SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 6, No. 1, pp. 374–396

Stochastic Least-Squares Petrov–Galerkin Method for Parameterized Linear Systems*

Kookjin Lee†, Kevin Carlberg‡, and Howard C. Elman§

Abstract. We consider the numerical solution of parameterized linear systems where the system matrix, the solution, and the right-hand side are parameterized by a set of uncertain input parameters. We explore spectral methods in which the solutions are approximated in a chosen finite-dimensional subspace. It has been shown that the stochastic Galerkin projection technique fails to minimize any measure of the solution error [A. Mugler and H.-J. Starkloff, *ESAIM Math. Model. Numer. Anal.*, 47 (2013), pp. 1237–1263]. As a remedy for this, we propose a novel stochastic least-squares Petrov–Galerkin (LSPG) method. The proposed method is optimal in the sense that it produces the solution that minimizes a weighted ℓ^2-norm of the residual over all solutions in a given finite-dimensional subspace. Moreover, the method can be adapted to minimize the solution error in different weighted ℓ^2-norms by simply applying a weighting function within the least-squares formulation. In addition, a goal-oriented seminorm induced by an output quantity of interest can be minimized by defining a weighting function as a linear functional of the solution. We establish optimality and error bounds for the proposed method, and extensive numerical experiments show that the weighted LSPG method outperforms other spectral methods in minimizing corresponding target weighted norms.

Key words. stochastic Galerkin, least-squares Petrov–Galerkin projection, residual minimization, spectral projection

AMS subject classifications. 35R60, 60H15, 60H35, 65C20, 65N30, 93E24

DOI. 10.1137/17M1110729

1. Introduction.

Forward uncertainty propagation for parameterized linear systems is important in a range of applications to characterize the effects of uncertainties on the output of computational models. Such parameterized linear systems arise in many important problems in science and engineering, including stochastic partial differential equations (SPDEs), where uncertain input parameters are modeled as a set of random variables (e.g., diffusion/ground water flow simulations where diffusivity/permeability is modeled as a random field [15, 30]). It has been shown [11] that intrusive methods (e.g., stochastic Galerkin [2, 10, 13, 19, 28]) for uncertainty propagation can lead to smaller errors for a fixed basis dimension, compared with

*Received by the editors January 6, 2017; accepted for publication (in revised form) November 1, 2017; published electronically March 29, 2018.

http://www.siam.org/journals/juq/6-1/M111072.html

Funding: This work was supported by the U.S. Department of Energy Office of Advanced Scientific Computing Research, Applied Mathematics program under award DEC-SC0009301, and by the U.S. National Science Foundation under grant DMS1418754. Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

†Department of Computer Science, University of Maryland, College Park, MD 20742 (klee@cs.umd.edu).

‡Sandia National Laboratories, Livermore, CA 94550 (ktcarlb@sandia.gov).

§Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 (elman@cs.umd.edu).

nonintrusive methods [27] (e.g., sampling-based methods [3, 14, 17], stochastic collocation [1, 21]).

The stochastic Galerkin method combined with generalized polynomial chaos (gPC) expansions [29] seeks a polynomial approximation of the numerical solution in the stochastic domain by enforcing a Galerkin orthogonality condition, i.e., the residual of the parameterized linear system is forced to be orthogonal to the span of the stochastic polynomial basis with respect to an inner product associated with an underlying probability measure. The Galerkin projection scheme is popular for its simplicity (i.e., the trial and test bases are the same) and its optimality in terms of minimizing an energy norm of solution errors when the underlying PDE operator is symmetric positive definite. In many applications, however, it has been shown that the stochastic Galerkin method does not exhibit any optimality property [20]. That is, it does not produce solutions that minimize any measure of the solution error. In such cases, the stochastic Galerkin method can lead to poor approximations and nonconvergent behavior.

To address this issue, we propose a novel optimal projection technique, which we refer to as the stochastic least-squares Petrov–Galerkin (LSPG) method. Inspired by the successes of LSPG methods in nonlinear model reduction [8, 9, 7], finite element methods [4, 5, 16], and iterative linear solvers (e.g., GMRES, GCR) [22], we propose, as an alternative to enforcing the Galerkin orthogonality condition, to directly minimize the residual of a parameterized linear system over the stochastic domain in a (weighted) ℓ^2-norm. The stochastic LSPG method produces an optimal solution for a given stochastic subspace and guarantees that the ℓ^2-norm of the residual monotonically decreases as the stochastic basis is enriched. In addition to producing monotonically convergent approximations as measured in the chosen weighted ℓ^2-norm, the method can also be adapted to target output quantities of interest (QoI); this can be accomplished by employing a weighted ℓ^2-norm used for least-squares minimization that coincides with the ℓ^2-(semi)norm of the error in the chosen QoI.

In addition to proposing the stochastic LSPG method, this study shows that specific choices of weighting functions lead to equivalence between the stochastic LSPG method and both the stochastic Galerkin method and the pseudospectral method [26, 27]. We demonstrate the effectiveness of this method with extensive numerical experiments on various SPDEs. The results show that the proposed LSPG technique significantly outperforms the stochastic Galerkin when the solution error is measured in different weighted ℓ^2-norms. We also show that the proposed method can effectively minimize the error in target QoIs.

An outline of the paper is as follows. Section 2 formulates parameterized linear algebraic systems and reviews conventional spectral approaches for computing numerical solutions. Section 3 develops a residual minimization formulation based on least-squares methods and its adaptation to the stochastic LSPG method. We also provide proofs of optimality and monotonic convergence behavior of the proposed method. Section 4 provides error analysis for stochastic LSPG methods. Section 5 demonstrates the efficiency and the effectiveness of the proposed methods by testing them on various benchmark problems. Finally, section 6 outlines some conclusions.

2. Spectral methods for parameterized linear systems.

We begin by introducing a mathematical formulation of parameterized linear systems and briefly reviewing the stochastic

Galerkin and the pseudospectral methods, which are spectral methods for approximating the numerical solutions of such systems.

2.1. Problem formulation. Consider a parameterized linear system

$$A(\xi)u(\xi) = b(\xi), \tag{1}$$

where $A : \Gamma \to \mathbb{R}^{n_x \times n_x}$, and $u, b : \Gamma \to \mathbb{R}^{n_x}$. The system is parameterized by a set of stochastic input parameters $\xi(\omega) \equiv \{\xi_1(\omega), \ldots, \xi_{n_\xi}(\omega)\}$. Here, $\omega \in \Omega$ is an elementary event in a probability space $(\Omega, \mathcal{F}, P)$ and the stochastic domain is denoted by $\Gamma \equiv \prod_{i=1}^{n_\xi} \Gamma_i$, where $\xi_i : \Omega \to \Gamma_i$. We are interested in computing solutions in finite-dimensional subspaces of $L^2(\Gamma)$ (defined below) using weak formulations of (1) corresponding to Galerkin and Petrov–Galerkin projections.

Let $\rho \equiv \rho(\xi)$ be a density function defining an underlying measure of the stochastic space Γ and

$$\langle g, h \rangle_\rho \equiv \int_\Gamma g(\xi)h(\xi)\rho(\xi)d\xi, \tag{2}$$

$$E[g] \equiv \int_\Gamma g(\xi)\rho(\xi)d\xi \tag{3}$$

define an inner product between scalar-valued functions $g(\xi)$ and $h(\xi)$ with respect to $\rho(\xi)$ and the expectation of $g(\xi)$, respectively. The inner product (2) also determines the Hilbert space $L^2(\Gamma)$. In addition, the ℓ^2-norm of a vector-valued function $v(\xi) \in \mathbb{R}^{n_x}$ is defined as

$$\|v\|_2^2 \equiv \sum_{i=1}^{n_x} \int_\Gamma v_i^2(\xi)\rho(\xi)d\xi = E[v^T v]. \tag{4}$$

We are interested in computing approximate solutions to (1) using spectral methods, that is, finding solutions in an n_ψ-dimensional subspace S_{n_ψ} spanned by a finite set of polynomials $\{\psi_i(\xi)\}_{i=1}^{n_\psi}$ such that $S_{n_\psi} \equiv \text{span}\{\psi_i\}_{i=1}^{n_\psi} \subseteq L^2(\Gamma)$. Then

$$u(\xi) \approx \tilde{u}(\xi) = \sum_{i=1}^{n_\psi} \bar{u}_i \psi_i(\xi) = (\psi^T(\xi) \otimes I_{n_x})\bar{u}, \tag{5}$$

where $\{\bar{u}_i\}_{i=1}^{n_\psi}$ with $\bar{u}_i \in \mathbb{R}^{n_x}$ are unknown coefficient vectors, $\bar{u} \equiv [\bar{u}_1^T \ \cdots \ \bar{u}_{n_\psi}^T]^T \in \mathbb{R}^{n_x n_\psi}$ is the vertical concatenation of these coefficient vectors, $\psi \equiv [\psi_1 \ \cdots \ \psi_{n_\psi}]^T \in \mathbb{R}^{n_\psi}$ is a concatenation of the polynomial basis, $\otimes$ denotes the Kronecker product, and I_{n_x} denotes the identity matrix of dimension n_x. Note that $\tilde{u} \in (S_{n_\psi})^{n_x}$. Typically, the "stochastic" basis $\{\psi_i\}$ consists of products of univariate polynomials: $\psi_i \equiv \psi_{\alpha(i)} \equiv \prod_{k=1}^{n_\xi} \pi_{\alpha_k(i)}(\xi_k)$, where $\{\pi_{\alpha_k(i)}\}_{k=1}^{n_\xi}$ are univariate polynomials, $\alpha(i) = (\alpha_1(i), \ldots, \alpha_{n_\xi}(i)) \in \mathbb{N}_0^{n_\xi}$ is a multi-index, and α_k represents the degree of a polynomial in ξ_k. The dimension of the stochastic subspace n_ψ depends on the number of random variables n_ξ, the maximum polynomial degree p, and a construction of the polynomial space (e.g., a total-degree space that contains polynomials with total degree up to p, $\sum_{k=1}^{n_\xi} \alpha_k(i) \leq p$). By substituting $u(\xi)$ with $\tilde{u}(\xi)$ in (1), the residual

can be defined as

$$r(\bar{u};\xi) \coloneqq b(\xi) - A(\xi)\sum_{i=1}^{n_\psi} \bar{u}_i\psi_i(\xi) = b(\xi) - (\psi^T(\xi)\otimes A(\xi))\bar{u}, \tag{6}$$

where $\psi^T(\cdot)\otimes A(\cdot) : \Gamma \to \mathbb{R}^{n_x\times n_\psi n_x}$.

