**Summary**

A well-known strategy for building effective preconditioners for higher-order discretizations of some PDEs, such as Poisson’s equation, is to leverage effective preconditioners for their low-order analogs. In this work, we show that high-quality preconditioners can also be derived for the Taylor-Hood discretization of the Stokes equations in much the same manner. In particular, we investigate the use of geometric multigrid based on the $\mathbb{Q}_1\text{iso}\mathbb{Q}_2/\mathbb{Q}_1$ discretization of the Stokes operator as a preconditioner for the $\mathbb{Q}_2/\mathbb{Q}_1$ discretization of the Stokes system. We utilize local Fourier analysis to optimize the damping parameters for Vanka and Braess-Sarazin relaxation schemes and to achieve robust convergence. These results are then verified and compared against the measured multigrid performance. While geometric multigrid can be applied directly to the $\mathbb{Q}_2/\mathbb{Q}_1$ system, our ultimate motivation is to apply algebraic multigrid within solvers for $\mathbb{Q}_2/\mathbb{Q}_1$ systems via the $\mathbb{Q}_1\text{iso}\mathbb{Q}_2/\mathbb{Q}_1$ discretization, which will be considered in a companion paper.

**KEYWORDS:**

Monolithic Multigrid, Stokes Equations, Braess-Sarazin, Additive Vanka, Local Fourier Analysis

---

**INTRODUCTION**

This paper focuses on developing efficient algorithms for the numerical approximation of solutions to the Stokes equations, which are used to simulate incompressible viscous flow and whose discretization results in saddle-point linear systems. Linear systems of saddle-point type appear in a variety of scientific and engineering applications. The block structure and indefiniteness of these systems often make it particularly challenging to construct efficient numerical schemes. The coupled physical fields, pressure and velocity, are often discretized using staggered grids or methods where degrees of freedom are not located at the same spatial point, further complicating the development of solvers such as monolithic multigrid for these systems.

Monolithic multigrid methods, that apply multigrid to the entire system in a coupled (or all-at-once) manner, have long demonstrated robust convergence for Stokes problems. Most of these approaches are geometric multigrid (GMG) methods, where the multigrid hierarchy is composed of a sequence of coarser discretizations on nested meshes, connected by canonical interpolation operators for each field. There are many fewer approaches to monolithic algebraic multigrid (AMG) for Stokes and similar systems, as algebraic coarsening of these matrices is complicated by the presence of higher-order finite-element bases, which are present in a majority of stable Stokes discretizations. This is in part due to two main obstacles. First, the stiffness matrices constructed with higher-order basis functions are no longer M-matrices, for which AMG methods were originally intended. Secondly, the coarsening of one field cannot be done independently of the other, due to inf-sup stability concerns. In order to address the first obstacle, we propose to coarsen with respect to the element order ($p$) and, then, to employ geometric multigrid on the low-order system by coarsening spatially in $h$. This does require that a low-order system be constructed, however, it is facilitated by the standard building blocks needed to build the high-order discretization. While the
approach presented here is one of geometric multigrid, we propose this primarily as a building block that can be leveraged in future AMG work that will aim to address the second obstacle.

Low-order preconditioning is a common approach when developing preconditioners for higher-order discretizations. These methods have been successfully used for preconditioning discretized Laplace operators in both geometric and algebraic multigrid contexts. In the case of the Stokes problem, non-nested low-order geometric multigrid methods have been shown to be effective preconditioners for higher-order discretizations. In this paper, we construct and use a stable lower-order \( \mathbb{Q}_1/\mathbb{Q}_1 \) discretization within a monolithic multigrid preconditioner for the \( \mathbb{Q}_2/\mathbb{Q}_1 \) discretized Stokes system. While “double discretization” or “defect correction” schemes and multigrid methods based on low-order discretizations have been considered before independently of each other, to our knowledge, these schemes have not yet been considered for the stable Taylor-Hood finite-element discretization. The goal of our work is to gauge the effectiveness of this type of preconditioning for this discretization.

A major complication in defect-correction algorithms is the plethora of algorithmic parameters that can arise, including relaxation parameters for both higher-order and lower-order discretizations, and additional parameters related to cycling strategies between the discretizations. Here, we consider both the choice of and sensitivity to these parameters in the context of geometric multigrid, which we use to approximate the inverse of the \( \mathbb{Q}_1/\mathbb{Q}_1 \) system within a standard defect-correction scheme. Using geometric multigrid offers an advantage in that parameter choices can be determined using local Fourier analysis. Local Fourier analysis has been studied and applied to many interesting problems, including monolithic multigrid methods for the Stokes equations. Indeed, early work in this direction focused on the use of both distributed and Vanka relaxation for the staggered marker-and-cell (MAC) finite-difference discretization scheme for the Stokes equations. More recent Fourier analysis has included that of multiplicative and additive Vanka relaxation for the \( \mathbb{Q}_2/\mathbb{Q}_1 \) discretization, as well as for Braess-Sarazin relaxation. Here, we make use of these tools to optimize multigrid parameters in stationary iterations for the proposed defect-correction scheme.

The remainder of the paper is structured in the following manner. In Section 2, we briefly introduce the \( \mathbb{Q}_2/\mathbb{Q}_1 \) and \( \mathbb{Q}_1/\mathbb{Q}_1 \) discretizations for the Stokes equations. Section 3 describes the monolithic multigrid framework with Braess-Sarazin and additive Vanka relaxation for the Stokes equations, while Section 4 introduces local Fourier analysis (LFA) as a tool to identify optimal GMG parameters. Section 5 exhibits the main contributions of this paper: optimized LFA two-grid convergence factors and the measured multigrid convergence for the Stokes problem. Section 6 presents the conclusions. The main contribution of this work is that of demonstrating that the principle of low-order preconditioning is also applicable to a mixed finite-element discretization such as the \( \mathbb{Q}_2/\mathbb{Q}_1 \) discretization of Stokes. In our view, this opens a new line of research into the development of truly algebraic multigrid methods for such discretizations, by providing a route to effectively precondition higher-order discretizations without directly applying AMG to them. A central question in that direction that we aim to address in future work is how to algebraically create a hierarchy of stable coarse-grid operators that allow effective multigrid cycling on problems such as the \( \mathbb{Q}_1/\mathbb{Q}_2/\mathbb{Q}_1 \) discretization.

