PROJECTION MULTILEVEL METHODS FOR QUASILINEAR PDES: V-CYCLE THEORY

STEPHEN F. MCCORMICK†

Abstract. The projection multilevel method can be an efficient solver for systems of nonlinear partial differential equations that, for certain classes of nonlinearities (including least-squares formulations of the Navier-Stokes equations), requires no linearization anywhere in the algorithm. This paper provides an abstract framework and establishes optimal V-cycle convergence theory for this method.

Key words. Multigrid, Least-Squares, Finite Elements, Nonlinear PDEs, Navier-Stokes

1. Introduction. The projection multilevel method (PML; cf. [8]) is designed to solve discretized nonlinear partial differential equations (PDEs) by formulating coarse levels in a way that is guided by the discretization methodology. The Rayleigh-Ritz form of PML is designed as an efficient multilevel method for minimizing nonquadratic functionals that come, for example, from variational principles for nonlinear PDEs. For some types of PDEs, including those with nonlinear terms of form $u\partial v$ (e. g., the Navier-Stokes equations), PML can be implemented with no need for linearization anywhere in the algorithm. The ability of PML to treat the nonlinearity directly on all levels is especially important for problems with small basins of attraction about the solution, as exemplified by Navier-Stokes equations with high Reynolds numbers.

Numerical performance of the PML method treated here is illustrated in [6] for Poisson's equation and so-called Kovasznay flow, which is a particular case of the Navier-Stokes equations with known solution. The results in [7] establish optimal two-grid convergence theory for PML in a general setting. The purpose of the present paper is to extend these results to optimal V-cycle theory. By virtue of the method's faithfulness to the minimization principle, the abstract theory is a subtle but otherwise relatively simple generalization of classical multigrid V-cycle results [9]. The subtlety comes primarily from our affine representation of the discrete problem and the error decomposition that we use in the analysis. Application of our abstract theory to the Navier-Stokes equations rests primarily on the properties of the least-squares functional established in [7].

We begin in Section 4 by developing an abstract framework for PML and a theory for optimal V-cycle convergence applied to a general minimization principle. We verify the applicability of two relaxation schemes in this general setting in Section 3. The applicability of this abstract PML method is illustrated by considering a least-squares formulation of the Navier-Stokes equation in Section 4.

2. Abstract Setting. Let $\mathcal{H}$ be an infinite-dimensional Banach space with norm $\|\cdot\|_1$ (suggesting but not necessarily requiring $\mathcal{H}$ to be an H^1-type Hilbert space). Suppose that F is a continuous functional with an isolated minimum at the center, $x_* \in \mathcal{H}$, of a ball of radius $r > 0$:

$$F : \mathcal{B}(x_*; r) := \left\{ x \in \mathcal{H} : \|x - x_*\|_1 < r \right\} \to \mathbb{R}$$

†Department of Applied Mathematics, Campus Box 526, University of Colorado at Boulder, Boulder, CO, 80309–0526. *Email*: `stevem@colorado.edu`. This work was sponsored by the Department of Energy under grant numbers DE-FC02-01ER25479 and DE-FG02-03ER25574, Lawrence Livermore National Laboratory under contract number B533502, Sandia National Laboratory under contract number 15268, and the National Science Foundation under VIGRE grant number DMS-9810751.

is continuous and has a unique minimum on $\mathcal{B}(x_*;r)$ at x_*. We write this last assumption as

$$x_* = \underset{x \in \mathcal{B}(x_*;r)}{\operatorname{argmin}} \; F(x). \tag{2.1}$$

As a general discretization of (2.1), consider fixed $x \in \mathcal{B}(x_*;r)$ and a finite-dimensional subspace, $S^h \subset \mathcal{H}$. For any fixed $x \in \mathcal{B}(x_*;r)$, assume that there exists a unique minimizer of $F(x+x^h)$ over $x^h \in S^h$ for which $x+x^h \in \mathcal{B}(x_*;r)$. More precisely, define

$$\mathcal{B}_x^h(x_*;r) := \left\{ x^h \in S^h : x + x^h \in \mathcal{B}(x_*;r) \right\}$$

and assume that

$$x_*^{(h)} := x + \underset{x^h \in \overline{\mathcal{B}_x^h(x_*;r)}}{\operatorname{argmin}} \; F(x + x^h) \tag{2.2}$$

(which exists by continuity and compactness) is unique and is $\mathcal{B}(x_*;r)$ (not its closure). Our aim is to develop a PML method that treats minimization problem (2.2) efficiently. (Normally, we would take $x = 0$ in (2.2) so that our target problem is to minimize $F(x^h)$, and this is indeed what we typically mean for the finest level. However, this more general affine form of (2.2) that we take here instead facilitates induction to coarser levels.)