It follows from (5) and (6) that our goal now is to compute the unknown coefficients $\{\bar{u}_i\}_{i=1}^{n_\psi}$ of the solution expansion. We briefly review two conventional approaches for doing so: the stochastic Galerkin method and the pseudospectral method. Typically, the polynomial basis is constructed to be orthogonal in the $\langle\cdot,\cdot\rangle_\rho$ inner product, i.e., $\langle\psi_i,\psi_j\rangle_\rho = \prod_{k=1}^{n_\xi}\langle\pi_{\alpha_k(i)},\pi_{\alpha_k(j)}\rangle_{\rho_k} = \delta_{ij}$, where δ_{ij} denotes the Kronecker delta.

2.2. Stochastic Galerkin method. The stochastic Galerkin method computes the unknown coefficients $\{\bar{u}_i\}_{i=1}^{n_\psi}$ of $\tilde{u}(\xi)$ in (5) by imposing orthogonality of the residual (6) with respect to the inner product $\langle\cdot,\cdot\rangle_\rho$ in the subspace S_{n_ψ}. This Galerkin orthogonality condition can be expressed as follows: Find $\bar{u}^{\mathrm{SG}} \in \mathbb{R}^{n_x n_\psi}$ such that

$$\langle r_i(\bar{u}^{\mathrm{SG}}),\psi_j\rangle_\rho = E[r_i(\bar{u}^{\mathrm{SG}})\psi_j] = 0, \qquad i = 1,\dots,n_x,\ j = 1,\dots,n_\psi, \tag{7}$$

where $r \equiv [r_1\ \cdots\ r_{n_x}]^T$. The condition (7) can be represented in matrix notation as

$$E[\psi\otimes r(\bar{u}^{\mathrm{SG}})] = 0. \tag{8}$$

From the definition of the residual (6), this gives a system of linear equations

$$E[\psi\psi^T\otimes A]\bar{u}^{\mathrm{SG}} = E[\psi\otimes b] \tag{9}$$

of dimension $n_x n_\psi$. This yields an algebraic expression for the stochastic-Galerkin approximation

$$\tilde{u}^{\mathrm{SG}}(\xi) = (\psi(\xi)^T\otimes I_{n_x})E[\psi\psi^T\otimes A]^{-1}E[\psi\otimes Au]. \tag{10}$$

If $A(\xi)$ is symmetric positive definite, the solution of linear system (9) minimizes the solution error $e(x) \equiv u - x$ in the $A(\xi)$-induced energy norm $\|v\|_A^2 \equiv E[v^TAv]$, i.e.,

$$\tilde{u}^{\mathrm{SG}}(\xi) = \underset{x\in(S_{n_\psi})^{n_x}}{\arg\min}\ \|e(x)\|_A^2. \tag{11}$$

In general, however, the stochastic-Galerkin approximation does not minimize any measure of the solution error.

2.3. Pseudospectral method. The pseudospectral method directly approximates the unknown coefficients $\{\bar{u}_i\}_{i=1}^{n_\psi}$ of $\tilde{u}(\xi)$ in (5) by exploiting orthogonality of the polynomial basis $\{\psi_i(\xi)\}_{i=1}^{n_\psi}$. That is, the coefficients $\bar{u}_i$ can be obtained by projecting the numerical solution $u(\xi)$ onto the orthogonal polynomial basis as

$$\bar{u}_i^{\mathrm{PS}} = E[u\psi_i], \quad i = 1,\dots,n_\psi, \tag{12}$$

which can be expressed as

$$\bar{u}^{\mathrm{PS}} = E[\psi \otimes A^{-1}b], \tag{13}$$

or equivalently

$$\tilde{u}^{\mathrm{PS}}(\xi) = (\psi(\xi)^T \otimes I_{n_x})E[\psi \otimes u]. \tag{14}$$

The associated optimality property of the approximation, which can be derived from optimality of orthogonal projection, is

$$\tilde{u}^{\mathrm{PS}}(\xi) = \underset{x \in (S_{n_\psi})^{n_x}}{\arg\min} \|e(x)\|_2^2. \tag{15}$$

In practice, the coefficients $\{\bar{u}_i^{\mathrm{PS}}\}_{i=1}^{n_\psi}$ are approximated via numerical quadrature as

$$\bar{u}_i^{\mathrm{PS}} = E[u\psi_i] = \sum_{k=1}^{n_q} u(\xi^{(k)})\psi_i(\xi^{(k)})w_k = \sum_{k=1}^{n_q} \left(A^{-1}(\xi^{(k)})b(\xi^{(k)})\right)\psi_i(\xi^{(k)})w_k, \tag{16}$$

where $\{(\xi^{(k)}, w_k)\}_{k=1}^{n_q}$ are the quadrature points and weights.

While stochastic Galerkin leads to an optimal approximation (11) under certain conditions and pseudospectral projection minimizes the ℓ^2-norm of the solution error (15), neither approach provides the flexibility to tailor the optimality properties of the approximation. This may be important in applications where, for example, minimizing the error in a quantity of interest is desired. To address this, we propose a general optimization-based framework for spectral methods that enables the choice of a targeted weighted ℓ^2-norm in which the solution error is minimized.

3. Stochastic least-squares Petrov–Galerkin method. As a starting point, we propose a residual-minimizing formulation that computes the coefficients $\bar{u}$ by directly minimizing the ℓ^2-norm of the residual, i.e.,

$$\tilde{u}^{\mathrm{LSPG}}(\xi) = \underset{x \in (S_{n_\psi})^{n_x}}{\arg\min} \|b - Ax\|_2^2 = \underset{x \in (S_{n_\psi})^{n_x}}{\arg\min} \|e(x)\|_{A^TA}^2, \tag{17}$$

where $\|v\|_{A^TA}^2 \equiv E[v^TA^TAv]$. Thus, the ℓ^2-norm of the residual is equivalent to a weighted ℓ^2-norm of the solution error. Using (5) and (6), we have

$$\bar{u}^{\mathrm{LSPG}} = \underset{\bar{x} \in \mathbb{R}^{n_x n_\psi}}{\arg\min} \|r(\bar{x})\|_2^2. \tag{18}$$

The definition of the residual (6) allows the objective function in (18) to be written in quadratic form as

$$\|r(\bar{x})\|_2^2 = \|b - (\psi^T \otimes A)\bar{x}\|_2^2 = \bar{x}^T E[\psi\psi^T \otimes A^TA]\bar{x} - 2E[\psi \otimes A^Tb]^T\bar{x} + E[b^Tb]. \tag{19}$$

Noting that the mapping $\bar{x} \mapsto \|r(\bar{x})\|_2^2$ is convex, the (unique) solution $\bar{u}^{\mathrm{LSPG}}$ to (18) is a stationary point of $\|r(\bar{x})\|_2^2$ and thus satisfies

$$E[\psi\psi^T \otimes A^TA]\bar{u}^{\mathrm{LSPG}} = E[\psi \otimes A^Tb], \tag{20}$$

which can be interpreted as the normal-equations form of the linear least-squares problem (18).

Consider a generalization of this idea that minimizes the solution error in a targeted weighted ℓ^2-norm by choosing a specific weighting function. Let us define a weighting function $M(\xi) \equiv M_\xi(\xi) \otimes M_x(\xi)$, where $M_\xi : \Gamma \to \mathbb{R}$ and $M_x : \Gamma \to \mathbb{R}^{n_x \times n_x}$. Then, the stochastic LSPG method can be written as

$$\tilde{u}^{\mathrm{LSPG}(M)}(\xi) = \underset{x \in (S_{n_\psi})^{n_x}}{\arg\min} \|M(b - Ax)\|_2^2 = \underset{x \in (S_{n_\psi})^{n_x}}{\arg\min} \|e(x)\|^2_{A^T M^T M A} \tag{21}$$

with $\|v\|^2_{A^T M^T M A} \equiv E[v^T A^T M^T M A v] = E[(M_\xi^T M_\xi \otimes (M_x A v)^T M_x A v]$. Algebraically, this is equivalent to

$$\begin{aligned} \bar{u}^{\mathrm{LSPG}(M)} &= \underset{\bar{x} \in \mathbb{R}^{n_x n_\psi}}{\arg\min} \|Mr(\bar{x})\|_2^2 = \underset{\bar{x} \in \mathbb{R}^{n_x n_\psi}}{\arg\min} \|(M_\xi \otimes M_x)(1 \otimes b - \left(\psi^T \otimes A\right) \bar{x})\|_2^2 \\ &= \underset{\bar{x} \in \mathbb{R}^{n_x n_\psi}}{\arg\min} \|M_\xi \otimes (M_x b) - \left((M_\xi \psi^T) \otimes (M_x A)\right) \bar{x}\|_2^2. \end{aligned} \tag{22}$$

We will restrict our attention to the case $M_\xi(\xi) = 1$ and denote $M_x(\xi)$ by $M(\xi)$ for simplicity. Now, the algebraic stochastic LSPG problem (22) simplifies to

$$\bar{u}^{\mathrm{LSPG}(M)} = \underset{\bar{x} \in \mathbb{R}^{n_x n_\psi}}{\arg\min} \|Mr(\bar{x})\|_2^2 = \underset{\bar{x} \in \mathbb{R}^{n_x n_\psi}}{\arg\min} \|Mb - (\psi^T \otimes MA)\bar{x}\|_2^2. \tag{23}$$

The objective function in (23) can be written in quadratic form as

$$\|Mr(\bar{x})\|_2^2 = \bar{x}^T E[(\psi\psi^T \otimes A^T M^T M A)]\bar{x} - 2(E[\psi \otimes A^T M^T M f])^T \bar{x} + E[b^T M^T M b]. \tag{24}$$

As before, because the mapping $\bar{x} \mapsto \|Mr(\bar{x})\|_2^2$ is convex, the unique solution $\bar{u}^{\mathrm{LSPG}(M)}$ of (23) corresponds to a stationary point of $\|Mr(\bar{x})\|_2^2$ and thus satisfies

$$E[\psi\psi^T \otimes A^T M^T M A]\bar{u}^{\mathrm{LSPG}(M)} = E[\psi \otimes A^T M^T M f], \tag{25}$$

which is the normal-equations form of the linear least-squares problem (23). This yields the following algebraic expression for the stochastic-LSPG approximation:

$$\tilde{u}^{\mathrm{LSPG}(M)}(\xi) = (\psi(\xi)^T \otimes I_{n_x}) E[\psi\psi^T \otimes A^T M^T M A]^{-1} E[\psi \otimes A^T M^T M A u]. \tag{26}$$

Petrov–Galerkin projection. Another way of interpreting the normal equations (25) is that the (weighted) residual $M(\xi)r(\bar{u}^{\mathrm{LSPG}(M)}; \xi)$ is enforced to be orthogonal to the subspace spanned by the optimal test basis $\{\phi_i\}_{i=1}^{n_\psi}$ with $\phi_i(\xi) \coloneqq \psi_i(\xi) \otimes M(\xi)A(\xi)$ and $\mathrm{span}\{\phi_i\}_{i=1}^{n_\psi} \subseteq L^2(\Gamma)$. That is, this projection is precisely the (least-squares) Petrov–Galerkin projection,

$$E[\phi^T (b - (\psi^T \otimes MA)\bar{u}^{\mathrm{LSPG}(M)})] = 0, \tag{27}$$

where $\phi(\xi) \equiv [\phi_1(\xi) \ \cdots \ \phi_{n_\psi}(\xi)]$.

