# LEAST-SQUARES PROPER GENERALIZED DECOMPOSITIONS FOR WEAKLY COERCIVE ELLIPTIC PROBLEMS*

THOMAS L. D. CROFT† AND TIMOTHY N. PHILLIPS†

**Abstract.** Proper generalized decompositions (PGDs) are a family of methods for efficiently solving high-dimensional PDEs, which seek to find a low-rank approximation to the solution of the PDE a priori. Convergence of PGD algorithms can only be proven for problems which are continuous, symmetric, and strongly coercive. In the particular case of problems which are only weakly coercive we have the additional issue that weak coercivity estimates are not guaranteed to be inherited by the low-rank PGD approximation. This can cause stability issues when employing a Galerkin PGD approximation of weakly coercive problems. In this paper we propose the use of PGD algorithms based on least-squares formulations which always lead to symmetric and strongly coercive problems and hence provide stable and provably convergent algorithms. Taking the Stokes problem as a prototypical example of a weakly coercive problem, we develop and compare rigorous least-squares PGD algorithms based on continuous least-squares estimates for two different reformulations of the problem. We show that these least-squares PGDs provide a much stabler algorithm than an equivalent Galerkin PGDs, and we provide proofs of convergence of the algorithms.

**Key words.** least-squares methods, proper generalized decomposition, weakly coercive

**AMS subject classifications.** 65N12, 65N30

**DOI.** 10.1137/15M1049269

**1. Introduction.** The systems of partial differential equations (PDEs) arising from many models in science and engineering are defined in high-dimensional spaces. Examples include the kinetic theory description of polymer dynamics, option pricing in financial mathematics, and quantum chemistry. The numerical solution of these systems represents a tremendous computational challenge since they exhibit the so-called curse of dimensionality when standard methods of discretization are used. This issue arises since many algorithms do not scale well with increasing dimensions, typically requiring computational effort (time or memory) that is exponential in the number of dimensions. Therefore, new algorithms are required to circumvent the curse of dimensionality to make the problems tractable.

One possibility lies in the use of sparse grids [12]. However, Achdou and Pironneau [1] argue that the use of sparse grids is restricted to models that possess moderate dimensionality. An alternative approach uses low-rank tensor methods to search for an approximation of the solution in a low-dimensional subset of the solution space. Classical low-rank subsets include canonical tensors and Tucker tensors and their variants [25].

Proper generalized decompositions (PGDs) are a relatively new family of methods which were introduced by Ammar et al. [4] for the efficient approximation of the solution to PDEs defined in high-dimensional spaces. The main concept underpinning all PGD algorithms is the approximation of the solution, $u$, to a $d$-dimensional PDE

†School of Mathematics, Cardiff University, Cardiff, CF24 4AG, Wales, UK (CroftT@cf.ac.uk, PhillipsTN@cf.ac.uk).

by the separated representation

$$(1.1) \qquad u_J(x_1, \ldots, x_d) = \sum_{j=1}^{J} F_{1,j}(x_1) \cdots F_{d,j}(x_d),$$

where $J$ is known as the rank of the PGD approximation. We say that

$$(1.2) \qquad u(x_1, \ldots, x_d) = \sum_{j=1}^{\infty} F_{1,j}(x_1) \cdots F_{d,j}(x_d)$$

if and only if

$$\lim_{J \to \infty} \|u - u_J\| = 0.$$

The importance of this separated representation in the numerical analysis of high-dimensional problems was first noted by Beylkin and Mohlenkamp [7]. In the case of the PGD the use of such an approximation can drastically reduce the number of unknowns required to solve high-dimensional PDEs. Indeed, given an $N$-node discretization, the number of unknowns in a standard mesh-based approximation of the solution to a $d$-dimensional PDE is $N^d$, whereas for the PGD there are $N \times J \times d$ unknowns. In other words, the complexity of a mesh-based approximation scales exponentially with increasing dimension, while the PGD scales linearly, clearly a vast improvement. Indeed, Ammar et al. [4] have used the PGD to accurately approximate the solution of a 100-dimensional Poisson equation, a truly phenomenal feat.

There are many variant algorithms based on the PGD (see Nouy [28]), but in this paper we consider only the simplest PGD algorithm: the progressive PGD, for which convergence is not guaranteed. In this algorithm the "best" (not necessarily optimal) rank-one separated representation of the true solution to the given PDE, $F_{1,j}(x_1) \cdots F_{d,j}(x_d)$, is found iteratively for each $j = 1, \ldots, J$. These rank-one separated representations are known as the PGD modes. At the beginning of each iteration, previously calculated PGD modes are simply moved to the right-hand side (RHS) of the equation before the next PGD mode is calculated. For Galerkin PGDs the best PGD mode is simply the one which satisfies Galerkin orthogonality with an appropriate choice of test function. This leads to a nonlinear system of equations which can be solved using an alternating directions fixed point linearization. We direct the interested reader to an in-depth description of a Galerkin progressive PGD algorithm for $d = 3$ by Chinesta et al. [18].

Le Bris, Lelièvre, and Maday [26] have recently associated progressive PGDs with greedy algorithms, the likes of which have been studied by Temlyakov [31]. The best choice of PGD mode for these greedy algorithms is the one which minimizes a certain functional whose minimizer is equivalent to the weak solution of the PDE. For variational problems which are continuous, symmetric, and strongly coercive this arises naturally as the minimization of an energy functional. The major advantage of considering PGD as a greedy algorithm is that it allows one to prove convergence of such algorithms. Unfortunately, this means that, for variational problems which are neither continuous and symmetric nor strongly coercive, no equivalent minimization principle exists, and hence it is not possible to prove convergence of the algorithm. Furthermore, stability conditions for problems which are only weakly coercive, such as the Ladyzhenskaya–Babuska–Brezzi (LBB) condition in the Stokes problem, are no longer guaranteed to hold in the PGD framework. It is for this reason that, in this paper, we investigate PGD algorithms based on least-squares formulations rather than those based on the more commonly used Galerkin PGD.

The main idea behind least-squares methods is the minimization of the residual of a differential operator in a carefully selected norm. This provides one with an "artificial" energy functional. Provided that certain coercivity estimates hold on the chosen norms, this can be used to prove convergence of an associated greedy algorithm. Billaud-Friess, Nouy, and Zahm [8] proposed an algorithm for generating low-rank approximations of solutions of weakly coercive problems based on a minimal residual method. The residual norm is constructed so that the resulting low-rank approximations are optimal with respect to particular norms. Convergence of the algorithm is proved under some conditions on the parameters of the algorithm. The normal equations resulting from the minimal residual approach are elliptic for many problems. The approach adopted in the present paper uses ADN (Agmon–Douglas–Nirenberg) theory [2, 3] to construct least-squares formulations that are homogeneous elliptic. The main advantage of this approach is that the ADN theory identifies least-squares reformulations of the problem that are practical in the sense of Bochev and Gunzberger [9], i.e., those which avoid the computation of dual norms. This approach generates symmetric and strongly coercive problems and hence leads to provably convergent algorithms. The strong coercivity of the least-squares formulation has the added benefit that it bypasses the stability conditions associated with weakly coercive problems. This is similar to the notion of the minimal residual PGD (e.g., [16, 28]). However, we use the terminology "least-squares PGD" to highlight the fact that we construct PGD algorithms based on rigorously defined least-squares principles. In this paper we focus on weakly coercive problems, taking the Stokes problem as a prototypical example. In particular, we highlight the advantages and disadvantages of this approach and investigate how the choice of least-squares formulation affects the convergence behavior of the algorithms.

This paper is structured as follows. In section 2 we briefly describe the PGD in the context of a pure greedy algorithm. In section 3 we show how stability conditions in weakly coercive problems are not guaranteed to be satisfied in the Galerkin PGD and provide some results for the Stokes problem. Section 4 provides a brief description of least-squares methods for an abstract problem, the ADN theory is summarized in section 5, and two proofs of convergence for least-squares PGDs are provided for this abstract problem in section 6. In section 7 we develop least-squares PGD algorithms for the Stokes problem and provide numerical results. Finally, in section 8, some conclusions are provided as well as some areas for future work.

**2. The pure greedy algorithm.** Consider the following variational problem: Find $u \in X$ such that

$$A(u,v) = L(v) \quad \forall v \in X \tag{2.1}$$

for some suitable tensor product Hilbert space $X$, linear form $L(\cdot)$, and continuous, symmetric, and strongly coercive bilinear form $A(\cdot,\cdot)$; i.e.,

$$|A(u,v)| \leq \alpha \|u\|_X \|v\|_X, \tag{2.2}$$

$$A(u,v) = A(v,u), \tag{2.3}$$

$$A(u,u) \geq \beta \|u\|_X^2, \tag{2.4}$$

for all $u, v \in X$ for some constants $\alpha, \beta > 0$. Under these assumptions the variational problem (2.1) is equivalent to the following minimization problem: Find $u \in X$ such that

$$u = \arg\min_{v \in X} \left( \frac{1}{2} A(v,v) - L(v) \right). \tag{2.5}$$

The pure greedy algorithm (see, e.g., Temlyakov [31]) can then be defined using this minimization problem in the following way.

---

**Algorithm 1.** Pure greedy algorithm.

---

**Require:** $(L(\cdot),\ \epsilon)$
 1: $n = 0$
 2: **while** $|L(\cdot)| \geq \epsilon$ **do**
 3:     $n \leftarrow n + 1$
 4:     $u = \arg \min_{v \in \mathcal{S}_1} \left( \frac{1}{2} A(v, v) - L(v) \right)$
 5:     $L(\cdot) \leftarrow L(\cdot) - A(u, \cdot)$
 6: **end while**

---

In this algorithm $\mathcal{S}_1$ is known as the dictionary, and it denotes the set of all rank-one tensors in the tensor product Hilbert space $X$:

$$\mathcal{S}_1 = \left\{ u \in X \ : \ \min_{J \in \mathbb{N}} \|u - u_J\|_X = 1 \right\}.$$

This pure greedy algorithm was first proven to converge in the context of a Galerkin formulation of the Poisson problem by Le Bris, Lelièvre, and Maday [26]. In this case if the variational problem (2.1) corresponds to the Galerkin formulation of the problem, then the minimization problem (and hence the pure greedy algorithm) can only be defined if the assumptions on the bilinear form (2.2)–(2.4) hold. For least-squares formulations of the problems, however, these assumptions are automatically satisfied, and hence a provably convergent pure greedy algorithm can always be defined, as will be shown in section 4.