2 DISCRETIZATION AND SOLUTIONS OF STOKES EQUATIONS

In this paper, we consider the two-dimensional Stokes equations, given by

\[
-\nabla^2 u + \nabla p = f \tag{1a}
\]

\[
-\nabla \cdot u = 0 \tag{1b}
\]

where \( u \) is the velocity of the fluid, \( p \) is the pressure, and \( f \) is a forcing term. We assume homogenous Dirichlet conditions on \( u \) over the boundary of a domain \( \Omega \) for simplicity. In addition, we consider a uniform grid of size \( h \) over \( \Omega \) and finite dimensional spaces of the form \( \mathcal{X} \subset H^1_0(\Omega) \) and \( \mathcal{M} \subset L_2(\Omega) \), where \( \mathcal{X} \) satisfies the appropriate homogeneous Dirichlet boundary conditions. The resulting discrete weak formulation of (1) is to find \( u \in \mathcal{X} \) and \( p \in \mathcal{M} \) such that

\[
a(u, v) + b(p, v) = F(v) \tag{2a}
\]

\[
b(q, u) = 0, \tag{2b}
\]

for all \( q \in \mathcal{M} \) and \( v \in \mathcal{X} \). Here, \( a(\cdot, \cdot) \) and \( b(\cdot, \cdot) \) are bilinear forms and \( F(\cdot) \) is a linear form given by

\[
a(u, v) = \int_\Omega \nabla u : \nabla v, \quad b(p, v) = -\int_\Omega p \nabla \cdot v, \quad F(v) = \int_\Omega f \cdot v.
\]
An inf-sup condition on the finite-dimensional spaces $\mathcal{X}^h$ and $\mathcal{M}^h$ is sufficient to guarantee the uniqueness of the solution up to a constant pressure.

Our focus is on two types of stable mixed finite-element discretizations for $\mathcal{X}^h$ and $\mathcal{M}^h$. The first is the $\mathbb{Q}_2/\mathbb{Q}_1$ discretization (also known as the Taylor-Hood discretization), which uses a bilinear representation of the pressure and a biquadratic representation for the velocity components. The second is the $\mathbb{Q}_1\text{iso}\mathbb{Q}_2/\mathbb{Q}_1$ discretization, which replaces the $\mathbb{Q}_2$ space for velocities with an $\mathbb{Q}_1$ approximation on a once-refined mesh. This is obtained by overlaying the higher-order nodes with a lower-order mesh — see Figure 1. The $\mathbb{Q}_1\text{iso}\mathbb{Q}_2/\mathbb{Q}_1$ discretization is not well-known due to its relatively low accuracy and computational efficiency. Yet, the pairing is known to be inf-sup stable, and we argue that it can be highly effective when used within a preconditioner for the $\mathbb{Q}_2/\mathbb{Q}_1$ discretization, by leveraging efficient first-order multigrid solvers.

The challenge presented by the discretization in (2) is the solution of the resulting saddle-point system of the form

$$
\begin{bmatrix}
A & B^T \\
B & 0
\end{bmatrix}
\begin{bmatrix}
u \\
p
\end{bmatrix} =
\begin{bmatrix}
f \\
0
\end{bmatrix},
$$

(3)

where matrix $A$ corresponds to the discrete vector-Laplacian, and $B$ represents the negative of the discrete divergence operator. Here, we overload the notation and use $u$ and $p$ to denote the discrete velocities and pressure for the remainder of the paper. We note that, for the boundary conditions described above, the matrix $K$ is singular, with a one-dimensional nullspace associated with constant shifts in the pressure. While many “fixes” to this nullspace are used in practice (e.g., fixing the value of the pressure at a point, or explicitly adding a constraint that the integral of the pressure is zero), we will leave the system in its singular form and implicitly treat the nullspace in our multigrid method.

### 3 | MONOLITHIC MULTIGRID

We study a monolithic geometric multigrid approach with standard coarsening — i.e., doubling $h$ on each level of the hierarchy. In the following, we assess the error reduction in multigrid when using a coupled relaxation method followed by a coarse-grid correction. Interpolation and restriction are constructed using the finite element basis. In the following, we use $\ell$ to denote the level of the multigrid hierarchy with $\ell = 0$ denoting the finest-level problem. The subscript may be dropped when context is clear.

For a linear system $K_\ell u_\ell = b_\ell$ on level $\ell$ of the multigrid hierarchy, we express the relaxation scheme as a fixed-point iteration

$$u_{\ell+1}^{j+1} = (I - \omega_\ell M_\ell^{-1} K_\ell) u_\ell^j + \omega_\ell M_\ell^{-1} b_\ell,$$

where $\omega_\ell$ is a relaxation parameter.
where $M_\ell$ is an inexpensive approximation to $K_\ell$ so that the action of its inverse is easy to apply. Here, $(I - \omega_\ell M_\ell^{-1} K_\ell)$ is the associated error-propagation operator for relaxation on $K_\ell$, where $\omega_\ell$ is a damping parameter. Using restriction $R_\ell$ and interpolation $P_\ell$, the coarse-grid correction operator projects the error onto grid $\ell' + 1$, where a correction is computed. Combining pre- and post-relaxation together with coarse grid correction we arrive at the two-grid error-propagation operator

$$G_\ell = (I - \omega_\ell M_\ell^{-1} K_\ell)(I - P_\ell R_{\ell+1} K_{\ell+1} P_\ell)(I - \omega_{\ell'} M_{\ell'}^{-1} K_{\ell'})$$

(4) when one sweep of pre- and one sweep of post-relaxation is applied within the multigrid cycle. The coarse-grid operators for the $Q_{\ell}iso Q_{2}/Q_{1}$ system are computed via the Galerkin product $K_{\ell+1} = R_{\ell} K_{\ell} P_{\ell}$, which is equivalent to rediscritization on the coarse grids. The correction on the coarsest grid is computed using a pseudoinverse, to account for the pressure nullspace.

### 3.1 Low-order Preconditioner

Lower-order preconditioning has been successfully used for a variety of problems, ranging from elliptic PDEs to saddle-point systems. For example, taking $A_0$ to be the matrix from a high-order spectral discretization of the Laplace operator and $A_1$ to be the matrix from a second-order finite-difference discretization on an auxiliary mesh constructed from the nodal points of the high-order mesh, then the condition number of $A_1^{-1} A_0$ is bounded asymptotically by $\pi^2/4$, establishing $A_1$ as an effective preconditioner for $A_0$.

Here, we consider a multigrid defect-correction method that incorporates this notion of high- and low-order operators. This is constructed by taking $K_0$ and $K_1$ to be the higher-order and lower-order Stokes operators assembled using the $Q_{2}/Q_{1}$ and $Q_{1}iso Q_{2}/Q_{1}$ finite element spaces, respectively. Note that the $Q_{1}iso Q_{2}/Q_{1}$ discretization is not a Galerkin coarsening of the $Q_{2}/Q_{1}$ discretization, as the finite-element spaces are not nested. That is, $K_1 \neq R_0 K_0 P_0$ within this multigrid hierarchy. Since the unknowns for levels 0 and 1 are, however, co-located, we transfer residuals and corrections between the low- and high-order discretizations using identity operators, effectively taking $P_0 = I$ and $R_0 = I$. On level 1, we then use $\gamma$ iterations of monolithic GMG based on the $K_1$ system, which is denoted as $G_1$. The level-1 multigrid cycle is represented by $\tau (I - G_1^{\gamma}) K_1^{-1}$, which leads to a generalization of (4) for $\ell' = 0$

$$E = (I - \omega_0 M_0^{-1} K_0) \gamma (I - \tau (I - G_1^{\gamma}) K_1^{-1} K_0) (I - \omega_0 M_0^{-1} K_0) \gamma,$$

(5) where $M_0$ is a relaxation operator based on the $K_0$ system, $\omega_0$ is the relaxation parameter, $\gamma$ is the number of level-1 $h$-multigrid cycles used to approximately solve the $K_1$ problem, $\tau$ is a damping parameter for the coarse correction, and the exponents $v_1$ and $v_2$ are the number of pre- and post-relaxation sweeps.

