A randomised finite element method for elliptic partial differential equations

Yue Wu[†], Dimitris Kamilis[†], and Nick Polydorides[†] [⋆]

[†]School of Engineering, University of Edinburgh, Edinburgh, UK
[⋆]The Alan Turing Institute, London, UK

18 March, 2019

Abstract

We consider a randomised implementation of the finite element method (FEM) for elliptic partial differential equations on high-dimensional models. This is motivated by the need to expedite the assembly of the associated stiffness matrices as indeed the solution of the resulting linear systems at the many-query context, where the numerical solution is sought for many different parameters. Our approach involves converting the symmetric positive definite FEM system into an over-determined least squares problem, and then projecting the solution onto a low-dimensional subspace. The low-dimensional system can be sketched as a product of two high-dimensional matrices, using some parameter-dependent non-uniform sampling distributions. In this context, we investigate how these distributions relate to those based on statistical leverage scores, and show that the adopted projection enables near optimal sampling. Further we derive a set of bounds for the error components affecting the quality of the proposed solution. Our algorithm is tested on numerical simulations for the elliptic diffusion boundary value problem with Dirichlet and Neumann conditions. Our results show that the randomised FEM approach has on average a ten-fold improvement on the computational times compared to the classical deterministic framework at the expense of a moderately small error.

1 Introduction

We consider the implementation of the Finite Element Method (FEM) in high-dimensional discrete models associated with elliptic partial differential equations, focusing in particular on the many-query context, where an approximate solution is sought for various inhomogeneous parameter fields. Owing to its versatility in handling models of realistic complexity, the method has been at the forefront of numerical computing and simulation for electromagnetic, mechanical, heat transfer and fluid dynamics systems [ESW14]. Beyond its appeal in applied engineering research, the method has led to several algorithmic advances in scientific computing such as matrix preconditioning, fast iterative algorithms and multigrid methods [Saa03].

Our work is motivated by the need to expedite model prediction, also referred to as forward problem evaluation, in the context of a FEM-based simulation in the cases where an approximate, yet fast solution is imperative. Realising fast, real-time simulation, with large three-dimensional models is a formidable task and yet it can be of critical importance in a number of instances like online calibration of a sensor network or the control of a manufacturing process, where accurate, expensive simulations are typically deferred offline on specialised high performance computing infrastructure. Reducing the computing time for forward evaluations has been a long-standing goal for model-order reduction in computational partial differential equations and the main bottleneck of statistical inference algorithms for inverse problems and Bayesian uncertainty quantification, where multiple model runs are sought in the many query or Monte Carlo simulation context [BOCW17], [LPS14]. It is worth emphasising however, that in practical applications involving experimental data contaminated with noise, an approximate evaluation to the respective forward problem suffices for the purpose of making model-based inferences with such data [BJMS15]. In this context, an approximate solution to an accurate model is preferable to an accurate solution to an oversimplified model as the former typically allows to quantify and control the model-induced error that's otherwise hard to estimate [CDSS18].

When the accuracy of the solution can be traded off against speed, algorithms based on randomised numerical linear algebra present a competitive alternative [Woo14]. The connection between this framework and numerical computing goes back to the sketching approach of Drineas and Mahoney for the Laplacian of a graph, where they coined the relationship between statistical leverage scoring and the so-called effective resistance of the graph [DM10]. Although the method is suitable to symmetric diagonally dominant (SDD) linear systems while the FEM systems are typically not SDD, there is an evocative similarity in the structure of the coefficient matrices in the respective systems, particularly in solving elliptic partial differential equations, where FEM leads to the so-called stiffness matrix, a generalisation of the Laplacian paradigm discretised on unstructured grids. The concept of effective resistances has since led to sketch-based preconditioners for SDD systems through sparsifier algorithms aimed at reducing the matrix fill-in and thus render the resulting systems solvable in a time that is asymptotically linear to their sparsity level [CKM$^+$14], [ST06]. More recently, Avron and Toledo have proposed a generalisation of this framework to the FEM context adapting the idea of effective resistance to that of the effective stiffness of an element in the grid [AT11], relaxing the restriction to SDD systems. In particular, for the FEM sparse symmetric positive definite (SSPD) matrices, they derive formulas for the effective stiffness and show their equivalence to the statistical leverage scores, claiming that sampling $O(n \log n)$ elements according to those can lead to a sparser preconditioner such that the resulting system is solvable, with high-probability, in a small number of iterations.

While the above approaches focus predominantly on the efficient preconditioning and assembling of such systems, randomised algorithms for large-scale linear systems have already been proposed and implemented. The framework of Gower and Richtarik for example randomises the row-action iterative methods by taking random projections onto convex sets [GR15]. Applied to the FEM-induced SSPD systems, the underpinning algorithm is equivalent to a stochastic gradient descent method with provable convergence, while the approach in [GR16] iteratively sketches the inverse of a matrix. Besides, there is a wealth of literature on sketching methods for least-squares problems, constrained or unconstrained, using data-oblivious subspace embeddings (randomised sketching transforms) that preserve some approximate isometry and orthogonality in the sketched systems. We refer the reader to the work of Woodruff [Woo14], Drineas and Mahoney [DMMS11], Pilanci and Wainwright [PW14], and Boutsidis and Drineas. [DB09].