In practice, progressive Galerkin PGD algorithms are performed by finding, at each iteration, the rank-one tensor which satisfies Galerkin orthogonality, and hence PGD algorithms can be defined for problems where assumptions (2.2)–(2.4) do not hold. However, the convergence behavior of these algorithms is not known since we cannot associate it with any greedy algorithm. In the case when the assumptions do hold, this is equivalent to solving the Euler–Lagrange equations associated with the minimization problem (2.5). Unfortunately, the set $\mathcal{S}_1$ over which we minimize is not a linear space, and hence solving the Euler–Lagrange equations is only a necessary condition for solving the minimization problem (2.5). The minimization problem itself is far too expensive to solve directly, and in this sense the progressive PGD can be viewed as a practical implementation of the pure greedy algorithm.

A further consideration relevant to the practical implementation of PGD algorithms is that the solution of the Euler–Lagrange equations associated with the minimization problem (2.5) leads to a nonlinear system. This system can be linearized by simply using an alternating directions fixed point algorithm (ADFPA), which was demonstrated to be particularly robust by Ammar et al. [4]. This linearization involves solving problems associated with each coordinate direction iteratively until some predefined convergence criterion is satisfied.

In the next section we highlight the further issue of numerical stability in progressive Galerkin PGD algorithms for the situation when the bilinear form $a(\cdot, \cdot)$ is assumed to be only weakly coercive.

**3. Galerkin PGD for weakly coercive problems.** Consider a continuous, weakly coercive problem: Find $u \in X$ such that

$$(3.1) \qquad A(u,v) = L(v) \quad \forall v \in Y$$

for suitable Hilbert spaces $X$ and $Y$ with a bilinear form that satisfies continuity:

$$(3.2) \qquad |A(u,v)| \leq \alpha \|u\|_X \|v\|_Y$$

for some constant $\alpha > 0$. We further assume that the bilinear form satisfies the following weak coercivity estimates:

$$(3.3) \qquad \inf_{v \in Y} \sup_{u \in X} \frac{A(u,v)}{\|u\|_X \|v\|_Y} \geq \beta, \qquad \inf_{u \in X} \sup_{v \in Y} \frac{A(u,v)}{\|u\|_X \|v\|_Y} \geq \gamma$$

for some constants $\beta, \gamma > 0$. In the case when $A(\cdot, \cdot)$ is symmetric, these two estimates are equivalent.

The coercivity estimates (3.3), together with continuity (3.2), are sufficient conditions for well-posedness of weakly coercive problems due to the Babuška–Lax–Milgram theorem [6]. Unlike strong coercivity estimates such as (2.4), weak coercivity estimates are not always guaranteed to be preserved on discrete subspaces $X^h \subset X$, $Y^h \subset Y$. This can lead to instabilities caused by spurious oscillations (see Brezzi and Fortin [11] in the particular case of saddle-point problems). However, in the case of a PGD approximation, it is not at all clear whether the weak coercivity estimates are preserved for the discrete problem or, indeed, how one can choose a suitable dictionary in order to ensure that they are.

To describe this issue further, and to provide some numerical experiments, we consider a particular weakly coercive problem: the Stokes problem. In this case the bilinear form $A(\cdot, \cdot)$ is defined by

$$A(\{\mathbf{u}, p\}, \{\mathbf{u}^*, p^*\}) := a(\mathbf{u}, \mathbf{u}^*) + b(\mathbf{u}, p^*) + b(\mathbf{u}^*, p), \qquad L(\{\mathbf{u}^*, p^*\}) := l(\mathbf{u}^*)$$

for all $\{\mathbf{u}, p\}, \{\mathbf{u}^*, p^*\} \in (H_0^1(\Omega))^d \times L^2(\Omega)$, where $d = 2, 3$ is the spatial dimension of the problem, and the bilinear forms $a(\cdot, \cdot)$, $b(\cdot, \cdot)$ and linear form $l(\cdot)$ are given by

$$a(\mathbf{u}, \mathbf{u}^*) = \int_\Omega \nabla \mathbf{u} : \nabla \mathbf{u}^* \ d\Omega, \qquad b(\mathbf{u}^*, p) = -\int_\Omega p(\nabla \cdot \mathbf{u}^*) d\Omega,$$

$$l(\mathbf{u}^*) = \int_\Omega \mathbf{f} \cdot \mathbf{u}^* \ d\Omega$$

for some given source term $\mathbf{f}$, where $\mathbf{u}$ denotes the velocity and $p$ denotes the pressure. The weak Galerkin formulation of the Stokes problem is: Find $\{\mathbf{u}, p\} \in (H_0^1(\Omega))^d \times L^2(\Omega)$ such that

$$(3.4) \qquad a(\mathbf{u}, \mathbf{u}^*) + b(\mathbf{u}^*, p) = l(\mathbf{u}^*),$$

$$(3.5) \qquad b(\mathbf{u}, p^*) = 0$$

for all $\{\mathbf{u}^*, p^*\} \in (H_0^1(\Omega))^d \times L^2(\Omega)$. The weak coercivity estimates (3.3) can be expressed in terms of these constituent bilinear forms.

The important estimate for the Stokes problem is the inf-sup condition. Its discrete analogue is given by

$$(3.6) \qquad \inf_{p^h \in Q^h} \sup_{\mathbf{u}^h \in V^h} \frac{b(\mathbf{u}^h, p^h)}{\|\mathbf{u}^h\|_{V^h} \|p^h\|_{Q^h}} \geq \beta_b^h$$

for some constant $\beta_b^h > 0$, where $V^h \subset (H_0^1(\Omega))^d$, $Q^h \subset L^2(\Omega)$ are some chosen discretization spaces. The discrete estimate (3.6) is known as the LBB condition and does not hold for all choices of $V^h$ and $Q^h$ [11]. However, it is possible to select discrete spaces which satisfy the LBB condition, e.g., the $P_N - P_{N-2}$ spectral element method of Maday, Patera, and Rønquist [27]. However, in the PGD, we no longer seek solutions in linear subspaces $V^h$ and $Q^h$ but instead in the nonlinear manifold of rank-one tensors. To this end, we require an additional assumption in the form of the weak estimate:

$$(3.7) \qquad \inf_{p \in \mathcal{S}_1^p} \sup_{\mathbf{u} \in \mathcal{S}_1^{\mathbf{u}}} \frac{b(\mathbf{u}, p)}{\|\mathbf{u}\|_{H^1} \|p\|_{L^2}} \geq \beta_b^*$$

for some constant $\beta_b^* > 0$, where $\mathcal{S}_1^{\mathbf{u}} \subset (H_0^1(\Omega))^d$ and $\mathcal{S}_1^p \subset L^2(\Omega)$ are the subsets of rank-one tensors in velocity and pressure, respectively. Unfortunately, it is not clear whether the sets $\mathcal{S}_1^{\mathbf{u}}$ and $\mathcal{S}_1^p$ can be chosen so that (3.7) is satisfied. Furthermore, in practice, we seek a solution in a discrete subset of $\mathcal{S}_1^p \times \mathcal{S}_1^{\mathbf{u}}$ which adds further uncertainty in the satisfaction of the discrete analogue of the LBB-like stability condition (3.7). This raises the question of stability of Galerkin PGD algorithms for the Stokes problem and, more generally, for weakly coercive problems.

We now provide a numerical example in which a progressive Galerkin PGD algorithm is applied to the Stokes problem. The discretization is based on a spectral element method [27] in which the discretization space for the pressure involves polynomial basis functions two degrees lower than that of the velocity. This scheme has been employed in an attempt to ensure the LBB-like stability of the algorithm. A detailed description of how spectral element methods, as well as the $P_N - P_{N-2}$ scheme, are incorporated into PGD algorithms can be found in the thesis of Croft [19].

Consider the Stokes problem (3.4)–(3.5) defined on the square domain $\Omega = [-1, 1]^2$ with homogeneous Dirichlet boundary conditions on the velocity, $\mathbf{u} = \mathbf{0}$ on $\partial\Omega$, and

$$(3.8) \qquad \mathbf{f}(x, y) = \begin{pmatrix} \pi y \cos(\pi x y) + 4\pi^2 \sin(2\pi y)(2\cos(2\pi x) - 1) \\ \pi x \cos(\pi x y) - 4\pi^2 \sin(2\pi x)(2\cos(2\pi y) - 1) \end{pmatrix}.$$

This problem has the following exact solution:

$$\mathbf{u} = \begin{pmatrix} -\sin(2\pi y)(\cos(2\pi x) - 1) \\ \sin(2\pi x)(\cos(2\pi y) - 1) \end{pmatrix}, \quad p = \sin(\pi x y).$$

Note that both velocity components possess a natural rank-one separated representation, whereas the pressure possesses an infinite-rank separated representation. A unique pressure is obtained by imposing a zero mean pressure constraint,

$$\int_\Omega p \, d\Omega = 0,$$

in a postprocessing step.

The $L^2$-norm of the error in the computed PGD approximation is plotted in Figure 1 against the rank of the separated representation. The rank-one separated representation of the velocity has been successfully captured by this algorithm despite the velocity's being coupled with an infinite-rank pressure solution. The error in the pressure is found to decrease rapidly until it reaches a level comparable with the discretization error. Note that the pressure approximation does not converge to a
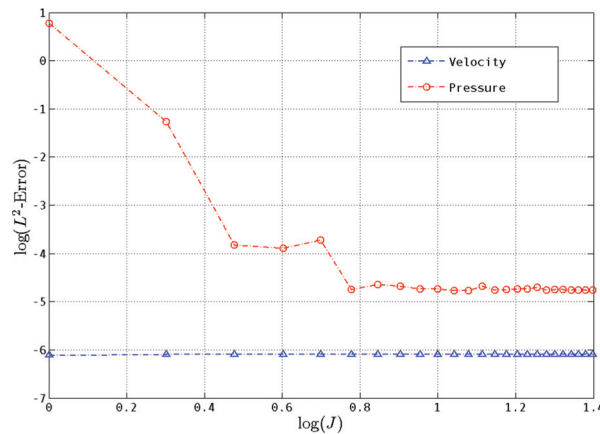
FIG. 1. *Convergence in increasing rank.*

level of accuracy comparable to that of the velocity, since the pressure approximation space comprises polynomial basis functions which are two degrees lower. Overall this convergence behavior seems very promising.

Although the progressive Galerkin PGD algorithm for the Stokes problem appears to perform reasonably well, appearances can be deceptive. The results in Figure 1 were generated using a discrete approximation with $N = 15$ on a single element, so essentially this is a high-order spectral method rather than a spectral element method. This choice of discretization was one of very few that yielded a converged approximation. For most choices of the discretization parameters the ADFPA failed to converge. This is a significant issue since there is no indication a priori which choices of discretization parameters will yield a convergent algorithm. This would be a serious disadvantage for more complex fully three-dimensional problems where we cannot rely on trial and error in order to produce a converged solution. It also does not rule out the possibility that for certain problems there may not be any choice of discretization parameter which yields a convergent algorithm.

