates will remain in the sublevel set $\mathcal{B}$, but the search region, and, e.g., the point $\xi_i + s_i$, may not be contained in $\mathcal{B}$. To be able to use Lipschitz continuity, we introduce an enlarged set

$$\mathcal{B}^+ := \{v \in \mathcal{V} \mid \operatorname{dist}(v, \mathcal{B}) \leq \sqrt{\chi}\}, \tag{24}$$

where χ is given by

$$\chi := \frac{2L^2}{\mu \min_i \mu_i^2}(E(u_0) - E(u)).$$

We then introduce an assumption on the Lipschitz continuity of E_i' with respect to the projection of $\mathcal{B}^+$:

(E4) (Lipschitz continuity of the first order derivative): There exists a constant $L_i > 0$, such that

$$\|E_i'(w) - E_i'(v)\|_{\mathcal{V}'} \leq L_i \|w - v\|_{\mathcal{V}}$$

for all $w, v \in \mathcal{B}_i := Q_i \mathcal{B}^+$.

Remark 4.5. Observe that we must assume that (E3) holds for (E4) to make sense. In other words, we cannot assume (E4) without first assuming (E3), since μ_i is involved in the definition of χ and, therefore, $\mathcal{B}^+$. Regarding $\mathcal{B}^+$, note that it is not a sublevel set. However, it is straightforward to verify that both $\mathcal{B}^+$ and $\mathcal{B}_i = Q_i \mathcal{B}^+$ are convex. The proofs are omitted for the sake of brevity.

Later, we will show that $\xi_i + s_i \in \mathcal{B}_i$ so that we can take advantage of the Lipschitz continuity of E_i' in our analysis. Notice that the Lipschitz continuity of E_i' is imposed for the set $Q_i \mathcal{B}^+$, which is related to $\mathcal{B}$ used in (E2). Interestingly, there is no relationship between E and E_i that is explicitly assumed for the moment. Indeed E and E_i are just related through the upper and lower bound of the first derivatives and norms. In general, based on the assumptions (E3) and (E4), we have the following lemma, which gives results analogous to those in Lemmas 2.2 and 2.4.

Lemma 4.6. *Assume E_i satisfies assumptions* (E3) *and* (E4). *For any $v, w \in \mathcal{B}_i$,*

$$\mu_i \|w - v\|_{\mathcal{V}}^2 \leq \langle E_i'(w) - E_i'(v), w - v\rangle \leq L_i \|w - v\|_{\mathcal{V}}^2$$

and

$$\frac{\mu_i}{2}\|w - v\|_{\mathcal{V}}^2 + \langle E_i'(v), w - v\rangle \leq E_i(w) - E_i(v) \leq \langle E_i'(v), w - v\rangle + \frac{L_i}{2}\|w - v\|_{\mathcal{V}}^2.$$

Though it is not required, if it happens that $u_i \in \mathcal{B}_i$, where $u_i \in \mathcal{V}_i$ is the global minimizer of E_i, then for all $w \in \mathcal{B}_i$,

$$\frac{\mu_i}{2}\|w - u_i\|_{\mathcal{V}}^2 \leq E_i(w) - E_i(u_i) \leq \frac{L_i}{2}\|w - u_i\|_{\mathcal{V}}^2.$$

The lower bounds above hold for all $w \in \mathcal{V}_i$, without restriction.

In order to better understand the choice of the step size, we introduce the scalar function f_i. See Figure 2 and equation (25).

Proposition 4.7. *Suppose that E satisfies assumption* (E1) *and the local energy E_i satisfies assumption* (E3). *Define the one dimensional energy section*

$$f_i(\alpha) := E(v_{i-1} + \alpha s_i). \tag{25}$$

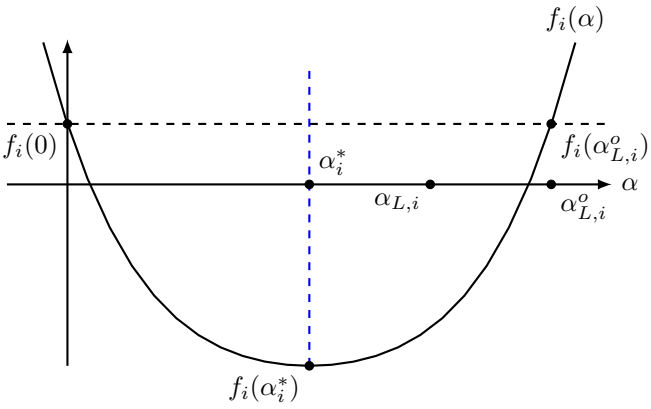

FIGURE 2. The function f_i defined in (25). f_i is a one dimensional energy section. It is straightforward to prove that its minimizer, α_i^*, is positive.

Then

$$f_i'(0) = \langle E'(v_{i-1}), s_i \rangle \leq -\mu_i \left\| s_i \right\|_{\mathcal{V}}^2 .$$

Furthermore, $\alpha_i^* > 0$*, and, for all* $\alpha \in (0, \alpha_i^*]$*,* $f_i(\alpha) < f_i(0)$*.*

Proof. Lemma 4.4 implies $f_i'(0) < 0$. As f_i' is continuous, we conclude that the minimizing point is positive, $\alpha_i^* > 0$, and for all $\alpha \in (0, \alpha_i^*]$, $f_i(\alpha) < f_i(0) = E(v_{i-1})$. □

Lemma 4.8. *Assume* E *satisfies assumptions* (E1) – (E2)*. Then* $f_i(\alpha)$*, defined in* (25)*, is differentiable and strongly convex in the following sense: for all* $\alpha, \beta \in \mathbb{R}$*,*

$$(f_i'(\alpha) - f_i'(\beta))(\alpha - \beta) \geq (\alpha - \beta)^2 \mu \|s_i\|_{\mathcal{V}}^2.$$

Furthermore, f_i' *is Lipschitz in the following sense: for all* $0 \leq \alpha, \beta \leq \alpha_{L,i}$*,*

$$|f_i'(\alpha) - f_i'(\beta)| \leq L \|s_i\|_{\mathcal{V}}^2 |\alpha - \beta|,$$

where $\alpha_{L,i} := (1 + \sqrt{\mu/L})\alpha_i^*$*.*

Proof. The proof is based on the following identity:

$$f_i'(\alpha) - f_i'(\beta) = \langle E'(v_{i-1} + \alpha s_i) - E'(v_{i-1} + \beta s_i), s_i \rangle.$$

Then, by assumption (E1),

$$\begin{aligned}(f_i'(\alpha) - f_i'(\beta))(\alpha - \beta) &= \langle E'(v_{i-1} + \alpha s_i) - E'(v_{i-1} + \beta s_i), \alpha s_i - \beta s_i \rangle \\ &\geq \mu \|(\alpha - \beta) s_i\|_{\mathcal{V}}^2.\end{aligned}$$

To use the Lipschitz inequality involving E', we need to ensure that the points of evaluation are inside the set $\mathcal{B}$, which imposes an upper bound on α and β. As $f_i'(0) < 0$ and $f_i'(\alpha_i^*) = 0$, by coercivity, there exists $\alpha_{L,i}^o > \alpha_i^*$, such that $f_i(0) = f_i(\alpha_{L,i}^o)$, and, for all $\alpha \in (0, \alpha_{L,i}^o)$, $f_i(\alpha) < f_i(0)$. This implies $v_{i-1} + \alpha s_i \in \mathcal{B}$ for all $\alpha \in (0, \alpha_{L,i}^o)$. See Figure 3.

We then show f_i' is Lipschitz with constant $L\|s_i\|_{\mathcal{V}}^2$ on the interval $[0, \alpha_{L,i}^o]$. For all $\alpha, \beta \in (0, \alpha_{L,i}^o)$, $\alpha \neq \beta$,

$$\begin{aligned}
|f_i'(\alpha) - f_i'(\beta)| &= |\langle E'(v_{i-1} + \alpha s_i) - E'(v_{i-1} + \beta s_i), s_i\rangle| \\
&= \frac{1}{|\alpha - \beta|} |\langle E'(v_{i-1} + \alpha s_i) - E'(v_{i-1} + \beta s_i), (\alpha - \beta) s_i\rangle| \\
&\leq \frac{1}{|\alpha - \beta|} L \|(\alpha - \beta) s_i\|_{\mathcal{V}}^2 \\
&= L\|s_i\|_{\mathcal{V}}^2 |\alpha - \beta|.
\end{aligned}$$

We now estimate $\alpha_{L,i}^o$. As $f_i'(\alpha_i^*) = 0$ and f_i' is Lipschitz in $(0, \alpha_{L,i}^o)$, we have

$$0 < f_i(\alpha_{L,i}^o) - f_i(\alpha_i^*) \leq (\alpha_{L,i}^o - \alpha_i^*)^2 \frac{L}{2} \|s_i\|_{\mathcal{V}}^2.$$

On the other hand, and again from Lemma 2.6,

$$f_i(\alpha_{L,i}^o) - f_i(\alpha_i^*) = f_i(0) - f_i(\alpha_i^*) \geq \frac{\mu(\alpha_i^*)^2}{2} \|s_i\|_{\mathcal{V}}^2.$$

The desired bound

$$\alpha_{L,i}^o \geq \alpha_{L,i} := \left(1 + \sqrt{\frac{\mu}{L}}\right) \alpha_i^* > \alpha_i^* > 0$$

then follows. Note that we need only f_i' is Lipschitz with the same constant on the smaller interval $[0, \alpha_{L,i}] \subseteq [0, \alpha_{L,i}^o]$. The proof is complete. □

To use the Lipschitz continuity of E_i, we require $\xi_i + s_i \in \mathcal{B}_i = Q_i \mathcal{B}^+$, which will be proved by a lower bound of the optimal step size.