To this end, consider a nested sequence of $m \geq 1$ subspaces of S^h:

$$S^{2^m h} \subset S^{2^{m-1} h} \subset \cdots \subset S^{2h} \subset S^h.$$

Assume that (2.2) holds for each of these spaces for any fixed $x \in \mathcal{B}(x_*;r)$:

$$x_*^{(2^k h)} := x + \underset{x^{2^k h} \in \overline{\mathcal{B}_x^{2^k h}(x_*;r)}}{\operatorname{argmin}} \; F(x + x^{2^k h}) \tag{2.3}$$

is unique and in $\mathcal{B}(x_*;r)$, $k = 0, 1, \ldots, m$.

A Notational Subtlety. It is important to keep in mind that no change in x on grid h can change $x_*^{(h)}$:

$$x + x^h + \underset{y^h \in \overline{\mathcal{B}_{x+x^h}^h(x_*;r)}}{\operatorname{argmin}} \; F(x + x^h + y^h) = x + \underset{y^h \in \overline{\mathcal{B}_x^h(x_*;r)}}{\operatorname{argmin}} \; F(x + y^h),$$

for any $x^h \in S^h$. This is also true on coarser levels:

$$x + x^{2h} + \underset{y^{2h} \in \overline{\mathcal{B}_{x+x^{2h}}^{2h}(x_*;r)}}{\operatorname{argmin}} \; F(x + x^{2h} + y^{2h}) = x + \underset{y^{2h} \in \overline{\mathcal{B}_x^{2h}(x_*;r)}}{\operatorname{argmin}} \; F(x + y^{2h}),$$

for any $x^{2h} \in S^{2h}$. Simply said, the solution does not depend on the initial guess. This observation implies that the coarse-level PML computations do not change the 'oscillatory' component of the error decomposition we introduce below. This property simplifies analysis of PML schemes that only use relaxation on the coarse-to-fine phase of the cycle. Note, however, that a change to x on a given level can (and should!) change coarser-level solutions:

$$x + x^h + \underset{x^{2h} \in \overline{\mathcal{B}_{x+x^h}^{2h}(x_*;r)}}{\operatorname{argmin}} \; F(x + x^h + x^{2h}) \neq x + \underset{x^{2h} \in \overline{\mathcal{B}_x^{2h}(x_*;r)}}{\operatorname{argmin}} \; F(x + x^{2h})$$

in general.

Our abstract PML method is based on an abstract relaxation scheme applied to (2.3), which we denote by

$$x^{2^k h} \leftarrow G_x^{2^k h}(x^{2^k h}), \quad k = 0, 1, \cdots, m-1.$$

(Note the dependence of $G_x^{2^k h}$ on x.) Assume that the coarsest grid uses an exact solver. (We only really need assume that the coarsest-grid relaxation process yields a fixed reduction in the error for (2.3) with $k = m$, but we make this assumption for convenience.) We use a V(1,0) cycle, which is represented recursively as follows: given *fixed* $x \in \mathcal{B}(x_*; r)$ and initial guess $x^h = 0$ to the argmin in (2.2), compute a correction by first solving the coarsest-grid problem according to

$$PML^{2^m h}(x) = \underset{x^{2^m h} \in \mathcal{B}_x^{2^m h}(x_*; r)}{\operatorname{argmin}} F(x + x^{2^m h}),$$

and define the successively-finer-grid relaxation steps by

$$PML^{2^{k-1} h}(x) = G_x^{2^{k-1} h}(PML^{2^k h}(x)), \quad k = m, m-1, \cdots, 1.$$

That is, we start on the coarsest grid with an exact solve, the result of which is used on the next finer grid as the initial guess for one relaxation step applied to determining the argmin in (2.3) with $k = m-1$; the result is used as the initial guess for one relaxation step applied to determining the argmin in (2.3) on the next finer level ($k = m-2$); and the process is repeated until the finest level is processed. Note that $PML^{2^k h}(x) \in S^{2^k h}$ for each k, that the result of the V-cycle ($PML^h(x)$) is just a correction to x, and that the final *corrected* x is obtained via

$$x \leftarrow x + PML^h(x).$$

Note also that the next cycle of PML would begin with this corrected x and, again, an initial guess of $x^h = 0$.