In order to try and minimize the possibility of these issues being caused by ill-conditioning of the linear systems propagating error throughout the ADFPA, a minimal residual (MINRES) iterative solver with a standard block preconditioner for the Stokes problem (see, e.g., Elman, Silvester, and Wathen [22]) was employed. This technique has been shown to be particularly effective for the solution of the Stokes problem. However, it did not lead to any improvement in the behavior of the PGD algorithm. Instead, we speculate that the convergence problem relates to a lack of LBB-like stability whereby these choices of discretization fail to satisfy the discrete analogue of the rank-one tensor inf-sup condition (3.7). Overall, the evidence suggests that the progressive Galerkin PGD algorithm for the Stokes problem is an unreliable and inefficient algorithm.

Clearly an alternative approach is needed for the efficient PGD approximation of weakly coercive problems. Consideration is now given to least-squares methods, since they are able to supply a strongly coercive setting for all elliptic problems.

**4. Abstract least-squares formulation.** Consider the following abstract boundary value problem:

$$\mathcal{L}u = f \quad \text{in } \Omega, \tag{4.1}$$

$$\mathcal{R}u = g \quad \text{on } \Gamma = \partial\Omega, \tag{4.2}$$

where $\mathcal{L}$ is a linear elliptic partial differential operator, $\mathcal{R}$ is a trace operator, and $f$ and $g$ are given functions. Furthermore, it is assumed that $\mathcal{L}$ is a first-order differential operator since any higher-order problem can be recast in terms of equivalent systems of first-order differential equations. In least-squares methods, first-order reformulations of the problem are required in order to obtain a practical approach (see, e.g., Cai et al. [13]). This is because the required differentiability can no longer be reduced by one order using Green's first integral identity, as in the case of Galerkin methods.

If the boundary value problem (4.1)–(4.2) is well-posed and there exists a homeomorphism $\{\mathcal{L}, \mathcal{R}\} : X \to Y \times Z$, where $X = X(\Omega)$, $Y = Y(\Omega)$, and $Z = Z(\Gamma)$ are some underlying Hilbert spaces with norms $\|\cdot\|_X$, $\|\cdot\|_Y$, and $\|\cdot\|_Z$, respectively, then there exist constants $C_1, C_2 > 0$ such that

$$C_1\|u\|_X \le \|\mathcal{L}u\|_Y + \|\mathcal{R}u\|_Z \le C_2\|u\|_X \quad \forall u \in X. \tag{4.3}$$

Let $\tilde{u}$ denote the unique solution of (4.1)–(4.2); then using the inequality (4.3), we can write

$$C_1\|u - \tilde{u}\|_X \le \|\mathcal{L}u - f\|_Y + \|\mathcal{R}u - g\|_Z \le C_2\|u - \tilde{u}\|_X \quad \forall u \in X. \tag{4.4}$$

This norm equivalence between the error in the the $X$-norm and the residual in the differential equation in the $Y \times Z$-norm is termed the coercivity estimate (or a priori estimate) and is the key ingredient in the analysis of least-squares methods. This is due to the fact that (4.4) implies that if there exists a sequence of functions $u_n \in X$ such that $\|\mathcal{L}u_n - f\|_Y \to 0$ and $\|\mathcal{R}u_n - g\|_Z \to 0$ as $n \to \infty$, then $\|u_n - \tilde{u}\|_X \to 0$ as $n \to \infty$, and vice versa. This means that the sequence $u_n$ converges to the true solution in the $X$-norm. Therefore, minimization of the convex functional

$$\mathcal{J}(u) = \|\mathcal{L}u - f\|_Y^2 + \|\mathcal{R}u - g\|_Z^2 \quad \forall u \in X \tag{4.5}$$

yields the unique solution $\tilde{u}$ to the boundary value problem (4.1)–(4.2). The functional $\mathcal{J}$ is known as the quadratic least-squares functional. A pure greedy algorithm for any least-squares formulated problem is defined by replacing line 4 of Algorithm 1 by the minimization

$$u = \arg \min_{v \in \mathcal{S}_1} \mathcal{J}(v). \tag{4.6}$$

In practice, it is the Euler–Lagrange equations associated with this minimization that are solved when implementing PGD algorithms: Find $u \in X$ such that

$$\lim_{\epsilon \to 0} \frac{d}{d\epsilon} \mathcal{J}(u + \epsilon v) = 0 \quad \forall v \in X.$$

The numerical results presented in this paper are generated using this approach. This leads to the following variational formulation: Find $u \in X$ such that

$$A(u, v) = L(v) \quad \forall v \in X, \tag{4.7}$$

where

$$A(u,v) = \langle \mathcal{L}u, \mathcal{L}v \rangle_Y + \langle \mathcal{R}u, \mathcal{R}v \rangle_Z, \qquad L(v) = \langle f, \mathcal{L}v \rangle_Y + \langle g, \mathcal{R}v \rangle_Z,$$

and $\langle \cdot, \cdot \rangle_Y$ and $\langle \cdot, \cdot \rangle_Z$ denote the $Y$ and $Z$ inner products, respectively. Note that since $A(\cdot, \cdot)$ was derived from the minimization of a continuous problem, it automatically satisfies assumptions (2.2)–(2.4). Further note that the minimization (2.5) is equivalent, but not identical, to the least-squares minimization (4.6).

**5. ADN theory.** The theory of Agmon, Douglis, and Nirenberg was developed in a series of two papers [2, 3]. Its importance in relation to least-squares methods was first highlighted by Aziz, Kellogg, and Stephens [5], and the elements of ADN theory relevant to this paper are summarized here. It is a particularly powerful tool since it reduces the verification of continuous estimates to the verification of some algebraic constraints. One can then use this to obtain the lower bound in the coercivity estimate (4.3), whereas the upper bound follows from the continuity of the operators $\mathcal{L}$ and $\mathcal{R}$. The main ingredient in the proof of convergence of least-squares PGD algorithms is the strong coercivity of the functional $\mathcal{J}$. In this paper we show for the first time that the coercivity condition is related to ADN theory.

The approach based on ADN theory presented in this paper has the same objective as the approach based on proper Riesz maps used by Billaud-Friess, Nouy, and Zahm [8], in that they both seek appropriate choices of spaces $Y$ and $Z$ so that (4.3) holds ideally with $C_1 = C_2 = 1$. The approach of [8] replaces the minimization of an ideal choice of residual norm by an approximate minimization problem, which is then solved using a perturbation of the ideal gradient-type algorithm. However, the approach described in this paper, in which ADN theory not only provides the appropriate residual norms but also allows one to identify an underlying first-order system that is strongly coercive and which avoids computation of dual norms, is novel and represents an important distinction between the two approaches.

Consider the abstract boundary value problem (4.1)–(4.2), where $\mathcal{L} = \mathcal{L}_{i,j}(D)$, $i,j = 1, \ldots, n$; $\mathcal{R} = \mathcal{R}_{l,j}(D)$, $l = 1, \ldots, m$, $j = 1, \ldots, n$; $n$ is the number of dependent variables; $m$ is the number of boundary conditions; $D = (\partial/\partial x_1, \ldots, \partial/\partial x_d)^T$; and $d$ is the number of independent variables (the dimension). The usual definition of ellipticity is that $\det(\mathcal{L}_{i,j}^p(\boldsymbol{\xi})) \neq 0$ for all real-valued $\boldsymbol{\xi} \neq \mathbf{0}$, where the principal part $\mathcal{L}^p$ of $\mathcal{L}$ is

$$\mathcal{L}^p \equiv \mathcal{L}_{i,j}^p(D) = \begin{cases} \mathcal{L}_{i,j}(D) & \text{if } \deg(\mathcal{L}_{i,j}(\boldsymbol{\xi})) = \max_{k,l} \deg(\mathcal{L}_{k,l}(\boldsymbol{\xi})), \\ 0 & \text{otherwise,} \end{cases}$$

where $\deg(\mathcal{L}_{i,j}(\boldsymbol{\xi}))$ is the degree of the polynomial $\mathcal{L}_{i,j}(\boldsymbol{\xi})$. The idea of ellipticity is extended to the more general idea of ADN ellipticity by introducing two sets of integer indices: the set $\{s_i\}$, $s_i \leq 0$, assigned to the $n$ equations, and the set $\{t_j\}$, $t_j \geq 0$, assigned to the $n$ dependent variables. These indices are chosen in such a way that for each $i, j = 1, \ldots, n$, $\deg(\mathcal{L}_{i,j}(\boldsymbol{\xi})) \leq s_i + t_j$. The principal part $\mathcal{L}^p$ is then defined to be

$$\mathcal{L}^p \equiv \mathcal{L}_{i,j}^p(D) = \begin{cases} \mathcal{L}_{i,j}(D) & \text{if } \deg(\mathcal{L}_{i,j}(\boldsymbol{\xi})) = s_i + t_j, \\ 0 & \text{otherwise.} \end{cases}$$

The principal part $\mathcal{R}^p$ can be defined analogously by introducing the set of indices $\{r_l\}$, $r_l \leq 0$, assigned to the $m$ boundary conditions such that $\deg(\mathcal{R}_{l,j}(\boldsymbol{\xi})) \leq r_l + t_j$. Note that the choice of indices is not in general unique, and hence problems can possess more than one principal part. We are now in a position to present the following definition of ADN ellipticity.

DEFINITION 5.1. *The linear differential operator $\mathcal{L}$ is called* ADN elliptic *if there exist integer sets $\{s_i\}$ and $\{t_j\}$ such that the following hold:*
  (i) $\deg(\mathcal{L}_{i,j}(\boldsymbol{\xi})) \leq s_i + t_j$,
 (ii) $\mathcal{L}_{i,j}(\boldsymbol{\xi}) \equiv 0$ *if $s_i + t_j < 0$,*
(iii) $\det(\mathcal{L}_{i,j}^p(\boldsymbol{\xi})) \neq 0$ *for all real-valued $\boldsymbol{\xi} \neq 0$.*
*Furthermore, $\mathcal{L}$ is ADN elliptic of order $2m$ if $\deg(\det(\mathcal{L}_{i,j}^p(\boldsymbol{\xi}))) = 2m$, and uniformly ADN elliptic of order $2m$ if there exists a constant $c_e > 0$ such that*

$$c_e^{-1}|\boldsymbol{\xi}|^{2m} \leq |\det(\mathcal{L}_{i,j}^p(\boldsymbol{\xi}))| \leq c_e|\boldsymbol{\xi}|^{2m},$$

*where $m$ is the number of prescribed boundary conditions.*

The following supplementary condition must be satisfied by $\mathcal{L}$.