Lemma 4.9. *Assume the energy E satisfies assumptions* (E1) – (E2) *and the local energy E_i satisfies the strong convexity assumption* (E3)*; then we have the lower bound*

$$\frac{\mu_i}{L} \leq \alpha_i^*.$$

Consequently,

$$\alpha_{L,i}^o \geq \alpha_{L,i} := \left(1 + \sqrt{\frac{\mu}{L}}\right) \alpha_i^* > \alpha_i^* \geq \frac{\mu_i}{L} > 0.$$

Proof. Recall that $\varepsilon_i = \alpha_i^* s_i \in \operatorname{span}\{s_i\}$, and, due to the line search, we still have an orthogonality property that can be utilized, namely,

$$\langle E'(v_i), w\rangle = 0 \quad \forall\, w \in \operatorname{span}\{s_i\}.$$

Thus $E'(v_{i-1} + \varepsilon_i) = 0$ in the dual of $\operatorname{span}\{s_i\}$. By step 2 in the FASD Algorithm 2,

$$-E'(v_{i-1}) = E_i'(\xi_i + s_i) - E_i'(\xi_i) \quad \text{in} \quad \mathcal{V}_i'.$$

The lower bound is obtained by the strong convexity of E_i and Lipschitz continuity of E':

$$\begin{aligned}\alpha_i^* L\|s_i\|_{\mathcal{V}}^2 = \frac{1}{\alpha_i^*} L\|\varepsilon_i\|_{\mathcal{V}}^2 &\geq \frac{1}{\alpha_i^*}\langle E'(v_{i-1}+\varepsilon_i) - E'(v_{i-1}), \varepsilon_i\rangle \\ &= \langle E'(v_{i-1}+\varepsilon_i) - E'(v_{i-1}), s_i\rangle \\ &= -\langle E'(v_{i-1}), s_i\rangle \\ &= \langle E_i'(\xi_i + s_i) - E_i'(\xi_i), s_i\rangle \\ &\geq \mu_i \|s_i\|_{\mathcal{V}}^2.\end{aligned}$$

Note that $v_{i-1} + \varepsilon_i \in \mathcal{B}$ by Lemma 4.8 so that we can use Lipschitz continuity of E'. □

Next we show the norm of s_i is bounded and thus $\xi_i + s_i \in \mathcal{B}_i$.

Lemma 4.10. *The point $\xi_i + s_i$ is in the set $\mathcal{B}_i$.*

Proof. To show that $\xi_i + s_i \in \mathcal{B}_i$, it suffices to show that $v_{i-1} + s_i \in \mathcal{B}^+$, since $\xi_i + s_i = Q_i(v_{i-1} + s_i)$. To start, we know that $v_{i-1} \in \mathcal{B}$; so by the definition of $\mathcal{B}^+$ in (24), it suffices to prove that $\|s_i\|_{\mathcal{V}}^2 \leq \chi$. By Theorem 4.3 and Lemma 4.9, we have

$$\frac{\mu_i^2}{L^2}\|s_i\|_{\mathcal{V}}^2 \leq (\alpha_i^*)^2\|s_i\|_{\mathcal{V}}^2 = \|\varepsilon_i\|_{\mathcal{V}}^2 \leq \frac{2}{\mu}(E(v_{i-1}) - E(v_i)) \leq \frac{2}{\mu}(E(u_0) - E(u)),$$

which implies

$$\|s_i\|_{\mathcal{V}}^2 \leq \frac{2L^2}{\mu \min_i \mu_i^2}(E(u_0) - E(u)) = \chi.$$

Therefore,

$$\operatorname{dist}(\mathcal{B}, v_{i-1} + s_i) \leq \|s_i\|_{\mathcal{V}} \leq \sqrt{\chi},$$

and the result is proven. □

4.5. **Upper bound.** With our estimates of α_i^* in place, we are now ready to establish an upper bound for the iterates in our FASD Algorithm 2.

Theorem 4.11. *Suppose the space decomposition satisfies* (SS1) *and* (SS2)*, the energy E satisfies* (E1) – (E2)*, and E_i satisfies* (E3) – (E4)*. Then we have the upper bound*

$$E(u^{k+1}) - E(u) \leq C_U \sum_{i=1}^N \|\varepsilon_i\|_{\mathcal{V}}^2,$$

where $C_U := C_A^2 \left[C_S + L\left(1 + \max_i\{L_i/\mu_i\}\right)\right]^2/(2\mu)$.

Proof. Note, for any $w \in \mathcal{V}$, we choose a stable decomposition $w = \sum_{i=1}^N w_i$; then

$$\begin{aligned}\langle E'(u^{k+1}), w\rangle &= \sum_{i=1}^N \langle E'(u^{k+1}), w_i\rangle \\ &= \sum_{i=1}^N \left(\langle E'(u^{k+1}) - E'(v_i), w_i\rangle + \langle E'(v_i), w_i\rangle\right) \\ &= \mathrm{I}_1 + \mathrm{I}_2,\end{aligned}$$

where

$$\mathrm{I}_1 := \sum_{i=1}^{N} \langle E'(u^{k+1}) - E'(v_i), w_i \rangle \quad \text{and} \quad \mathrm{I}_2 := \sum_{i=1}^{N} \langle E'(v_i), w_i \rangle.$$

Using the stability of the decomposition (SS1) and the strengthened Cauchy Schwartz inequality (SS2), I_1 can be estimated in exactly the same way as in Theorem 3.7. Therefore,

$$\mathrm{I}_1 \leq C_S C_A \left(\sum_{i=1}^{N} \|\varepsilon_i\|_{\mathcal{V}}^2 \right)^{1/2} \|w\|_{\mathcal{V}}.$$

For I_2, we insert $\tau_i - E_i'(\xi_i + s_i)$, which is zero in $\mathcal{V}_i'$, use the Lipschitz continuities, the standard Cauchy Schwartz inequality, to get

$$\begin{aligned}
\mathrm{I}_2 &= \sum_{i=1}^{N} \langle E'(v_i) - E'(v_{i-1}) - E_i'(\xi_i + s_i) + E_i'(\xi_i), w_i \rangle \\
&\leq \sum_{i=1}^{N} \left(L\|\varepsilon_i\|_{\mathcal{V}} + L_i \|s_i\|_{\mathcal{V}} \right) \|w_i\|_{\mathcal{V}} \\
&\leq L \sum_{i=1}^{N} \left(1 + \frac{L_i}{\mu_i} \right) \|\varepsilon_i\|_{\mathcal{V}} \|w_i\|_{\mathcal{V}} \\
&\leq L C_A \left(1 + \max_{1 \leq i \leq N} \frac{L_i}{\mu_i} \right) \left(\sum_{i=1}^{N} \|\varepsilon_i\|_{\mathcal{V}}^2 \right)^{1/2} \|w\|_{\mathcal{V}}.
\end{aligned}$$

In the last estimate, we used the relation $s_i = {\alpha_i^*}^{-1} \varepsilon_i$ and the lower bound of α_i^* given in Lemma 4.9.

Putting the estimates together, we have,

$$\left\| E'(u^{k+1}) \right\|_{\mathcal{V}'}^2 \leq C_A^2 \left[C_S + L \left(1 + \max_{1 \leq i \leq N} \frac{L_i}{\mu_i} \right) \right]^2 \sum_{i=1}^{N} \|\varepsilon_i\|_{\mathcal{V}}^2.$$

Using inequality (8) in Lemma 2.5 with $v = u^{k+1}$, the result follows. □

Remark 4.12. Our theory suggests we can simply choose

$$E_i(w) = \frac{1}{2}\|w - \xi_i\|_{\mathcal{V}}^2 = \frac{1}{2}\|w - Q_i v_{i-1}\|_{\mathcal{V}}^2 \quad \forall\, w \in \mathcal{V}_i; \tag{26}$$

for then (E3) and (E4) hold with $L_i = \mu_i = 1$. Moreover, the local problem becomes like that of the linear preconditioned gradient descent method:

$$(\eta_i - \xi_i, w)_{\mathcal{V}_i} = -\langle R_i E'(v_{i-1}), w \rangle \quad \forall\, w \in \mathcal{V}_i. \tag{27}$$

In this case (23) has the closed form solution

$$\eta_i - \xi_i =: s_i = -\mathfrak{R}_i R_i E'(v_{i-1}),$$

where $\mathfrak{R}_i$ is the Riesz map $\mathcal{V}_i' \to \mathcal{V}_i$ and its realization is the inverse of an SPD matrix of size $\dim \mathcal{V}_i$. In fact, we can even use, for any fixed $g_i \in \mathcal{V}_i$ that we like,

$$E_i(w) = \frac{1}{2}\|w - g_i\|_{\mathcal{V}}^2 \quad \forall\, w \in \mathcal{V}_i,$$

and the same basic result is true (by linearity): $s_i = -\mathfrak{R}_i R_i E'(v_{i-1})$.

In any case, solving a linear local problem can dramatically reduce the computational cost of the FASD. See Section 7 for a practical discussion of this point. In this setting, FASD is closely related to the coordinate descent methods analyzed in [20]. See also, for example, [10]. Another advantage of using (26) is that one does not need to worry about the particular choice of $\mathcal{B}_i$. The quadratic energy in (26) is globally Lipschitz with Lipschitz constant 1.