The multigrid scheme in (5) is depicted in Figure 2. If $\gamma = 0$, then the method relaxes only on the $Q_{2}/Q_{1}$ problem and the $Q_{1}iso Q_{2}/Q_{1}$ discretization is unused. At the same time, if $G_1^0 \equiv 0$ in (5), then the $Q_{1}iso Q_{2}/Q_{1}$ problem is solved exactly. The left image in Figure 2 illustrates the case where the overall multigrid cycle employs two-level (blue) and multilevel (green) $h$-multigrid schemes for the $Q_{1}iso Q_{2}/Q_{1}$ part of the cycle with $\gamma = 1$. The right image shows $\gamma = 2$ for a two-level scheme. In the following tests, we study multigrid convergence over a range of parameters, possibly omitting relaxation on different levels. For the remainder of the paper, the terms two-level, and multilevel always refer to the number of levels associated with only the GMG (or $h$-multigrid) part of the overall cycle defined by (5).

### 3.2 Relaxation

The relaxation method $M_\ell^{-1}$ applied to the saddle-point system $K_\ell$ is either chosen as an additive Vanka or Braess-Sarazin coupled relaxation scheme. In this section, we drop the operator subscript $\ell$, since the relaxation operator construction is the same on all levels of the multigrid hierarchy.

Letting $n_p$ be the number of pressure degrees of freedom (DoFs), Vanka relaxation partitions the $n \times n$ system $K$ into $n_p$ overlapping saddle-point problems, consisting of $2 \times 2$ patches of elements around each nodal pressure DoF. An important feature of the construction of Vanka relaxation (in contrast to so-called “star” relaxation) is that each patch consists of all velocity degrees of freedom that are on the closure of the elements adjacent to the nodal pressure DoF, and not just those in the interior of mesh entities (elements and faces) adjacent to the node. Each patch contains a single pressure DoF, however. Figure 3 shows the construction of a Vanka patch for both the $Q_{2}/Q_{1}$ and $Q_{1}iso Q_{2}/Q_{1}$ discretizations; it is important to note that the construction of the Vanka patches in the $Q_{1}iso Q_{2}/Q_{1}$ discretization is based on the (coarse) pressure elements and not the (fine) velocity elements. For each patch, indexed by pressure DoF $i$, we form a (binary) restriction operator, $V_i$, which selects those entries in a
global vector that appears on patch $i$. The system matrix is then projected onto the patch DoFs by a triple matrix product, $V_i K V_i^T$. A single iteration of Vanka is then given by

$$M_V^{-1} = \sum_{i=1}^{n_p} V_i^T W_i (V_i K V_i^T)^{-1} V_i,$$

where the diagonal weighting matrix $W_i$ is defined such that each diagonal entry is equal to the reciprocal of the number of patches that contain the associated degree of freedom.

In contrast to Vanka relaxation, Braess-Sarazin relaxation retains the block structure of the Stokes system to produce a suitable relaxation method. A single application of Braess-Sarazin relaxation is formulated as the (approximate) solution to the following

FIGURE 2 Multilevel scheme for the $\mathbb{Q}_2/\mathbb{Q}_1$ operator. Coarse levels are refinements in $h$ on the $\mathbb{Q}_1$iso $\mathbb{Q}_2/\mathbb{Q}_1$ discretization, resulting in either a two-level scheme (blue) or a multilevel scheme (green). On the left, a single V-cycle is executed on the $\mathbb{Q}_1$iso $\mathbb{Q}_2/\mathbb{Q}_1$ problem (with $\gamma = 1$); on the right, two V-cycles are executed (with $\gamma = 2$). The exact, coarsest level solves are marked in red.

FIGURE 3 At left, construction of a typical Vanka patch for the $\mathbb{Q}_2/\mathbb{Q}_1$ discretization. At right, construction of a typical Vanka patch for the $\mathbb{Q}_1$iso $\mathbb{Q}_2/\mathbb{Q}_1$ discretization. Note that the patch in (b) includes one pressure DoF and all the velocity DoFs that geometrically lie within (or on the boundary) of the sub-domain associated with the elements that include that pressure DoF.
Such operators can be formally diagonalized by the Fourier modes

\[ \tilde{A} \delta u = \tilde{A}^{-1} r_u - \alpha r_p, \]

Thus, we instead focus here on the two-grid LFA convergence factors, which offer sharp predictions of actual multigrid performance. Unfortunately, the simpler LFA smoothing factor can provide poor predictions when used on complicated or higher-order operators. In particular, LFA is often used to guide parameter choices for multigrid components, including those within relaxation methods and for the MG cycle defined by Equations (4) and (5).

Both Vanka and IBSR relaxation include several parameters that must be chosen. While parameter choices are known in some cases, our defect-correction framework is somewhat unique and so we employ LFA to inform our parameter choices for the relaxation methods and for the MG cycle defined by Equations (4) and (5).

4. LOCAL FOURIER ANALYSIS

LFA is a well-known and valuable tool in predicting and analyzing algorithmic performance for the solution of discretized PDEs. In particular, LFA is often used to guide parameter choices for multigrid components, including those within relaxation schemes and grid-transfer operators. Here, we apply LFA to the monolithic multigrid methods in Section 3 and aim to optimize the spectral radius of the two-grid error propagation operator (4) over the possible choices of parameters.

4.1 Definition and notations

For completeness, we give a brief introduction to LFA. First, consider a two-dimensional infinite uniform grid

\[ G = \{ x = (x_1, x_2) = (k_1, k_2)h, \quad (k_1, k_2) \in \mathbb{Z}^2 \}, \]

with uniform grid size \( h \). Let \( A \) be a scalar Toeplitz operator defined by entries in a “stencil”, \( s_\kappa \in \mathbb{R} \), where \( \kappa \in V \subset \mathbb{Z}^2 \) is a finite index set over which the stencil is nonzero, that acts on a vector \( u(x) \in l^2(G) \) as follows:

\[ Au(x) = \sum_{\kappa \in V} s_\kappa u(x + \kappa h). \]