Monotonic convergence. The stochastic least-squares Petrov-Galerkin is monotonically convergent. That is, as the trial subspace S_{n_ψ} is enriched (by adding polynomials to the basis),

the optimal value of the convex objective function $\|Mr(\bar{u}^{\mathrm{LSPG}(M)})\|_2^2$ monotonically decreases. This is apparent from the LSPG optimization problem (21): Defining

$$\tilde{u}^{\mathrm{LSPG}'(M)}(\xi) = \underset{x\in(S_{n_\psi+1})^{n_x}}{\arg\min} \|M(b - Ax)\|_2^2, \tag{28}$$

we have $\|M(b - A\tilde{u}^{\mathrm{LSPG}'(M)})\|_2^2 \leq \|M(b - A\tilde{u}^{\mathrm{LSPG}(M)})\|_2^2$ (and $\|u - u^{\mathrm{LSPG}'(M)}\|_{A^TM^TMA} \leq \|u - u^{\mathrm{LSPG}(M)}\|_{A^TM^TMA}$) if $S_{n_\psi} \subseteq S_{n_\psi+1}$.

Weighting strategies. Different choices of weighting function $M(\xi)$ allow LSPG to minimize different measures of the error. We focus on four particular choices:

1. $M(\xi) = C^{-1}(\xi)$, where $C(\xi)$ is a Cholesky factor of $A(\xi)$, i.e., $A(\xi) = C(\xi)C^T(\xi)$. This decomposition exists if and only if A is symmetric positive semidefinite. In this case, LSPG minimizes the energy norm of the solution error $\|e(x)\|_A^2 \equiv \|C^{-1}r(\bar{x})\|_2^2$ $(= \|e((\Psi^T \otimes I_{n_x})\bar{x})\|_A^2)$ and is mathematically equivalent to the stochastic Galerkin method described in section 2.2, i.e., $\tilde{u}^{\mathrm{LSPG}(C^{-1})} = \tilde{u}^{\mathrm{SG}}$. This can be seen by comparing (11) and (21) with $M = C^{-1}$, as $A^TM^TMA = A$ in this case.
2. $M(\xi) = I_{n_x}$, where I_{n_x} is the identity matrix of dimension n_x. In this case, LSPG minimizes the ℓ^2-norm of the residual $\|e(x)\|_{A^TA} \equiv \|r(\bar{x})\|_2^2$.
3. $M(\xi) = A^{-1}(\xi)$. In this case, LSPG minimizes the ℓ^2-norm of solution error $\|e(x)\|_2^2 \equiv \|A^{-1}r(\bar{x})\|_2^2$. This is mathematically equivalent to the pseudospectral method described in section 2.3, i.e., $\tilde{u}^{\mathrm{LSPG}(A^{-1})} = \tilde{u}^{\mathrm{PS}}$, which can be seen by comparing (15) and (21) with $M = A^{-1}$.
4. $M(\xi) = F(\xi)A^{-1}(\xi)$, where $F : \Gamma \to \mathbb{R}^{n_o\times n_x}$ is a linear functional of the solution associated with a vector of n_o output quantities of interest. In this case, LSPG minimizes the ℓ^2-norm of the error in the output quantities of interest $\|Fe(x)\|_2^2 \equiv \|FA^{-1}r(\bar{x})\|_2^2$.

We again emphasize that two particular choices of the weighting function $M(\xi)$ lead to equivalence between LSPG and existing spectral-projection methods (stochastic Galerkin and pseudospectral projection), i.e.,

$$\tilde{u}^{\mathrm{LSPG}(C^{-1})} = \tilde{u}^{\mathrm{SG}}, \qquad \tilde{u}^{\mathrm{LSPG}(A^{-1})} = \tilde{u}^{\mathrm{PS}}, \tag{29}$$

where the first equality is valid (i.e., the Cholesky decomposition $A(\xi) = C(\xi)C^T(\xi)$ can be computed) if and only if A is symmetric positive semidefinite. Table 1 summarizes the target quantities to minimize (i.e., $\|e(x)\|_\Theta^2 \equiv E[e(x)^T\Theta e(x)]$), the corresponding LSPG weighting functions, and the method names LSPG(Θ).

Table 1
Different choices for the LSPG weighting function.

Quantity minimized by LSPG		Weighting function	Method name
Quantity	Expression		
Energy norm of error	$\|e(x)\|_A^2$	$M(\xi) = C^{-1}(\xi)$	LSPG(A)/SG
ℓ^2-norm of residual	$\|e(x)\|_{A^TA}^2$	$M(\xi) = I_{n_x}$	LSPG(A^TA)
ℓ^2-norm of solution error	$\|e(x)\|_2^2$	$M(\xi) = A^{-1}(\xi)$	LSPG(2)/PS
ℓ^2-norm of error in quantities of interest	$\|Fe(x)\|_2^2$	$M(\xi) = F(\xi)A^{-1}(\xi)$	LSPG(F^TF)

Table 2
Stability constant C in (32).

	$\Theta' = A$	$\Theta' = A^T A$	$\Theta' = 2$	$\Theta' = F^T F$
$\Theta = A$	1	$\sigma_{\max}(A)$	$\frac{1}{\sigma_{\min}(A)}$	$\frac{\sigma_{\max}(F)^2}{\sigma_{\min}(A)}$
$\Theta = A^T A$	$\frac{1}{\sigma_{\min}(A)}$	1	$\frac{1}{\sigma_{\min}(A)^2}$	$\frac{\sigma_{\max}(F)^2}{\sigma_{\min}(A)^2}$
$\Theta = 2$	$\sigma_{\max}(A)$	$\sigma_{\max}(A)^2$	1	$\sigma_{\max}(F)^2$
$\Theta = F^T F$	$\frac{\sigma_{\max}(A)}{\sigma_{\min}(F)^2}$	$\frac{\sigma_{\max}(A)^2}{\sigma_{\min}(F)^2}$	$\frac{1}{\sigma_{\min}(F)^2}$	1

4. Error analysis. If an approximation satisfies an optimal-projection condition

$$\tilde{u} = \underset{x \in (S_{n_\psi})^{n_x}}{\arg\min} \|e(x)\|_\Theta^2, \tag{30}$$

then

$$\|e(\tilde{u})\|_\Theta^2 = \min_{x \in (S_{n_\psi})^{n_x}} \|e(x)\|_\Theta^2. \tag{31}$$

Using norm equivalence

$$\|x\|_{\Theta'}^2 \leq C \|x\|_\Theta^2, \tag{32}$$

we can characterize the solution error $e(\tilde{u})$ in any alternative norm Θ' as

$$\|e(\tilde{u})\|_{\Theta'}^2 \leq C \min_{x \in (S_{n_\psi})^{n_x}} \|e(x)\|_\Theta^2. \tag{33}$$

Thus, the error in an alternative norm Θ' is controlled by the optimal objective-function value $\min_{x \in (S_{n_\psi})^{n_x}} \|e(x)\|_\Theta^2$ (which can be made small if the trial space admits accurate solutions) and the stability constant C.

Table 2 reports norm-equivalence constants for the norms considered in this work. Here, we have defined

$$\sigma_{\min}(M) \equiv \inf_{x \in (L^2(\Gamma))^{n_x}} \|Mx\|_2 / \|x\|_2, \quad \sigma_{\max}(M) \equiv \sup_{x \in (L^2(\Gamma))^{n_x}} \|Mx\|_2 / \|x\|_2. \tag{34}$$

This exposes several interesting conclusions. First, if the number of output quantities of interest n_o is less than n_x, then the null space of F is nontrivial and so $\sigma_{\min}(F) = 0$. This implies that LSPG($F^T F$), for which $\Theta = F^T F$, will have an undefined value of C when the solution error is measured in other norms, i.e., for $\Theta' = A$, $\Theta' = A^T A$, and $\Theta' = 2$. It will have controlled errors only for $\Theta' = F^T F$, in which case $C = 1$. Second, note that for problems with small $\sigma_{\min}(A)$, the ℓ^2 norm in the quantities of interest may be large for the LSPG(A)/SG, or LSPG($A^T A$), while it will remain well behaved for LSPG(2)/PS and LSPG($F^T F$).

5. Numerical experiments. This section explores the performance of the LSPG methods for solving elliptic SPDEs parameterized by one random variable (i.e., $n_\xi = 1$). The maximum polynomial degree used in the stochastic space S_{n_ψ} is p; thus, the dimension of S_{n_ψ} is $n_\psi = p + 1$. In physical space, the SPDE is defined over a two-dimensional rectangular bounded domain D, and it is discretized using the finite element method with bilinear (Q_1) elements as

implemented in the Incompressible Flow and Iterative Solver Software package [24]. Sixteen elements are employed in each dimension, leading to $n_x = 225 = 15^2$ degrees of freedom excluding boundary nodes. All numerical experiments are performed on an Intel 3.1 GHz i7 CPU, 16 GB RAM, using MATLAB R2015a.

Measuring weighted ℓ^2-norms. For all LSPG methods, the weighted ℓ^2-norms can be measured by evaluating the expectations in the quadratic form of the objective function shown in (24). This requires evaluation of three expectations

$$\|Mr(\bar{x})\|_2^2 := \bar{x}^T T_1 \bar{x} - 2T_2^T \bar{x} + T_3 \tag{35}$$

with

$$T_1 := E[(\psi\psi^T \otimes A^T M^T M^T A)] \in \mathbb{R}^{n_x n_\psi \times n_x n_\psi}, \tag{36}$$

$$T_2 := E[\psi \otimes A^T M^T M b] \in \mathbb{R}^{n_x n_\psi}, \tag{37}$$

$$T_3 := E[b^T M^T M b] \in \mathbb{R}. \tag{38}$$

Note that T_3 does not depend on the stochastic-space dimension n_ψ. These quantities can be evaluated by numerical quadrature or analytically if closed-form expressions for those expectations exist. Unless otherwise specified, we compute these quantities using the `integral` function in MATLAB, which performs adaptive numerical quadrature based on the 15-point Gauss–Kronrod quadrature formula [23].