Remark 4.13. We can also choose the local quadratic energy

$$E_i(w) = \frac{1}{2}\|w-\xi_i\|^2_{E''(\xi_i)} := \frac{1}{2}\langle E''(\xi_i)(w-\xi_i), w-\xi_i\rangle \quad \forall w \in \mathcal{V}_i. \tag{28}$$

Here, recall that $\xi_i = Q_i v_{i-1}$ and $E''(\xi_i)$ should be understood as the restriction of the bilinear form $E''(\xi_i)$ on subspace $\mathcal{V}_i \times \mathcal{V}_i$. Then the local problem becomes one damped Newton's iteration in subspace $\mathcal{V}_i$

$$s_i = -(R_i E''(\xi_i) I_i)^{-1} R_i E'(v_{i-1}).$$

In this setting, the block Newton's method proposed in [18] can be interpreted as a FASD with an appropriate space decomposition. We will investigate the randomized version in a future paper.

Corollary 4.14. *In addition to the hypotheses of the last theorem, let us assume that E_i is quadratic, chosen as in* (26)*. Then,*

$$E(u^{k+1}) - E(u) \leq \frac{C_A^2 (C_S + 2L)^2}{2\mu} \sum_{i=1}^{N} \|\varepsilon_i\|^2_{\mathcal{V}}.$$

4.6. **Convergence.** Using Theorems 4.3 and 4.11, and Corollary 3.4, we obtain the following linear convergence result.

Corollary 4.15. *Let u^k be the kth iteration and $u^{k+1} = \mathrm{FASD}(u^k)$. Suppose that the space decomposition satisfies assumptions* (SS1) *and* (SS2)*, the energy E satisfies assumption* (E1) – (E2)*, and the energy E_i satisfies assumption* (E3) – (E4)*; then we have*

$$E(u^{k+1}) - E(u) \leq \rho(E(u^k) - E(u)),$$

with

$$\rho = \frac{C_A^2 \left[C_S + L\left(1 + \max_i\{L_i/\mu_i\}\right)\right]^2}{C_A^2 \left[C_S + L\left(1 + \max_i\{L_i/\mu_i\}\right)\right]^2 + \mu^2}.$$

Furthermore if E_i is quadratic, chosen as in (26)*, then*

$$\rho = \frac{C_A^2 \left(C_S + 2L\right)^2}{C_A^2 \left(C_S + 2L\right)^2 + \mu^2}.$$

5. Fast subspace descent method with approximate line search

In this section, we consider the FASD algorithm with approximated line search. The method is detailed in Algorithm 3. The key difference between this algorithm and Algorithm 2 is that a fixed step size α_i is employed rather than computing α_i^* via a line search. In this case, there is no need to repeatedly evaluate E and its derivatives in the subspace. We need only compute $R_i E'(v_{i-1})$ once for the local problem (for use in the computation of τ_i and α_i).

In the next section, Section 6, we shall also consider the original FAS, which corresponds to FASD with the step size $\alpha_i = 1$. We prove its convergence based on an additional approximation property.

Algorithm: $u^{k+1} = \text{FASD-ALS}(u^k)$
$v_0 = u^k$;
for $i = 1 : N$ **do**
Compute the subspace τ-*perturbation*: let $\xi_i = Q_i v_{i-1}$ and

$$\tau_i := E_i'(\xi_i) - R_i E'(v_{i-1}) \in \mathcal{V}_i'; \tag{29}$$

Solve the subspace residual problem: Find $\eta_i \in \mathcal{V}_i$, such that

$$\langle E_i'(\eta_i), w\rangle = \langle \tau_i, w\rangle \ \forall w \in \mathcal{V}_i. \tag{30}$$

Compute the search direction and the quadratic step size:

$$s_i := \eta_i - \xi_i \in \mathcal{V}_i, \tag{31}$$

$$\alpha_i^q := -\frac{\langle R_i E'(v_{i-1}), s_i\rangle}{L\|s_i\|_{\mathcal{V}}^2}. \tag{32}$$

Apply the subspace correction:

$$v_i := v_{i-1} + \alpha_i^q s_i. \tag{33}$$

end
$u^{k+1} := v_N$

Algorithm 3: FASD algorithm with approximate line search (ALS).

Recall the scalar function $f_i(\alpha) := E(v_{i-1} + \alpha s_i)$, with $f_i(0) = E(v_{i-1})$, $f_i'(0) = \langle E'(v_{i-1}), s_i\rangle < 0$. Using $f_i(0)$ and $f_i'(0)$, we define the quadratic function

$$q_i(\alpha) := f_i(0) + f_i'(0)\alpha + \frac{L\|s_i\|_{\mathcal{V}}^2}{2}\alpha^2. \tag{34}$$

See Figure 3. The optimal step size for FASD is, of course, $\alpha_i^* = \text{argmin}_{\alpha\in\mathbb{R}} f_i(\alpha)$. Our choice for this new algorithm is

$$\alpha_i^q = \text{argmin}_{\alpha\in\mathbb{R}}\, q_i(\alpha) = -\frac{f_i'(0)}{L\|s_i\|_{\mathcal{V}}^2} = -\frac{\langle R_i E'(v_{i-1}), s_i\rangle}{L\|s_i\|_{\mathcal{V}}^2},$$

which satisfies the following estimate.

Lemma 5.1. *Assume the energy E satisfies the Lipschitz continuity assumption* (E2) *and the local energy E_i satisfies the strong convexity assumptions* (E3)*; then*

$$\frac{\mu_i}{L} \le \alpha_i^q \le \alpha_i^*.$$

Proof. The lower bound is obtained by the definition of α_i^q and Lemma 4.4. To prove the upper bound, we notice that, due to line search,

$$f_i'(\alpha_i^*) = \langle E'(v_{i-1} + \alpha_i^* s_i), s_i\rangle = 0$$

and thus

$$\alpha_i^q L\|s_i\|_{\mathcal{V}}^2 = -\langle R_i E'(v_{i-1}), s_i\rangle = \langle E'(v_{i-1} + \alpha_i^* s_i) - E'(v_{i-1}), s_i\rangle \le \alpha_i^* L\|s_i\|_{\mathcal{V}}^2.$$

□

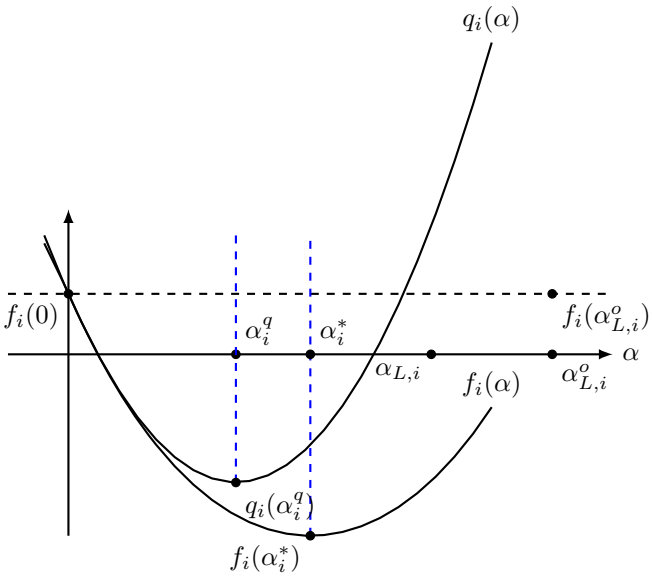

FIGURE 3. The functions f_i, defined in (25), and its quadratic approximation q_i, defined in (34). The quadratic minimizer α_i^q is always to the left of α_i^* by construction.

Now, since the optimal linear search procedure is broken, the orthogonality conditions with respect to the corrections are broken, and establishing the lower bound is a little more complicated.

Theorem 5.2. *Let u^k be the kth iteration and $u^{k+1} = \text{FASD-ALS}(u^k)$. Suppose that E satisfies assumption* (E1) – (E2) *and the local energy E_i is strongly convex, satisfying assumption* (E3)*. Then, we have*

$$E(u^k) - E(u^{k+1}) \geq C_L \sum_{i=1}^{N} \|\alpha_i^q s_i\|_{\mathcal{V}}^2, \quad C_L = \frac{L}{2}.$$

Proof. It suffices to prove

$$E(v_{i-1}) - E(v_i) = f(0) - f(\alpha_i^q) \geq \frac{L}{2}\|\alpha_i^q s_i\|^2.$$

By Lemma 4.8, for $\alpha \in [0, \alpha_{L,i}]$, f_i' is Lipschitz continuous with constant $L\|s_i\|_{\mathcal{V}}^2$. Then for $\alpha \in [0, \alpha_{L,i}]$,

$$f_i(\alpha) - q_i(\alpha) = f_i(\alpha) - f_i(0) - \alpha f_i'(0) - \frac{L\|s_i\|_{\mathcal{V}}^2}{2}\alpha^2 \leq 0.$$

Namely $f_i(\alpha) \leq q_i(\alpha)$ for all $\alpha \in [0, \alpha_{L,i}]$. As $\alpha_i^q = \operatorname{argmin}_{\alpha\in\mathbb{R}} q_i(\alpha)$, and $\alpha_i^q \leq \alpha_i^*$, we get

$$f_i(\alpha_i^q) \leq q_i(\alpha_i^q) = \min_{\alpha\in\mathbb{R}} q_i(\alpha) = f_i(0) - \frac{1}{2L\|s_i\|_{\mathcal{V}}^2}|f_i'(0)|^2 = f_i(0) - \frac{L}{2}\|\alpha_i^q s_i\|_{\mathcal{V}}^2.$$

In the last step, we have used the definition of α_i^q and this completes the proof. □

Since α_i^q has the same lower bound as α_i^*, we can derive the upper bound in exactly the same way as the proof of Theorem 4.11, only replacing $\varepsilon_i = \alpha_i^* s_i$ by $\alpha_i^q s_i$ and using the lower bound from Lemma 5.1. Thus, we only state the theorem below, and the proof is omitted.