To analyze this algorithm, we first introduce some additional notation. Suppose an approximation in $\mathcal{B}_x^h(x_*; r)$ to the argmin of (2.2) is given:

$$x^h \approx \operatorname{argmin}\{F(x + x^h) : x^h \in \mathcal{B}_x^h(x_*; r)\}.$$

This approximation could come from combining the effects of relaxation on a combination of grid levels, as our PML scheme does. In practice, we would typically keep x and its correction on coarser grids separate to avoid the severe expense of interpolating the correction to the finest grid every time we relax on a coarser one. But, to monitor the evolving error theoretically, it is helpful to combine x and x^h even while x^h is evolving on a coarse level (e.g., $x^h = PML^{2^k h}(x)$). We thus denote the evolving error by $e^h = x + x^h - x_*^{(h)}$. Note that e^h and x^h are in S^h even though x and $x_*^{(h)}$ may not be (hence, the parentheses in $x_*^{(h)}$). To measure the size of this error in $x + x^h$ for fixed x and given approximation x^h, we use what we call the *F-metric* given by $F(x + x^h) - F(x_*^{(h)})$, which is zero *when* and, by assumption, *only when* $\operatorname{argmin}\{F(x + x^h + y^h) : y^h \in \mathcal{B}_{x+x^h}^h(x_*; r)\}$ (i.e., $x + x^h = x_*^{(h)}$ or $e^h = 0$).

To analyze the error in $x + x^h$, we decompose it into coarse-level and fine-level components as follows:

$$e^h = s^h + t^h, \quad s^h = \underset{-s^h \in \mathcal{B}_x^{2h}(x_*; r)}{\operatorname{argmin}} F(x + x^h - s^h), \quad t^h = x + x^h - (x_*^{(h)} + s^h). \quad (2.4)$$

Note that this also decomposes the evolving approximation: $x + x^h = x_*^{(h)} + s^h + t^h$. Moreover, s^h and t^h are in S^h and, with $x + x^h = x_*^{(h)} + s^h + t^h$ fixed, we have $x_*^{(2h)} = x_*^{(h)} + t^h$. These relations yield the following F-metric decomposition (obtained by simply adding and subtracting $F(x_*^{(2h)}) = F(x_*^{(h)} + t^h)$):

$$F(x + x^h) - F(x_*^{(h)}) = [F(x + x^h) - F(x_*^{(2h)})] + [F(x_*^{(h)} + t^h) - F(x_*^{(h)})]. \tag{2.5}$$

Both bracketed terms are nonnegative because, by definition,

$$F(x + x^h) \geq F(x_*^{(2h)}) = F(x_*^{(h)} + t^h) \geq F(x_*^{(h)}).$$

The first term on the right of (2.5) represents coarse-grid error, which we assume by induction is reduced by PML coarse-grid cycling. The second term can be thought of as oscillatory error that relaxation presumably reduces. More precisely, the Smoothing Property assumed below can be interpreted as bounding error reduction in proportion to $F(x_*^{(h)} + t^h) - F(x_*^{(h)})$, which is the size of t^h as measured by the F-metric. Our first theorem in effect shows that these reductions do not substantially conflict and thus combine to yield optimal convergence of the PML cycle.

To confirm optimal V-cycle convergence, we impose what we show later is a natural smoothing property on relaxation. In fact, it is a straightforward generalization of the condition introduced in [9]. We describe this property here only on grid h, although we assume that the identical relation holds on all coarser levels.

Smoothing Property. Assume that there exists an $r_1 \leq r$ such that

$$G_x^h : \mathcal{B}_x^h(x_*; r_1) \rightarrow \mathcal{B}_x^h(x_*; r),$$

for any fixed $x \in \mathcal{B}(x_*; r)$, and that there exists a $\gamma < 1$ such that

$$F(x + G_x^h(x^h)) - F(x_*^{(h)}) \leq [F(x + x^h) - F(x_*^{(2h)})] + \gamma[F(x_*^{(h)} + t^h) - F(x_*^{(h)})], \tag{2.6}$$

for any fixed $x \in \mathcal{B}(x_*; r)$ and for all $x^h \in \mathcal{B}_x^h(x_*; r_1)$. (Compare with (2.5).)