Error measures. In the experiments, we assess the error in approximate solutions computed using various spectral-projection techniques using four relative error measures (see Table 1):

$$\eta_r(x) := \frac{\|e(x)\|^2_{A^TA}}{\|b\|_2^2}, \quad \eta_e(x) := \frac{\|e(x)\|_2^2}{\|u\|_2^2}, \quad \eta_A(x) := \frac{\|e(x)\|_A^2}{\|u\|_A^2}, \quad \eta_Q(x) := \frac{\|Fe(x)\|_2^2}{\|Fu\|_2^2}. \tag{39}$$

5.1. Stochastic diffusion problems. Consider the steady-state stochastic diffusion equation with homogeneous boundary conditions,

$$\begin{cases} -\nabla \cdot (a(x,\xi)\nabla u(x,\xi)) = f(x,\xi) & \text{in } D \times \Gamma, \\ u(x,\xi) = 0 & \text{on } \partial D \times \Gamma, \end{cases} \tag{40}$$

where the diffusivity $a(x,\xi)$ is a random field and $D = [0,1] \times [0,1]$. The random field $a(x,\xi)$ is specified as an exponential of a truncated Karhunen–Loève (KL) expansion [18] with covariance kernel, $C(x,\, y) \equiv \sigma^2 \exp\left(-\frac{|x_1-y_1|}{c} - \frac{|x_2-y_2|}{c}\right)$, where c is the correlation length, i.e.,

$$a(x,\xi) \equiv \exp(\mu + \sigma a_1(x)\xi), \tag{41}$$

where $\{\mu, \sigma^2\}$ are the mean and variance of the KL expansion and $a_1(x)$ is the first eigenfunction in the KL expansion. After applying the spatial (finite-element) discretization, the problem can be reformulated as a parameterized linear system of the form (1), where $A(\xi)$ is a pa-

rameterized stiffness matrix obtained from the weak form of the problem whose (i,j)-element is $[A(\xi)]_{ij} = \int_D \nabla a(x,\xi)\varphi_i(x)\cdot\varphi_j(x)dx$ (with $\{\varphi_i\}$ standard finite element basis functions) and $b(\xi)$ is a parameterized right-hand side whose ith element is $[b(\xi)]_i = \int_D f(x,\xi)\varphi_i(x)dx$. Note that $A(\xi)$ is symmetric positive definite for this problem; thus LSPG(A)/SG is a valid projection scheme (the Cholesky factorization $A(\xi) = C(\xi)C(\xi)^T$ exists) and is equal to stochastic Galerkin projection.

Output quantities of interest. We consider n_o output quantities of interest ($F(\xi)u(\xi) \in \mathbb{R}^{n_o}$) that are random linear functionals of the solution and $F(\xi)$ is of dimension $n_o \times n_x$ having the following form:

(1) $F_1(\xi) \coloneqq g(\xi) \times G$ with $G \in [0,1]^{n_o \times n_x}$ a constant matrix: The elements of G are drawn from a uniform distribution (note that this is independent of the distribution $\rho(\xi)$) and $g(\xi)$ is a scalar-valued function of ξ. The resulting output QoI, $F_1(\xi)u(\xi)$, is a vector-valued function of dimension n_o.

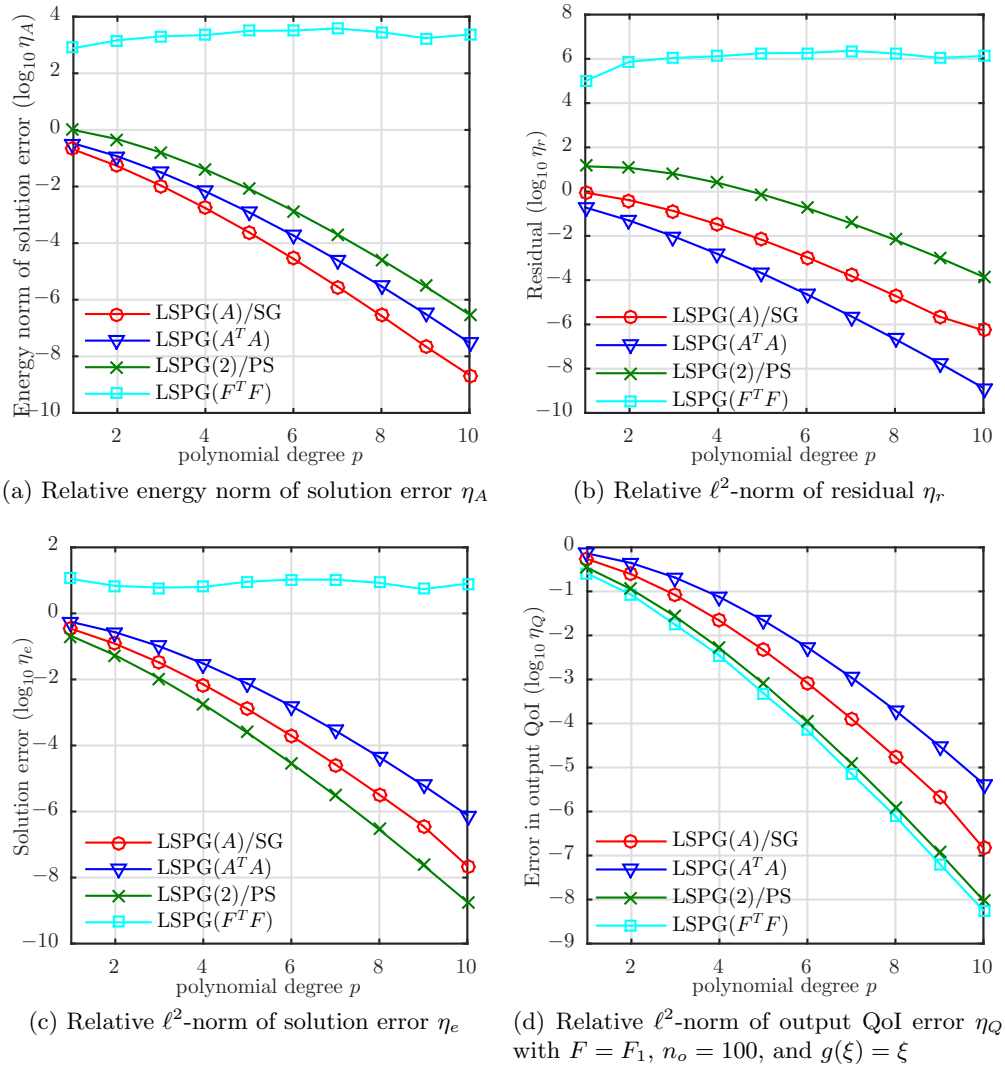
(2) $F_2(\xi) \coloneqq b(\xi)^T\bar{M}$: $\bar{M}$ is a mass matrix defined via $[\bar{M}]_{ij} \equiv \int_D \varphi_i(x)\varphi_j(x)dx$. The output QoI is a scalar-valued function $F_2(\xi)u(\xi) = b(\xi)^T\bar{M}u(\xi)$, which approximates a spatial average $\frac{1}{|D|}\int_D f(x,\xi)u(x,\xi)dx$.

5.1.1. Diffusion problem 1: Lognormal random coefficient and deterministic forcing.

In this example, we take ξ in (41) to follow a standard normal distribution (i.e., $\rho(\xi) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\xi^2}{2}\right)$ and $\xi \in (-\infty,\infty)$) and $f(x,\xi) = 1$ is deterministic. Because ξ is normally distributed, normalized Hermite polynomials (orthogonal with respect to $\langle\cdot,\cdot\rangle_\rho$) are used as polynomial basis $\{\psi_i(\xi)\}_{i=1}^{n_\psi}$.

Figure 1 reports the relative errors (39) associated with solutions computed using four LSPG methods (LSPG(A)/SG, LSPG(A^TA), LSPG(2)/PS, and LSPG(F^TF)) for varying polynomial degree p. Here, we consider the random output QoI, i.e., $F = F_1$, $n_o = 100$, and $g(\xi) = \xi$. This result shows that three methods (LSPG(A)/SG, LSPG(A^TA), and LSPG(2)/PS) monotonically converge in all four error measures, whereas LSPG(F^TF) does not. This is an artifact of rank deficiency in F_1, which leads to $\sigma_{\min}(F_1) = 0$; as a result, all stability constants C for which $\Theta = F^TF$ in Table 2 are unbounded, implying lack of error control. Figure 1 also shows that each LSPG method minimizes its targeted error measure for a given stochastic-subspace dimension (e.g., LSPG minimizes the ℓ^2-norm of the residual); this is also evident from Table 2, as the stability constant realizes its minimum value ($C = 1$) for $\Theta = \Theta'$. Table 3 shows actual values of the stability constant of this problem and well explains the behaviors of all LSPG methods. For example, the first column of Table 3 shows that the stability constant is increasing in the order (LSPG(A)/SG, LSPG(A^TA), LSPG(2)/PS, and LSPG(F^TF)), which is represented in Figure 1(a).

The results in Figure 1 do not account for computational costs. This point is addressed in Figure 2, which shows the relative errors as a function of CPU time. As we would like to devise a method that minimizes both the error and computational time, we examine a Pareto front (black dotted line), that is, a curve identifying the methods that minimize the two competing objectives considered in the figure. This typically corresponds to LSPG(2)/PS. This is because this method does not require solution of a coupled system of linear equations of dimension $n_x n_\psi$, which is required by the other three LSPG methods (LSPG(A)/SG, LSPG(A^TA), and LSPG(F^TF)). As a result, pseudospectral projection (LSPG(2)/PS) generally yields the best



(a) Relative energy norm of solution error η_A

(b) Relative ℓ^2-norm of residual η_r

(c) Relative ℓ^2-norm of solution error η_e

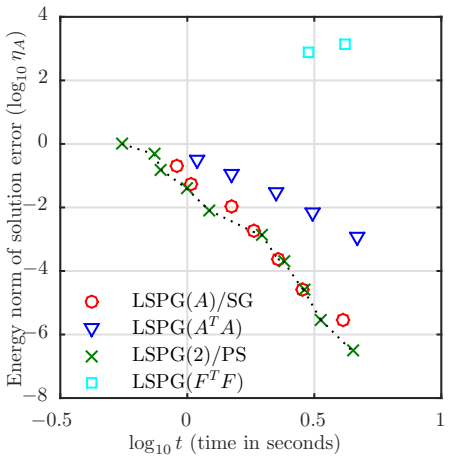
(d) Relative ℓ^2-norm of output QoI error η_Q with $F = F_1$, $n_o = 100$, and $g(\xi) = \xi$

Figure 1. *Relative error measures versus polynomial degree for diffusion problem* 1: *lognormal random coefficient and deterministic forcing. Note that each LSPG method performs best in the error measure it minimizes.*

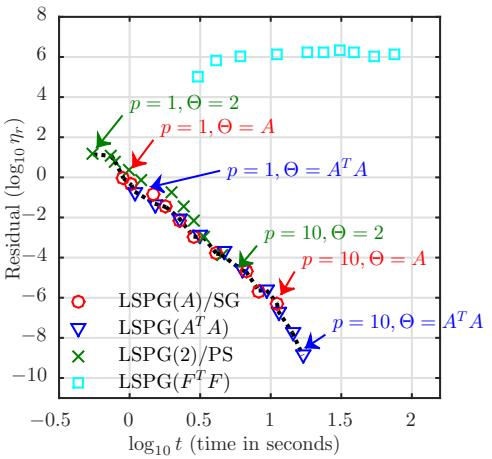
overall performance in practice, even when it produces larger errors than other methods for a fixed value of p. Also, for a fixed value of p, LSPG(A)/SG is faster than LSPG(A^TA) because the weighted stiffness matrix $A(\xi)$ obtained from the finite element discretization is sparser than $A^T(\xi)A(\xi)$. That is, the number of nonzero entries to be evaluated for LSPG(A)/SG in numerical quadrature is smaller than the ones for LSPG(A^TA), and exploiting this sparsity structure in the numerical quadrature causes LSPG(A)/SG to be faster than LSPG(A^TA). Also, note that there are cases (Figure 2(b)) where the Pareto front does not correspond to a single method; this outcome will occur with other benchmark problems considered below.