Such operators can be formally diagonalized by the Fourier modes \( \psi(\theta, x) = e^{i \theta \cdot x / h} = e^{i \theta_1 x_1 / h} e^{i \theta_2 x_2 / h} \), where \( \theta = (\theta_1, \theta_2) \) and \( r^2 = -1 \). Thus, we use \( \psi(\theta, x) \) as a Fourier basis with \( \theta \in \left[ \frac{-\pi}{2}, \frac{3\pi}{2} \right]^2 \). Considering standard coarsening by a factor of 2 in each direction, the relevant spaces of low and high frequencies are given by

\[ \theta \in T^{\text{low}} = \left[ \frac{-\pi}{2}, \frac{\pi}{2} \right]^2, \quad \theta \in T^{\text{high}} = \left[ \frac{-\pi}{2}, \frac{3\pi}{2} \right]^2 \setminus \left[ \frac{-\pi}{2}, \frac{\pi}{2} \right]^2. \]

LFA provides two possible predictions of multigrid performance through the so-called “smoothing” and “two-grid convergence” factors, which often offer sharp predictions of actual multigrid performance. Unfortunately, the simpler LFA smoothing factor can provide poor predictions when used on complicated or higher-order operators. Thus, we instead focus here on the two-grid LFA convergence factor, which includes the full details of the coarse-grid correction process. To do this, we define the following harmonic modes

\[ \theta^0 = (\theta^0_1, \theta^0_2) = \theta + \pi \cdot \xi, \quad \theta^0 :\in T^{\text{low}}, \]

where \( \xi = (\xi_1, \xi_2) \in \{0, 0\}, (1, 0), (0, 1), (1, 1) \} \).
Note that for all Fourier modes, \( \psi(\theta, x) \), we have
\[
A\psi(\theta, x) = \tilde{A}(\theta)\psi(\theta, x).
\]
If the multigrid relaxation scheme, \( M \), is also represented by a Toeplitz operator (or its inverse), then the symbol for the error-propagation operator \( \tilde{S}(\theta) = I - \tilde{M}^{-1}(\theta)\tilde{A}(\theta) \) provides information on how the relaxation scheme damps errors at each Fourier frequency. The LFA smoothing factor arises from computing the maximum (absolute) value of \( \tilde{S}(\theta) \) over the high-frequency set, \( T^{\text{high}} \).

To get a better picture of multigrid convergence, we study how the relaxation scheme represented by \( \tilde{S}(\theta) \) interacts with the coarse-grid correction. Here, we must account for the coupling of fine-grid errors in the coarse-grid correction process. For each low-frequency mode, \( \theta \in T^{\text{low}} \), we define a four-dimensional harmonic space,
\[
F(\theta) = \text{span}\left\{ \psi(\theta^\xi, \cdot) : \xi \in \{(0,0), (1,0), (0,1), (1,1)\} \right\},
\]
which is invariant for standard full-coarsening two-grid algorithms under certain assumptions on their grid-transfer operators. The symbol of the two-grid algorithm is defined analogously to Definition 1, accounting for how interpolation and restriction map between grids. For the given interpolation and restriction operators, \( P \) and \( R \), we can define their symbols by how they map harmonic modes between grids. For interpolation, a single coarse-grid mode with frequency \( 2\theta \) is mapped onto \( F(\theta) \), resulting in a \( 4 \times 1 \) symbol for \( P \), denoted \( \tilde{P}_{2g} \), with entries corresponding to each harmonic frequency. The restriction is similarly mapped into a \( 1 \times 4 \) symbol, \( \tilde{R}_{2g} \), representing how each fine-grid harmonic frequency is mapped onto the coarse-grid mode with frequency \( 2\theta \). As a result, the symbol of the two-grid error propagation operator is a \( 4 \times 4 \) matrix given by
\[
\tilde{E}_{2g}(\theta) = \tilde{S}_{2g}(\theta) \left( I - \tilde{P}_{2g}(\theta)(\tilde{A}_c(2\theta))^{-1}\tilde{R}_{2g}(\theta)\tilde{A}_{2g}(\theta) \right) \tilde{S}_{2g}^{-1}(\theta),
\]
where
\[
\tilde{A}_{2g}(\theta) = \text{diag}\left\{ \tilde{A}(\theta^{00}), \tilde{A}(\theta^{01}), \tilde{A}(\theta^{10}), \tilde{A}(\theta^{11}) \right\},
\]
\[
\tilde{R}_{2g}(\theta) = \begin{bmatrix} \tilde{R}(\theta^{00}), & \tilde{R}(\theta^{01}), & \tilde{R}(\theta^{10}), & \tilde{R}(\theta^{11}) \end{bmatrix},
\]
\[
\tilde{P}_{2g}(\theta) = \begin{bmatrix} \tilde{P}(\theta^{00}), & \tilde{P}(\theta^{01}), & \tilde{P}(\theta^{10}), & \tilde{P}(\theta^{11}) \end{bmatrix}^T.
\]
\[
\tilde{S}_{2g}(\theta) = \text{diag}\left\{ \tilde{S}(\theta^{00}), \tilde{S}(\theta^{01}), \tilde{S}(\theta^{10}), \tilde{S}(\theta^{11}) \right\}.
\]
Here, \( \text{diag}\{A_1, A_2, A_3, A_4\} \) is the diagonal matrix with diagonal entries, \( A_1, A_2, A_3, \) and \( A_4 \), and \( \tilde{A}_c \) is the symbol of the coarse-grid operator.

**Definition 2.** The two-grid LFA convergence factor, \( \hat{\rho} \), is defined as
\[
\hat{\rho} = \sup_{\theta \in T^{\text{low}}} \rho\left( \tilde{E}_{2g}(\theta) \right),
\]
where \( \rho(\tilde{E}_{2g}(\theta)) \) denotes the spectral radius of matrix \( \tilde{E}_{2g}(\theta) \).

In this work, we will apply LFA to the monolithic multigrid algorithms described above, and optimize the two-grid convergence factor using a robust optimization framework recently developed for cases such as this with multiple parameters to be optimized\[13\]. While we do not compute \( \hat{\rho} \) from (9) exactly, we use this to denote the values obtained from the optimization algorithm that, necessarily, samples the spectral radius of \( \tilde{E}_{2g}(\theta) \) at only a finite number of points.