Theorem 5.3. *Let u^k be the kth iteration and $u^{k+1} = \text{FASD-ALS}(u^k)$. Suppose the space decomposition satisfies* (SS1) *and* (SS2)*, the energy E satisfies* (E1) – (E2)*, and E_i satisfies* (E3) – (E4)*. Then we have the upper bound*

$$E(u^{k+1}) - E(u) \leq C_U \sum_{i=1}^{N} \|\alpha_i^q s_i\|_{\mathcal{V}}^2,$$

where $C_U := C_A^2 \left[C_S + L\left(1 + \max_i\{L_i/\mu_i\}\right)\right]^2 / (2\mu)$.

We summarize the linear convergence result below.

Corollary 5.4. *Let u^k be the kth iteration and $u^{k+1} = \text{FASD-ALS}(u^k)$. Suppose that the space decomposition satisfies assumptions* (SS1) *and* (SS2)*, the energy E satisfies assumption* (E1) – (E2)*, and the energy E_i satisfies assumption* (E3) – (E4)*; then we have*

$$E(u^{k+1}) - E(u) \leq \rho(E(u^k) - E(u)),$$

with

$$\rho = \frac{C_A^2 \left[C_S + L\left(1 + \max_i\{L_i/\mu_i\}\right)\right]^2}{C_A^2 \left[C_S + L\left(1 + \max_i\{L_i/\mu_i\}\right)\right]^2 + L\mu}.$$

The Lipschitz constant L is used in the step size α_i which can be replaced by a local Lipschitz constant for the scalar function $f_i(\alpha)$ for $\alpha \in (0, \alpha_i^*)$ and popular line search algorithms can be used.

Remark 5.5. Consider a special case that $\mathcal{V} := \mathbb{R}^n$ with an orthogonal decomposition $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \cdots \oplus \mathcal{V}_N$, $\mathcal{V}_i \subset \mathcal{V}$. If we simply choose $E_i(w) = \frac{1}{2}\|w - \xi_i\|^2$ $\forall\ w \in \mathcal{V}_i$, where $\|\cdot\|$ denotes the standard ℓ^2-norm induced by the standard ℓ^2-inner product $(\cdot,\cdot)$ defined on $\mathbb{R}^n$, then we have $s_i = -R_i E'(v_{i-1})$ and the FASD algorithm (Algorithm 3) becomes the block coordinated descent method discussed in [2, 20]. Therefore, Corollary 5.4 gives a convergence analysis of the cyclic variant of the block coordinate descent method. To the best of our knowledge, the only convergence results for the cyclic block coordinated descent method was presented in [2]. Here, we give a linear convergence result from a subspace decomposition point of view for the strongly convex case and our result can be generalized to other related methods as well, for example, the preconditioned block coordinated descent method.

6. Original FAS method: FASD without line search

Notice that the original FAS does not have the last line search step. Traditional FAS, listed as Algorithm 4, applies the subspace correction via

$$v_i := v_{i-1} + \alpha_i^{\text{FAS}} s_i, \quad \alpha_i^{\text{FAS}} = 1.$$

Algorithm: $u^{k+1} = \mathrm{FAS}(u^k)$

$v_0 = u^k$;

for $i = 1 : N$ **do**

Compute the subspace τ-*perturbation*: let $\xi_i = Q_i v_{i-1}$ and

$$\tau_i := E_i'(\xi_i) - R_i E'(v_{i-1}) \in \mathcal{V}_i'; \tag{35}$$

Solve the subspace residual problem: Find $\eta_i \in \mathcal{V}_i$, such that

$$\langle E_i'(\eta_i), w\rangle = \langle \tau_i, w\rangle \ \forall w \in \mathcal{V}_i; \tag{36}$$

Compute the correction

$$s_i := \eta_i - \xi_i \in \mathcal{V}_i; \tag{37}$$

Apply the subspace correction:

$$v_i := v_{i-1} + s_i. \tag{38}$$

end

$u^{k+1} := v_N$

Algorithm 4: Traditional FAS: FASD with no line search.

Previously, our choice of step size was motivated by the choice of step size in the gradient descent method [21]. We shall prove $\alpha_i^{\mathrm{FAS}} = 1$ is also allowed – that is to say, it leads to a convergent algorithm – provided that the following approximation property is satisfied.

(AP) Both E and E_i are twice Fréchet differentiable. Furthermore, there exists a constant $0 < \epsilon < \mu/2$ so that for all $w \in \mathcal{B}, \eta_i \in \mathcal{V}_i$ and all $u_i, v_i \in \mathcal{V}_i$

$$|\langle E''(w+\eta_i)u_i, v_i\rangle - \langle E_i''(Q_i w + \eta_i)u_i, v_i\rangle| \le \epsilon \|u_i\|_{\mathcal{V}} \|v_i\|_{\mathcal{V}}.$$

For quadratic energy, $R_i E'' I_i$ is the coarse matrix on $\mathcal{V}_i$ formed by the triple product, via the so-called Galerkin method, and E_i'' is the matrix obtained using the bilinear form associated to the local energy E_i. They should be close in a certain norm.

The original FAS is to choose $E_i = E|_{\mathcal{V}_i}$ so that is $E_i'' = E''$ on $\mathcal{V}_i \times \mathcal{V}_i$. Assume furthermore E'' is also Lipschitz continuous. Then (AP) can be verified

$$\|E''(w+\eta_i) - E''(Q_i w + \eta_i)\| \le C\|w - Q_i w\|. \tag{39}$$

Note that in this case the local problem $E'(\eta_i) = \tau_i$ is cheaper than solving the Euler equation $E'(v_{i-1} + e_i) = 0$ in SSO.

Lemma 6.1. *Assume the energy E satisfies the assumptions* (E1) *and* (E2)*, and the approximation assumption* (AP) *holds. Then, E_i' satisfies the Lipschitz condition and strongly convexity condition as follows:*

$$(\mu - \epsilon)\|v - w\|_{\mathcal{V}}^2 \le \langle E_i'(v) - E_i'(w), v - w\rangle \le (L + \epsilon)\|v - w\|_{\mathcal{V}}^2 \quad \forall\ v, w \in \mathcal{B} \cap \mathcal{V}_i.$$

Proof. For any $v, w \in \mathcal{V}_i$, by Taylor's theorem with integral reminder, we have

$$\begin{aligned}
&\langle E_i'(v) - E_i'(w), v - w\rangle \\
&\quad = \int_0^1 \langle E_i''(z(t))(v-w), v-w\rangle \mathrm{d}t \\
&\quad = \int_0^1 \langle E''(z(t))(v-w), v-w\rangle \mathrm{d}t \\
&\qquad + \int_0^1 \langle E_i''(z(t))(v-w), v-w\rangle - \langle E''(z(t))(v-w), v-w\rangle \mathrm{d}t \\
&\quad = \langle E'(v) - E'(w), v - w\rangle \\
&\qquad + \int_0^1 \langle E_i''(z(t))(v-w), v-w\rangle - \langle E''(z(t))(v-w), v-w\rangle \mathrm{d}t,
\end{aligned}$$

where $z(t) = tv + (1-t)w \in \mathcal{V}_i$, and thus $Q_i z = z$. When $v, w \in \mathcal{B} \cap \mathcal{V}_i$, using assumptions (E2) and (AP), we have

$$\langle E_i'(v) - E_i'(w), v - w\rangle \leq L\|v-w\|_{\mathcal{V}}^2 + \epsilon\|v-w\|_{\mathcal{V}}^2 = (L+\epsilon)\|v-w\|_{\mathcal{V}}^2.$$

On the other hand, when $v, w \in \mathcal{B} \cap \mathcal{V}_i$, using assumptions (E1) and (AP), we have

$$\langle E_i'(v) - E_i'(w), v - w\rangle \geq \mu\|v-w\|_{\mathcal{V}}^2 - \epsilon\|v-w\|_{\mathcal{V}}^2 = (\mu-\epsilon)\|v-w\|_{\mathcal{V}}^2.$$

□

Theorem 6.2. *Let u^k be the kth iteration and $u^{k+1} = \mathrm{FAS}(u^k)$, as in Algorithm 4, with local step size $\alpha_i^{\mathrm{FAS}} = 1$. Suppose that E satisfies assumption* (E1) *and the approximation assumption* (AP) *holds with $\epsilon < \mu/2$. Then, we have*

$$E(u^k) - E(u^{k+1}) \geq C_L \sum_{i=1}^N \|s_i\|_{\mathcal{V}}^2, \quad C_L = \left(\frac{\mu}{2} - \epsilon\right).$$

Proof. Recall that $\xi_i = Q_i v_{i-1}$ and $Q_i s_i = s_i$. Using equation (36) and Taylor's theorem with integral remainder, we first estimate $|\langle E'(v_i), s_i\rangle|$ by

$$\begin{aligned}
|\langle E'(v_i), s_i\rangle| &= |\langle E'(v_{i-1}+s_i), s_i\rangle - \langle E'(v_{i-1}), s_i\rangle - [\langle E_i'(\xi_i + s_i), s_i\rangle - \langle E_i'(\xi_i), s_i\rangle]| \\
&= \left| \int_0^1 \langle E''(y(t))\ s_i, s_i\rangle - \langle E_i''(Q_i y(t))\ s_i, s_i\rangle \,\mathrm{d}t \right| \\
&\leq \int_0^1 |\langle E''(y(t))\ s_i, s_i\rangle - \langle E_i''(Q_i y(t))\ s_i, s_i\rangle| \,\mathrm{d}t \\
&\leq \epsilon \|s_i\|_{\mathcal{V}}^2,
\end{aligned}$$

where $y(t) := (1-t)v_{i-1} + t(v_{i-1}+s_i) = v_{i-1} + ts_i$. Note that $v_{i-1} \in \mathcal{B}$ and $s_i \in \mathcal{V}_i$ which allows us to use assumption (AP) in the last step.