Table 3
Stability constant C of diffusion problem 1.

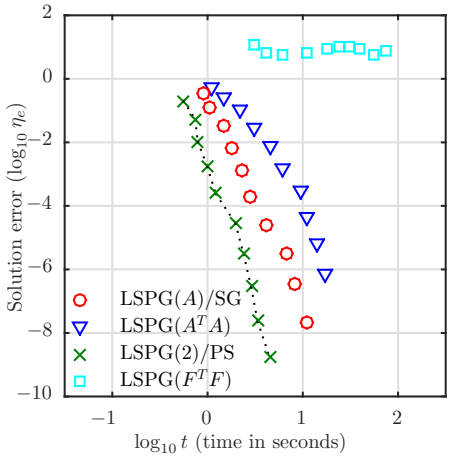
	$\Theta' = A$	$\Theta' = A^T A$	$\Theta' = 2$	$\Theta' = F^T F$
$\Theta = A$	1	26.43	2.06	11644.22
$\Theta = A^T A$	2.06	1	4.25	24013.48
$\Theta = 1$	26.43	698.53	1	5646.32
$\Theta = F^T F$	∞	∞	∞	1



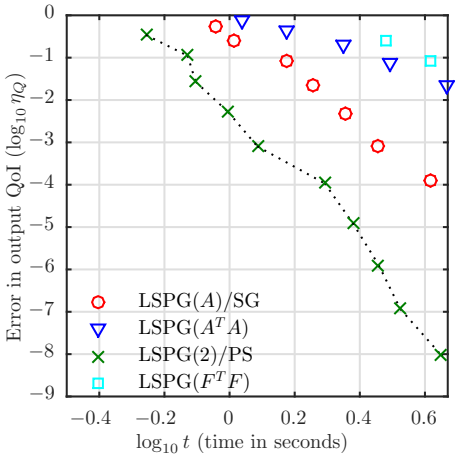
(a) Relative energy norm of solution error η_A



(b) Relative ℓ^2-norm of residual η_r



(c) Relative ℓ^2-norm of solution error η_e



(d) Relative ℓ^2-norm of output QoI error η_Q with $F = F_1$, $n_o = 100$, and $g(\xi) = \xi$

Figure 2. *Pareto front of relative error measures versus wall time for varying polynomial degree p (p varies from 1 to 10 in increments of 1 going from left to right) for diffusion problem 1: lognormal random coefficient and deterministic forcing.*

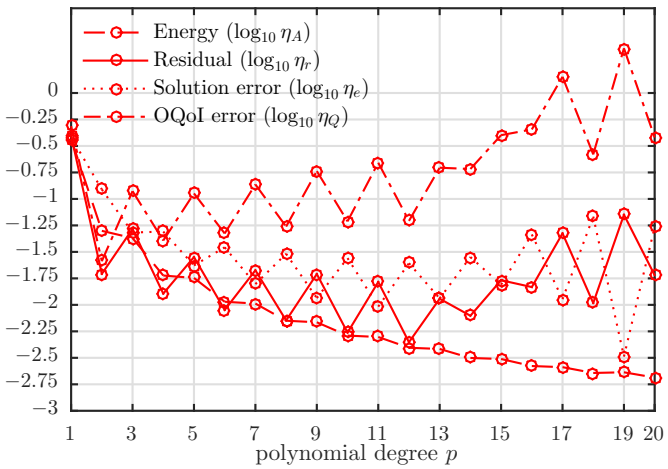
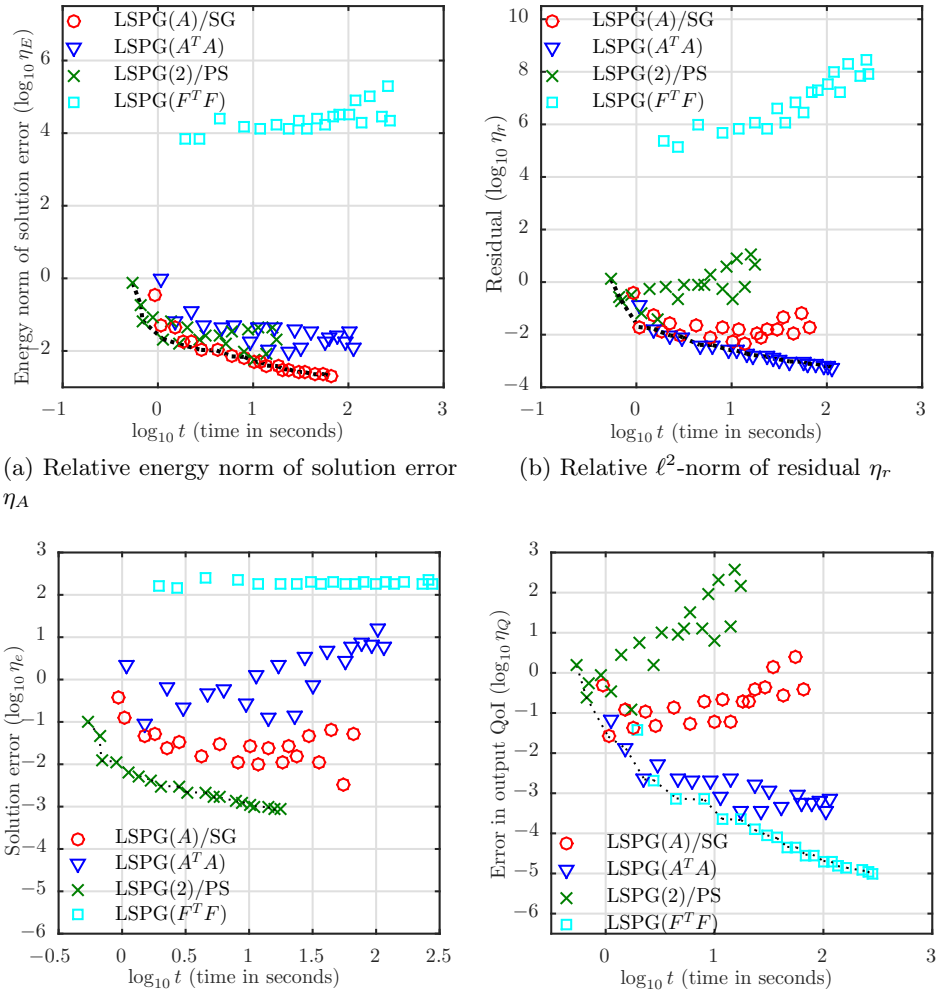


Figure 3. *Relative errors versus polynomial degree for stochastic Galerkin (i.e., LSPG(A)/SG) for diffusion problem* 2: *lognormal random coefficient and random forcing. Note that monotonic convergence is observed only in the minimized error measure* η_A.

5.1.2. Diffusion problem 2: Lognormal random coefficient and random forcing. This example uses the same random field $a(x,\xi)$ (41), but instead employs a random forcing term[1] $f(x,\xi) = \exp(\xi)|\xi - 1|$. Again, ξ follows a standard normal distribution and normalized Hermite polynomials are used as polynomial basis. We consider the second output QoI, $F = F_2$. As shown in Figure 3, the stochastic Galerkin method fails to converge monotonically in three error measures as the stochastic polynomial basis is enriched. In fact, it exhibits monotonic convergence only in the error measure it minimizes (for which monotonic convergence is guaranteed).

Figure 4 shows that this trend applies to other methods as well when effectiveness is viewed with respect to CPU time; each technique exhibits monotonic convergence in its tailored error measure only. Moreover, the Pareto fronts (black dotted lines) in each subgraph of Figure 4 show that the LSPG method tailored for a particular error measure is Pareto optimal in terms of minimizing the error and computational wall time. In the next experiments, we examine goal-oriented LSPG(F^TF) for a varying number of output quantities of interest n_o and its effect on the stability constant C. Figure 5 reports three error measures computed using all four LSPG methods. For LSPG(F^TF), the first linear function $F = F_1$ is applied with $g(\xi) = \sin(\xi)$ and a varying number of outputs $n_o = \{100, 150, 200, 225\}$. When $n_o = 225$, LSPG(F^TF) and LSPG(2)/PS behave similarly in all three weighted ℓ^2-norms. This is because when $n_0 = 225 = n_x$, then $\sigma_{\min}(F) > 0$, so the stability constants C for $\Theta = F^TF$ in Table 2 are bounded. Figure 6 reports relative errors in the quantity of interest η_Q associated with linear functionals $F = F_1$ for two different functions $g(\xi)$, $g_1(\xi) = \sin(\xi)$ and $g_2(\xi) = \xi$. Note that LSPG(A)/SG and LSPG(A^TA) fail to converge, whereas LSPG(2)/PS and LSPG(F^TF) converge, which can be explained by the stability constant C in Table 2, where $\sigma_{\max}(A) =$

[1] In [20], it was shown that stochastic Galerkin solutions of an analytic problem $a(\xi)u(\xi) = f(\xi)$ with this type of forcing are divergent in the ℓ^2-norm of solution errors as p increases.