We need to impose an additional property on F that holds, for example, when the associated metric is equivalent to $\| \cdot \|_1^2$ (cf. [7]).

Quasi-Monotonicity Property. Assume that there exists an $r_0 \leq r_1$ such that $F(x) < F(y)$ for all $x \in \mathcal{B}(x_*; r_0)$ and $y \in \mathcal{B}(x_*; r) - \mathcal{B}(x_*; r_1)$.

THEOREM 1. *Assume that the Smoothing Property holds on all levels and that the Quasi-Monotonicity Property also holds. Then, for any $x \in \mathcal{B}(x_*; r_0)$, the corrected x remains in $\mathcal{B}(x_*; r_1)$ and converges optimally to $x_*^{(h)}$ in the F-metric:*

$$F(x + PML^h(x)) - F(x_*^{(h)}) \leq \gamma(F(x) - F(x_*^{(h)})).$$

Proof. We first argue that if $x \in \mathcal{B}(x_*; r_0)$, then all PML iterates must be such that the corrected x remains in $\mathcal{B}(x_*; r_1)$. First note that $x + PML^{2^m h}(x) \in \mathcal{B}(x_*; r)$ and $F(x + PML^{2^m h}(x)) \leq F(x)$ by definition, so the Quasi-Monotonicity Property confirms that $x + PML^{2^m h}(x) \in \mathcal{B}(x_*; r_1)$. The Smoothing Property thus allows us to conclude that

$$x + PML^{2^{m-1}h}(x) = x + G^{2^{m-1}h}(PML^{2^m h}(x)) \in \mathcal{B}(x_*; r)$$

and

$$F(x + PML^{2^{m-1}h}(x)) \le F(x),$$

so the Quasi-Monotonicity Property also confirms that $x + PML^{2^{m-1}h} \in \mathcal{B}(x_*; r_1)$. Continuing in this way to finer levels shows that $x + PML^h(x)$ (the corrected x) is in $\mathcal{B}(x_*; r_1)$.

It remains to prove the convergence bound, which we do by induction on the number of levels, m. This bound clearly follows when $m = 1$ because the first term on the right of (2.6) is zero by the assumption that the coarsest-grid uses an exact solver. Assuming that the bound is true for $m = m_0 - 1$ for some $m_0 \ge 2$, we prove now that it must be true for the case of $m = m_0$ levels. Since stopping the V-cycle at level $2h$ for this case actually corresponds to the case of $m_0 - 1$ levels, our induction hypothesis translates to the assumption that

$$F(x + PML^{2h}(x)) - F(x_*^{(2h)}) \le \gamma(F(x) - F(x_*^{(2h)})). \tag{2.7}$$

But the definition of $PML^h(x)$ and property (2.6) imply that

$$\begin{aligned} &F(x + PML^h(x)) - F(x_*^{(h)}) \\ &\quad = F(x + G_x^h(PML^{2h}(x))) - F(x_*^{(h)}) \\ &\quad \le [F(x + PML^{2h}(x)) - F(x_*^{(2h)})] + \gamma[F(x_*^{(h)} + t^h) - F(x_*^{(h)})]. \end{aligned}$$

Appealing now to induction hypothesis (2.7) and decomposition (2.5) proves the result. $\square$

Nearness Assumption. For a typical fine-grid problem, x would be in S^h and would in fact represent the current approximation to the minimizer of $F(x^h)$. To ask that x be close to x_* for this case, then, implicitly assumes that h must be so small that $\mathcal{B}_x^h(x_*; r_0) \ne \emptyset$ or, equivalently, that $\mathcal{B}_0^h(x_*; r_0) = \mathcal{B}(x_*; r_0) \cap S^h \ne \emptyset$.

3. Relaxation. The objective here is to establish the Smoothing Property for two specific relaxation schemes. We limit this development to grid h only because treatment of the coarser levels is identical.