In [BY09], Bertsekas and Yu present an alternative approach for simulating an approximate solution to linear fixed-point equations and least squares problems, in the context of evaluating the cost of stationary policies in a Markovian decision. This is based on approximate dynamic programming algorithms that solve a projected form of Bellman's equation in a low-dimensional subspace, using sample-based approximations. Subsequently this framework was extended and coupled with importance sampling schemes by Polydorides et al. [PWB12] in solving linear inverse problems associated with Fredholm integral equations of the first kind, exploiting the characteristic smooth structure of the integral kernels.

In the many-query context one faces two computational challenges, namely the fast assembly of the large FEM system for each query (parameter vector), and the efficient solution of the resulting FEM system to some level of accuracy. We begin by transforming the linear SSPD FEM system into an over-determined least squares problem, and then apply a deterministically chosen orthogonal projection onto a low-dimensional subspace. Our efforts then focus on the efficient randomisation of the projected least-squares equations for every parameter query, by extending ideas from [DM10] and [BY09]. In this context, our contributions are in the development of the projected randomisation algorithm, the analysis of the impact of the projection on the approximation of the leverage scores, and the derivation of error bounds for the sketched projected solution. Further, we implement the proposed algorithm on Dirichlet and Neumann problems for the elliptic diffusion partial differential equation.

Our paper is organised as follows: In the next section we provide a brief introduction to the Galerkin FEM formulation, starting from the definition of the variational form of the elliptic boundary value problem. We then derive the subspace-projected formulation that yields a low-dimensional system and in the next section we provide a description of the sketching algorithm. Subsequently, we investigate the conditions under which the row-norm based sampling distribution approximates the optimal leverage score distribution, with emphasis on the impact of the subspace projection in this respect. Thereafter, we present an error analysis to quantify the errors imparted to the solution due to the subspace approximation, the sketching process and the regression misfit. We end our report with numerical results.

2 Finite element method preliminaries

We consider the elliptic partial differential equation

$$-\nabla\cdot p(x)\nabla u(x) = f(x) \quad \text{in } \Omega\subset\mathbb{R}^3, \tag{1}$$

and associated boundary conditions

$$u = g^{(D)} \text{ on } \partial\Omega_D \quad\text{and}\quad \nabla u\cdot\hat{n} = g^{(N)} \text{ on } \partial\Omega_N, \tag{2}$$

on a bounded and simply connected domain $\Omega\subset\mathbb{R}^d$, $d = 2, 3$ with a Lipschitz smooth boundary $\partial\Omega = \partial\Omega_D\cup\partial\Omega_N$, and $\hat{n}$ the unit normal on the boundary. Further let $p(x)$ be a real, scalar and positive parameter function supported over the closure of the domain

$$0 < p_{\min} \le p(x) \le p_{\max} < \infty, \quad x\in\Omega, \tag{3}$$

where $x \doteq (x_1,\dots,x_d)$ denotes the spatial coordinate vector. In this work we consider primarily the three-dimensional case ($d = 3$) but whenever possible we keep the notation general for the adaptation of our methodology to $d = 2$. Multiplying (1) by an appropriate test function v, then integrating over the domain and invoking the divergence theorem yields

$$\int_\Omega \mathrm{d}x\,\nabla u\cdot p\nabla v = \int_\Omega \mathrm{d}x\, fv + \int_{\partial\Omega}\mathrm{d}s\, g^{(N)}v, \tag{4}$$

where dx and ds are volume and surface integration elements respectively. Using the standard definition of the Sobolev space on this domain as

$$\mathcal{H}^1(\Omega) \doteq \Big\{u(\Omega)\Big| u, \frac{\partial u}{\partial x_q}, q = 1,\dots,d\in L^2(\Omega)\Big\}, \tag{5}$$

where $L^2(\Omega)$ is the space of square-integrable functions on Ω we can define the solution and test function spaces as

$$\mathcal{H}^1_U \doteq \left\{ u \in \mathcal{H}^1(\Omega) \middle| u = g^{(D)} \text{ on } \partial\Omega_D \right\}, \quad \mathcal{H}^1_0 \doteq \left\{ v \in \mathcal{H}^1(\Omega) \middle| v = 0 \text{ on } \partial\Omega_D \right\} \tag{6}$$

respectively. Assuming $p \in L^\infty(\bar{\Omega})$ and $f \in L^2(\Omega)$ are in the Banach spaces of real functions defined on the closure of the domain $\bar{\Omega}$ and its interior respectively, and similarly $g^{(D)} \in H^{\frac{1}{2}}(\partial\Omega_D)$, $g^{(N)} \in H^{-\frac{1}{2}}(\partial\Omega_N)$, the weak form of the boundary value problem (1)-(2) is to find a function $u \in \mathcal{H}^1_U$ such that

$$\int_\Omega \mathrm{d}x\, \nabla u \cdot p \nabla v = \int_\Omega \mathrm{d}x\, f v + \int_{\partial\Omega} \mathrm{d}s\, g^{(N)} v, \qquad \forall v \in \mathcal{H}^1_0. \tag{7}$$

In these conditions the existence and uniqueness of the weak solution is guaranteed by Lax-Milgram's theorem [ESW14].