### 4.2 Fourier representation for \( Q_2/Q_1 \) and \( Q_1 \text{iso} Q_2/Q_1 \) discretizations

LFA for coupled systems discretized using mixed finite-element approaches is necessarily more complicated than in the scalar case. While the staggered finite-difference case can be successfully handled using slight generalizations of the scalar case\[7,10,12\], greater adaptation is required for higher-order or otherwise non-nodal finite-element discretization\[13,14,20,22\]. The key step is in realizing that the resulting discrete operators (including the infinite-grid analog of (3)) can be reordered into block structured linear systems with Toeplitz blocks. Once suitably reordered, each block can be diagonalized using the classical Fourier approach. In this way, LFA of more complicated operators in a two-grid method results in block Fourier symbols that can be assembled as in (8) and maximized to yield a two-grid LFA convergence factor, as in (9).
For the $Q_2/Q_1$ and $Q_1iso Q_2/Q_1$ systems considered here, the natural structure is as $9 \times 9$ block systems. Consider the velocity DoFs as pictured in either Figure 1 or Figure 2; both the $Q_2$ and $Q_1iso Q_2$ spaces have basis functions associated with the nodes of the quadrilateral mesh, as well as along horizontal and vertical mesh edges, and at the center of the mesh cells. Since there are 2 components (horizontal and vertical) to the velocity, this gives eight types of DoFs associated with the velocity discretization. The $Q_1$ pressure gives the ninth set of DoFs in the block $9 \times 9$ structure. A key realization necessary for the LFA is that, when the system matrices (or their infinite-grid analogs) are reordered blockwise, each block in the reordered system has Toeplitz structure. Thus, we can diagonalize each block in the system by a Fourier transform, as discussed in Section 4.4. After diagonalizing the blocks, the system can be reformed frequency-wise, into a block operator with $9 \times 9$ blocks, by collecting all rows/columns corresponding to each Fourier frequency into adjacent rows/columns of the operator. If we further order the frequencies in harmonic sets, we get a block $36 \times 36$ structure associated with each harmonic space, $F(\theta)$. While analysis of the relaxation operators is most natural using the $9 \times 9$ block structure, complete analysis of the two-grid convergence factor requires use of the $36 \times 36$ block structure.

The block symbol for the $Q_2/Q_1$ Stokes discretization was derived in earlier work as symbols for the $Q_2$ Laplacian and the canonical finite-element interpolation and restriction operators for the $Q_2$ finite-element space. Symbols for both Braess-Sarazin and additive Vanka relaxation for this discretization were also computed in earlier work. There is, in principle, very little different to computing these symbols for the $Q_1iso Q_2/Q_1$ discretization. The reordered block structure of the $Q_1iso Q_2/Q_1$ saddle point matrix is the same as that of the $Q_2/Q_1$ matrix. Once again, each individual block of the $9 \times 9$ system is a Toeplitz matrix, though the specific entries are now different than for the $Q_2/Q_1$ case. For example, a block row associated with one type of velocity DoF would have five nonzero blocks. If this type corresponds to horizontal velocity DoFs located at pressure cell centers (denoted by a diamond marker in Figure 1), then four of these nonzero blocks correspond to types associated with horizontal velocity DoFs and the fifth block corresponds to the gradient operator acting on the pressure. The diagonal block within this block row is itself diagonal, reflecting the fact that there are no matrix connections between different velocity types in the $Q_1iso Q_2$ Laplacian matrix. Referring to the right side of Figure 1 this corresponds to the fact that two adjacent velocity DoFs in the velocity mesh always have different markers. Two of the other nonzero velocity blocks within this same block row are Toeplitz matrices with only 2 nonzeros per row. Again referring to the right side of Figure 1 the nonzeros in one of these blocks correspond to stencil entries associated with the North and South horizontal velocity DoFs, located on the horizontal edges of the mesh, which share the same marker. The two nonzeros in the other block correspond to stencil entries associated with the East and West horizontal velocity DoFs, now on vertical edges of the mesh, which again share the same marker. The final nonzero velocity block in this same block row would have four nonzeros that correspond to stencil entries associated with the corner horizontal velocity DoFs, located at the nodes of the mesh, which all share the same marker. The Toeplitz block for the gradient operator mapping onto these velocity DoFs would have four nonzero entries per row, associated with the four nodal pressure DoFs for the cell. It should be noted that the number of nonzeros per row in a gradient Toeplitz block depends on the specific block row that is being considered, i.e. the particular velocity marker associated with the block row in Figure 1. For the stencil representation of each Toeplitz matrix, standard Fourier techniques, as described above, give the block-structured symbol of the operator. Similarly, the symbols for interpolation and restriction can be computed in the $Q_1iso Q_2$ setting, and those for Braess-Sarazin and additive Vanka relaxation as well. Since the full calculations are rather tedious, we omit the details, noting that there are existing software packages that could be used to compute the symbols automatically.

5 | NUMERICAL RESULTS

To evaluate the effectiveness of $Q_1iso Q_2/Q_1$ multigrid preconditioners for the $Q_2/Q_1$ system, we start by comparing LFA predicted two-grid convergence factors against measured convergence factors for the Stokes system with Dirichlet boundary conditions on a uniform square mesh with $h = \frac{1}{64}$. For convenience, the $Q_2/Q_1$ discretization matrices are assembled using FireDrake, while the $Q_1iso Q_2/Q_1$ discretization matrices are assembled by forming a finer-mesh $Q_2/Q_1$ discretization, then applying coarsening-in-order to the velocity DoFs and coarsening-in-space to the pressure DoFs. The defect-correction preconditioner is implemented as a GMG variant taken from the PyAMG library, with our own implementation of the relaxation schemes. The measured convergence factors are then computed by running the solver 100 times using a zero right-hand side and a different random initial guess each time. The vector $\ell^2$ norm of the last $\min(m_s, 10)$ residuals are collected for each of the 100 runs, where $m_s$ is the number of iterations required to reduce the residual by a factor of $10^{-10}$ for the $s^{th}$ solve. When this criteria is not met in 100 iterations, we take $m_s = 100$. Collected residual norms are then fit by a linear least-squares model of
the form $\xi_0 + \xi_1 j = \ln r_j$, where $r_j$ is the $j^{th}$ residual within the $\min(m_r, 10)$ collected residuals from the $s^{th}$ run. The averaged asymptotic convergence factor is then taken as $\rho = e^{\bar{\xi}_1}$.

Within the multigrid cycle, the number of pre- and post-relaxation sweeps is always 1 for $\ell' \geq 1$. To distinguish between different preconditioners based on (5), we focus on the triple $(v_1, v_2, \gamma)$, where $v_1$ and $v_2$ are the number of pre-/post- relaxation sweeps on the $K_0$ system, and $\gamma$ is the number of $h$-multigrid cycles on the $K_1$ system. For fixed values of these parameters, we use the LFA optimization framework of Brown et al.\textsuperscript{[5]} to choose the remaining parameters in the method. Referring to (5), these parameters include a damping parameter for the $Q_1$-iso $Q_2/Q_1$ cycling, $\tau$, outer relaxation damping parameters for both the $Q_2/Q_1$ and $Q_1$-iso $Q_2/Q_1$ relaxation schemes and, if IBSR relaxation is used, inner relaxation parameters for that relaxation. The notation for these parameters is summarized in Table 1.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_0$</td>
<td>saddle point system for $Q_2/Q_1$</td>
</tr>
<tr>
<td>$K_1$</td>
<td>saddle point system for $Q_1$-iso $Q_2/Q_1$</td>
</tr>
<tr>
<td>$K_{\ell'}$</td>
<td>coarse saddle point $Q_1$-iso $Q_2/Q_1$ system for $\ell' \geq 2$</td>
</tr>
<tr>
<td>$M_{\ell'}$</td>
<td>relaxation operator on level $\ell'$</td>
</tr>
<tr>
<td>$v_1$</td>
<td>number of pre-relaxation sweeps used on the $K_0$ system</td>
</tr>
<tr>
<td>$v_2$</td>
<td>number of post-relaxation sweeps used on the $K_0$ system</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>number of $h$-multigrid cycles used to solve the $K_1$ system</td>
</tr>
<tr>
<td>$\tau$</td>
<td>$Q_1$-iso $Q_2/Q_1$-cycle solution damping parameter on level 1</td>
</tr>
<tr>
<td>$a_0$</td>
<td>Inexact Braess-Sarazin (IBS) relaxation parameter on level 0</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>weighted Jacobi relaxation parameter in IBSR on level 0</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>relaxation global update parameter on level 0</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>Braess-Sarazin (BS) relaxation parameter for all $\ell' \geq 1$</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>BS weighted Jacobi relaxation parameter for all $\ell' \geq 1$</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>relaxation global update parameter for all $\ell' \geq 1$</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>LFA predicted two-grid convergence factor</td>
</tr>
<tr>
<td>$\rho$</td>
<td>measured convergence factor for Dirichlet BC problem</td>
</tr>
<tr>
<td>$m$</td>
<td>measured iteration count for multigrid or preconditioned FGMRES</td>
</tr>
</tbody>
</table>