(a) Relative energy norm of solution error η_A

(b) Relative ℓ^2-norm of residual η_r

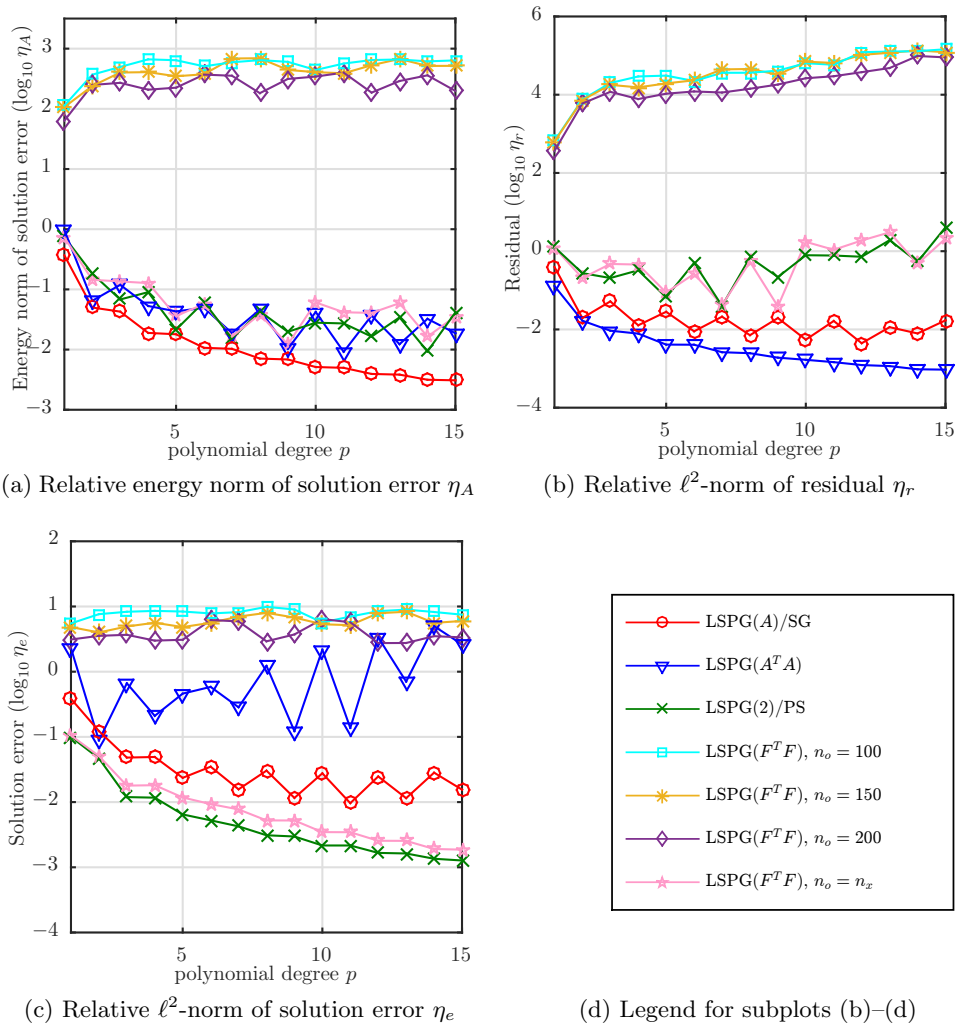
(c) Relative ℓ^2-norm of solution error η_e

(d) Relative ℓ^2-norm of output QoI error η_Q with $F = F_2$

Figure 4. *Pareto front of relative error measures versus wall time for varying polynomial degree p (p varies from 1 to 20 in increments of 1 going from left to right) for diffusion problem 2: lognormal random coefficient and random forcing.*

26.43 and $\sigma_{\min}(A) = 0.48$ for the linear operator $A(\xi)$ of this problem. LSPG(F^TF) converges monotonically and produces the smallest error (for a fixed polynomial degree p) of all the methods as expected.

5.1.3. Diffusion problem 3: Gamma random coefficient and random forcing. This section considers a stochastic diffusion problem parameterized by a random variable that has a Gamma distribution, where $a(x,\xi) \equiv \exp(1 + 0.25a_1(x)\xi + 0.01\sin(\xi))$ with density $\rho(\xi) \equiv \frac{\xi^\alpha \exp(-\xi)}{\bar{\Gamma}(\alpha+1)}$, $\bar{\Gamma}$ is the Gamma function, $\xi \in [0,\infty)$, and $\alpha = 0.5$. Normalized Laguerre polynomials (which are orthogonal with respect to $\langle\cdot,\cdot\rangle_\rho$) are used as the polynomial basis. We



(a) Relative energy norm of solution error η_A

(b) Relative ℓ^2-norm of residual η_r

(c) Relative ℓ^2-norm of solution error η_e

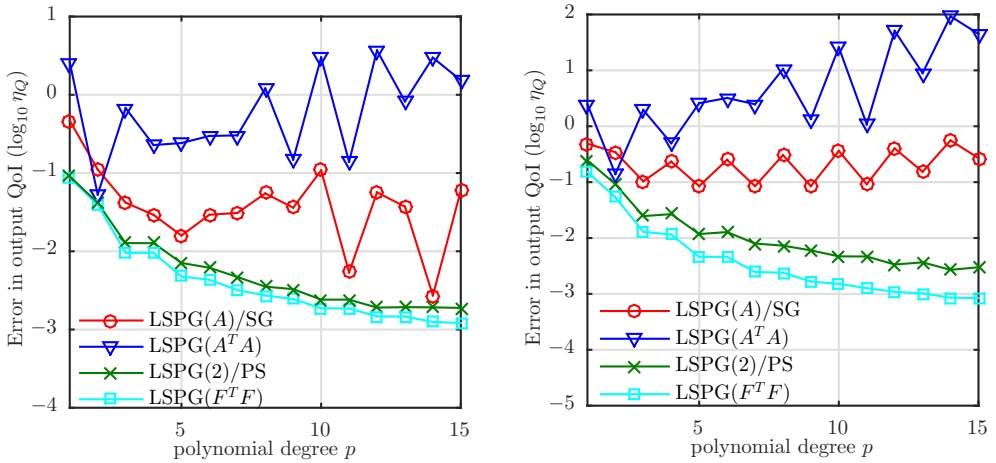
(d) Legend for subplots (b)–(d)

Figure 5. *Relative error measures versus polynomial degree for a varying dimension n_o of the output matrix $F = F_1$ for diffusion problem* 2*: lognormal random coefficient and random forcing. Note that LSPG(F^TF) has controlled errors only when $n_o = n_x$, in which case $\sigma_{min}(F) > 0$.*

consider a random forcing term $f(x, \xi) = \log_{10}(\xi)|\xi - 1|$ and the second QoI $F(\xi) = F_2(\xi) = b(\xi)^T \bar{M}$. Note that numerical quadrature is the only option for computing expectations arise in this problem.

Figure 7 shows the results of solving the problem with the four different LSPG methods. Again, each version of LSPG monotonically decreases its corresponding target weighted ℓ^2-norm as the stochastic basis is enriched. Further, each LSPG method is Pareto optimal in terms of minimizing its targeted error measure and the computational wall time.

5.2. Stochastic convection-diffusion problem: Lognormal random coefficient and deterministic forcing. We now consider a non-self-adjoint example, the steady-state convection-



(a) Relative ℓ^2-norm of output QoI error η_Q with $F = F_1$, $n_o = 100$, and $g(\xi) = g_1(\xi) = \sin(\xi)$

(b) Relative ℓ^2-norm of output QoI error η_Q with $F = F_1$, $n_o = 100$, and $g(\xi) = g_2(\xi) = \xi$

Figure 6. *Plots of the error norm of output QoI for diffusion problem* 2: *lognormal random coefficient and random forcing when a linear functional is* (a) $F(\xi) \equiv \sin(\xi) \times [0,1]^{100\times n_x}$ *and* (b) $F(\xi) = \xi \times [0,1]^{100\times n_x}$ *for varying* p *and varying* n_o.

diffusion equation

$$\begin{cases} -\epsilon\nabla\cdot(a(x,\xi)\nabla u(x,\xi)) + \vec{w}\cdot\nabla u(x,\xi) = f(x,\xi) & \text{in } D\times\Gamma, \\ u(x,\xi) = g_D(x) & \text{on } \partial D\times\Gamma, \end{cases} \tag{42}$$

where $D = [-1,1]\times[-1,1]$, ϵ is the viscosity parameter, and u satisfies inhomogeneous Dirichlet boundary conditions

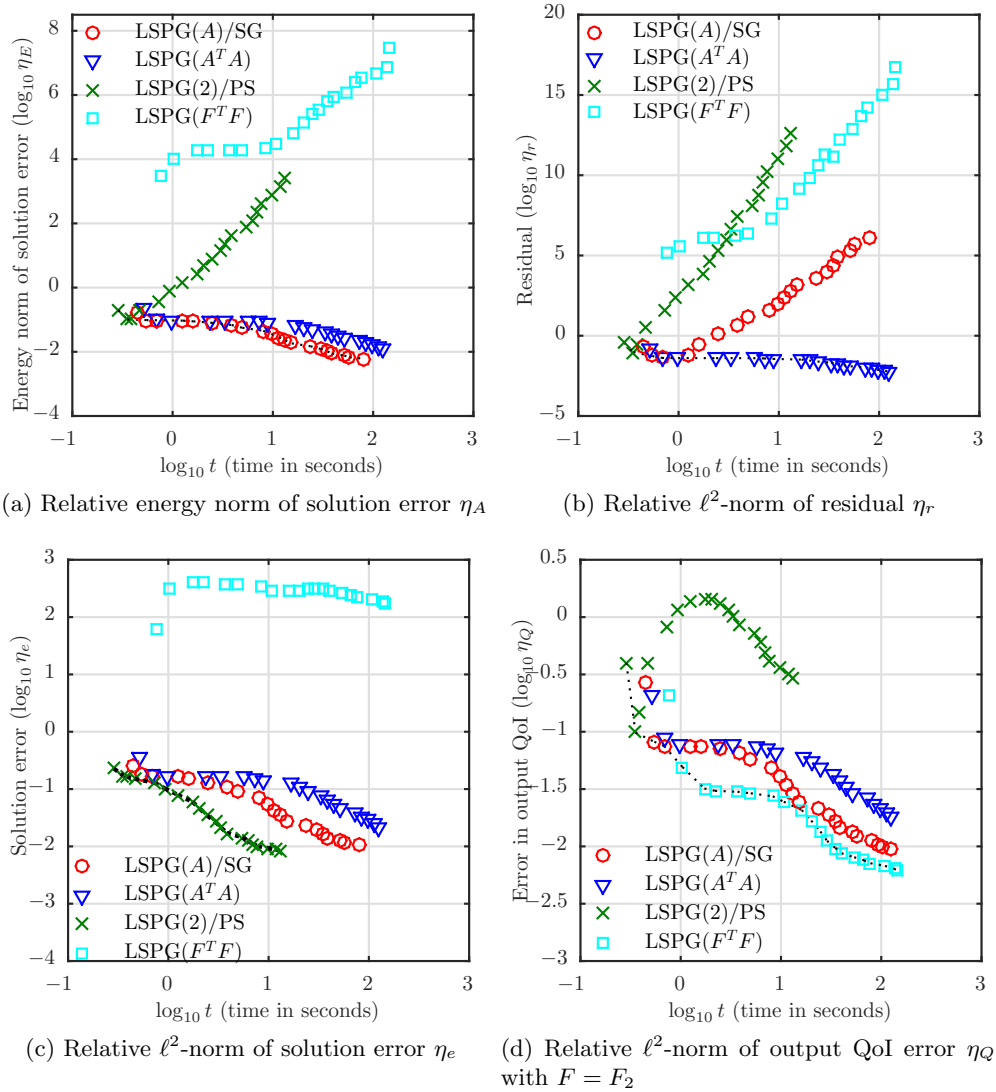
$$g_D(x) = \begin{cases} g_D(x,1) = 0 & \text{for } [-1,y]\cup[x,1]\cup[-1\le x\le 0,-1], \\ g_D(1,y) = 1 & \text{for } [1,y]\cup[0\le x\le 1,-1]. \end{cases} \tag{43}$$

The inflow boundary consists of the bottom and the right portions of ∂D, $[x,-1]\cup[1,y]$ [12]. We consider a zero forcing term $f(x,\xi) = 0$ and a constant convection velocity $\vec{w} \equiv (-\sin\frac{\pi}{6}, \cos\frac{\pi}{6})$. We consider the convection-dominated case (i.e., $\epsilon = \frac{1}{200}$).