DEFINITION 5.2 (supplementary condition). *Let $\mathcal{L}$ be ADN elliptic of order $2m$; then the operator $\mathcal{L}$ is said to satisfy the supplementary condition if for all pairs of linearly independent real-valued vectors $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ the polynomial in $\tau$ given by $\det(\mathcal{L}_{i,j}^p(\boldsymbol{\xi} + \tau\boldsymbol{\xi}'))$ has exactly $m$ roots with positive imaginary part.*

*Remark* 5.1. The supplementary condition is automatically satisfied when $d > 2$ and only needs to be verified for $d = 2$ [3].

Let $\tau_k^+(\boldsymbol{\xi})$, $k = 1, \ldots, m$, denote the $m$ positive roots of $\det(\mathcal{L}_{i,j}^p(\boldsymbol{\xi} + \tau\boldsymbol{\xi}'))$ whose existence is ensured by the supplementary condition, and define $M^+(\boldsymbol{\xi}, \tau)$ to be the polynomial in $\tau$ for a given $\boldsymbol{\xi}$:

$$M^+(\boldsymbol{\xi}, \tau) = \prod_{k=1}^m \left( \tau - \tau_k^+(\boldsymbol{\xi}) \right).$$

Note that for a problem to be well-posed the operators $\mathcal{L}$ and $\mathcal{R}$ cannot be chosen independently but in such a way that their principal parts $\mathcal{L}^p$ and $\mathcal{R}^p$ "complement" one another. Accordingly, Agmon, Douglas, and Nirenberg [3] introduced a so-called complementing condition, as given next.

DEFINITION 5.3 (complementing condition). *For any point $P \in \Gamma$, let $\mathbf{n}$ be the outward unit normal vector to $\Gamma$ at $P$. For any real-valued $\boldsymbol{\xi} \neq \mathbf{0}$ tangent to $\Gamma$ at $P$ consider the matrix with the entries*

$$(5.1) \qquad \sum_{j=1}^n \mathcal{R}_{l,j}^p(\boldsymbol{\xi} + \tau\mathbf{n})\mathcal{L}_{j,k}'(\boldsymbol{\xi} + \tau\mathbf{n}),$$

*which are polynomials in $\tau$ and where $\mathcal{L}'$ denotes the adjoint matrix of $\mathcal{L}^p$. The operators $\mathcal{L}$ and $\mathcal{R}$ are said to satisfy the complementing condition if the rows of the matrix defined by (5.1) are linearly independent modulo $M^+(\boldsymbol{\xi}, \tau)$. In other words,*

$$\sum_{l=1}^m c_l \sum_{j=1}^n \mathcal{R}_{l,j}^p \mathcal{L}_{j,k}' \equiv 0 \pmod{M^+} \quad \forall k = 1, \ldots, n$$

*if and only if $c_l = 0$ for all $l = 1, \ldots, m$.*

The key theorem for least-squares methods in the ADN theory [3] applied to (4.1)–(4.2) provides the required a priori estimates and associated functional spaces. In the following, $\| \cdot \|_i$ and $\| \cdot \|_{i,\Gamma}$ are norms on the spaces $H^i(\Omega)$ and $H^i(\Gamma)$, respectively, and $\langle \cdot, \cdot \rangle_i$ and $\langle \cdot, \cdot \rangle_{i,\Gamma}$ are their equivalent inner products.

THEOREM 5.4. *Let $\mathcal{L}$ be a uniformly ADN elliptic operator of order $2m$ which satisfies the supplementary condition and, together with the trace operator $\mathcal{R}$, satisfies the complementing condition. Then assume that for some $q \geq 0$, $u \in \prod_{j=1}^{n} H^{q+t_j}(\Omega)$, $f \in \prod_{i=1}^{n} H^{q-s_i}(\Omega)$, and $g \in \prod_{l=1}^{m} H^{q-r_l-1/2}(\Gamma)$. Then there exists a constant $C > 0$ such that*

$$(5.2) \qquad \sum_{j=1}^{n} \|u_j\|_{q+t_j} \leq C\left( \sum_{i=1}^{n} \|f_i\|_{q-s_i} + \sum_{l=1}^{m} \|g_l\|_{q-r_l-1/2,\Gamma} + \sum_{j=1}^{n} \|u_j\|_0 \right),$$

*where $u = (u_1, \ldots, u_n)^T$, $f = (f_1, \ldots, f_n)^T$, and $g = (g_1, \ldots, g_m)^T$. Moreover, if the problem (4.1)–(4.2) has a unique solution, then the term on the RHS of (5.2) involving the $L^2$-norm can be omitted.*

It can be ensured that the problem always has a unique solution (and hence that the $L^2$-norm term can be omitted) by including additional constraints (e.g., the zero mean pressure constraint in the Stokes problem). Indeed, if there are $k$ additional constraints $\ell(u) = c$, where $\ell : X \to \mathbb{R}^k$, then the quadratic least-squares functional (4.5) can be modified in the following way:

$$(5.3) \qquad \mathcal{J}(u) = \|\mathcal{L}u - f\|_Y^2 + \|\mathcal{R}u - g\|_Z^2 + |\ell(u) - c|^2 \quad \forall u \in X.$$

It follows that the Euler–Lagrange equations associated with the minimization of (5.3) are: Find $u \in X$ such that

$$A(u, v) = L(v) \quad \forall v \in X,$$

where

$$A(u, v) = \langle \mathcal{L}u, \mathcal{L}v \rangle_Y + \langle \mathcal{R}u, \mathcal{R}v \rangle_Z + \ell(u) \cdot \ell(v),$$
$$L(v) = \langle f, \mathcal{L}v \rangle_Y + \langle g, \mathcal{R}v \rangle_Z + c \cdot \ell(v).$$

Note that the ADN theory can be extended to cover inclusion of constraints by assuming that there exists a homeomorphism $\{\mathcal{L}, \mathcal{R}, \ell\} : X \to Y \times Z \times \mathbb{R}^k$.

Equipped with the knowledge that a unique solution can always be found, it follows that Theorem 5.4 has provided us with the lower bound in the a priori estimate (4.3). To see this we let $X = \prod_{j=1}^{n} H^{q+t_j}(\Omega)$, $Y = \prod_{i=1}^{n} H^{q-s_i}(\Omega)$, and $Z = \prod_{l=1}^{m} H^{q-r_l-1/2}(\Gamma)$ with corresponding norms $\|\cdot\|_X = \sum_{j=1}^{n} \|\cdot\|_{q+t_j}$, $\|\cdot\|_Y = \sum_{i=1}^{n} \|\cdot\|_{q-s_i}$, and $\|\cdot\|_Z = \sum_{l=1}^{m} \|\cdot\|_{q-r_l-1/2,\Gamma}$. The inequality (5.2) then reduces to

$$\|u\|_X \leq C(\|f\|_Y + \|g\|_Z) = C(\|\mathcal{L}u\|_Y + \|\mathcal{R}u\|_Z).$$

The upper bound of the a priori estimate (4.3) follows directly from the continuity of the operator $\{\mathcal{L}, \mathcal{R}\}$, and combining this with the above lower bound, we obtain the required a priori estimate and, in particular, the appropriate choices of the Hilbert spaces $X$, $Y$, and $Z$.

There are several issues that need to be addressed when employing least-squares methods. These include the requirement to reformulate the problem as a first-order system and the imposition of boundary conditions. Imposing boundary conditions weakly (as in (4.7)) requires the evaluation of the trace norm $\|\cdot\|_Z$, which cannot be computed easily using finite/spectral element methods and which can lead to poorly conditioned systems [29]. To avoid this issue, only strongly imposed Dirichlet boundary conditions are considered in this paper. Finally, the most practical setting for

a least-squares method is when $Y$ is a Cartesian product of $L^2$ spaces so that only first-order derivatives appear in the system (4.7). Systems which satisfy this property are known as homogeneous elliptic. It is often possible to reformulate problems in terms of an equivalent first-order system that is homogeneous elliptic. Homogeneous ellipticity of the system was observed to be a key factor to constructing efficient least-squares PGD algorithms in the thesis of Croft [19] in the case of the Poisson and Stokes problems. Nonhomogeneous elliptic systems can be handled using weighted $L^2$-norms [10].

**6. Convergence of least-squares PGD algorithms.** In this section we provide two distinct proofs of convergence of pure greedy algorithms based on a least-squares formulation of the abstract problem presented in section 4. We shall consider a proof based on minimization of energies by Cancès, Ehrlacher, and Lelièvre [15] and comment on the proof based on a functional Eckart–Young theorem by Falcó and Nouy [23]. The convergence results presented here apply to the algorithm based on the minimization formulation rather than the version based on the iterative solution of the Euler–Lagrange equations.

**6.1. Energy minimization.** As we previously stated, the pure greedy algorithm for a least-squares formulated problem involves the minimization of the quadratic least-squares functional:

$$(6.1) \qquad \mathcal{J}(u) = \|\mathcal{L}u - f\|_Y^2.$$

Cancès, Ehrlacher, and Lelièvre [15] provided a proof of convergence of a pure greedy algorithm for the minimization of a general functional $\mathcal{J}(u)$ which satisfies the following two assumptions:

1. $\mathcal{J}$ is strongly convex for $\| \cdot \|_X$; i.e., there exists a constant $\alpha > 0$ such that for $t \in [0, 1]$

$$\mathcal{J}(tu + (1-t)v) \leq t\mathcal{J}(u) + (1-t)\mathcal{J}(v) - \frac{\alpha}{2}t(1-t)\|u-v\|_X^2 \quad \forall u, v \in X.$$

   We then say that $\mathcal{J}$ is $\alpha$-convex [24].

2. $\mathcal{J}$ is differentiable and its Fréchet derivative is Lipschitz continuous; i.e., there exists a constant $L \geq 0$ such that

$$\|\mathcal{J}'(u) - \mathcal{J}'(v)\|_X \leq L\|u-v\|_X \quad \forall u, v \in X,$$

   where $\mathcal{J}'$ denotes the Fréchet derivative of $\mathcal{J}$.

To prove that least-squares formulated pure greedy algorithms converge we need to verify these assumptions for the general quadratic least-squares functional (6.1).