superscript $P$ problem with periodic boundary conditions

### TABLE 1 Notation Reference Table

#### 5.1 Two-grid Results

We start by verifying that the measured two-grid convergence factors, $\rho$, agree with the LFA predictions, $\hat{\rho}$. Table 2 demonstrates that there is generally good agreement. As expected, we also see that increasing the number of $K_0$ relaxation sweeps improves convergence. For example, increasing the number of IBSR sweeps from $(v_1, v_2, \gamma) = (1, 0, 1)$ to $(v_1, v_2, \gamma) = (2, 2, 1)$ improves the measured convergence factor from $0.16$ to $0.05$. Similar V-cycles for Vanka relaxation demonstrate an analogous trend, where the convergence factor decreases from $0.31$ to $0.15$. Interestingly, there is some notable variation in both the damping parameter, $\tau$, and the outer relaxation weights $\omega_\gamma$. In results not reported here, the optimal value for $\tau$ was found to be $1/2$ for a simpler preconditioner with error-propagation operator $I - \tau K_0^{-1} K_0$ (i.e., taking $v_1 = v_2 = 0$ and $\gamma \rightarrow \infty$). Here, we see that adding relaxation on the $K_0$ system and using a more practical approximation to $K_{\ell'}^{-1}$ result in values for $\tau$ much closer to $1$. Depending on the $(v_1, v_2, \gamma)$ values, we see values for $\omega_\gamma$ varying in the range from roughly $1/2$ to $1$, and similar variations in the inner IBSR parameters.

In order to test whether relaxation on both the $K_0$ and $K_1$ systems is necessary for robust convergence, we eliminate the relaxation on one of these systems by setting either $\omega_0 = 0$ or $\omega_1 = 0$. The results in Table 3 suggest that removing the $K_0$
TABLE 2 LFA predictions ($\hat{\beta}$) compared with the measured convergence factors ($\rho$) for inexact Braess-Sarazin (BS) and Vanka (V) relaxation. As $\alpha \gamma$ and $\beta \gamma$ are only needed for IBSR relaxation, we do not record values of these parameters for Vanka relaxation.

<table>
<thead>
<tr>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$\gamma$</th>
<th>$\tau$</th>
<th>$\omega_0$</th>
<th>$\omega_1$</th>
<th>$\alpha_0$</th>
<th>$\beta_0$</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
<th>$\hat{\beta}$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>.87</td>
<td>1.02</td>
<td>.90</td>
<td>1.20</td>
<td>.79</td>
<td>.74</td>
<td>.88</td>
<td>.18</td>
<td>.16</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>.86</td>
<td>.75</td>
<td>1.04</td>
<td>.99</td>
<td>.70</td>
<td>1.15</td>
<td>.88</td>
<td>.20</td>
<td>.18</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.97</td>
<td>.91</td>
<td>1.04</td>
<td>1.02</td>
<td>.93</td>
<td>.78</td>
<td>.86</td>
<td>.09</td>
<td>.11</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1.04</td>
<td>.55</td>
<td>.59</td>
<td>.72</td>
<td>.99</td>
<td>.85</td>
<td>1.30</td>
<td>.04</td>
<td>.05</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>.86</td>
<td>.78</td>
<td>1.01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.37</td>
<td>.31</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>.90</td>
<td>.98</td>
<td>.83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.21</td>
<td>.20</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>.85</td>
<td>1.02</td>
<td>.82</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.20</td>
<td>.19</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.99</td>
<td>.67</td>
<td>.93</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.29</td>
<td>.27</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1.05</td>
<td>.74</td>
<td>.78</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.12</td>
<td>.09</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.98</td>
<td>.71</td>
<td>1.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.16</td>
<td>.15</td>
</tr>
</tbody>
</table>

relaxation is more detrimental to convergence than removing relaxation on the $K_1$ system. Specifically, we compare $(v_1, v_2, \gamma) = (0, 0, 1)$ and $\omega_0 = 0$ with $(v_1, v_2, \gamma) = (1, 1, 1)$ and $\omega_1 = 0$. In this case, there is a single pre- and post-relaxation in the cycle, on either the $Q_1isoQ_2/Q_1$ or $Q_2/Q_1$ system. Similarly, we compare $(v_1, v_2, \gamma) = (0, 0, 2)$ and $\omega_0 = 0$ with $(v_1, v_2, \gamma) = (2, 2, 1)$ and $\omega_1 = 0$. Here, there are two pre- and post-relaxation sweeps in the cycle on either the $Q_1isoQ_2/Q_1$ or $Q_2/Q_1$ system, although using $\gamma = 2$ results in one additional coarse-grid correction in comparison to the $\gamma = 1$ case. We emphasize that setting $\omega_1 = 0$ omits only relaxation on the “fine grid” $K_1$ system in the two-level method. In the case of Braess-Sarazin with $\omega_0 = 0$ and $\gamma = 2$, the predicted convergence factor is 0.52 compared with 0.11 for relaxation only on $K_0$ using $(v_1, v_2, \gamma) = (2, 2, 1)$. Similar V-cycles for Vanka relaxation demonstrate less of a convergence factor reduction from 0.57 to 0.29. We also note that some of the measured convergence factors deviate from the LFA predictions in the case of Braess-Sarazin, while measured Vanka results agree closely with the LFA predictions. The two-grid convergence factors for periodic boundary condition problems are expected to match the predicted values exactly$^{30, 48, 49}$, while sometimes a gap between the LFA predicted factors and the measured factors is observed for the Dirichlet boundary condition case$^{30, 54}$.

TABLE 3 Effect of omitting relaxation on $K_0$ and $K_1$. $\omega_0 = 0$ corresponds only relaxation on the $Q_1isoQ_2/Q_1$ system, while $\omega_1 = 0$ corresponds to only relaxation on the $Q_2/Q_1$ system.