For the spatial discretization, we essentially use the same finite element as above (bilinear Q_1 elements) applied to the weak formulation of (42). In addition, we use the streamline-diffusion method [6] to stabilize the discretization in elements with large mesh Peclet number. (See [12, Chap. 8 for details].) Such spatial discretization leads to a parameterized linear system of the form (1) with

$$A(\xi) = \epsilon D(a(\xi);\xi) + C(\xi) + S(\xi), \tag{44}$$

where $D(a(\xi);\xi)$, $C(\xi)$, and $S(\xi)$ are the diffusion term, the convection term, and the streamline-diffusion term, respectively, and $[b(\xi)]_i = \int_D f(x,\xi)\varphi_i(x)dx$. For this numerical experi-



(a) Relative energy norm of solution error η_A

(b) Relative ℓ^2-norm of residual η_r

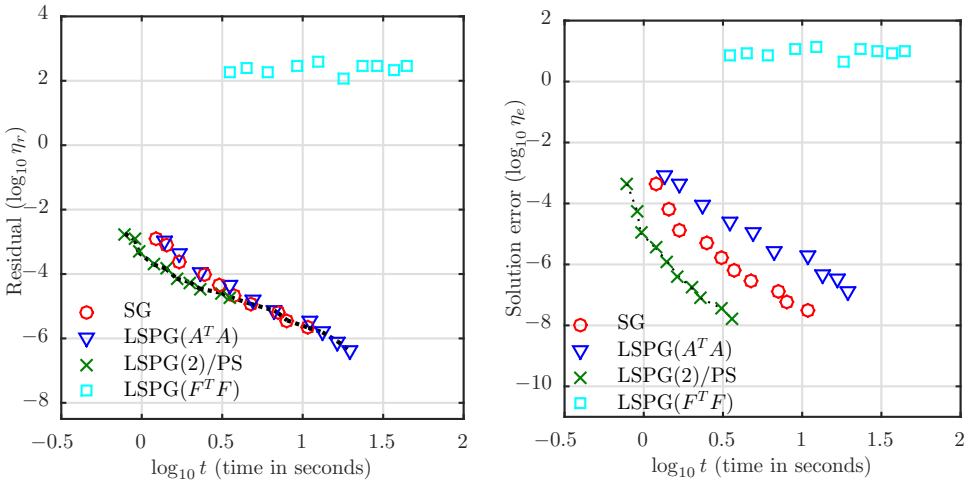
(c) Relative ℓ^2-norm of solution error η_e

(d) Relative ℓ^2-norm of output QoI error η_Q with $F = F_2$

Figure 7. *Pareto front of relative error measures versus wall time for varying polynomial degree p (p varies from 1 to 20 in increments of 1 going from left to right) for diffusion problem 3: gamma random coefficient and random forcing. Note that each method is Pareto optimal in terms of minimizing its targeted error measure and computational wall time.*

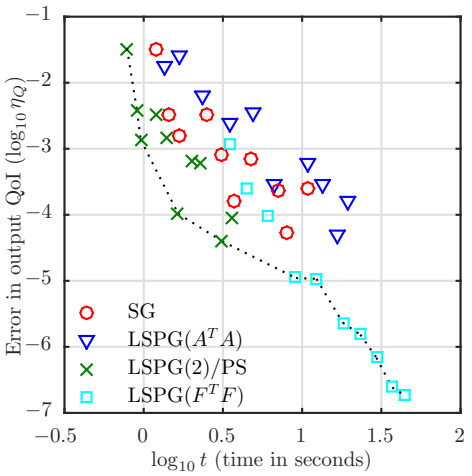
ment, the number of degrees of freedom in spatial domain is $n_x = 225$ (15 nodes in each spatial dimension) excluding boundary nodes. For LSPG(F^TF), the first linear function $F = F_1$ is applied with $n_o = 100$ outputs and $g(\xi) = \exp(\xi)|\xi - 1|$.

Figure 8 shows the numerical results computed using the stochastic Galerkin method and three LSPG methods (LSPG(A^TA), LSPG(2)/PS, LSPG(F^TF)). Note that the operator $A(\xi)$ is not symmetric positive-definite in this case; thus LSPG(A) is not a valid projection scheme (the Cholesky factorization $A(\xi) = C(\xi)C(\xi)^T$ does not exist and the energy norm of the solution error $\|e(x)\|_A^2$ cannot be defined) and stochastic Galerkin does not minimize any



(a) Relative ℓ^2-norm of residual η_r

(b) Relative ℓ^2-norm of solution error η_e



(c) Relative ℓ^2-norm of output QoI error η_Q with $F = F_1$, $n_o = 100$, $g(\xi) = \exp(\xi)|\xi - 1|$

Figure 8. *Pareto front of relative error measures versus wall time for varying polynomial degree p (p varies from* 1 *to* 10 *in increments of* 1 *going from left to right) for stochastic convection-diffusion problem: lognormal random coefficient and deterministic forcing term.*

measure of the solution error. These results show that pseudospectral projection is Pareto optimal for achieving relatively larger error measures; this is because of its relatively low cost since, in contrast to the other methods, it does not require the solution of a coupled linear system of dimension $n_x n_\psi$. In addition, the stochastic Galerkin projection is not Pareto optimal for any of the examples; this is caused by the lack of optimality of stochastic Galekin in this case and highlights the significant benefit of optimal spectral projection, which is offered by the stochastic LSPG method. In addition, the residual η_r and solution error η_e incurred by LSPG(F^TF) are uncontrolled, because $n_o < n_x$ and thus $\sigma_{\min}(F) = 0$. Finally, note that each LSPG method is Pareto optimal for small errors in its targeted error measure.

5.3. Numerical experiment with analytic computations. For the results presented above, expected values were computed using numerical quadrature (using the MATLAB function `integral`). This is a practical and general approach for numerically computing the required integrals of (36)–(38) and is the only option when analytic computations are not available (as in section 5.1.3). In this section, we briefly discuss how the costs change if analytic methods based on closed-form integration exist and are used for these integrals. Note that in general, however, analytic computation are unavailable, for example, if the random variables have a finite support (e.g., truncated Gaussian random variables as shown in [25]).

Computing T_1. Analytic computation of T_1 is possible if either $E[A^T M M A \psi_l]$ or $E[MA\psi_l]$ can be evaluated analytically. For LSPG(A)/SG and LSPG(A^TA), if $E[A\psi_l]$ can be evaluated so that the following gPC expansion can be obtained analytically,

$$A(\xi) = \sum_{l=1}^{\infty} A_l \psi_l(\xi), \quad A_l \equiv E[A\psi_l], \tag{45}$$

where $A_l \in \mathbb{R}^{n_x \times n_x}$, then T_1 can be computed analytically. Replacing $A(\xi)$ with the series of (45) for LSPG(A)/SG ($M(\xi) = C^{-1}(\xi)$) and LSPG(A^TA) ($M(\xi) = I_{n_x}$) yields

$$T_1^{\mathrm{LSPG}(A)} = \sum_{l=1}^{n_a} E\left[\psi\psi^T \otimes (A_l\psi_l)\right] = \sum_{l=1}^{n_a} E[\psi\psi^T\psi_l \otimes A_l], \tag{46}$$

and

$$T_1^{\mathrm{LSPG}(A^TA)} = E[\psi\psi^T \otimes \sum_{k=1}^{n_a}\sum_{l=1}^{n_a} (A_k\psi_k)^T (A_l\psi_l)] = \sum_{k=1}^{n_a}\sum_{l=1}^{n_a} E[\psi\psi^T\psi_k\psi_l \otimes A_k^T A_l], \tag{47}$$

where the expectations of triple or quadruple products of the polynomial basis (i.e., $E[\psi_i\psi_j\psi_k]$ and $E[\psi_i\psi_j\psi_k\psi_l]$) can be computed analytically. For LSPG(2)/PS, an analytic computation of T_1 is straightfoward because $M(\xi)A(\xi) = I_{n_x}$ and, thus,