Assume now that $\mathcal{H}$ is an H^1 Hilbert space equipped with L^2 inner product $< \cdot, \cdot >$ and H^1 norm $\| \cdot \|_1$, and that F is twice-continuously differentiable on $\mathcal{B}(x_*; r)$. Let $\nabla^h F(x)$ denote the *discrete gradient* of F at $x \in \mathcal{B}(x_*; r)$, defined as the unique element of S^h satisfying

$$F'(x)[y^h] = < \nabla^h F(x), y^h >, \quad \forall y^h \in S^h.$$

Define the discrete L^2 norm of $F''(x)$, $x \in \mathcal{B}(x_*; r)$, by

$$\|F''(x)\|_h = \sup_{0 \ne y^h \in S^h} \frac{\left|F''[y^h, y^h]\right|}{<y^h, y^h>}.$$

Our two relaxation schemes are defined in terms of G_x^h as follows.

Steepest Descent Iteration. This method uses the discrete gradient as the descent direction and the optimal step size for which the iterate remains in $\mathcal{B}(x_*; r)$:

$$G_x^h(x^h) \equiv x^h - s\nabla^h F(x + x^h),$$

where

$$s = \operatorname{argmin}\{F(x + x^h - s\nabla^h F(x + x^h)) : s \in \mathbb{R} \ni x + x^h - s\nabla^h F(x + x^h) \in \mathcal{B}(x_*; r)\}.$$

Nonlinear Richardson Iteration. This method also uses the discrete gradient as the descent direction, but chooses a fixed step size based on the discrete norm of F'' and a sufficiently small fixed damping parameter, $\omega > 0$:

$$G_x^h(x^h) \equiv x^h - \frac{\omega}{\|F''(x + x^h)\|_h}\nabla^h F(x + x^h).$$

Similar to what we imposed on the discretization, here we assume that these relaxation schemes are well defined and remain in $\mathcal{B}_x^h(x_*; r)$ provided $x^h \in \mathcal{B}_x^h(x_*; r_1)$. Specifically, we assume, for steepest descent, that its argmin is attained so that $x + x^h - s\nabla^h F(x + x^h)$ stays in $\mathcal{B}_x^h(x_*; r)$ and, for nonlinear Richardson, that its iterate remains in $\mathcal{B}_x^h(x_*; r)$. (By imposing continuity on F''', the restriction that $\omega > 0$ be sufficiently small could be relaxed to the assumption that $\omega < 1$.)

Next, we introduce general conditions on our problem and finite element space that allow us to establish the Smoothing Property. Rather than being specific here, we simply assume that the property holds in the bilinear case.

To this end, let $a : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a bilinear form that satisfies the following H^1-equivalence property:

$$c\|y\|_1^2 \leq a(y, y) \leq C\|y\|_1^2, \tag{3.1}$$

for all $y \in \mathcal{H}$ and for some constants c and C. (We use c and C as generic constants that may change meaning with each occurrence, but are independent of h and other obvious quantities.) Assume further that, for every $y^h \in S^h$, there exists a $y^{2h} \in S^{2h}$ such that

$$a(y^h - y^{2h}, y^h - y^{2h}) \leq \delta \frac{\|\nabla^h a(y^h, y^h)\|^2}{\|a''(y^h, y^h)\|_h}, \tag{3.2}$$

where δ is a constant that does not depend on y^h or h. We emphasize that $a''(y^h, y^h)$ here denotes the second Fréchet derivative of $a(y^h, y^h)$ as a function of $y^h \in S^h$. This bound follows from standard finite element theory (cf. [3, 4, 5]) for the case that a corresponds to a linear H^2-regular PDE and S^h is a conventional finite element space associated with a quasi-uniform grid.

In particular, we assume that (3.2) holds when a is the bilinear form whose discrete gradient agrees with $\nabla^h F(x)$ for fixed $x \in \mathcal{B}(x_*; r)$. To articulate this assumption, first note that optimality of $x_*^{(h)}$ implies that $\nabla^h F(x_*^{(h)}) = 0$, so we can write

$$< \nabla^h F(x), y^h > = < \nabla^h F(x) - \nabla^h F(x_*^{(h)}), y^h > = F''(\tilde{x})[x - x_*^{(h)}, y^h],$$

for some $\tilde{x} \in \mathcal{B}(x_*; r)$ and for all $y^h \in S^h$. Now define

$$a(y^h, y^h) = \frac{1}{2} F''(\tilde{x})[y^h, y^h].$$

Remembering that $x - x_*^{(h)} \in S^h$, we may then take $y^h = x - x_*^{(h)}$ and conclude that

$$\nabla^h a(x - x_*^{(h)}, x - x_*^{(h)}) = \nabla^h F(x),$$

as intended.