<table>
<thead>
<tr>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$\gamma$</th>
<th>$\tau$</th>
<th>$\omega_0$</th>
<th>$\omega_1$</th>
<th>$\alpha_0$</th>
<th>$\beta_0$</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
<th>$\hat{\beta}$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.93</td>
<td>.24</td>
<td>-</td>
<td>-</td>
<td>.67</td>
<td>1.56</td>
<td>.60</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>.63</td>
<td>0</td>
<td>.40</td>
<td>-</td>
<td>-</td>
<td>.52</td>
<td>1.35</td>
<td>.52</td>
<td>.51</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.77</td>
<td>.45</td>
<td>.46</td>
<td>1.37</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.32</td>
<td>.52</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.99</td>
<td>.85</td>
<td>.90</td>
<td>1.12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.11</td>
<td>.41</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>.64</td>
<td>.98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.63</td>
<td>.65</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>.66</td>
<td>0</td>
<td>.63</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.57</td>
<td>.55</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.48</td>
<td>.81</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.56</td>
<td>.51</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.80</td>
<td>.85</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.29</td>
<td>.25</td>
</tr>
</tbody>
</table>

Considering the information in Tables 2 and 3, we compare whether relaxation on both $K_0$ and $K_1$ is necessary or relaxation on $K_0$ system is sufficient. For example, Vanka relaxation with $(v_1, v_2, \gamma) = (1, 1, 1)$ in Table 2 relaxes on both systems, while taking $(v_1, v_2, \gamma) = (2, 2, 1)$ with $\omega_1 = 0$ in Table 3 use the same total number of relaxation sweeps, but only on the $K_0$ system. Their convergence factors are comparable, giving 0.27 and 0.25, respectively. For Braess-Sarazin, the LFA predicted
performance for the same cycle types is also comparable, giving 0.09 and 0.11, respectively. However, the measured Braess-Sarazin convergence factors deviate quite significantly from these predictions. Since there appears to be little benefit to omitting this relaxation even in the ideal setting of periodic boundary conditions, and potentially significant degradation in performance when used with Dirichlet boundary conditions, we do not consider the methods from Table 3 further and, instead, focus on methods where we allow relaxation on both $K_0$ and $K_1$.

Remark 1. Comparison with direct application of GMG to the $\mathbb{Q}_2/\mathbb{Q}_1$ discretization. The $hp$-multigrid convergence rates that we report are quite similar to the $\mathbb{Q}_2/\mathbb{Q}_1$ $h$-multigrid results reported in earlier work\cite{15, 34}. For Vanka relaxation, our two-grid cycle $(1, 1, 1)$ has the same total number of fine-level relaxation sweeps (considering both $K_0$ and $K_1$) as a $(2, 2)$ $h$-multigrid cycle, where a total of two pre- and two post- relaxation sweeps are performed. The measured convergence factor for a unit square domain with periodic boundary conditions is reported as 0.29 for the $h$-multigrid solver\cite{25}, which exactly matches our results. For IBSR based GMG, a measured convergence factor of 0.09 is reported\cite{15} with an $h$-multigrid $(2, 2)$ cycle for a problem on a unit square domain with Dirichlet boundary conditions, which closely matches our rate of .11 for a $(1, 1, 1)$ cycle in Table 3. Here, small differences could potentially arise due to the different number of weighted-Jacobi sweeps in the computation of the pressure correction (two sweeps for the $h$-multigrid benchmark versus one for our method). Overall, however it is clear that comparable convergence rates are obtained when either GMG is applied directly to the $\mathbb{Q}_2/\mathbb{Q}_1$ discretization or when it is applied indirectly via an intermediate $\mathbb{Q}_1$ iso-$\mathbb{Q}_2/\mathbb{Q}_1$ system.

In order to study the convergence rate sensitivity to parameter choices, we consider the two-grid convergence factor for the same Dirichlet boundary value problem as in the previous section. In case of Vanka relaxation, we consider the V(1, 1, 2) cycle from Table 2 where we fix $\tau$ and vary $\omega_0$ and $\omega_1$ from 0.02 to 1.0. In the left plot of Figure 4, we observe poor performance for the majority of $\omega_1$ values when $\omega_0$ is “too small”. However, for larger values of $\omega_0$, the performance becomes less sensitivity to perturbation in $\omega_1$. Multigrid solvers using inexact Braess-Sarazin relaxation have many more parameters than Vanka, which makes the sensitivity analysis slightly more complicated. For simplicity, we fix $\tau$, $\alpha_i$, and $\beta_i$ to be the same as in cycle BS(2, 2, 1) in Table 2 while varying $\omega_0$ and $\omega_1$. As seen in the right plot of Figure 4, there is a large region where we see near-optimal BS(2, 2, 1) convergence without having to re-optimize the $\alpha_i$ and $\beta_i$. This suggests that the performance of BS(2, 2, 1) is not too sensitive to its parameters. We note that, for both relaxation schemes, there is a clear advantage to an informed choice of the relaxation parameters over naive choices, such as $\omega_0 = \omega_1 = 1$, but the sensitivity to their choices is not so severe as to be a practical drawback to using these approaches.

![FIGURE 4](image)

**FIGURE 4** Sensitivity analysis of the measured two-grid convergence factors for V(1, 1, 2) (left) and BS(2, 2, 1) (right) cycles.

### 5.2 Multilevel results

We now consider five-level algorithms, where the coarsest grid is a $4 \times 4$ mesh. The test problem is the same as in the previous subsection, with a uniform grid on the unit square and $h = \frac{1}{64}$. We use the same parameters from Table 2 and compare the
multilevel convergence results against both the LFA predicted two-grid convergence factors and the two-level performance reported above. The multilevel algorithm is formed by recursively extending the $h$-multigrid hierarchy in operator $G_1$. For completeness, Table 4 measures both V- and W-cycle convergence factors for stationary multigrid iterations, as well as the number of iterations required to reduce the norm of the absolute residual based on $K_1$ to $10^{-10}$, using both stationary GMG iterations and multigrid preconditioned FGMRES. The measured convergence factors for W-cycles presented in Table 4 agree with the LFA predicted results above. The V-cycle convergence rates, on the other hand, tend to underperform for both periodic and Dirichlet boundary conditions. Even though the W-cycle preconditioned FGMRES solver tends to converge in the fewest number of iterations, the difference in iteration counts is not large, suggesting that V-cycle preconditioned FGMRES may be the more computationally efficient approach, particularly in a parallel setting. We note that the large discrepancy between the number of iterations required for convergence using stationary V-cycles and V-cycle preconditioned FGMRES suggests that there are only a few modes that are not well-captured by the GMG operators.

**TABLE 4** Predicted two-grid LFA convergence factors and measured convergence estimates for the multilevel method applied to the unit-square test problem. A $P$ superscript denotes use of periodic boundary conditions, while no superscript denotes Dirichlet boundary conditions. The last four columns list iteration counts to achieve a relative residual reduction tolerance using either stationary GMG iterations or GMG as a preconditioner for FGMRES.