LEMMA 6.1. *The least-squares functional $\mathcal{J}$ defined by* (6.1) *is strongly convex, and its Fréchet derivative is Lipschitz continuous.*

*Proof.* The key ingredient to proving that these two assumptions hold for the least-squares functional is the coercivity relation arising from the ADN theory,

$$(6.2) \qquad C_1\|u\|_X \leq \|\mathcal{L}u\|_Y \leq C_2\|u\|_X \quad \forall u \in X.$$

In other words, it is the continuity and strong coercivity of the least-squares formulation which is key to proving convergence of an associated greedy algorithm. Indeed, since we know that

$$\|u - v\|_X^2 \leq \frac{1}{C_1^2}\|\mathcal{L}(u - v)\|_Y^2 = \frac{1}{C_1^2}\|\mathcal{L}u - \mathcal{L}v\|_Y^2,$$

then proving strong convexity amounts to proving that for $t \in [0,1]$

$$(6.3) \quad \mathcal{J}(tu + (1-t)v) \leq t\mathcal{J}(u) + (1-t)\mathcal{J}(v) - \frac{\alpha}{2C_1^2}t(1-t)\|\mathcal{L}u - \mathcal{L}v\|_Y^2 \quad \forall u, v \in X.$$

Indeed, if we consider the LHS of (6.3), we see

$$\begin{aligned}
\mathcal{J}(tu + (1-t)v) &= \|t\mathcal{L}u + (1-t)\mathcal{L}v - f\|_Y^2 \\
&= t^2\|\mathcal{L}u\|_Y^2 + (1-t)^2\|\mathcal{L}v\|_Y^2 + \|f\|_Y^2 - 2t\langle \mathcal{L}u, f\rangle_Y \\
&\quad - 2(1-t)\langle \mathcal{L}v, f\rangle_Y + 2t(1-t)\langle \mathcal{L}u, \mathcal{L}v\rangle_Y \\
&= t(\|\mathcal{L}u\|_Y^2 - 2\langle \mathcal{L}u, f\rangle_Y + \|f\|_Y^2) \\
&\quad + (1-t)(\|\mathcal{L}v\|_Y^2 - 2\langle \mathcal{L}v, f\rangle_Y + \|f\|_Y^2) \\
&\quad - t(1-t)(\|\mathcal{L}u\|_Y^2 - 2\langle \mathcal{L}u, \mathcal{L}v\rangle_Y + \|\mathcal{L}v\|_Y^2) \\
&= t\|\mathcal{L}u - f\|_Y^2 + (1-t)\|\mathcal{L}v - f\|_Y^2 - t(1-t)\|\mathcal{L}u - \mathcal{L}v\|_Y^2,
\end{aligned}$$

which is the RHS of (6.3) with $\alpha = 2C_2^1$. Hence $\mathcal{J}$ is $2C_1^2$-convex.

For the second part of the proof we use

$$\langle \mathcal{J}'(u), w\rangle_X = \langle \mathcal{L}w, \mathcal{L}u - f\rangle_Y \quad \forall w \in X.$$

For all $u, v, w \in X$ we have

$$\begin{aligned}
|\langle \mathcal{J}'(u) - \mathcal{J}'(v), w\rangle_X| &= |\langle \mathcal{J}'(u), w\rangle_X - \langle \mathcal{J}'(v), w\rangle_X| \\
&= |\langle \mathcal{L}w, \mathcal{L}u - f\rangle_Y - \langle \mathcal{L}w, \mathcal{L}v - f\rangle_Y| \\
&= |\langle \mathcal{L}w, \mathcal{L}u - \mathcal{L}v\rangle_Y|,
\end{aligned}$$

and by the Cauchy–Schwarz equation we have that

$$\begin{aligned}
|\langle \mathcal{L}w, \mathcal{L}u - \mathcal{L}v\rangle_Y| &\leq \|\mathcal{L}w\|_Y\|\mathcal{L}u - \mathcal{L}v\|_Y \\
&\leq C_2^2\|w\|_X\|u - v\|_X,
\end{aligned}$$

using the coercivity relation (6.2). Hence we have

$$|\langle \mathcal{J}'(u) - \mathcal{J}'(v), w\rangle_X| \leq C_2^2\|w\|_X\|u - v\|_X \quad \forall u, v, w \in X.$$

In particular, taking $w = \mathcal{J}'(u) - \mathcal{J}'(v)$ yields

$$\|\mathcal{J}'(u) - \mathcal{J}'(v)\|_X \leq C_2^2\|u - v\|_X.$$

Therefore $\mathcal{J}'$ is Lipschitz continuous.                    $\square$

*Remark* 6.1. Note that the above proof can be extended trivially to include the least-squares functional with weakly imposed boundary conditions,

$$\mathcal{J}(u) = \|\mathcal{L}u - f\|_Y^2 + \|\mathcal{R}u - g\|_Z^2.$$

There are two additional conditions that must also be satisfied in order for the proof of convergence given in [15] to hold. If we let $\Sigma$ denote the set of all rank-one tensors, then the following conditions must be satisfied:

1. $\text{Span}(\Sigma)$ is a dense subset of $X$ for $\|\cdot\|_X$.
2. $\Sigma$ is weakly closed in $(X, \|\cdot\|_X)$.

The ADN theory supplies us with a functional space $X$ that is simply a Sobolev space depending on the set of indices defining the principal part of the differential operator $\mathcal{L}$. As a result these two conditions will hold for a least-squares formulated problem. Indeed, a proof of this for the simple case of $H^1$ spaces can be found in the paper by Cancès, Ehrlacher, and Lelièvre [15] in the context of a high-dimensional Poisson equation.

The four conditions are therefore satisfied by a least-squares formulated problem. This means that the convergence of the pure greedy algorithm associated with any least-squares PGD algorithm is guaranteed since it is covered by the general proof provided by Cancès, Ehrlacher, and Lelièvre [15].

**6.2. Functional Eckart–Young approach.** A proof of convergence for least-squares PGD algorithms was also given by Falcó and Nouy [23] based on their generalized Eckart–Young theorem approach. This approach is more abstract but provides an analogy with the classic result of Eckart and Young [21] for the error in the truncated proper orthogonal decomposition.

Since $\|u\|_{\mathcal{L}}^2 = \|\mathcal{L}u\|_Y^2$, norm equivalence between $\|\cdot\|_{\mathcal{L}}$ and $\|\cdot\|_X$ follows directly from the coercivity estimate (4.3). Hence under the assumption that $\Sigma$ is weakly closed in $(X, \|\cdot\|_X)$ we have that it is also weakly closed in $(X, \|\cdot\|_{\mathcal{L}})$, since equivalent norms induce the same weak topology. For a given $z \in X$ we define an associated rank-one projector $\Pi_{\mathcal{L}}(z)$ (see [23] for a definition of this projector) with which we can define the optimal progressive rank-$J$ separated representation of the solution $u = \mathcal{L}^{-1}f$ by

$$u_J = \sum_{j=1}^{J} u^{(j)}, \qquad u^{(j)} \in \Pi_{\mathcal{L}}(u - u_{j-1}).$$

The generalized Eckart–Young theorem in [23] then ensures that this sequence converges as $J \to \infty$. The additional assumption that $\mathrm{Span}(\Sigma)$ is a dense subset of $X$ for $\|\cdot\|_X$ ensures that it converges to the solution $u = \mathcal{L}^{-1}f$.

**7. Least-squares PGD algorithms for the Stokes problem.** We now consider a prototypical example of a weakly coercive problem: the Stokes problem. Let $\Omega$ be a bounded connected domain in $\mathbb{R}^2$ with boundary $\Gamma$. The Stokes problem in its classical form is given by

$$\begin{align} -\nabla^2 \mathbf{u} + \nabla p &= \mathbf{f} \quad \text{in } \Omega, \tag{7.1}\\ \nabla \cdot \mathbf{u} &= 0 \quad \text{in } \Omega, \tag{7.2}\\ \mathbf{u} &= \mathbf{g} \quad \text{on } \Gamma, \tag{7.3} \end{align}$$

where $\mathbf{u}$ is the velocity, $p$ is the pressure, and $\mathbf{f}$ is the body force. The inhomogeneous Dirichlet data $\mathbf{g}$ must satisfy

$$\int_{\Gamma} \mathbf{g} \cdot \mathbf{n} \, ds = 0,$$

where $\mathbf{n}$ is the unit outward normal to $\Gamma$. There are several possible equivalent first-order systems for the Stokes problem, and a wide selection of these have been documented in the thesis of Proot [29]. In this paper we consider two of these reformulations, both a nonhomogeneous elliptic (section 7.1) and homogeneous elliptic (section 7.2) reformulation.

**7.1. VVP system.** To derive the velocity-vorticity-pressure (VVP) formulation of the Stokes problem we first define the vorticity in two dimensions by $\omega = \nabla \times \mathbf{u}$.

Then, using the identity $\nabla^{\perp}(\nabla \times \mathbf{u}) = -\nabla^2 \mathbf{u} + \nabla(\nabla \cdot \mathbf{u})$ together with incompressibility $\nabla \cdot \mathbf{u} = 0$, we can write $-\nabla^2 \mathbf{u} = \nabla^{\perp}(\nabla \times \mathbf{u}) = \nabla^{\perp}\omega$. Hence the VVP system is given by

$$(7.4) \qquad \nabla^{\perp}\omega + \nabla p = \mathbf{f} \quad \text{in } \Omega,$$

$$(7.5) \qquad \omega - \nabla \times \mathbf{u} = 0 \quad \text{in } \Omega.$$

$$(7.6) \qquad \nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega.$$

$$(7.7) \qquad \mathbf{u} = \mathbf{g} \quad \text{on } \Gamma.$$

The following coercivity estimate for this formulation can be derived from ADN theory (see Bochev and Gunzburger [10], for example):

$$(7.8) \quad \|\mathbf{u}\|_{q+2} + \|\omega\|_{q+1} + \|p\|_{q+1} \leq C_q\big(\|\nabla^{\perp}\omega + \nabla p\|_q + \|\omega - \nabla \times \mathbf{u}\|_{q+1} + \|\nabla \cdot \mathbf{u}\|_{q+1}\big)$$

for some constant $C_q > 0$, $q \in \mathbb{N}$, where we have employed the notation that $\|\cdot\|_q$ defines the norm on the Sobolev space $H^q(\Omega)$. This can, in fact, be extended to all $q \in \mathbb{R}$ (see Bochev and Gunzburger [9]). Hence we can choose $q = -1$ to overcome practical implementation issues related to the differentiability requirements of the relevant function spaces. This yields the following coercivity estimate:

$$(7.9) \qquad \|\mathbf{u}\|_1 + \|\omega\|_0 + \|p\|_0 \leq C_{-1}\big(\|\nabla^{\perp}\omega + \nabla p\|_{-1} + \|\omega - \nabla \times \mathbf{u}\|_0 + \|\nabla \cdot \mathbf{u}\|_0\big).$$

Note that the coercivity estimates (7.8)–(7.9) rely on the assumption that there exists a unique solution. Since the pressure can only be evaluated up to a constant, we need to include an additional constraint in the quadratic least-squares functionals to ensure uniqueness. For the Stokes problem we use the zero mean pressure constraint $\ell(p) = \int_{\Omega} p \, d\Omega = 0$. The $H^{-1}$-norm in (7.9) is problematic to work with; in practice it is replaced with a weighted $L^2$-norm. In this paper we consider two such weightings. First, we take the simplest case in which the $H^{-1}$-norm is replaced by an unweighted $L^2$-norm. Second, we replace the $H^{-1}$-norm by a mesh parameter–scaled $L^2$-norm. This leads to the following two quadratic least-squares functionals for the VVP system:

$$(7.10) \qquad \mathcal{J}_1(\mathbf{u},\omega,p) = \|\nabla^{\perp}\omega + \nabla p - \mathbf{f}\|_0^2 + \|\omega - \nabla \times \mathbf{u}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 + \mu|\ell(p)|^2,$$

$$(7.11) \qquad \mathcal{J}_2(\mathbf{u},\omega,p) = h^2\|\nabla^{\perp}\omega + \nabla p - \mathbf{f}\|_0^2 + \|\omega - \nabla \times \mathbf{u}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 + \mu|\ell(p)|^2,$$

where $h$ denotes the mesh-width and $\mu > 0$ is an adjustable constant. The discrete analogue of the coercivity estimate (7.9) related to the minimization of $\mathcal{J}_1$ can be shown to depend on the mesh-width $h$ (see [10], for example), and hence the validity of the estimate as $h \to 0$ is questionable. On the other hand, for $\mathcal{J}_2$, the discrete estimate is independent of $h$, and hence, in theory, it should provide a more robust least-squares algorithm.