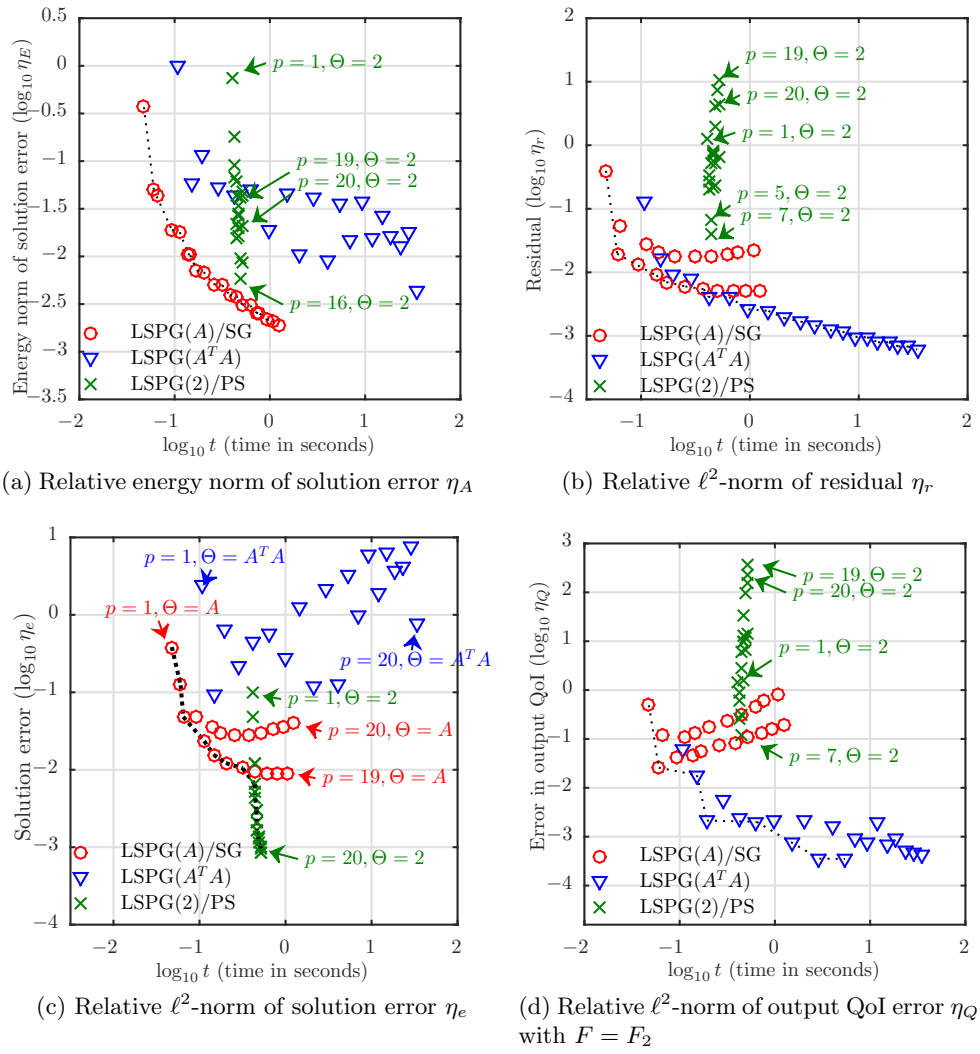
$$T_1^{\mathrm{LSPG}(2)} = E[\psi\psi^T \otimes I_{n_x}] = I_{n_x n_\psi}. \tag{48}$$

Similarly, analytic computation of T_1 is possible for LSPG(F^TF) if there exists a closed formulation for $E[F\psi_l]$ or $E[F^TF\psi_l]$, which is again in general not available.

Computing T_2. Analytic computation of T_2 can be performed in a similar way. If the random function $b(\xi)$ can be represented using a gPC expansion,

$$b(\xi) = \sum_{l=1}^{n_b} b_l \psi_l(\xi), \quad b_l \equiv E[b\psi_l], \tag{49}$$

then, for LSPG(A)/SG and LSPG(A^TA), T_2 can be evaluated analytically by computing expectations of bi or triple products of the polynomial bases (i.e., $E[\psi_i\psi_j]$ and $E[\psi_i\psi_j\psi_k]$). For LSPG(2)/PS and LSPG(F^TF), however, an analytic computation of T_2 is typically unavailable because a closed-form expression for $A^{-1}(\xi)$ does not exist.



(a) Relative energy norm of solution error η_A

(b) Relative ℓ^2-norm of residual η_r

(c) Relative ℓ^2-norm of solution error η_e

(d) Relative ℓ^2-norm of output QoI error η_Q with $F = F_2$

Figure 9. *Pareto front of relative error measures versus wall time for varying polynomial degree p (p varies from 1 to 20 in increments of 1 going from left to right) for diffusion problem 2: lognormal random coefficient and random forcing. Analytic computations are used as much as possible to evaluate expectations.*

We examine the impact of these observations on the cost of solution of the problem studied in section 5.1.2, the steady-state stochastic diffusion equation (40) with lognormal random field $a(x,\xi)$ as in (41), and random forcing $f(x,\xi) = \exp(\xi)|\xi - 1|$.

Figure 9 reports results for this problem for analytic computation of expectations. For LSPG(A)/SG, analytic computation of the expectations $\{T_i\}_{i=1}^3$ requires fewer terms than for LSPG(A^TA). In fact, comparing (46) and (47) shows that computing $T_1^{\mathrm{LSPG}(A^TA)}$ requires computing and assembling n_a^2 terms, whereas computing $T_1^{\mathrm{LSPG}(A)}$ involves only n_a terms. Additionally the quantities $\{A_k^TA_l\}_{k,l=1}^{n_a}$ appearing in the terms of $T_1^{\mathrm{LSPG}(A^TA)}$ in (47) are typically denser than the counterparts $\{A_k\}_{k=1}^{n_a}$ appearing in (46), as the sparsity pattern of

$\{A_k\}_{k=1}^{n_a}$ is identical to that of the finite element stiffness matrices. As a result, LSPG(A)/SG is Pareto optimal for small computational wall times when any error metric is considered. When the polyomial degree p is small, LSPG(A)/SG is computationally faster than LSPG(2)/PS, as LSPG(2)/PS requires the solution of $A(\xi^{(k)})u(\xi^{(k)}) = f(\xi^{(k)})$ at each quadrature point and cannot exploit analytic computation. As the stochastic basis is enriched, however, each tailored LSPG method outperforms other LSPG methods in minimizing its corresponding target error measure.

6. Conclusion. In this work, we have proposed a general framework for optimal spectral projection wherein the solution error can be minimized in weighted ℓ^2-norms of interest. In particular, we propose two new methods that minimize the ℓ^2-norm of the residual (LSPG(A^TA)) and the ℓ^2-norm of the error in an output quantity of interest (LSPG(F^TF)). Further, we showed that when the linear operator is symmetric positive definite, stochastic Galerkin is a particular instance of the proposed methodology for a specific choice of weighted ℓ^2-norm. Similarly, pseudospectral projection is a particular case of the method for a specific choice of weighted ℓ^2-norm.

Key results from the numerical experiments include the following:

- For a fixed stochastic subspace, each LSPG method minimizes its targeted error measure (Figure 1).
- For a fixed computational cost, each LSPG method often minimizes its targeted error measure (Figures 4, 7). However, this does not always hold, especially for smaller computational costs (and smaller stochastic-subspace dimensions) when larger errors are acceptable. In particular pseudospectral projection (LSPG(2)/PS) is often significantly less expensive than other methods for a fixed stochastic subspace, as it does not require solving a coupled linear system of dimension $n_x n_\psi$ (Figures 2, 8). Alternatively, when analytic computations are possible, stochastic Galerkin (LSPG(A)/SG)) may be significantly less expensive than other methods for a fixed stochastic subspace (Figure 9).
- Goal-oriented LSPG(F^TF) can have uncontrolled errors in error measures that deviate from the output-oriented error measure η_Q when the linear operator F has more columns n_x than rows n_o (Figure 5). This is because the minimum singular value is zero in this case (i.e., $\sigma_{\min}(F) = 0$)), which leads to unbounded stability constants in other error measures (Table 2).
- Stochastic Galerkin often leads to divergence in different error measures (Figure 3). In this case, applying LSPG with the appropriate targeted error measure can significantly improve accuracy (Figure 4).

Future work includes developing efficient sparse solvers for the stochastic LSPG methods and extending the methods to parameterized nonlinear systems.

REFERENCES

[1] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034.

[2] I. Babuška, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.

[3] A. Barth, C. Schwab, and N. Zollinger, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.
[4] P. B. Bochev and M. D. Gunzburger, *Finite element methods of least-squares type*, SIAM Rev., 40 (1998), pp. 789–837.
[5] P. B. Bochev and M. D. Gunzburger, *Least-Squares Finite Element Methods*, Appl. Math. Sci. 166, Springer, New York, 2009.
[6] A. N. Brooks and T. J. Hughes, *Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations*, Comput. Methods Appl. Mech. Engrg., 32 (1982), pp. 199–259.
[7] K. Carlberg, M. Barone, and H. Antil, *Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction*, J. Comput. Phys., 330 (2017), pp. 693–734.
[8] K. Carlberg, C. Farhat, and C. Bou-Mosleh, *Efficient nonlinear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations*, Internat. J. Numer. Methods Engrg., 86 (2011), pp. 155–181.
[9] K. Carlberg, C. Farhat, J. Cortial, and D. Amsallem, *The GNAT method for nonlinear model reduction: Effective implementation and application to computational fluid dynamics and turbulent flows*, J. Comput. Phys., 242 (2013), pp. 623–647.
[10] M. K. Deb, I. M. Babuška, and J. T. Oden, *Solution of stochastic partial differential equations using Galerkin finite element techniques*, Comput. Methods Appl. Mech. Engrg., 190 (2001), pp. 6359–6372.
[11] H. C. Elman, C. W. Miller, E. T. Phipps, and R. S. Tuminaro, *Assessment of collocation and Galerkin approaches to linear diffusion equations with random data*, Int. J. Uncertain. Quantif., 1 (2011), pp. 19–33.
[12] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Oxford University Press, Oxford, 2014.
[13] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Dover, New York, 2003.
[14] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan, *Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications*, J. Comput. Phys., 230 (2011), pp. 3668–3694.
[15] H. Holden, B. Øksendal, J. Ubøe, and T. Zhang, *Stochastic Partial Differential Equations*, in Stochastic Partial Differential Equations, Springer, New York, 1996, pp. 141–191.
[16] B.-N. Jiang and L. A. Povinelli, *Least-squares finite element method for fluid dynamics*, Comput. Methods Appl. Mech. Engrg., 81 (1990), pp. 13–37.
[17] F. Y. Kuo, C. Schwab, and I. H. Sloan, *Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 3351–3374.
[18] M. Loéve, *Probability Theory*, Vol. II, Grad. Texts in Math. 46, Springer, New York, 1978.
[19] H. G. Matthies and A. Keese, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1295–1331.
[20] A. Mugler and H.-J. Starkloff, *On the convergence of the stochastic Galerkin method for random elliptic partial differential equations*, ESAIM Math. Model. Numer. Anal., 47 (2013), pp. 1237–1263.
[21] F. Nobile, R. Tempone, and C. G. Webster, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345.
[22] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2003.
[23] L. F. Shampine, *Vectorized adaptive quadrature in MATLAB*, J. Comput. Appl. Math., 211 (2008), pp. 131–140.
[24] D. Silvester, H. Elman, and A. Ramage, *Incompressible Flow and Iterative Solver Software (IFISS) version* 3.4, 2015, http://www.manchester.ac.uk/ifiss/.
[25] E. Ullmann, H. C. Elman, and O. G. Ernst, *Efficient iterative solvers for stochastic Galerkin discretizations of log-transformed random diffusion problems*, SIAM J. Sci. Comput., 34 (2012), pp. A659–A682.
[26] D. Xiu, *Efficient collocational approach for parametric uncertainty analysis*, Commun. Comput. Phys., 2 (2007), pp. 293–309.
[27] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, Princeton, NJ, 2010.

[28] D. Xiu and G. E. Karniadakis, *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 4927–4948.
[29] D. Xiu and G. E. Karniadakis, *The Wiener–Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.
[30] D. Zhang, *Stochastic Methods for Flow in Porous Media: Coping with Uncertainties*, Academic Press, New York, 2001.