<table>
<thead>
<tr>
<th>$\nu_1$</th>
<th>$\nu_2$</th>
<th>$\gamma$</th>
<th>Cycle</th>
<th>$\hat{\rho}$</th>
<th>$\rho^P$</th>
<th>$\rho$</th>
<th>GMG</th>
<th>w/ FGMRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>V</td>
<td>.18</td>
<td>.17</td>
<td>.20</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>V</td>
<td>.09</td>
<td>.25</td>
<td>.55</td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>W</td>
<td>.17</td>
<td>.17</td>
<td>.17</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>W</td>
<td>.08</td>
<td>.08</td>
<td>.08</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>V</td>
<td>.21</td>
<td>.47</td>
<td>.20</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>W</td>
<td>.20</td>
<td>.19</td>
<td>.20</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>V</td>
<td>.12</td>
<td>.18</td>
<td>.18</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>W</td>
<td>.10</td>
<td>.10</td>
<td>.09</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

We note that these results use damping parameters obtained from the two-level analysis. As the two-level analysis assumes exact coarse-grid solves and W-cycles are expected to provide more accurate coarse solution than V-cycles, the W-cycle case is somewhat closer to the assumptions underlying the two-level analysis. From this viewpoint, the better performance of the W-cycles is not surprising. Given the observed success using V-cycles as preconditioners for FGMRES, we have not undertaken a multilevel LFA to see if there might possibly be better parameter choices for the V-cycle.

### 5.3 Stokes backward-facing step problem

In this section, we consider the performance of the preconditioned FGMRES algorithms as we increase the fine-scale problem size and number of levels in the multigrid hierarchy. For this test, we consider a backward-facing step domain with the solution strategies discussed in Section 5.2 and verify that the resulting iteration counts are relatively stable with respect to mesh refinement. For all tests, the Stokes problem is defined on a backward-facing step domain with a parabolic inflow and natural outflow boundary conditions. A Poiseuille flow profile is imposed on the inflow boundary ($x = 0; 0 \leq y \leq 2$). A homogeneous Neumann boundary condition is imposed on the outflow boundary ($x = 2; 0 \leq y \leq 2$), thereby fixing the mean outflow pressure to be zero. No-slip (zero velocity) boundary conditions are imposed on all other boundary faces. More details on this type of problem can be found, for example, in Elman et al.

Table 5 summarizes the results, presenting iteration counts for multigrid preconditioned FGMRES. Using both IBSR and Vanka as relaxation within the preconditioner, we perform V- and W-cycle convergence tests with respect to problem size. The relaxation parameters are picked based on the most successful multigrid convergence results in Section 5.2. The parameter values for these two multigrid hierarchies can be found in Tables 2 and 3 under the corresponding ($\nu_1, \nu_2, \gamma$) values: (1, 1, 1)
for Braess-Sarazin and (1, 1, 2) for Vanka relaxation. For both IBSR and Vanka relaxation, the V-cycle results exhibit modest iteration growth as the problem size is increased. The convergence results for the W-cycle tests show constant iteration counts.

**TABLE 5** FGMRES iterations to convergence for backward-facing step problem. Braess-Sarazin relaxation parameters are based on the results in Table 2. Vanka relaxation parameters are based on the results in Table 4.

<table>
<thead>
<tr>
<th>DoFs</th>
<th>GMG levels</th>
<th>Method, ((v_1, v_2, \gamma))</th>
<th>BS, (1, 1, 1)</th>
<th>V, (1, 1, 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>515</td>
<td>2</td>
<td>V-cycle</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>1891</td>
<td>3</td>
<td></td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>7235</td>
<td>4</td>
<td></td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>28291</td>
<td>5</td>
<td></td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>111875</td>
<td>6</td>
<td></td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>444931</td>
<td>7</td>
<td></td>
<td>16</td>
<td>10</td>
</tr>
</tbody>
</table>

In addition to the number of FGMRES iterations in Table 5, we present the solve phase timings for Vanka and Braess-Sarazin based solvers in Figure 5. Both solvers’ timings scale proportionally with the number of DoFs in the system and the total number of FGMRES iterations. While a single W-cycle is more computationally expensive per iteration than a V-cycle, the degradation in performance when using V-cycles makes W-cycles the faster option, due to the constant number of iterations to convergence. This is true for both Braess-Sarazin and Vanka-based multigrid cycles. While we see generally faster times for the Braess-Sarazin cycles here, we refrain from drawing conclusions about the relative performance between the two relaxation schemes, because timings are highly dependent on the specifics of their implementation and optimization.

6 | CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrate that the low-order, \(Q_1\) iso \(Q_2/Q_1\), finite-element discretization can be used to construct effective preconditioners for the higher-order \(Q_2/Q_1\) finite-element discretization of the Stokes equations. To achieve effective performance, we use LFA in combination with robust optimization algorithms to compute relaxation parameters that optimize the resulting two-grid convergence factors for various two-grid algorithms based on the \(Q_1\) iso \(Q_2/Q_1\) hierarchy of grids. The measured two-grid convergence factors for both periodic and Dirichlet boundary conditions generally agree quite closely with the LFA-predicted two-grid convergence factors. For multilevel convergence, we observe close agreement between the measured W-cycle convergence factors and the LFA predictions. The V-cycle multilevel convergence factors, however, can deviate significantly more from the LFA predictions than the W-cycle. Both V- and W-cycles, however, lead to effective preconditioners for FGMRES for some parameter choices. When used on a more challenging backward-facing step problem, W-cycle preconditioned FGMRES leads to no growth in iteration counts as the mesh is refined, while modest growth in iterations is seen with V-cycle preconditioners. Hence for the large problem sizes, W-cycle preconditioned FGMRES solvers are faster.

An immediate next step in this research is to see whether these results can be leveraged in the development of monolithic AMG algorithms for the \(Q_2/Q_1\) discretization of the Stokes equations. Despite a long history of research effort, there has been little success in developing true algebraic MG approaches for saddle-point systems such as these. Preliminary numerical results, that we intend to report in a future manuscript, suggest that it is easier to develop AMG algorithms for the \(Q_1\) iso \(Q_2/Q_1\) discretization as we consider here. If successful, further work is possible for higher-order discretizations, such as the Scott-Vogelius discretization of the (Navier-) Stokes equations, or for coupled systems of equations of saddle-point type, such as viscoresistive magnetohydrodynamics.
FIGURE 5 Timing of Braess-Sarazin and Vanka relaxation-based V- and W-cycles. The dashed line represents the scaling with problem size in the form of $c \cdot \text{DoFs} \cdot m$, for some constant $c$ and where $m$ is the number of FGMRES iterations.

The source code used to collect data for this paper is publicly available at https://github.com/lexeyV/Stokes_isoQ2Q1. It is implemented in Python 3 under the 3-Clause BSD License. Version 1.0 is used in this paper, under commit 6ba1fa393dc0c113a01145363695afa49c9d7d.

ACKNOWLEDGMENTS

The work of SPM was partially supported by an NSERC Discovery Grant. RT was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under grant DE-NA-0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References