In the case of least-squares PGDs we are interested in the convergence of these algorithms as the rank, $J$, of the PGD approximation is increased rather than in the refinement of the mesh. Ideally we would like to provide a weighted norm based on the rank of the approximation. Unfortunately, due to the nonlinear nature of the PGD approximation, it is unclear how to derive such a weighting and furthermore how the discrete estimates will depend on rank. This can lead to undesirable convergence rates in the rank for nonhomogeneous elliptic systems, as was observed in the thesis of Croft [19].

We now derive the Euler–Lagrange equations associated with the minimization of the functionals (7.10)–(7.11): Find $\boldsymbol{v} = (\mathbf{u},\omega,p) \in \mathbf{H}_{\mathbf{g}}^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ such that

$$A_k(\boldsymbol{v}, \boldsymbol{v}^*) = L_k(\boldsymbol{v}^*) \quad \forall \boldsymbol{v}^* = (\mathbf{u}^*, \omega^*, p^*) \in (H_0^1(\Omega))^2 \times L^2(\Omega) \times L^2(\Omega)$$

for $k = 1, 2$, where

$$\begin{aligned}
A_1(\boldsymbol{v}, \boldsymbol{v}^*) &= \langle \nabla^\perp \omega + \nabla p, \nabla^\perp \omega^* + \nabla p^* \rangle_0 + \langle \omega - \nabla \times \mathbf{u}, \omega^* - \nabla \times \mathbf{u}^* \rangle_0 \\
&\quad + \langle \nabla \cdot \mathbf{u}, \nabla \cdot \mathbf{u}^* \rangle_0 + \mu \ell(p) \ell(p^*),
\end{aligned}$$

$$\begin{aligned}
A_2(\boldsymbol{v}, \boldsymbol{v}^*) &= h^2 \langle \nabla^\perp \omega + \nabla p, \nabla^\perp \omega^* + \nabla p^* \rangle_0 + \langle \omega - \nabla \times \mathbf{u}, \omega^* - \nabla \times \mathbf{u}^* \rangle_0 \\
&\quad + \langle \nabla \cdot \mathbf{u}, \nabla \cdot \mathbf{u}^* \rangle_0 + \mu \ell(p) \ell(p^*),
\end{aligned}$$

$$L_1(\boldsymbol{v}^*) = \langle \mathbf{f}, \nabla^\perp \omega^* + \nabla p^* \rangle_0, \qquad L_2(\boldsymbol{v}^*) = h^2 \langle \mathbf{f}, \nabla^\perp \omega^* + \nabla p^* \rangle_0,$$

and $\mathbf{H}_{\mathbf{g}}^1(\Omega) = \{\mathbf{v} : v_i \in H^1(\Omega), i = 1, 2, \mathbf{v} = \mathbf{g} \text{ on } \Gamma\}$.

**7.2. Extended VGVP system.** We now consider the velocity gradient–velocity pressure (VGVP) formulation of the Stokes problem. Define the velocity gradient by

$$\underline{\mathbf{V}} = (\nabla \mathbf{u})^T = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \begin{pmatrix} V_1 & V_2 \\ V_3 & V_4 \end{pmatrix}.$$

If we then define the divergence of a tensor to be the divergence of its rows, we obtain the identity $\nabla \cdot \underline{\mathbf{V}} = \nabla^2 \mathbf{u}$. Hence we can rewrite the Stokes problem (7.1)–(7.3) as the following first-order VGVP system:

$$\begin{aligned}
(7.12) && -\nabla \cdot \underline{\mathbf{V}} + \nabla p &= \mathbf{f} && \text{in } \Omega, \\
(7.13) && \nabla \cdot \mathbf{u} &= 0 && \text{in } \Omega, \\
(7.14) && \underline{\mathbf{V}} - (\nabla \mathbf{u})^T &= \underline{\mathbf{0}} && \text{in } \Omega, \\
(7.15) && \mathbf{u} &= \mathbf{g} && \text{on } \Gamma.
\end{aligned}$$

Unfortunately, it has been shown by Cai, Manteuffel, and McCormick [14] that this does not lead to a homogeneous elliptic system. However, we can include additional redundant equations to provide us with a problem which is homogeneous elliptic. Indeed, this leads to the following extended VGVP (XVGVP) system:

$$\begin{aligned}
(7.16) && -\nabla \cdot \underline{\mathbf{V}} + \nabla p &= \mathbf{f} && \text{in } \Omega, \\
(7.17) && \nabla \cdot \mathbf{u} &= 0 && \text{in } \Omega, \\
(7.18) && \underline{\mathbf{V}} - (\nabla \mathbf{u})^T &= \underline{\mathbf{0}} && \text{in } \Omega, \\
(7.19) && \nabla(\text{Tr} \, \underline{\mathbf{V}}) &= \mathbf{0} && \text{in } \Omega, \\
(7.20) && \nabla \times \underline{\mathbf{V}} &= \mathbf{0} && \text{in } \Omega, \\
(7.21) && \mathbf{u} &= \mathbf{0} && \text{on } \Gamma, \\
(7.22) && \mathbf{n} \times \underline{\mathbf{V}} &= \mathbf{0} && \text{on } \Gamma.
\end{aligned}$$

The additional boundary condition holds since from (7.18) we have that $\mathbf{n} \times \underline{\mathbf{V}} = \mathbf{n} \times (\nabla \mathbf{u})^T = 0$, since the boundary condition on $\mathbf{u}$ implies that its tangential derivatives vanish on the boundary. Note that for simplicity we have considered homogeneous velocity Dirichlet boundary conditions (7.21). For the nonhomogeneous case, $\mathbf{u} = \mathbf{g}$ on $\Gamma$, the additional boundary condition (7.22) should be replaced by $\mathbf{n} \times \underline{\mathbf{V}} = \mathbf{n} \times (\nabla \mathbf{g})^T$.

The first redundant equation, (7.19), is satisfied since $\text{Tr} \, \underline{\mathbf{V}} = V_1 + V_4 = \nabla \cdot \mathbf{u} = 0$. The second redundant equation, (7.20), is satisfied since if we define the curl of a

tensor to be the curl of its rows, then we have

$$\nabla \times \underline{\mathbf{V}} = \begin{pmatrix} \frac{\partial^2 u}{\partial x \partial y} - \frac{\partial^2 u}{\partial y \partial x} \\ \frac{\partial^2 v}{\partial x \partial y} - \frac{\partial^2 u}{\partial y \partial x} \end{pmatrix} = \mathbf{0}.$$

This system has been proven to be homogeneous elliptic in an ad hoc manner by Cai et al. [14], but it can also be verified using ADN theory (see Croft [19]), showing that the XVGVP system satisfies the supplementary and complementing conditions.

   This yields the following estimate:

$$\|\mathbf{u}\|_1 + \|p\|_1 + \|\underline{\mathbf{V}}\|_1 \leq C\big(\| - \nabla \cdot \underline{\mathbf{V}} + \nabla p\|_0 + \|\nabla \cdot \mathbf{u}\|_0 + \|\underline{\mathbf{V}} - (\nabla \mathbf{u})^T\|_0$$

(7.23)
$$+ \|\nabla(\mathrm{Tr}\,\underline{\mathbf{V}})\|_0 + \|\nabla \times \underline{\mathbf{V}}\|_0\big)$$

for some constant $C > 0$. This leads to the following quadratic least-squares functional:

(7.24)
$$\mathcal{J}_3(\mathbf{u}, p, \underline{\mathbf{V}}) = \| - \nabla \cdot \underline{\mathbf{V}} + \nabla p - \mathbf{f}\|_0^2 + \|\nabla \cdot \mathbf{u}\|_0^2 + \|\underline{\mathbf{V}} - (\nabla \mathbf{u})^T\|_0^2$$
$$+ \|\nabla(\mathrm{Tr}\,\underline{\mathbf{V}})\|_0^2 + \|\nabla \times \underline{\mathbf{V}}\|_0^2 + \mu|\ell(p)|^2.$$

We are then able to derive the Euler–Lagrange equations associated with the minimization of (7.24): Find $\boldsymbol{v} = (\mathbf{u}, p, \underline{\mathbf{V}}) \in \mathbf{H_0^1}(\Omega) \times H^1(\Omega) \times \underline{\mathbf{H}}_\times^1(\Omega)$ such that

$$A_3(\boldsymbol{v}, \boldsymbol{v}^*) = L_3(\boldsymbol{v}^*) \quad \forall \boldsymbol{v}^* = (\mathbf{u}^*, p^*, \underline{\mathbf{V}}^*) \in \mathbf{H_0^1}(\Omega) \times H^1(\Omega) \times \underline{\mathbf{H}}_\times^1(\Omega),$$

where

$$\underline{\mathbf{H}}_\times^1(\Omega) = \{\underline{\mathbf{V}} \in \underline{\mathbf{H}}^1(\Omega) \mid \mathbf{n} \times \underline{\mathbf{V}} = \mathbf{0} \quad \text{on } \Gamma\}$$

and where

$$A_3(\boldsymbol{v}, \boldsymbol{v}^*) = \langle -\nabla \cdot \underline{\mathbf{V}} + \nabla p, -\nabla \cdot \underline{\mathbf{V}}^* + \nabla p^* \rangle_0 + \langle \underline{\mathbf{V}} - (\nabla \mathbf{u})^T, \underline{\mathbf{V}}^* - (\nabla \mathbf{u}^*)^T \rangle_0$$
$$+ \langle \nabla \cdot \mathbf{u}, \nabla \cdot \mathbf{u}^* \rangle_0 + \langle \nabla(\mathrm{Tr}\,\underline{\mathbf{V}}), \nabla(\mathrm{Tr}\,\underline{\mathbf{V}}^*) \rangle_0 + \langle \nabla \times \underline{\mathbf{V}}, \nabla \times \underline{\mathbf{V}}^* \rangle_0$$
$$+ \mu\ell(p)\ell(p^*)$$

and

$$L_3(\boldsymbol{v}^*) = \langle \mathbf{f}, -\nabla \cdot \underline{\mathbf{V}}^* + \nabla p^* \rangle_0.$$

We have included the zero mean pressure constraint $\ell(p) = \int_\Omega p \, d\Omega = 0$ to ensure uniqueness of the solution and where, as before, $\mu > 0$ is an adjustable constant.