Using assumption (E1) – specifically estimate (4) of Theorem 2.1 – we get

$$E(v_{i-1}) - E(v_{i-1}+s_i) \geq -\langle E'(v_{i-1}+s_i), s_i\rangle + \frac{\mu}{2}\|s_i\|_{\mathcal{V}}^2 \geq \left(\frac{\mu}{2} - \epsilon\right)\|s_i\|_{\mathcal{V}}^2. \tag{40}$$

□

The upper bound for FAS (where $\alpha_i^{\mathrm{FAS}} = 1$) is easy, as there is now no need to have a lower bound of the step size.

Theorem 6.3. *Let u^k be the kth iteration and $u^{k+1} = \mathrm{FAS}(u^k)$ with local step size $\alpha_i = 1$. Suppose the space decomposition satisfies* (SS1) *and* (SS2)*, the energy E satisfies* (E1) – (E2)*, and assumption* (AP) *holds. Then we have the upper bound*

$$E(u^{k+1}) - E(u) \le C_U \sum_{i=1}^{N} \|s_i\|_{\mathcal{V}}^2,$$

where $C_U = C_A^2(C_S + \epsilon)^2/(2\mu)$.

Proof. For any $w \in \mathcal{V}$, we choose a stable decomposition $w = \sum_{i=1}^N w_i$; then

$$\begin{aligned}\langle E'(u^{k+1}), w\rangle &= \sum_{i=1}^{N} \langle E'(u^{k+1}), w_i\rangle \\ &= \sum_{i=1}^{N} \langle E'(u^{k+1}) - E'(v_i), w_i\rangle + \sum_{i=1}^{N} \langle E'(v_i), w_i\rangle \\ &= \mathrm{I}_1 + \mathrm{I}_2.\end{aligned}$$

The first term is bounded as before. Therefore,

$$\mathrm{I}_1 \le C_S C_A \left(\sum_{i=1}^{N} \|s_i\|_{\mathcal{V}}^2\right)^{1/2} \|w\|_{\mathcal{V}}.$$

For the second term, we insert $\tau_i - E_i'(\xi_i + s_i) = -E'(v_{i-1}) + E_i'(\xi_i) - E_i'(\xi_i + s_i)$, which is zero in $\mathcal{V}_i'$, and use Taylor's theorem with integral remainder, followed by assumption (AP), to get

$$\begin{aligned}\mathrm{I}_2 &= \sum_{i=1}^{N} \langle E'(v_i) - E'(v_{i-1}) - \left[E_i'(\xi_i + s_i) - E_i'(\xi_i)\right], w_i\rangle \\ &\le \epsilon \sum_{i=1}^{N} \|s_i\|_{\mathcal{V}} \|w_i\|_{\mathcal{V}} \\ &\le \epsilon C_A \left(\sum_{i=1}^{N} \|s_i\|_{\mathcal{V}}^2\right)^{1/2} \|w\|_{\mathcal{V}}.\end{aligned}$$

□

Corollary 6.4. *Let u^k be the kth iteration and $u^{k+1} = \mathrm{FAS}(u^k)$. Suppose that the space decomposition satisfies assumptions* (SS1) *and* (SS2)*, the energy E satisfies assumption* (E1) – (E2)*, and the energy E_i satisfies assumption* (AP) *with $\epsilon < \mu/2$; then we have*

$$E(u^{k+1}) - E(u) \le \rho(E(u^k) - E(u)),$$

with

$$\rho = \frac{(C_S + \epsilon)^2 C_A^2}{(C_S + \epsilon)^2 C_A^2 + \mu(\mu - 2\epsilon)}.$$

7. Application and numerical experiments

In this section we shall apply our theory to a model nonlinear problem with polynomial nonlinearity and provide numerical examples to illustrate the efficiency of a variant of FAS (Algorithm 4) with a local quadratic energy.

7.1. **A model nonlinear problem.** Suppose that $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a star-shaped polytope, i.e., a polygon in 2-D or a polyhedron in 3-D. Suppose that $2 \leq p < \infty$, when $d = 2$, and $2 \leq p \leq 6$, when $d = 3$. We consider the following problem: given $f \in L^2(\Omega)$, find $u \in H_0^1(\Omega)$ such that

$$\left(|u|^{p-2}u, \xi\right) + \varepsilon^2 \left(\nabla u, \nabla \xi\right) = (f, \xi) \quad \forall\, \xi \in H_0^1(\Omega), \tag{41}$$

where $\varepsilon > 0$ is parameter. One can show that the unique solution of (41) is the unique minimizer of a certain strictly convex energy.

Theorem 7.1. *Suppose that $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a star-shaped polytope, i.e., a polygon in* 2-D *or a polyhedron in* 3-D. *Suppose that $2 \leq p < \infty$, when $d = 2$, and $2 \leq p \leq 6$, when $d = 3$. For any $\nu \in H_0^1(\Omega)$, define the energy*

$$E(\nu) := \frac{1}{p} \|\nu\|_{L^p}^p + \frac{\varepsilon^2}{2} \|\nabla \nu\|^2 - (f, \nu)\,. \tag{42}$$

The energy functional E defined in (42) *is twice Fréchet differentiable and satisfies assumptions* (E1) *and* (E2) *with respect to the space $\mathcal{V} = H_0^1(\Omega)$, equipped with the norm $\|\nabla v\|$ for $v \in \mathcal{V}$. Therefore E has a unique global minimizer in $H_0^1(\Omega)$. Furthermore, $u \in H_0^1(\Omega)$ is the unique minimizer of* (42) *iff it is the solution of* (41).

Proof. We verify that E satisfies our assumptions. The first Fréchet derivative of E at a point ν may be calculated as follows: for any $\xi \in H_0^1(\Omega)$,

$$\frac{\mathrm{d}}{\mathrm{d}t} E(\nu + t\xi)\bigg|_{t=0} = \langle E'(\nu), \xi \rangle = \left(|\nu|^{p-2}\nu, \xi\right) + \varepsilon^2 \left(\nabla \nu, \nabla \xi\right) - (f, \xi)\,.$$

The second Fréchet derivative exists for $p \geq 2$ and is a continuous bilinear operator. Given a fixed $\nu \in H_0^1(\Omega)$, the action of the second variation on the arbitrary pair $(\xi, \eta) \in H_0^1(\Omega) \times H_0^1(\Omega)$ is given by

$$\langle E''(\nu)\xi, \eta \rangle = (p-1) \left(|\nu|^{p-2}\xi, \eta\right) + \varepsilon^2 \left(\nabla \xi, \nabla \eta\right).$$

Without loss of generality, we choose $u_0 = 0$, so that $E(u_0) = 0$. Recall that $\mathcal{B} = \{v \in \mathcal{V} \mid E(v) \leq E(u_0)\}$. Observe that $\mathcal{B}$ is convex, since E is convex. For $v \in \mathcal{B}$, $E(v) \leq 0$, and we have

$$\frac{1}{p} \|v\|_{L^p}^p + \frac{\varepsilon^2}{2} \|\nabla v\|^2 \leq (f, v) \leq \|f\| \|v\| \leq C_0(\varepsilon, C_{\mathrm{P}}) \|f\|^2 + \frac{\varepsilon^2}{4} \|\nabla v\|^2,$$

where $C_{\mathrm{P}} = C_{\mathrm{P}}(\Omega) > 0$ is the constant in the Poincare inequality:

$$\|v\| \leq C_{\mathrm{P}}(\Omega)\, \|\nabla v\| \quad \forall\, v \in H_0^1(\Omega).$$

Thus, for $v \in \mathcal{B}$, the follow norms are bounded:

$$\|v\|_{L^p} + \|\nabla v\| \leq C_1 = C_1(u_0, \varepsilon, p, f). \tag{43}$$

By the mean value theorem, there exists a $z = tv + (1-t)w$, for some $t \in [0, 1]$, such that

$$\langle E'(w), \xi \rangle - \langle E'(v), \xi \rangle = \langle E''(z)\xi, w - v \rangle \quad \forall\, \xi \in H_0^1(\Omega).$$

If $w, v \in \mathcal{B}$, then, since $\mathcal{B}$ is convex, $z \in \mathcal{B}$. By (43) $\|z\|_{L^p} \leq C_1$. Using Hölder's inequality, we have

$$\begin{aligned} |\langle E''(\nu)\xi, \eta \rangle| &\leq (p-1) \|\nu\|_{L^p}^{p-2} \|\xi\|_{L^p} \|\eta\|_{L^p} + \varepsilon^2 \|\nabla \xi\| \cdot \|\nabla \eta\| \\ &\leq \left[(p-1) C_{\mathrm{P}}^2 \|\nu\|_{L^p}^{p-2} + \varepsilon^2\right] \|\nabla \xi\| \cdot \|\nabla \eta\|\,. \end{aligned}$$

Therefore,

$$\begin{aligned}\left|\langle E'(w),\xi\rangle - \langle E'(v),\xi\rangle\right| &= \left|\langle E''(z)\xi, w-v\rangle\right| \\ &\le \left[(p-1)C_{\mathrm{P}}^2 C_1^{p-2} + \varepsilon^2\right] \|\nabla\xi\| \cdot \|\nabla(w-v)\|.\end{aligned}$$

Namely (E2) holds with $L := (p-1)C_{\mathrm{P}}^2 C_1^{p-2} + \varepsilon^2$.

To see that E is uniformly elliptic, for any $w, v \in \mathcal{V}$, there is an $\eta \in \mathcal{V}$,

$$\begin{aligned}\langle E'(w) - E'(v), w-v\rangle &= \langle E''(z)(w-v), w-v\rangle, \\ &= (p-1)\left(|\nu|^{p-2}(w-v), w-v\right) + \varepsilon^2\left(\nabla(w-v), \nabla(w-v)\right) \\ &\ge \varepsilon^2 \|\nabla(w-v)\|^2.\end{aligned}$$

(E1) holds with $\mu = \varepsilon^2$.