To derive the Galerkin finite element approximation method from the weak form (7), we consider $\mathcal{T}_\Omega \doteq \{\Omega_1, \dots, \Omega_k\}$ a tetrahedral mesh and $\mathcal{S}^1_\Omega \subset \mathcal{H}^1_0$ the conforming finite dimensional space associated with the chosen finite element basis defined on $\mathcal{T}_\Omega$. Let us explicitly quote also $\Delta_\Omega \doteq \{\partial\Omega_1, \dots, \partial\Omega_\tau\}$ the set of τ triangular faces (resp. straight edges in $d = 2$) spanning the outer surface of the discrete domain so that $\bigcup_{\ell=1}^k \Omega_\ell \approx \Omega$ and $\bigcup_{\ell=1}^\tau \partial\Omega_\ell \approx \partial\Omega$. The notations $|\Omega|$ and $|\partial\Omega|$ are used to express the volume (resp. area) and boundary area (resp. length) of the domain respectively. In particular, we denote the subset of Δ_Ω on the Neumann boundary as Δ^N_Ω. If $\mathcal{S}^1_\Omega \doteq \mathrm{span}\{\phi_1(x), \dots, \phi_n(x), \dots, \phi_{n+n_\partial}(x)\}$ comprises of piecewise linear shape functions with local support over the elements in $\mathcal{T}_\Omega$ then we can express the FEM approximation of the potential as

$$u_h = \sum_{i=1}^n u_i \phi_i + \sum_{i=n+1}^{n+n_\partial} u_i \phi_i, \tag{8}$$

separating the expansion between the functions defined on the n interior and n_∂ boundary nodes. From this, the finite element formulation of the boundary value problem is to find $u_h \in \mathcal{S}^1_\Omega$ such that

$$\sum_{\Omega_\ell \in \mathcal{T}_\Omega} \int_{\Omega_\ell} \mathrm{d}x\, \nabla u_h \cdot p \nabla u_h = \sum_{\Omega_\ell \in \mathcal{T}_\Omega} \int_{\Omega_\ell} \mathrm{d}x\, f v_h + \sum_{\Omega_{\partial\ell} \in \Delta^N_\Omega} \int_{\Omega_{\partial\ell}} \mathrm{d}s\, g^{(N)} v_h, \forall v_h \in \mathcal{S}^1_\Omega, \tag{9}$$

where $g^{(N)}$ is the Neumann function on $\partial\Omega_N$. Further we select a piecewise constant basis of characteristic functions $\{\chi_1, \dots, \chi_k\}$, where $\chi_\ell = 1$ over Ω_ℓ and zero elsewhere, so that the parameter and forcing terms[1] are expressed as

$$p_h = \sum_{\ell=1}^k p_\ell \chi_\ell, \quad \text{and} \quad f_h = \sum_{\ell=1}^k f_\ell \chi_\ell. \tag{10}$$

[1] This choice of basis is not restrictive although it simplifies the notation and the cal-

We then write the Galerkin system of equations for the vector $\{u_1, \ldots, u_{n+n_\partial}\}$

$$\sum_{j=1}^{n+n_\partial} u_j \sum_{\Omega_\ell \in \mathcal{T}_\Omega} \int_{\Omega_\ell} \mathrm{d}x \, \nabla\phi_i \cdot p_\ell \nabla\phi_j = \sum_{\Omega_\ell \in \mathcal{T}_\Omega} \int_{\Omega_\ell} \mathrm{d}x \, f_\ell \phi_i + \sum_{\Omega_{\partial\ell} \in \Delta_\Omega^N} \int_{\Omega_{\partial\ell}} \mathrm{d}s \, {g^{(N)}}_\ell \phi_i, \tag{11}$$

for $i = 1, \ldots, n + n_\partial$. Note that in the instance of the Dirichlet problem where $\partial\Omega_N = \emptyset$, the surface integral vanishes and the coefficients $\{u_{n+1}, \ldots, u_{n+n_\partial}\}$ are fixed through $g^{(D)}$, hence the Galerkin system of equations has n degrees of freedom, while for the Neumann problem u has dimension $n+n_\partial-1$, after applying the uniqueness condition. The assembly of (11) over the elements in the domain yields a system

$$Au = b, \tag{12}$$

where $A \in \mathbb{R}^{n+n_\partial \times n+n_\partial}$, the so-called FEM stiffness matrix, that is sparse, symmetric and positive-definite. The FEM construction guarantees that $b \in \mathbb{R}^{n+n_\partial}$ is in the column space of A therefore the system (12) admits a unique solution $u^* = A^{-1}b$. The focus of our work is the efficient approximation of u^* in the many p query context, such as the one used in Monte-Carlo approaches for inverse problems [BJMS15]. As such our approach will be faced with two main challenges: the efficient assembly of the stiffness matrix, and thereafter the speedy solution of the resulted FEM problem. For completeness, we define our target problem as follows.

Definition 2.1. *If $p^{(1)}(x), \ldots, p^{(N)}(x)$ are parameter functions corresponding to the boundary value problem (1)-(2) with fixed boundary and forcing conditions and $A^{(1)}, \ldots, A^{(N)}$ the respective FEM stiffness matrices, compute the approximate solutions $u^{(i)}$ of*

$$A^{(i)}u^{(i)} = b, \quad for \quad i = 1, \ldots, N,$$

where N and the dimensions of A are large.