   **7.3. Numerical results.** To test and compare these least-squares PGD algorithms we revisit the problem introduced in section 3 defined on the square domain $\Omega = [-1, 1]^2$ with the source term, $\mathbf{f}$, given by (3.8).

   Figure 2 shows convergence in the rank for the velocity and pressure for all three least-squares PGD algorithms for the Stokes problem. We have only compared the convergence of the approximation of these dependent variables, since they are the only variables common to both the VVP and VGVP formulations. It is clear from this that the algorithm based on the homogeneous elliptic XVGVP formulation displays the superior rate of convergence.

   Unfortunately, none of the algorithms captured the natural rank-one separated form of the true solution to the velocity. This is particularly disappointing when we note that the Galerkin PGD algorithm for the same problem was able to do so. However, there is no guarantee that PGD algorithms will always achieve this.
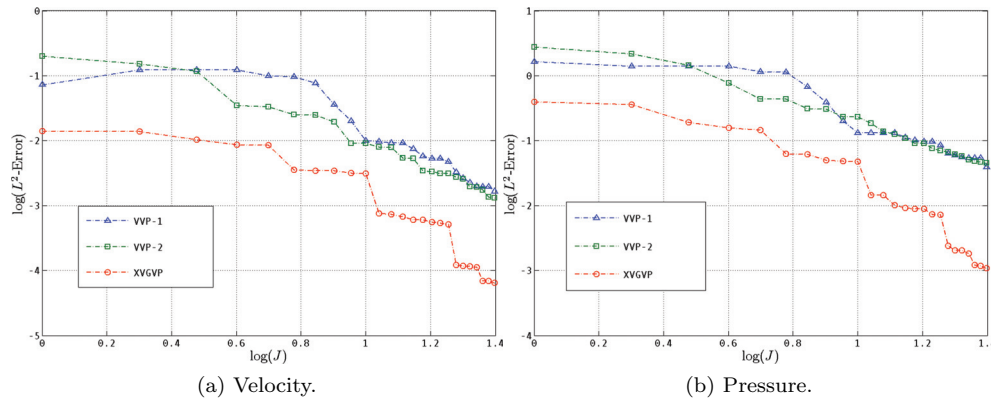
(a) Velocity.                    (b) Pressure.

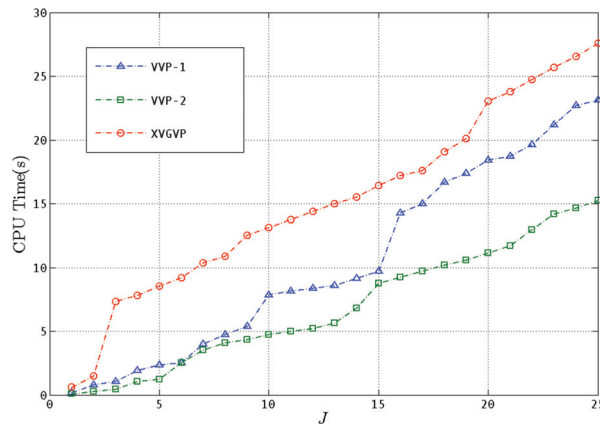FIG. 2. *Convergence in the rank for least-squares PGD algorithms.*



FIG. 3. *CPU time for the least-squares PGD algorithms.*

In Figure 3, we plot the CPU times for each of the three algorithms. The XVGVP algorithm is only slightly slower than the VVP algorithms, despite the fact that the XVGVP algorithm involves the solution of linear systems almost twice the size of those in the VVP algorithms. Certainly this slight increase in computational cost is insignificant when we consider how much faster the XVGVP algorithm converges in terms of the rank. Furthermore, we notice that the majority of the CPU expense comes from single PGD iterations taking a significantly longer time to run than other iterations (e.g., the third iteration in the XVGVP algorithm) the reason for this is that the linearization step of the PGD (i.e., the ADFPA) takes a long time to converge. This could be remedied by introducing a more sophisticated adaptive linearization or by limiting the number of iterations in the linearization.

Note that the discretization used in Figures 2 and 3 was a spectral element method using degree 8 polynomial basis functions over three elements in each coordinate direction. We further note that all three of these algorithms yielded reasonably accurate approximations for all choices of discretization parameters. To demonstrate this we have plotted the convergence rates for a variety of discretization parameters for the the XVGVP algorithm in Figure 4.
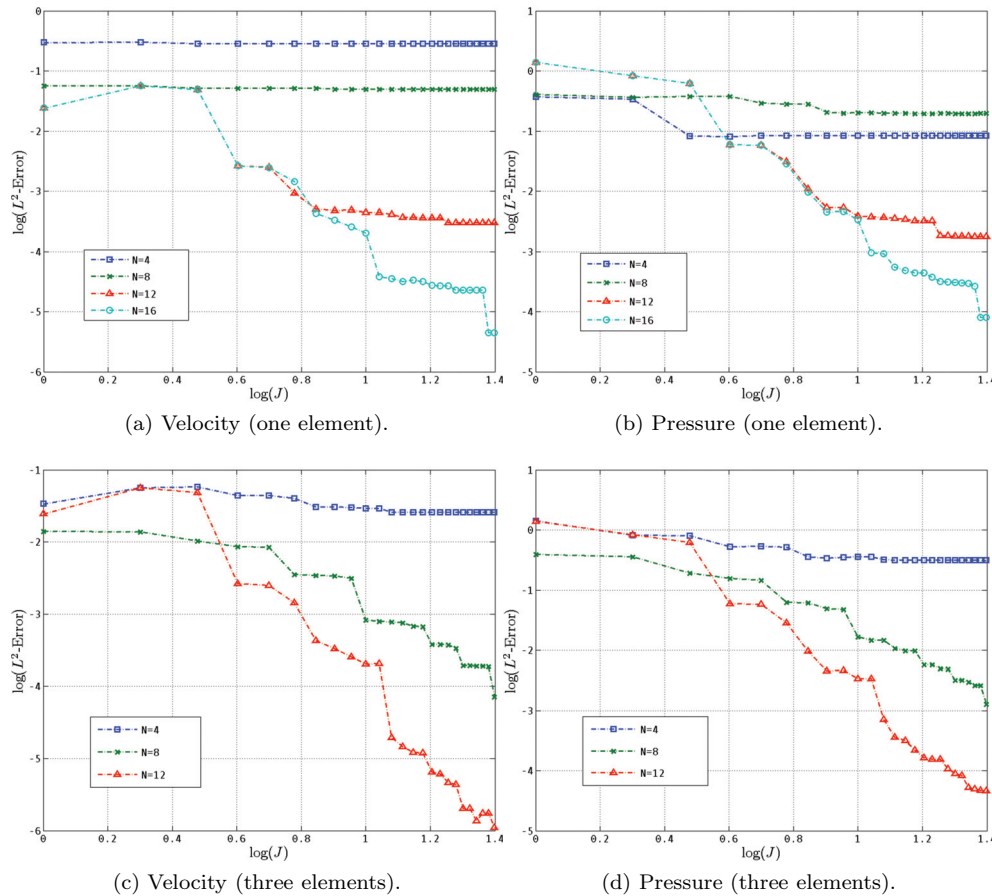
(a) Velocity (one element).

(b) Pressure (one element).

(c) Velocity (three elements).

(d) Pressure (three elements).

FIG. 4. *Convergence in the rank for XVGVP algorithm.*

In Figure 4 we present the convergence behavior of the PGD algorithm as the underlying approximation is refined in terms of polynomial order ($N$) for $K = 1$ and $K = 3$, where $K$ is the number of spectral elements in each coordinate direction. From this we can see that, other than the very coarse discretization with $N = 4$ and $N = 8$ for $K = 1$, we obtain good rates of convergence which improve with further refinement of the discretization space, as we would expect. We also observe the level of convergence plateauing for coarser discretizations, which is due to the convergence in increasing rank $J$ being limited by the discretization error. This coincides with observations of the behavior of spectral elements in the PGD in the thesis of Croft [19], and we direct the interested reader there for more information. Similar results can also be obtained for the VVP algorithms.

In the case of the two-dimensional Stokes problem we compare the performance of the least-squares PGD with a standard spectral element solver. The standard spectral element method employs a sparse direct solver, whereas the least-squares PGD method uses an ADFPA. In Figure 5 the ratio of CPU times for the standard and least-squares PGD spectral element solvers is plotted for different discretizations for $K = 3$. It can be clearly seen for small values of $N$ that the standard solver is more efficient. However, as $N$ is increased beyond 20, the least-squares PGD solver becomes
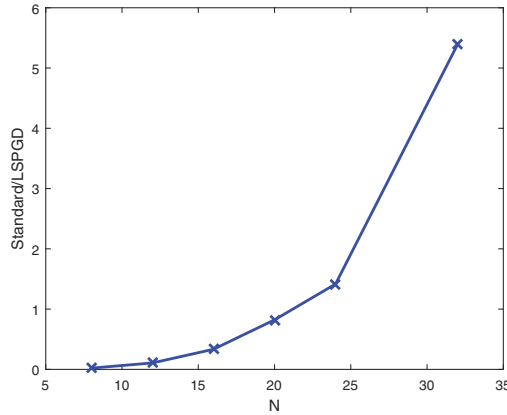
Fig. 5. *Influence of the ratio of CPU times for the standard and least-squares PGD solvers on $N$.*

increasingly more efficient compared with the standard spectral element solver.