It follows that there is a unique global minimizer of the energy (42):

$$u := \mathrm{argmin}_{\nu \in H_0^1(\Omega)} E(\nu).$$

Consequently, there is a unique solution to the Euler problem which is equation (41). □

Now, suppose that $\Omega \subset \mathbb{R}^2$ is a polygonal domain and $\mathcal{T}_H$ is a conforming triangulation of Ω. Let $\mathcal{T}_h$ be the triangulation obtained by quadri-secting $\mathcal{T}_H$. Specifically, if $K_i \in \mathcal{T}_h$ is one of the four daughter triangles ($i = 1, \cdots, 4$) obtained by quadri-secting $K \in \mathcal{T}_H$ – that is by connecting the midpoints of K – then $h_{K_i} = H_K/2$, $i = 1, \cdots, 4$. A family of meshes constructed in this way is known to be globally quasi-uniform.

Define

$$S_h := \left\{v \in C(\Omega) \cap H_0^1(\Omega) \middle| \, v|_K \in \mathcal{P}_1(K) \ \forall K \in \mathcal{T}_h\right\}$$

with a similar definition for S_H. Then, $S_H \subset S_h$, and the containment is proper.

We shall consider the minimization of energy E restricted to S_h which is a subspace of $H_0^1(\Omega)$

$$\min_{v \in S_h} E(v),$$

and thus now $\mathcal{V} = S_h$ with norm $|v|_1 = \|\nabla v\|$. Notice that (E1) and (E2) still hold, as $S_h \subset H_0^1(\Omega)$.

Next we give a two-level space decomposition of $\mathcal{V}$ as follows. Let $\mathcal{N} = \{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^2$ be the set of *interior* nodes of $\mathcal{T}_h$ and define the Lagrange nodal basis

$$B_h = \{\psi_i \in S_h,\ 1 \le i \le N \mid \psi_i(\mathbf{x}_j) = \delta_{i,j},\ 1 \le i, j \le N\}.$$

B_h is a *bona fide* basis for S_h, and we may use the following decomposition:

$$\mathcal{V} = \sum_{i=0}^{N} \mathcal{V}_i = S_h, \tag{44}$$

where $\mathcal{V}_0 = S_H$, $\mathcal{V}_i = \mathrm{span}(\{\psi_i\})$, $1 \le i \le N$. (Note that we give the coarse space the index 0.)

The fact that this forms a stable decomposition is well known, i.e., assumption (SS1) holds.

Lemma 7.2. *The decomposition of the finite element space S_h described in* (44) *satisfies assumption* (SS1).

Proof. Let $Q_H : L^2(\Omega) \to S_H$ be the L^2-projection into S_H:

$$(Q_H v, w) = (v, w) \quad \forall\ w \in S_H.$$

For any $v \in S_h$, let $\tilde{v} = (I - Q_H)v \in S_h$ denote the error, and suppose that $\tilde{v} = \sum_{i=1}^N \tilde{v}_i$ is the nodal decomposition of the error in S_h. By the standard approximation property of Q_H on quasi-uniform grids, an inverse inequality, and the stability of nodal decompositions in the L^2-norm, we have

$$\sum_{i=1}^N |\tilde{v}_i|_1^2 \leq C \sum_{i=1}^N h^{-2} \|\tilde{v}_i\|^2 \leq C h^{-2} \|\tilde{v}\|^2 \leq C |v|_1^2.$$

By the H^1-stability of Q_H on quasi-uniform grids, we also have $|Q_H v|_1 \lesssim |v|_1$. In conclusion, assumption (SS1) holds if, for $v \in S_h$, we use the decomposition

$$v = Q_H v + (v - Q_H v) = Q_H v + \sum_{i=1}^N \tilde{v}_i.$$

□

Lemma 7.3. *Let E be defined as in* (42)*, and let $\mathcal{V} = S_h$ be decomposed into subspaces as in* (44)*. Then assumption* (SS2) *holds.*

Proof. Suppose that $w_{i,j} \in \mathcal{B}$, $u_i \in \mathcal{V}_i$, $v_j \in \mathcal{V}_j$, with $w_{i,j} + u_i \in \mathcal{B}$. By Taylor's theorem,

$$\begin{aligned}
\sum_{i=0}^N \sum_{j=i+1}^N \langle E'(w_{i,j} + u_i) - E'(w_{i,j}), v_j \rangle & \\
&= \sum_{i=0}^N \sum_{j=i+1}^N \langle E''(z_{i,j}) v_j, u_i \rangle \\
&\leq \sum_{i=0}^N \sum_{j=i+1}^N \left| (p-1) \left(|z_{i,j}|^{p-2} u_i, v_j \right) + \varepsilon^2 \left(\nabla u_i, \nabla v_j \right) \right|,
\end{aligned}$$

for some $z_{i,j} \in \mathcal{B}$ between $w_{i,j} \in \mathcal{B}$ and $w_{i,j} + u_i \in \mathcal{B}$, which satisfies the bound (43). The functions u_i, $1 \leq i, j \leq N$, are local, though u_0 may have global support. The support of v_i, $1 \leq i \leq N$, denoted S_i, is exactly equal to the union of those triangles that have the node $\mathbf{x}_i$ as a vertex. Define

$$\mathcal{N}(i) := \{ j > i \mid S_j \cap S_i \neq \emptyset \}.$$

Observe that $\#(\mathcal{N}(i))$ is bounded by an integer that is much smaller than N. We have, using the continuous and discrete Cauchy Schwartz inequalities,

$$\begin{aligned}
\sum_{i=0}^{N}\sum_{j=i+1}^{N}(\nabla u_i,\nabla v_j) &= \sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}(\nabla u_i,\nabla v_j)_{S_i\cap S_j}\\
&\le \sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\|\nabla u_i\|_{S_i\cap S_j}\|\nabla v_j\|_{S_i\cap S_j}\\
&\le \left(\sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\|\nabla u_i\|^2_{S_i\cap S_j}\right)^{\frac12}\left(\sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\|\nabla v_j\|^2_{S_i\cap S_j}\right)^{\frac12}\\
&\le \left(C_{\mathcal{T}}\sum_{i=0}^{N}\|\nabla u_i\|^2\right)^{\frac12}\left(C_{\mathcal{T}}\sum_{j=0}^{N}\|\nabla v_j\|^2\right)^{\frac12},
\end{aligned}$$

where $C_{\mathcal{T}}>0$ is a mesh-structure-dependent parameter. Since our mesh is shape regular and quasi-uniform, $C_{\mathcal{T}}$ is independent of N and h.

Similarly,

$$\begin{aligned}
\sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\left(|z_{i,j}|^{p-2}u_i,v_j\right) &= \sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\left(|z_{i,j}|^{p-2}u_i,v_j\right)_{S_i\cap S_j}\\
&\le \sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\|z_{i,j}\|^{p-2}_{L^p(S_i\cap S_j)}\|u_i\|_{L^p(S_i\cap S_j)}\|v_j\|_{L^p(S_i\cap S_j)}\\
&\le \sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\|z_{i,j}\|^{p-2}_{L^p(\Omega)}\|u_i\|_{L^p(S_i\cap S_j)}\|v_j\|_{L^p(S_i\cap S_j)}\\
&\le \sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}C_1^{p-2}\|u_i\|_{L^p(S_i\cap S_j)}\|v_j\|_{L^p(S_i\cap S_j)}\\
&\le C_1^{p-2}\left(\sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\|u_i\|^2_{L^p(S_i\cap S_j)}\right)^{\frac12}\\
&\quad\times\left(\sum_{i=0}^{N}\sum_{j\in\mathcal{N}(i)}\|v_j\|^2_{L^p(S_i\cap S_j)}\right)^{\frac12}\\
&\le C_1^{p-2}\left(C_{\mathcal{T}}\sum_{i=0}^{N}\|u_i\|^2_{L^p}\right)^{\frac12}\left(C_{\mathcal{T}}\sum_{i=0}^{N}\|v_i\|^2_{L^p}\right)^{\frac12}\\
&\le C_1^{p-2}C_{\mathcal{T}}\left(\sum_{i=0}^{N}C_{\mathrm{P}}^2\|\nabla u_i\|^2\right)^{\frac12}\left(\sum_{i=0}^{N}C_{\mathrm{P}}^2\|\nabla v_i\|^2_{L^p}\right)^{\frac12}\\
&= C_1^{p-2}C_{\mathcal{T}}C_{\mathrm{P}}^2\left(\sum_{i=0}^{N}\|\nabla u_i\|^2\right)^{\frac12}\left(\sum_{i=0}^{N}\|\nabla v_i\|^2_{L^p}\right)^{\frac12}.
\end{aligned}$$

Therefore, there is a $C_S > 0$ such that

$$\sum_{i=0}^{N}\sum_{j=i+1}^{N}\langle E'(w_{i,j}+u_i)-E'(w_{i,j}),v_j\rangle \le C_S\left(\sum_{i=0}^{N}\|\nabla u_i\|^2\right)^{\frac{1}{2}}\left(\sum_{j=0}^{N}\|\nabla v_j\|^2\right)^{\frac{1}{2}}.$$

In particular, $C_S := LC_{\mathcal{T}}$. Assumption (SS2) holds. □

For our FAS-like algorithm, we apply SSO (nonlinear Gauss-Seidel) to each subspace $\mathcal{V}_i$ on the fine level, which is equivalent to, according to Remark 4.2, using $E_i(\eta) := E(v_{i-1} - Q_i v_{i-1} + \eta)$. On the coarse space, we use $E_H = E|_{\mathcal{V}_H}$. The assumption (AP) can be verified by (39) and a standard approximation property of the projection Q_H, i.e., $\|w - Q_H w\| \le CH\|w\|_1$. We then have $\epsilon = CH$ for this case and, therefore, the condition $\epsilon < \mu/2 = \varepsilon^2/2$ in Theorem 6.2 holds when H is small enough. For finite element functions $w \in V_h$, when near the minimizer, we could expect $w \in H^{3/2-\delta}$ for any $0 < \delta \ll 1$ and thus a higher-order approximation $\|w - Q_H w\| \le CH^{3/2-\delta}\|w\|_{3/2-\delta}$ may hold.