2.1 Notation

In a discrete model $\mathcal{T}_\Omega$ with k elements we express as p_ℓ the ℓth element of the positive parameter vector $p \in \mathbb{R}_+^k$, $|\Omega_\ell|$ the volume or area of the ℓth element in d dimensions, $\omega \doteq \{|\Omega_\ell|\}_{\ell=1}^k$ and $u \in \mathbb{R}^{n+n_\partial}$ the FEM solution coefficients in (12). For a matrix X, $X_{(\ell)*}$ and $X_{*(\ell)}$ denote the ℓth row and

culations. Alternatively, one could take for example

$$f_\ell = \frac{1}{|\Omega_\ell|} \int_{\Omega_\ell} \mathrm{d}x \, f$$

and compute the volume integrals encountered in the Galerkin formulation (11) using numerical quadrature rules.

column of X respectively, X_{ij} its i,jth element, $X_{i,(j:j+n)}$ the elements on the ith row between columns j and $j+n$ and $X_{*(i:j)}$ the part of the matrix in columns i to j. We use X^{-1} and $X^{\dagger}$ for matrix inverse and pseudo-inverse. Further we write $\|\cdot\|$ for the Euclidean norm for a vector or the spectral norm of a matrix, $\|\cdot\|_{\max}$ the max norm and $\|\cdot\|_F$ the Frobenius norm of a matrix. Moreover we express by $\kappa(X)$ the condition number of X, $\sigma_i(X)$ its ith singular value, and $\lambda_i(X)$ its corresponding eigenvalue. For $X \in \mathbb{R}^{m\times n}$ with $m \geq n$ we denote its singular value decomposition as $X = U_X \Sigma_X V_X^T$ where $U_X \in \mathbb{R}^{m\times n}$, $\Sigma_X \in \mathbb{R}^{n\times n}$ and $V_X \in \mathbb{R}^{n\times n}$. Unless otherwise stated, singular values and eigenvalues are ordered in descending order, therefore for a matrix $X \in \mathbb{R}^{n\times n}$, $\lambda_1(X) = \lambda_{\max}(X)$ is the largest eigenvalue and $\lambda_n(X) = \lambda_{\min}(X)$ the smallest. For two matrices X and Y with the same number of rows we write $(X|Y)$ to denote the augmented matrix formed by column concatenation. We express as I_n the identity matrix in dimension n, and $[n]$ the set of integers from 1 to n inclusive. The notation $\mathbb{E}_\xi[\cdot]$ stands for the expectation of a random variable or matrix under probability ξ, and $\mathrm{Var}_\xi[\cdot]$ denotes the variance of an estimator under ξ. Finally, for two scalar quantities a and b, $a \vee b$ stands for the supremum of a and b.

2.2 Assembly of the stiffness matrix

From the definitions of the shape functions in (8), forcing terms and Neumann boundary conditions, the element of the stiffness matrix is given by

$$A_{ij} = \sum_{\Omega_\ell \in \mathcal{T}_\Omega} |\Omega_\ell|\, p_\ell\, \nabla\phi_i \cdot \nabla\phi_j, \quad i,j \in \mathcal{I}_\ell, \tag{13}$$

where $\mathcal{I}_\ell$ is the index set of the $d+1$ vertices of the ℓth element. Forming the sparse matrices $D_\ell \in \mathbb{R}^{d\times(n+n_\partial)}$ of the gradients $\nabla\phi$ associated with Ω_ℓ, and subsequently stacking them for all k elements of the model in a matrix $D \in \mathbb{R}^{kd\times(n+n_\partial)}$ we can then define

$$Y = Z^{\frac{1}{2}} D \tag{14}$$

for a positive diagonal $Z \in \mathbb{R}^{kd\times kd}$ such that

$$Z = z \otimes I_d, \quad \text{for a vector} \quad z = \omega \odot p, \tag{15}$$

where $\otimes$ and $\odot$ denote the Kronecker and Hadamard products respectively. The FEM construction intrinsically allows forming the stiffness matrix either as a high-dimensional sum

$$A = \sum_{\ell=1}^{k} Y_\ell^T Y_\ell, \quad \text{where} \quad Y_\ell = \sqrt{z_\ell} D_\ell, \tag{16}$$

or a matrix product

$$A = Y^T Y = \sum_{\ell=1}^{k} \sum_{j=0}^{d-1} \sum_{j'=0}^{d-1} Y^T_{*(3\ell-j)} Y_{(3\ell-j')*}, \tag{17}$$

both of which typically require an efficient assembly using reference elements and geometry mappings [KL07]. For notational simplicity we demonstrate our methodology for the Dirichlet problem where A has dimensions $n \times n$, and revisit the trivial extensions for the Neumann problem in the numerical results section. In our approach we follow the product construction (17), for which the spectrum of the matrix A and thus that of Y are of critical importance, hence we quote two relevant bounds from [KHX14].