Here, we implicitly assume that this particular choice for a satisfies (3.1). This H^1-equivalence property is exhibited for the class of PDEs that we have in mind, as confirmed for the Navier-Stokes example of the next section by Theorem 1 in [7]. The additional assumption expressed in (3.2) holds when the PDE is H^2 regular, which follows for our Navier-Stokes example by what we imposed on the domain.

We collect these assumptions on our functional as follows.

Functional Properties. Assume that

$$c\|y\|_1^2 \leq F''(x)[y, y] \leq C\|y\|_1^2, \tag{3.3}$$

for all $x \in \mathcal{B}(x_*; r)$ and $y \in \mathcal{H}$, and that there exists a constant, δ, such that

$$F''(\tilde{x})[x - x_*^{(h)} - y^{2h}, x - x_*^{(h)} - y^{2h}] \leq \delta \frac{\|\nabla^h F(x)\|^2}{\|F''(\tilde{x})\|_h}, \tag{3.4}$$

for all $x \in \mathcal{B}(x_*; r)$, for some $\tilde{x} \in \mathcal{B}(x_*; r)$, and for some $y^{2h} \in S^{2h}$.

We can now easily establish the following simpler estimate.

Lemma 3.1. *Assume that the Functional Properties hold. Then there exists a constant, $\eta > 0$, independent of h and $x \in \mathcal{B}(x_*; r)$, such that*

$$\frac{\|\nabla^h F(x)\|^2}{\|F''(x)\|_h} \geq \eta \Big(F(x_*^{(h)} + t^h) - F(x_*^{(h)}) \Big).$$

Proof. Let y^{2h} be the element of S^{2h} guaranteed to satisfy (3.4). Then optimality of $x_*^{(h)} + t^h$ with respect to coarse-grid correction of x and a Taylor series expansion using the fact that $\nabla^h F(x_*^{(h)}) = 0$ yield

$$\begin{aligned} F(x_*^{(h)} + t^h) - F(x_*^{(h)}) &\leq F(x - y^{2h}) - F(x_*^{(h)}) \\ &= \frac{1}{2} F''(\hat{x})[x - x_*^{(h)} - y^{2h}, x - x_*^{(h)} - y^{2h}], \end{aligned} \tag{3.5}$$

for some $\hat{x} \in \mathcal{B}(x_*; r)$. Equivalence bound (3.3) shows that

$$F''(\hat{x})[x - x_*^{(h)} - y^{2h}, x - x_*^{(h)} - y^{2h}] \leq \frac{C}{c} F''(\tilde{x})[x - x_*^{(h)} - y^{2h}, x - x_*^{(h)} - y^{2h}] \tag{3.6}$$

and, similarly, that

$$\|F''(x)\|_h \leq \frac{C}{c} \|F''(\tilde{x})\|_h. \tag{3.7}$$

The lemma now follows from (3.4)-(3.7) with $\eta = \frac{2c^2}{\delta C^2}$. $\square$

This result sets the stage for establishing optimal convergence of PML based on either relaxation scheme.

THEOREM 2. *Assume that the Functional Properties hold. Then both steepest descent and nonlinear Richardson satisfy the Smoothing Property.*

Proof. It is enough to prove that nonlinear Richardson iteration satisfies this property because, by definition, steepest descent cannot give a larger functional value. To this end, first note that we may assume that $x^h = 0$ without loss of generality, for otherwise we simply replace x by $x + x^h$. We now use Taylor series to conclude that

$$\begin{aligned} F(x + G_x^h(x^h)) &- F(x_*^{(h)}) \\ &= F\Big(x - \frac{\omega}{\|F''(x)\|_h}\nabla^h F(x)\Big) - F(x_*^{(h)}) \\ &= F(x) - F(x_*^{(h)}) - \frac{\omega}{\|F''(x)\|_h} F'(x)[\nabla^h F(x)] \\ &\quad + \frac{\omega^2}{\|F''(x)\|_h^2} F''(\tilde{x})[\nabla^h F(x), \nabla^h F(x)], \end{aligned}$$

for some $\tilde{x} \in \mathcal{B}(x_*; r)$. Now, using bound (3.7) with x and $\tilde{x}$ reversed, choosing $\omega \leq c/(2C)$, and appealing to relation $F'(x)[\nabla^h F(x)] = \|\nabla^h F(x)\|^2$, we thus have

$$F(x + G_x^h(x^h)) - F(x_*^{(h)}) \leq F(x) - F(x_*^{(h)}) - \frac{\omega \|\nabla^h F(x)\|^2}{2\|F''(x)\|_h}.$$