7.2. **Numerical examples.** In this subsection, we present some numerical results for the nonlinear problems described in the previous two subsections to illustrate our theoretical results. For both problems, we will use piecewise linear finite elements to define S_h, and we use different versions of FAS to solve the discretized nonlinear equations. Our algorithms are implemented in MATLAB based on the software package *i*FEM [6]. The numerical experiments are conducted on a System76 Galago with an Intel Core i7-8550U CPU and 32GB RAM.

We mainly focus on three different implementations of FAS (Algorithm 4), based on different choices of space decomposition and local energy. The geometric multigrid setting is considered here, i.e., we have a set of uniformly refined meshes and nested linear finite element spaces $\mathcal{V}^1 \subset \mathcal{V}^2 \subset \cdots \subset \mathcal{V}^J$, where $\mathcal{V}^\ell = \operatorname{span}\{\phi_1^\ell, \phi_2^\ell, \cdots, \phi_{N_\ell}^\ell\}$, with ϕ_i^ℓ being the ith nodal linear finite element basis element on level ℓ.

(1) The first implementation is the original FAS. We consider standard multilevel nodal-based space decomposition $\mathcal{V} = \sum_{\ell=1}^{J}\sum_{i=1}^{N_\ell}\operatorname{span}\{\phi_i^\ell\}$ and the local energy E_i is defined as the restriction of E on the subspace $\operatorname{span}\{\phi_i^\ell\}$. Newton's method is used to solve the local nonlinear problem and we set the tolerance to be 10^{-10} and at most 100 iterations are allowed (in general, less than 5 iterations are needed for solving the local problems in all of our numerical tests). We use a small tolerance to make sure each local problem is solved exactly in order to be consistent with our theoretical analysis.

(2) The second implementation is a simplified version of FAS based on Remark 4.12 and we refer to it as "FASq1". We again consider the multilevel nodal-based space decomposition $\mathcal{V} = \sum_{\ell=1}^{J}\sum_{i=1}^{N_\ell}\operatorname{span}\{\phi_i^\ell\}$ but quadratic energy E_i defined as in (26) is used, which requires that we solve a linear system for each local correction. In fact, since nodal-based space decomposition is used here, we solve a scalar linear equation on each subspace.

(3) The third implementation is a further simplified version and we refer it as "FASq2". In this case, we use space decomposition $\mathcal{V} = \sum_{\ell=1}^{J}\mathcal{V}^\ell$ and consider quadratic energy (26). As mentioned in Remark 4.12, this involves the Riesz map which can be computed by inverting an SPD matrix defined on $\mathcal{V}^\ell$. For our example, this is equivalent to solving a discrete Laplacian

matrix on each level, which is still expensive. Therefore, we solve the discrete Laplacian matrix approximately by just applying one step of the symmetric Gauss-Seidel (SGS) method. This is because we use multilevel space decomposition here and the SGS method is usually used as a smoother in multigrid methods for solving discrete Laplacian matrix. Of course, other types of iterative methods can also be used here, such as Richardson's method or Jacobi's method. For the sake of simplicity, we only consider SGS method here.

In all of our numerical experiments, we use Newton's method to solve the nonlinear problem on the coarsest level. We use 10^{-10} as the tolerance and the maximal number of iterations is 100, which means that the coarse problem is solved exactly. Moreover, we use $\alpha_i = 1$ in the tests to make sure our implementation is simple and practical. The overall stopping criterion of FAS is 10^{-10}.

TABLE 1. Numerical results of FAS (varying p and ε, fix $h = 1/64$)

FAS	$\varepsilon^2 = 1$	$\varepsilon^2 = 1/2$	$\varepsilon^2 = 1/4$	$\varepsilon^2 = 1/8$	$\varepsilon^2 = 10^{-1}$	$\varepsilon^2 = 10^{-2}$	$\varepsilon^2 = 10^{-3}$
$p = 4$	15 (0.195)	15 (0.193)	14 (0.189)	14 (0.186)	14 (0.186)	12 (0.164)	10 (0.133)
$p = 5.5$	14 (0.195)	14 (0.192)	14 (0.189)	14 (0.189)	14 (0.189)	12 (0.166)	11 (0.162)
$p = 6$	15 (0.195)	15 (0.192)	14 (0.190)	14 (0.190)	14 (0.189)	13 (0.167)	11 (0.167)
$p = 8$	15 (0.196)	15 (0.193)	15 (0.192)	14 (0.191)	14 (0.190)	13 (0.176)	12 (0.173)
$p = 10$	15 (0.198)	15 (0.196)	15 (0.194)	15 (0.192)	14 (0.191)	13 (0.178)	12 (0.170)
$p = 20$	16 (0.216)	16 (0.221)	16 (0.210)	15 (0.197)	15 (0.194)	14 (0.182)	13 (0.178)
$p = 40$	18 (0.267)	18 (0.273)	17 (0.248)	16 (0.209)	16 (0.204)	14 (0.188)	13 (0.180)
$p = 80$	21 (0.333)	21 (0.338)	20 (0.304)	18 (0.243)	17 (0.226)	15 (0.192)	14 (0.200)

TABLE 2. Numerical results of FASq1 (varying p and ε, fix $h = 1/64$)

FASq1	$\varepsilon^2 = 1$	$\varepsilon^2 = 1/2$	$\varepsilon^2 = 1/4$	$\varepsilon^2 = 1/8$	$\varepsilon^2 = 10^{-1}$	$\varepsilon^2 = 10^{-2}$	$\varepsilon^2 = 10^{-3}$
$p = 4$	15 (0.193)	15 (0.189)	14 (0.185)	14 (0.180)	13 (0.179)	23 (0.331)	-
$p = 5.5$	15 (0.192)	15 (0.189)	14 (0.186)	14 (0.184)	14 (0.183)	-	-
$p = 6$	15 (0.192)	15 (0.189)	14 (0.187)	14 (0.185)	14 (0.183)	-	-
$p = 8$	15 (0.193)	15 (0.190)	14 (0.190)	14 (0.191)	14 (0.186)	-	-
$p = 10$	15 (0.195)	15 (0.193)	14 (0.191)	14 (0.192)	14 (0.187)	-	-
$p = 20$	16 (0.211)	16 (0.215)	16 (0.215)	16 (0.216)	16 (0.220)	-	-
$p = 40$	18 (0.260)	18 (0.281)	19 (0.298)	21 (0.334)	23 (0.367)	-	-
$p = 80$	21 (0.342)	23 (0.383)	25 (0.407)	109 (0.844)	-	-	-

TABLE 3. Numerical results of FASq2 (varying p and ε, fix $h = 1/64$)

FASq2	$\varepsilon^2 = 1$	$\varepsilon^2 = 1/2$	$\varepsilon^2 = 1/4$	$\varepsilon^2 = 1/8$	$\varepsilon^2 = 10^{-1}$	$\varepsilon^2 = 10^{-2}$	$\varepsilon^2 = 10^{-3}$
$p = 4$	14 (0.190)	14 (0.187)	14 (0.183)	14 (0.181)	14 (0.181)	-	-
$p = 5.5$	14 (0.189)	14 (0.189)	14 (0.183)	14 (0.185)	14 (0.187)	-	-
$p = 6$	14 (0.188)	14 (0.186)	14 (0.185)	14 (0.188)	14 (0.190)	-	-
$p = 8$	14 (0.190)	14 (0.190)	14 (0.188)	14 (0.193)	15 (0.196)	-	-
$p = 10$	15 (0.191)	15 (0.191)	15 (0.193)	15 (0.199)	15 (0.202)	-	-
$p = 20$	15 (0.211)	16 (0.223)	17 (0.239)	18 (0.265)	20 (0.290)	-	-
$p = 40$	18 (0.264)	19 (0.300)	21 (0.334)	29 (0.452)	49 (0.643)	-	-
$p = 80$	21 (0.350)	24 (0.393)	32 (0.504)	-	-	-	-

In Tables 1, 2, and 3, we report the numerical results of FAS, FASq1, and FASq2, respectively. Here, we fix the finest mesh size $h = 1/64$ and the coarsest mesh size is $1/4$ but change p and ε to adjust the nonlinearity. In this case, bigger p and/or smaller ε lead to stronger nonlinearity.