Lemma 2.2. *For $Y = Z^{\frac{1}{2}} D$ with a singular value decomposition (SVD) $Y = U_Y \Sigma_Y V_Y^T$, then the largest eigenvalue of the stiffness matrix A is $\lambda_1(\Sigma_Y)^2$ and it is bounded by*

$$\max_i A_{ii} \leq \lambda_1(\Sigma_Y)^2 \leq (d+1) \max_i A_{ii} \tag{18}$$

Proof. The proof is in lemma 4.1 of [KHX14]. □

Lemma 2.3. *For $Y = Z^{\frac{1}{2}} D$ with SVD $Y = U_Y \Sigma_Y V_Y^T$, then the smallest eigenvalue of the stiffness matrix A is bounded from below by*

$$\lambda_n(\Sigma_Y)^2 \geq C p_{\min} \frac{1}{k} \begin{cases} \left(1 + \log \frac{\bar{\omega}}{\omega_{\min}}\right)^{-1}, & d = 2 \\ \left(\frac{1}{k} \sum_{\Omega_\ell \in \mathcal{T}_\Omega} \left(\frac{\bar{\omega}}{\omega_\ell}\right)^{\frac{1}{2}}\right)^{-\frac{2}{3}}, & d = 3 \end{cases} \tag{19}$$

where $\bar{\omega}$ is the average element size, and $\omega_{\min}$ is the minimum element size in the mesh. C is a generic constant, k the total number of elements in the mesh and p_{min} the minimum value in the parameter vector.

Proof. The proof is in lemma 5.1 of [KHX14]. □

2.3 Dimensionality reduction

Let us recall from (17) and the definition $Y = Z^{\frac{1}{2}} D$ that the dependence of the stiffness matrix A on the parameter vector p is restricted to the diagonal Z. It can thus be shown that the solution of the consistent system of the FEM equations $Au = b$ can be alternatively obtained by solving the overdetermined least squares problem

$$\hat{u}_{\text{ls}} = \arg\min_{u \in \mathbb{R}^n} \|Yu - Z^{-\frac{1}{2}} (D^T)^\dagger b\|^2, \tag{20}$$

Assuming that the inverse of A exists, this is immediately obvious by evaluating the estimator

$$\hat{u}_{\text{ls}} = (Y^T Y)^{-1} Y^T Z^{-\frac{1}{2}} (D^T)^\dagger b = A^{-1} D^T Z^{\frac{1}{2}} Z^{-\frac{1}{2}} (D^T)^\dagger b = A^{-1} u = u^*.$$

Inspired by [YB10] we consider projecting u^* onto a low-dimensional subspace $\mathcal{S}_\rho$, spanned by a basis of ρ linearly independent functions and thereafter attempt to simulate an approximate solution within this subspace. Given an $n \times \rho$ matrix Ψ with $\rho \ll n$ orthonormal columns, the projection operator $\Pi \doteq \Psi\Psi^T$ maps vectors $u \in \mathbb{R}^n$ to the subspace

$$\mathcal{S}_\rho \doteq \{\Psi r \,|\, r \in \mathbb{R}^\rho\}, \tag{21}$$

such that for $u = \Pi u + (I - \Pi)u$ there is a unique, optimal low-dimensional solution r^* satisfying

$$\Psi r^* = \Pi u^*. \tag{22}$$

Letting $X = Y\Psi$, then if the basis Ψ is chosen so that the projection error $(I - \Pi)u$ is sufficiently small, the task at hand is to evaluate a low dimensional vector $r \in \mathbb{R}^\rho$ that approximates r^* and respectively $\Psi r \in \mathbb{R}^n$ that approximates Πu^*, in the least squares sense

$$r = \arg\min_{r \in \mathbb{R}^\rho} \left\| Xr - (Y^T)^\dagger b \right\|^2, \tag{23}$$

whose solution is

$$\begin{aligned} r &= (X^T X)^{-1} X^T (Y^T)^\dagger b \\ &= (X^T X)^{-1} X^T Y (Y^T Y)^{-1} b \\ &= (X^T X)^{-1} \Psi^T b \\ &= (\Psi^T A \Psi)^{-1} \Psi^T A u \\ &= (\Psi^T A \Psi)^{-1} \Psi^T A (\Pi u + (I - \Pi) u) \\ &= \Psi^T u + (\Psi^T A \Psi)^{-1} \Psi^T A (I - \Pi) u. \end{aligned} \tag{24}$$

Notice that, despite the reduction in the dimension of the solution, this problem turns out to be computationally more expensive than the original (12) as it requires the pseudo-inverse of the large, parameter dependent Y^T matrix. However, it can be shown that (23) admits a more efficient formulation, stated in the form of the following lemma.

Lemma 2.4. *The solution of the least-squares problem* (23) *can be computed via the alternative formulation*

$$r = \arg\min_{r \in \mathbb{R}^\rho} \left\| Xr - Z^{-\frac{1}{2}} (D^T)^\dagger b \right\|^2. \tag{25}$$

Proof. Developing the squared norm and introducing the expression of $(D^T)^\dagger$

we have

$$\begin{aligned}
(X^TX)^{-1}X^TZ^{-\frac{1}{2}}(D^T)^\dagger b &= (X^TX)^{-1}\Psi^TD^TZ^{\frac{1}{2}}Z^{-\frac{1}{2}}(D^T)^\dagger b \\
&= (X^TX)^{-1}\Psi^TD^T(D^T)^\dagger b \\
&= (X^TX)^{-1}\Psi^TD^TD(D^TD)^{-1}b \\
&= (X^TX)^{-1}\Psi^Tb \\
&= (\Psi^TA\Psi)^{-1}\Psi^TAu \\
&= (\Psi^TA\Psi)^{-1}\Psi^TA(\Pi u + (I-\Pi)u) \\
&= \Psi^Tu + (\Psi^TA\Psi)^{-1}\Psi^TA(I-\Pi)u = r.
\end{aligned}$$