The result now follows with $\gamma = 1 - \eta\omega/2$ from Lemma 3.1 and decomposition (2.5). $\square$

4. Navier-Stokes Example. We introduce the least-squares formulation of the Navier-Stokes equations here to provide a concrete example for the abstract setting of the previous sections. Consider the first-order velocity-flux formulation of these equations (see [1] and [2]) represented by

$$\mathcal{L}(\mathbf{x}) = \mathrm{g} := \begin{cases} \nabla \mathbf{u}^t - \mathbf{U} = \mathbf{0} & \text{in } \Omega, \\ -(\nabla \cdot \mathbf{U})^t + \nabla p + Re\ \mathbf{U}^t \mathbf{u} = \mathbf{f} & \text{in } \Omega, \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega, \\ \nabla \times \mathbf{U} = \mathbf{0} & \text{in } \Omega, \\ \nabla \big(tr \mathbf{U}\big) = \mathbf{0} & \text{in } \Omega, \end{cases} \tag{4.1}$$

where $\mathbf{f} \in L^2(\Omega)^n$ and domain $\Omega \subset \mathbb{R}^n$ $(n = 2, 3)$ is either convex polygonal or has a $C^{1,1}$ boundary, $\partial\Omega$. Without loss of generality, we take the boundary conditions to be $\mathbf{u} = \mathbf{0}$ and $\mathbf{n} \times \mathbf{U} = \mathbf{0}$ on $\partial\Omega$, where $\mathbf{n}$ is the outward unit normal on $\partial\Omega$. Writing the unknowns as $\mathbf{x} = (\mathbf{u}, \mathbf{U}, p)$, a minimization principle can then be obtained by taking the L^2 norm of each interior equation:

$$\mathcal{F}(\mathbf{x}; \mathrm{g}) = \|\mathcal{L}(\mathbf{x}) - \mathrm{g}\|_{0,\Omega}^2, \qquad \mathbf{x} \in \boldsymbol{\mathcal{V}}, \tag{4.2}$$

where $\mathrm{g} = (\mathbf{0}, \mathbf{f}, 0, \mathbf{0}, \mathbf{0})^T$ and the space is defined by

$$\boldsymbol{\mathcal{V}} = H_0^1(\Omega)^n \times \mathcal{V}_0 \times (H^1(\Omega)/\mathbb{R}),$$

with

$$\mathcal{V}_0 = \{\mathbf{U} \in H^1(\Omega)^{n^2} : \mathbf{n} \times \mathbf{U} = \mathbf{0} \ \text{ on } \ \partial\Omega\}.$$

Our target differential problem is then to minimize $\mathcal{F}(\mathbf{x}; \mathrm{g})$ in (4.2) over $\boldsymbol{\mathcal{V}}$.

In many practical examples, most solutions of (4.1) are isolated in the sense that neighborhoods exist in which the solutions are unique. Assume, therefore, that we are in a closed neighborhood, $\overline{\mathcal{B}}(\mathbf{x}_*, r)$, of an isolated solution, $\mathbf{x}_* \in \boldsymbol{\mathcal{V}}$, to (4.1), that is, a global minimum of (4.2), for which $\mathcal{F}(\mathbf{x}_*; \mathrm{g}) = 0$. The neighborhood is taken to be an H^1-ball around $\mathbf{x}_*$ of radius $r > 0$ defined as

$$\mathcal{B}(\mathbf{x}_*; r) := \left\{\mathbf{x} \in \boldsymbol{\mathcal{V}} : \|\mathbf{x} - \mathbf{x}_*\|_{1,\Omega} < r\right\},$$

where

$$\|\mathbf{x}\|^2_{1,\Omega} \equiv \|\mathbf{u}\|^2_{1,\Omega} + \|\mathbf{U}\|^2_{1,\Omega} + \|p\|^2_{1,\Omega}.$$