The final example considered is the regularized lid-driven cavity problem [30]. This problem does not possess an analytical solution, and so the velocity components and pressure cannot be written in terms of a separated expansion. The Stokes problem is solved in the domain $[-0.5, 0.5]^2$ with $\mathbf{u} = \mathbf{0}$ on all sides of the boundary except along the top boundary where

$$\mathbf{u} = (\mathcal{F}(x; \delta), 0)^T,$$

with

$$\mathcal{F}(x; \delta) = \begin{cases} \frac{1}{2}(1 + \cos[\frac{\pi}{\delta}(x + \frac{1}{2}(1 - 2\delta))]) & \text{if} \quad x \in [-0.5, -0.5 + \delta], \\ 1 & \text{if} \quad x \in [-0.5 + \delta, 0.5 - \delta], \\ \frac{1}{2}(1 + \cos[\frac{\pi}{\delta}(x + \frac{1}{2}(1 - 2\delta))]) & \text{if} \quad x \in [0.5 - \delta, -0.5]. \end{cases}$$

Although the velocity is continuous along the boundary, there are steep gradients near the top corners of the cavity. The XVGVP least-squares algorithm is used to solve the problem with a spectral element discretization characterized by $K = 4$, $N = 8$, and regularization parameter $\delta = 0.05$. In two dimensions, the stream function is defined by

$$\mathbf{u} = \left( \frac{\partial \psi}{\partial y}, -\frac{\partial \psi}{\partial x} \right).$$

Since the problem does not possess an analytical solution, the convergence of the least-squares PGD algorithm is demonstrated by tabulating the model error using the approximation for $J = 40$ as the benchmark. The $L^2$-norm of $\psi_J - \psi_{40}$ is tabulated in Table 1 as a function of the number of PGD modes, $J$. This example demonstrates the convergence of the least-squares PGD algorithm for a problem whose solution possesses steep gradients.

We conclude this section with a few remarks concerning mass conservation of least-squares algorithms. Mass conservation is a constraint in Galerkin methods, whereas in least-squares methods the continuity equation contributes one term in a quadratic functional which is minimized. Although mass is not conserved strongly, Proot [29] has shown that when spectral element methods are used to discretize the

TABLE 1

*Convergence of the $L^2$-norm of the model error $\|\psi_J - \psi_{40}\|_2$ with number of PGD modes, $J$.*

| $J$ | $\|\psi_J - \psi_{40}\|_2$ |
|-----|------------------------------|
| 10  | 0.0613 |
| 15  | 0.0248 |
| 20  | 0.0108 |
| 25  | 0.0067 |
| 30  | 0.0059 |
| 35  | 0.0030 |

least-squares formulation, it is satisfied weakly as $N \to \infty$. They also argue that a number of attempts to improve mass conservation for least-squares methods introduce additional problems. For example, in the weighted least-squares method [20] an improvement in mass conservation comes at the expense of momentum conservation and the overall quality of the solution. The constrained least-squares method [17] sacrifices the positive definiteness of the formulation by employing Lagrange multipliers to enforce mass conservation, and requires an additional compatibility condition to guarantee a unique solution.

**8. Conclusions and further work.** The crucial observation from this work is that least-squares PGD algorithms are robust in terms of their performance with respect to the choice of the discretization parameters. This was contrary to our experience of the performance of the Galerkin PGD algorithm, where a significant amount of trial and error was undertaken in order to obtain converged results, an issue associated with the lack of LBB-like stability in a Galerkin formulation of the Stokes problem originating from the weak coercivity of the problem. In this sense the least-squares PGD algorithms are significantly more reliable. Furthermore, in sections 4 and 5, we provided theoretical results that guarantee the convergence of greedy algorithms based on least-squares formulated problems. The theoretical underpinning of this approach is a significant advantage for using least-squares methods for these types of problems, as there are no corresponding convergence results for the Galerkin formulation of the same problem. For these reasons we strongly advocate using least-squares PGD algorithms for solving weakly coercive problems, particularly for those which are defined in high-dimensional spaces, since this is where the maximum benefits of PGD will be realized.

However, least-squares PGD algorithms are not without pitfalls. Indeed, we observed rather disappointing rates of convergence and CPU times for all three algorithms, an observation that was also made by Nouy [28] when employing minimal residual PGDs for convection-diffusion equations. The use of a more a sophisticated linearization could lead to improvements in the rate of convergence. Some improvement in convergence rates might also be obtained if one were able to derive weighted norms based on the rank of the approximation. Recent work of Billaud-Friess, Nouy, and Zahm [8] on highly efficient minimal residual PGD algorithms might also be applied and extended to weakly coercive problems such as the Stokes problem, which could lead to significantly improved rates of convergence.

There are a number of possible future directions for this work in addition to the improvements to efficiency that we have already mentioned. One could investigate the performance of least-squares algorithms for different weakly coercive problems as well as nonsymmetric problems. It also remains to prove the convergence of the particular implementation of the PGD algorithm used here. The idealized pure greedy

algorithm does not take into consideration the fact that we solve the Euler–Lagrange equations for the problem, or the particular linearization and discretization that are used. While the nonlinearity of the PGD approximation makes this a very daunting task, it would be extremely illuminating to understand how all these elements of PGD algorithms are combined. Furthermore, results on expected convergence rates do not presently exist; the availability of such results would not only provide a great deal of insight into how these algorithms work but also lead to practical developments such as extremely accurate error estimators, which could be used as global stopping criteria or used to design efficient adaptive strategies.

## REFERENCES

[1] Y. Achdou and O. Pironneau, *Computational Methods for Option Pricing*, Frontiers in Appl. Math. 30, SIAM, Philadelphia, 2005.

[2] S. Agmon, A. Douglis, and L. Nirenberg, *Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions* I, Commun. Pur. Appl. Math., 12 (1959), pp. 623–727.

[3] S. Agmon, A. Douglis, and L. Nirenberg, *Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions* II, Commun. Pur. Appl. Math., 17 (1964), pp. 35–92.

[4] A. Ammar, B. Mokdad, F. Chinesta, and R. Keunings, *A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids*, J. Non-Newton. Fluid Mech., 139 (2006), pp. 153–176, https://doi.org/10.1016/j.jnnfm.2006.07.007.

[5] A. K. Aziz, R. B. Kellogg, and A. B. Stephens, *Least squares methods for elliptic systems*, Math. Comput., 44 (1985), pp. 53–70.

[6] I. Babuška, *Error bounds for finite element method*, Numer. Math., 16 (1971), pp. 322–333.

[7] G. Beylkin and M. J. Mohlenkamp, *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 10246–10251, https://doi.org/10.1073/pnas.112329799.

[8] M. Billaud-Friess, A. Nouy, and O. Zahm, *A tensor approximation method based on ideal minimal residual formulations for the solution of high-dimensional problems*, ESAIM-Math. Model. Numer., 48 (2014), pp. 1777–1806, https://doi.org/10.1051/m2an/2014019.

[9] P. B. Bochev and M. D. Gunzburger, *Analysis of least-squares finite element methods for the Stokes equations*, Math. Comput., 63 (1994), pp. 479–506.

[10] P. B. Bochev and M. D. Gunzburger, *Least-Squares Finite Element Methods*, Springer-Verlag, New York, 2009.

[11] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York, 1991.

[12] H. J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer., 13 (2004), pp. 1–123, https://doi.org/10.1017/S0962492904000182.

[13] Z. Cai, R. Lazarov, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part I*, SIAM J. Numer. Anal., 31 (1994), pp. 1785–1799, https://doi.org/10.1137/0731091.

[14] Z. Cai, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for the Stokes equations, with applications to linear elasticity*, SIAM J. Numer. Anal., 34 (1997), pp. 1727–1741, https://doi.org/10.1137/S003614299527299X.

[15] E. Cancès, V. Ehrlacher, and T. Lelièvre, *Convergence of a greedy algorithm for high-dimensional convex nonlinear problems*, Math. Models Methods Appl. Sci., 21 (2011), pp. 2433–2467, https://doi.org/10.1142/S0218202511005799.

[16] E. Cancès, V. Ehrlacher, and T. Lelièvre, *Greedy algorithms for high-dimensional non-symmetric linear problems*, in ESAIM Proc. 41, Soc. Math. Appl. Indust., Paris, 2013, pp. 95–131, https://doi.org/10.1051/proc/201341005.

[17] C. L. Chang and J. J. Nelson, *Least-squares finite element method for the Stokes problem with zero residual of mass conservation*, SIAM J. Numer. Anal., 34 (1997), pp. 480–489, https://doi.org/10.1137/S0097539794273368.

[18] F. CHINESTA, A. AMMAR, A. LEYGUE, AND R. KEUNINGS, *An overview of the proper generalized decomposition with applications in computational rheology*, J. Non-Newton. Fluid Mech., 166 (2011), pp. 578–592, https://doi.org/10.1016/j.jnnfm.2010.12.012.

[19] T. L. D. CROFT, *Proper Generalised Decompositions: Theory and Applications*, Ph.D. thesis, School of Mathematics, Cardiff University, Cardiff, Wales, 2015.

[20] J. M. DEANG AND M. D. GUNZBERGER, *Issues related to least-squares finite element methods for the Stokes equations*, SIAM J. Sci. Comput., 20 (1998), pp. 878–906, https://doi.org/10.1137/S1064827595294526.

[21] C. ECKART AND G. YOUNG, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.

[22] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Finite Elements and Fast Iterative Solvers*, Oxford University Press, Oxford, UK, 2005.

[23] A. FALCÓ AND A. NOUY, *A proper generalized decomposition for the solution of elliptic problems in abstract form by using a functional Eckart-Young approach*, J. Math. Anal. Appl., 376 (2011), pp. 469–480, https://doi.org/10.1016/j.jmaa.2010.12.003.

[24] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimisation Algorithms I*, Springer-Verlag, Berlin, 1993.

[25] T. G. KOLDA, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243–255, https://doi.org/10.1137/S0895479800368354.

[26] C. LE BRIS, T. LELIÈVRE, AND Y. MADAY, *Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations*, Constr. Approx., 30 (2009), pp. 621–651, https://doi.org/10.1007/s00365-009-9071-1.

[27] Y. MADAY, A. T. PATERA, AND E. M. RØNQUIST, *The $P_N - P_{N-2}$ Method for the Approximation of the Stokes Problem*, Technical Report 92025, Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, Paris, 1992.

[28] A. NOUY, *A priori model reduction through proper generalized decomposition for solving time-dependent partial differential equations*, Comput. Methods Appl. Mech. Engrg., 199 (2010), pp. 1603–1626, https://doi.org/10.1016/j.cma.2010.01.009.

[29] M. M. J. PROOT, *The Least-Squares Spectral Element Method*, Ph.D. thesis, Department of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands, 2003.

[30] P. N. SHANKAR, *Slow Viscous Flows: Qualitative Features and Quantitative Analysis Using Complex Eigenfunction Expansions*, Imperial College Press, London, 2007.

[31] V. N. TEMLYAKOV, *The best m-term approximation and greedy algorithms*, Adv. Comput. Math., 8 (1998), pp. 249–265, https://doi.org/10.1023/A:1018900431309.