□

Developing the formulation (25) we arrive at the projected normal equations for the FEM system as[2]

$$\Psi^TA\Psi r = \Psi^Tb. \tag{26}$$

Following the approach of Drineas et al. [DM10] we consider the randomisation of the projected coefficients matrix (the Hessian of the residual in (25))

$$X^TSS^TX\hat{r} = \Psi^Tb, \tag{27}$$

noticing that this can be deduced from (26)

$$X^TSS^TXr + X^T(I-SS^T)Xr = \Psi^Tb,$$

by neglecting the sketching error term $X^T(I-SS^T)X$. In the next sections we discuss how we randomise the computation of X^TSS^TX using a sketching matrix S that depends on the parameter vector p. In the analysis that

[2] We emphasise the contrast between the projected equations in (26) and the projected variable equations $\Psi^TA^TA\Psi r' = \Psi^TA^Tb$ which correspond to the LS problem

$$r' = \arg\min_{r\in\mathbb{R}^\rho}\left\|A\Psi r - b\right\|^2,$$

the solution of which is

$$\begin{aligned}
r' &= (\Psi^TA^2\Psi)^{-1}\Psi^TAb \\
&= (\Psi^TA^2\Psi)^{-1}\Psi^TA^2u \\
&= (\Psi^TA^2\Psi)^{-1}\Psi^TA^2(\Pi u + (I-\Pi)u) \\
&= \Psi^Tu + (\Psi^TA^2\Psi)^{-1}\Psi^TA^2(I-\Pi)u,
\end{aligned}$$

that incurs a subspace regression error term that is quadratic in A. Moreover, note that the right hand side vector in the normal equations $\Psi^TA^TA\Psi r' = \Psi^TA^Tb$ has dependence on the parameter through A.

follows we focus our attention on the various sources of errors affecting the induced sketched approximation of u and in bounding the overall error.

So far we have discussed the projection of the high-dimensional system without providing explicit details on how the basis Ψ is selected. A desired property of the appropriate basis is to sustain a small projection error $\|u - \Pi u\|$ for all admissible p choices under the constraint $\rho \ll n$. Suitable options include parameter-specific bases such as a subset of the right singular vectors of A obtained through a randomised decomposition or Krylov-subspace bases which are orthogonalised via a Gram-Schmidt process [HMT11]. Here we opt for a generic basis exploiting the smoothness of the solution in Lipschitz domains. In particular, we select the basis among the eigenfunctions of the discrete Laplacian operator

$$\Delta = D^T D, \tag{28}$$

and we denote by Q and Λ the eigenvectors and eigenvalues of Δ respectively such that $\Delta Q = Q\Lambda$. We further split Q as

$$Q = \big(Q_{*(1:n-\rho-1)} | \Psi\big),$$

such that Ψ corresponds to the last ρ columns of the orthonormal matrix Q and the ρ smallest eigenvalues $\{\lambda_{n-\rho-1}(\Delta), \ldots, \lambda_n(\Delta)\}$. This arrangement implies that the columns of Ψ are ordered in decreasing variation

$$\|D\Psi_{*(i)}\| > \|D\Psi_{*(j)}\| > 0, \quad \text{for} \quad \rho \geq i > j \geq 1,$$

where D is the discrete first-order gradient operator in the definition $Y = Z^{\frac{1}{2}} D$. Clearly, the decomposition of Δ is computationally expensive but this can be performed offline, once and for all instances of the parameter vector. From (26), the existence of r requires a basis matrix Ψ such that $X^T S S^T X$ is invertible, hence it suffices to show that $SS^T \to I$ as $c \to \infty$ with probability 1.

3 Simulating the reduced system

In this section we focus attention to the randomised simulation of the reduced problem (27). In what follows we assume that all mesh-dependent quantities, including the basis Ψ are readily available through offline computations, and their entries can be traced directly from memory on demand. Our focus is to estimate the low-dimensional system matrix in (27) so that it maintains a minimal Frobenius norm from its deterministic counterpart in (26)

$$G \doteq X^T X = \sum_{\ell=1}^{kd} X^T_{(\ell)*} X_{(\ell)*} \, .$$

We assume a sampling distribution $\xi \doteq \{\xi_\ell\}_{\ell=1}^{kd}$ with $\sum_{\ell=1}^{kd} \xi_\ell = 1$ such that each index ℓ in the set $[kd]$ can be drawn with probability ξ_ℓ. Then collecting $c \ll kd$ independent and identically distributed index samples $\{r_1, r_2, \dots, r_c\}$ according to ξ we can approximate G as

$$\hat{G} \doteq X^T S S^T X = \frac{1}{c} \sum_{t=1}^{c} \frac{1}{\xi_{r_t}} X_{r_t *}^T X_{r_t *} \tag{29}$$

for a sketching matrix $S = BC$, where C is a $c \times c$ diagonal matrix and B is a tall $kd \times c$ sparse matrix with entries

$$C = \frac{1}{\sqrt{c}} \mathrm{diag}\big(\xi_{r_1}^{-\frac{1}{2}}, \dots, \xi_{r_c}^{-\frac{1}{2}}\big), \quad \text{and} \quad B = \big(1_{r_1}, \dots, 1_{r_c}\big), \tag{30}$$

and 1_i is the ith canonical vector. Indeed, $SS^T = BC^2B^T$ returns a $kd \times kd$ diagonal matrix with non-negative entries. It is important to observe that the above construction preserves the semi-definiteness and symmetric structure of $\hat{G}$, though it involves significantly fewer operations compared to computing G. The estimator $\hat{G}$ can be shown to be an unbiased estimator of G through probabilistic arguments.