TABLE 4. Computational complexity comparison with $\varepsilon = 1$ and $p = 6$

	FAS		FASq2	
h	#iter	CPU time	#iter	CPU time
1/32	15	1.65	14	0.03
1/64	15	7.86	14	0.05
1/128	16	45.60	14	0.16
1/256	16	391.08	15	0.49
1/512	16	>1,000	15	1.67
1/1024	16	>1,000	15	7.12

The number of iterations and convergence rates (in the parenthesis) are listed in Tables 1, 2, and 3. Notation "-" means that the method stagnates or diverges. As we can see, FAS is the most robust one and converges for all the choices of our parameters. The number of iterations are quite stable, ranging from $10 - 21$ iterations, and the convergence rate is about 0.2. This is consistent with our theoretical results presented in Section 6. For FAS, the local energy E_i is defined as the restriction of E on the subspace. Then assumption (AP) holds with $\epsilon < \mu/2$. Therefore, according to Corollary 6.4, FAS converges robustly. For FASq1 and FASq2, both implementations perform well when p is relatively small and/or ε is relatively large. We can clearly see that the number of iterations grows when p gets larger or ε gets smaller. Both implementations fail to converge when nonlinearity is strong, while FASq1 seems to be slightly more robust than FASq2 since it converges for a slightly larger set of parameters. This observation is also consistent with Corollary 6.4. For both FASq1 and FASq2, the local energy E_i is the quadratic energy (26). When p is relatively small and/or ε is relatively large, the nonlinearity of the model problem is relatively weak, and the quadratic energy provides a good approximation in the sense that assumption (AP) holds with $\epsilon < \mu/2$. According to Corollary 6.4, the methods should converge. However, when p gets larger and/or ε gets smaller, the problem becomes more nonlinear and the quadratic energy is not a good approximation of the original energy E any more. Then assumption (AP) does not hold with $\epsilon < \mu/2$ and, according to Corollary 6.4, the method may not converge. Although FASq1 and FASq2 might not converge for strongly nonlinear problems, the advantage of using quadratic energy on local subspaces is that we only need to solve linear problems locally, which could save computational cost considerably.

Next, we compare the CPU time of FAS and FASq2. The reason we choose FASq2 to compare is that FASq2 only involves a symmetric Gauss-Seidel smoother on each level, which basically has the same cost as the multigrid method for solving linear problems. This could dramatically improve the computational complexity for solving our model problem (41). The results are shown in Table 4.

In Table 4, we fix $\varepsilon = 1$ and $p = 6$ and change h. As we can see, for these choices of p and ε, the quadratic energy provides a good approximation of the global energy restricted to the subspace, therefore, the number of iterations of FASq2 is similar with the number of iterations of FAS and remains robust with respect to the mesh size h. The CPU time of FAS grows faster than linear, which is due to the inefficiency of large `for` loops in our current MATLAB implementation.

In contrast, FASq2 is significantly faster than FAS and scales linearly. This demonstrates that, when nonlinearity is mild, we can use a simple quadratic energy and save considerable computational cost.

On the other hand, we want to point out that FAS is more robust than FASq2, as shown before. We have also tested the quadratic energy defined by the Hessian at the previous iteration, cf., (28), which is more or less equivalent to using one approximated Newton's iteration, and the results are similar. Therefore, in practice, we should consider the trade-off between robustness and efficiency in order to decide which kind of local energy should be used on each subspace.

Acknowledgments

The authors wish to thank the anonymous referees for their incisive comments, which helped to improve the paper greatly. The authors also thank Hao Luo, Jea-Hyun Park, and Abner Salgado for carefully reading the revised manuscript and for suggesting several improvements.

References

[1] K. Atkinson and W. Han, *Theoretical Numerical Analysis*: *A functional analysis framework*, 3rd ed., 3rd ed., Texts in Applied Mathematics, vol. 39, Springer, Dordrecht, 2009. MR2511061

[2] A. Beck and L. Tetruashvili, *On the convergence of block coordinate descent type methods*, SIAM J. Optim. **23** (2013), no. 4, 2037–2060, DOI 10.1137/120887679. MR3116649

[3] A. Brandt, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp. **31** (1977), no. 138, 333–390, DOI 10.2307/2006422. MR431719

[4] A. Brandt and O. E. Livne, *Multigrid techniques—1984 guide with applications to fluid dynamics*, Classics in Applied Mathematics, vol. 67, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011. Revised edition of the 1984 original [MR0772748]. MR3396211

[5] L. Chen, *Mesh smoothing schemes based on optimal Delaunay triangulations*, 13th International Meshing Roundtable, pages 109–120, Williamsburg, VA, 2004. Sandia National Laboratories.

[6] L. Chen, *iFEM: An Integrated Finite Element Methods Package in MATLAB*, Technical Report, University of California at Irvine, 2009.

[7] L. Chen and J. Xu, *Optimal Delaunay triangulations*, J. Comput. Math. **22** (2004), no. 2, 299–308. Special issue dedicated to the 70th birthday of Professor Zhong-Ci Shi. MR2058939

[8] P. G. Ciarlet, *Introduction to numerical linear algebra and optimisation*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 1989. With the assistance of Bernadette Miara and Jean-Marie Thomas; Translated from the French by A. Buttigieg. MR1015713

[9] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, 1976. Translated from the French; Studies in Mathematics and its Applications, Vol. 1. MR0463994

[10] W. Feng, A. J. Salgado, C. Wang, and S. M. Wise, *Preconditioned steepest descent methods for some nonlinear elliptic equations involving p-Laplacian terms*, J. Comput. Phys. **334** (2017), 45–67, DOI 10.1016/j.jcp.2016.12.046. MR3606217

[11] E. Gelman and J. Mandel, *On multilevel iterative methods for optimization problems*, Math. Programming **48** (1990), no. 1, (Ser. B), 1–17, DOI 10.1007/BF01582249. MR1049769

[12] S. Gratton, A. Sartenaer, and P. L. Toint, *Recursive trust-region methods for multiscale nonlinear optimization*, SIAM J. Optim. **19** (2008), no. 1, 414–444, DOI 10.1137/050623012. MR2403039

[13] W. Hackbusch, *Multigrid Methods and Applications*, Springer Series in Computational Mathematics, vol. 4, Springer-Verlag, Berlin, 1985. MR814495

[14] V. E. Henson, *Multigrid methods for nonlinear problems: an overview*, submitted to the conference proceedings of the SPIE 15th Annual Symposium on Electronic Imaging, 2005.

[15] Z. Hu, S. M. Wise, C. Wang, and J. S. Lowengrub, *Stable and efficient finite-difference nonlinear-multigrid schemes for the phase field crystal equation*, J. Comput. Phys. **228** (2009), no. 15, 5323–5339, DOI 10.1016/j.jcp.2009.04.020. MR2541456

[16] J. Huang, L. Chen, and H. Rui, *Multigrid methods for a mixed finite element method of the Darcy-Forchheimer model*, J. Sci. Comput. **74** (2018), no. 1, 396–411, DOI 10.1007/s10915-017-0466-z. MR3742884

[17] R. M. Lewis and S. G. Nash, *Model problems for the multigrid optimization of systems governed by differential equations*, SIAM J. Sci. Comput. **26** (2005), no. 6, 1811–1837, DOI 10.1137/S1064827502407792. MR2196577

[18] Z. Lu, *Randomized block proximal damped Newton method for composite self-concordant minimization*, SIAM J. Optim. **27** (2017), no. 3, 1910–1942, DOI 10.1137/16M1082767. MR3693609

[19] S. G. Nash, *A multigrid approach to discretized optimization problems*, Optim. Methods Softw. **14** (2000), no. 1-2, 99–116, DOI 10.1080/10556780008805795. International Conference on Nonlinear Programming and Variational Inequalities (Hong Kong, 1998). MR1809605

[20] Y. Nesterov, *Efficiency of coordinate descent methods on huge-scale optimization problems*, SIAM J. Optim. **22** (2012), no. 2, 341–362, DOI 10.1137/100802001. MR2968857

[21] Y. Nesterov, *Introductory Lectures on Convex Optimization*, A basic course. Applied Optimization, vol. 87, Kluwer Academic Publishers, Boston, MA, 2004. MR2142598

[22] A. Reusken, *Convergence of the multigrid full approximation scheme for a class of elliptic mildly nonlinear boundary value problems*, Numer. Math. **52** (1988), no. 3, 251–277, DOI 10.1007/BF01398879. MR929572

[23] A. Reusken, *Convergence of the multilevel full approximation scheme including the V-cycle*, Numer. Math. **53** (1988), no. 6, 663–686, DOI 10.1007/BF01397135. MR955979

[24] R. M. Spitaleri, *Full-FAS multigrid grid generation algorithms*, Appl. Numer. Math. **32** (2000), no. 4, 483–494, DOI 10.1016/S0168-9274(99)00064-1. Numerical grid generation-technologies for advanced simulations (Berlin, 1997). MR1759311

[25] X.-C. Tai, *Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities*, Numer. Math. **93** (2003), no. 4, 755–786, DOI 10.1007/s002110200404. MR1961887

[26] X.-C. Tai and J. Xu, *Subspace correction methods for convex optimization problems*, Eleventh International Conference on Domain Decomposition Methods (London, 1998), DDM.org, Augsburg, 1999, pp. 130–139. MR1827418

[27] X.-C. Tai and J. Xu, *Global and uniform convergence of subspace correction methods for some convex optimization problems*, Math. Comp. **71** (2002), no. 237, 105–124, DOI 10.1090/S0025-5718-01-01311-4. MR1862990

[28] S. Wise, J. Kim, and J. Lowengrub, *Solving the regularized, strongly anisotropic Cahn-Hilliard equation by an adaptive nonlinear multigrid method*, J. Comput. Phys. **226** (2007), no. 1, 414–446, DOI 10.1016/j.jcp.2007.04.020. MR2356365

[29] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Rev. **34** (1992), no. 4, 581–613, DOI 10.1137/1034116. MR1193013

[30] I. Yavneh and G. Dardyk, *A multilevel nonlinear method*, SIAM J. Sci. Comput. **28** (2006), no. 1, 24–46, DOI 10.1137/040613809. MR2219286

Department of Mathematics, University of California at Irvine, Irvine, California 92697

Email address: chenlong@math.uci.edu

Department of Mathematics, Tuffs University, Medford, Massachussetts 02155

Email address: Xiaozhe.Hu@tufts.edu

Department of Mathematics, The University of Tennessee, Knoxville, Tennessee 37996

Email address: swise1@utk.edu