For the discretization, consider a quasi-uniform finite element partition of Ω with approximate mesh size h and let $H^h(\Omega)$ be the corresponding conforming finite element subspace of $H^1(\Omega)$ consisting of piecewise polynomials: a function in $H^h(\Omega)$ is continuous on Ω and polynomial within each element. Let $H^h_0(\Omega)$ denote the subspace of $H^h(\Omega)$ of functions that are zero on $\partial\Omega$. Then define

$$\mathcal{S}^h = H^h_0(\Omega)^n \times \mathcal{V}^h_0 \times (H^h(\Omega)/\mathbb{R}) \subset \boldsymbol{\mathcal{V}},$$

with

$$\mathcal{V}^h_0 = \{\mathbf{U}^h \in H^h(\Omega)^{n^2} : \mathbf{n} \times \mathbf{U}^h = \mathbf{0} \ \text{ on } \ \partial\Omega\}.$$

The discrete problem is now to minimize $\mathcal{F}(\mathbf{x}; \mathrm{g})$ in (4.2) over S^h.

Results established in [7] support the assumptions made in the previous sections on this functional. In particular. letting $\mathcal{F}'(\mathbf{x}; \mathrm{g})$ and $\mathcal{F}''(\mathbf{x}; \mathrm{g})$ denote the respective first and second Fréchet derivatives of $\mathcal{F}(\mathbf{x}; \mathrm{g})$, then Theorem 1 of [7] establishes continuity and coercivity of $\mathcal{F}''(\mathbf{x}; \mathrm{g})$: for small enough $r > 0$, there exist positive constants, c and C, which depend only on Re, r, and Ω, such that

$$\mathcal{F}''(\mathbf{x}; \mathrm{g})[\mathbf{y}, \mathbf{z}] \ \leq \ C\|\mathbf{y}\|_{1,\Omega}\|\mathbf{z}\|_{1,\Omega} \tag{4.3}$$

and

$$c\|\mathbf{y}\|^2_{1,\Omega} \ \leq \ \mathcal{F}''(\mathbf{x}; \mathrm{g})[\mathbf{y}, \mathbf{y}], \tag{4.4}$$

for any $\mathbf{x} \in \overline{\mathcal{B}}(\mathbf{x}_*; r)$ and all $\mathbf{y} \in \boldsymbol{\mathcal{V}}$. Theorem 2 of [7] establishes continuity and coercivity of $\mathcal{F}(\mathbf{x}; \mathrm{g})$: for small enough $r > 0$, there exist positive constants, c and C, which depend only on Re, r, and Ω, such that

$$c\|\mathbf{x}-\mathbf{x}_*\|^2_{1,\Omega} \leq \mathcal{F}(\mathbf{x}; \mathrm{g}) \leq C\|\mathbf{x}-\mathbf{x}_*\|^2_{1,\Omega}, \tag{4.5}$$

for all $\mathbf{x} \in \overline{\mathcal{B}}(\mathbf{x}_*, r)$. Results in [7] also establish that the discrete problem has a unique solution in $\mathcal{B}(\mathbf{x}_*; r)$ provided r is small enough and that relaxation applied to the discrete problem stays in $\mathcal{B}(\mathbf{x}_*; r)$ provided the initial guess is close enough to the discrete solution. These results confirm that this important example satisfies all of the assumptions that we make in the abstract setting above.

REFERENCES

[1] P. Bochev, Z. Cai, T. Manteuffel, and S. McCormick, *Analysis of velocity-flux first-order system least-squares principles for the Navier-Stokes equations: Part I*, SIAM J. Numer. Anal, 35 (1998), pp. 990–1009.

[2] ———, *Analysis of velocity-flux least-squares principles for the Navier-Stokes equations: Part II*, SIAM J. Numer. Anal., 36 (1999), pp. 1125–1144.

[3] D. Braess, *Finite elements*, Cambridge University Press, Cambridge, 2001.

[4] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, vol. 15 of Texts in Applied Mathematics, Springer-Verlag, New York, 2002.

[5] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, vol. 40 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[6] T. A. Manteuffel, S. F. McCormick, O. Roehrle, and J. Ruge, *Projection multilevel methods for quasilinear elliptic partial differential equations: numerical results*, manuscript.

[7] T. A. Manteuffel, S. F. McCormick, and O. Roehrle, *Projection multilevel methods for quasilinear elliptic partial differential equations: theoretical results*, manuscript.

[8] S. F. McCormick, *Multilevel Projection Methods for Partial Differential Equations*, vol. 62 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[9] S. F. McCormick, *Multigrid methods for variational problems: general theory for the V-cycle*, SIAM J. Numer. Anal. 22 (l985), pp. 634-643.