Proposition 3.1. *The matrix $\hat{G}$ constructed in (29) is an unbiased estimator for G in the sense of $\mathbb{E}_\xi[\hat{G}] = G$. In effect, when $c \to \infty$, $\mathrm{Var}_\xi[\|G - \hat{G}\|_F] \to 0$ with probability 1.*

Corollary 3.2. *Define $S = BC$. Then SS^T is an unbiased estimator for I the identity matrix, i.e. $\mathbb{E}_\xi[SS^T] = I$ under probability ξ.*

An optimal choice for ξ_ℓ in the sense of minimising the Frobenius norm of the simulation error $G - \hat{G}$, can be made according to the parameter vector p as shown next.

Proposition 3.3. *The optimal sampling probability ξ for $\hat{G}$ in (29) in the sense of minimising the error $\mathbb{E}_\xi[\|G - \hat{G}\|_F^2]$ is given by*

$$\xi_\ell = \frac{\|X_{(\ell)*}\|^2}{\|X\|_F^2}, \quad \textit{for all} \;\; 1 \le \ell \le kd, \tag{31}$$

for which the corresponding variance is bounded by

$$\mathrm{Var}_\xi[\|G - \hat{G}\|_F] \le \mathbb{E}_\xi[\|G - \hat{G}\|_F^2] \le \frac{1}{c} \Big(\sum_{\ell=1}^{kd} \|X_{(\ell)*}\| \Big)^2 \le \frac{d^2}{c} p_\Omega^2 \|D\|^2, \tag{32}$$

where $p_\Omega = \sum_{\ell=1}^{k} p_\ell |\Omega_\ell|$ is the discretised integral of $p(x)$ over Ω.

Note that for an arbitrary sampling distribution ξ, the singular values of $\hat{G}$ can be shown to be bounded by the product $d\, p_\Omega$ and further bounded in terms of the sample budget c and the corresponding singular values of G.

Proposition 3.4. *Assume the randomised sampling procedure in* (29) *for approximating G with sampling probabilities as in* (31)*, then the spectrum of $\hat{G}$ is bounded from above as*

$$\sigma_1(\hat{G}) \leq \sum_{\ell=1}^{kd} z_\ell \|(D\Psi)_{(\ell)*}\|^2 \leq d\, p_\Omega \|D\|^2.$$

Complementary to Proposition 3.4, the positive singular values of $\hat{G}$ can be bounded by the corresponding singular values of G and its minimum singular value.

Proposition 3.5. *Assume that $\hat{G}$ is full rank, then for any $\gamma \in (0, 1)$*

$$\sigma_i(\hat{G}) \geq \sigma_i(G) - \gamma\sigma_{\min}(G) \quad \textit{for} \quad 1 \leq i \leq \rho, \tag{33}$$

holds with probability at least

$$1 - \min\left\{1, \frac{\mathbb{E}_\xi[\|G - \hat{G}\|_F]}{\gamma\sigma_{\min}(G)}\right\}.$$

Due to the positive semi-definiteness of G and $\hat{G}$, their singular values coincide with their eigenvalues, while for $c \ll \rho$ then almost surely $\hat{G}$ is full rank. These results, in conjunction with lemmas 2.2 and 2.3 will be used in calculating the simulation error in Section 4.1. The total computational cost for obtaining (29) is at most $\mathcal{O}(c\rho^2) + \mathcal{O}(k)$.

Algorithm 1 Randomised simulation algorithm

```
Input: Sparse matrix $D \in \mathbb{R}^{kd\times n}$, orthonormal basis $\Psi \in \mathbb{R}^{n\times\rho}$, load vector $b \in \mathbb{R}^n$, and element size vector $\omega \in \mathbb{R}^k$.
Compute offline $D\Psi \in \mathbb{R}^{kd\times\rho}$ and vector $\Psi^T b \in \mathbb{R}^\rho$
for $i = 1, 2, \ldots, N$ do
    input parameters vector $p \in \mathbb{R}^k$
    Compute $z = p \odot \omega$ and $Z = z \otimes I_d$
    Form $\xi$ based on (31)
    Sample $\{r_1, \ldots, r_c\}$ iid from $\xi$ to form $S$
    Form $X = Z^{\frac{1}{2}}(D\Psi)$
    Compute $\hat{G} = X^T S S^T X$
    Output: $\hat{r} = \hat{G}^{-1}(\Psi^T b)$ and $\hat{u} = \Psi\hat{r}$.
end
```

3.1 Statistical leverage score sampling

As proved in [DM10] for the graph Laplacian paradigm and later in [AT11] for the FEM stiffness matrix, the optimal sampling probabilities are derived from the leverage scores of the respective matrices. As these scores are typically impractical to compute, it is reasonable to consider approximating them in a computationally efficient way. In this context, we compute the discrepancy between the leverage score probability and the one in (31), in order to investigate how this is affected by the subspace projection. In fact we argue that this discrepancy shrinks in norm when simulating the product $G = (Y\Psi)^T(Y\Psi)$ instead of $A = Y^TY$ when $\rho < n$. To show this, consider that for a matrix B with kd rows we can define the statistical leverage score and row norm sampling probabilities as

$$\xi^{l(B)} = \frac{l(B)}{\sum_{\ell=1}^{kd} l_\ell(B)} \quad \text{and} \quad \xi^{r(B)} = \frac{r(B)}{\sum_{\ell=1}^{kd} r_\ell(B)}, \tag{34}$$

respectively, where the ℓth leverage score and row-squared-norm for $B = U_B\Sigma_BV_B^T$ are

$$l_\ell(B) = (U_BU_B^T)_{\ell\ell} \quad \text{and} \quad r_\ell(B) := \|B_{\ell*}\|^2 = (BB^T)_{\ell\ell} \tag{35}$$

with $\ell = 1, \dots, kd$. Based on this we seek to show that the projection onto the low-dimensional subspace induces

$$\|\xi^{l(X)} - \xi^{r(X)}\| < \|\xi^{l(Y)} - \xi^{r(Y)}\|. \tag{36}$$

and

$$\|l(X) - r(X)\|_{\max} < \|l(Y) - r(Y)\|_{\max}. \tag{37}$$

For clarity we address first the simple case where Z is uniform, i.e. when p and ω are constant vectors.

3.2 Simple case: homogeneous model

For the $kd \times n$ matrix D in $Y = Z^{\frac{1}{2}}D$ with $kd > n$ we have $D = U_D\Sigma_DV_D^T$ where $U_D \in \mathbb{R}^{kd\times n}$ and $\Sigma_D \in \mathbb{R}^{n\times n}$ with nonzero diagonal values denoted by $\lambda_1(\Sigma_D) \le \lambda_2(\Sigma_D) \le \cdots \le \lambda_n(\Sigma_D)$.

Lemma 3.6. *In the homogeneous model, $Z = zI$ with $z > 0$, and we have that*

$$\|\xi^{r(X)} - \xi^{l(X)}\|_{\max} \le \Big(\frac{\lambda_{n-\rho+1}(\Sigma_D)^2}{\sum_{i=n-\rho+1}^{n} \lambda_i(\Sigma_D)^2} - \frac{1}{\rho}\Big) \vee \Big(\frac{1}{\rho} - \frac{\lambda_n(\Sigma_D)^2}{\sum_{i=n-\rho+1}^{n} \lambda_i(\Sigma_D)^2}\Big), \tag{38}$$

and

$$\|\xi^{r(X)} - \xi^{l(X)}\| \le \sqrt{\sum_{i=n-\rho+1}^{n} \Big(\frac{\lambda_i(\Sigma_D)^2}{\sum_{i=n-\rho+1}^{n} \lambda_i(\Sigma_D)^2} - \frac{1}{n}\Big)^2}. \tag{39}$$

Proof. Based on the above definitions, from the discrete Laplacian $\Delta = D^T D = V_D \Sigma_D^2 V_D^T$ we form the $n \times \rho$ basis Ψ by partitioning as

$$\Sigma_D = \begin{pmatrix} \bar{\Sigma}_D & 0 \\ 0 & \bar{\Sigma}_\rho \end{pmatrix}, \quad V_D = (\bar{V}_D | \Psi), \tag{40}$$

where $\bar{\Sigma}_\rho$ is $\rho \times \rho$, and clearly $\mathrm{Trace}(\bar{\Sigma}_D) > \mathrm{Trace}(\bar{\Sigma}_\rho)$. We can now write the decomposition $Y = \sqrt{z} U_D \Sigma_D V_D^T$ and

$$X = \sqrt{z} U_D \Sigma_D \begin{pmatrix} 0 \\ I_\rho \end{pmatrix} = \sqrt{z} U_D \begin{pmatrix} 0 \\ \bar{\Sigma}_\rho \end{pmatrix} = \sqrt{z} (U_D)_{*(n-\rho+1:n)} \bar{\Sigma}_\rho,$$

where $(U_D)_{*(n-\rho+1:n)}$ is the submatrix of U_D with its $n-\rho+1$ to n columns. We can then formulate the leverage scores for X as

$$l_i(X) = \mathrm{diag}\big((U_D)_{*(n-\rho+1:n)} (U_D)_{*(n-\rho+1:n)}^T\big)_i, \tag{41}$$

and the probabilities associated with leverage scores of X are therefore

$$\xi^{l(X)} = \frac{1}{\rho} \mathrm{diag}\big((U_D)_{*(n-\rho+1:n)} (U_D)_{*(n-\rho+1:n)}^T\big).$$

Similarly,

$$r_i(X) = z \mathrm{diag}\Big((U_D)_{*(n-\rho+1:n)} \bar{\Sigma}_\rho^2 (U_D)_{*(n-\rho+1:n)}^T\Big)_i, \tag{42}$$

with associated probabilities

$$\xi^{r(X)} = \frac{1}{\mathrm{Trace}(\bar{\Sigma}_\rho^2)} \mathrm{diag}\Big((U_D)_{*(n-\rho+1:n)} \bar{\Sigma}_\rho^2 (U_D)_{*(n-\rho+1:n)}^T\Big).$$

It is now apparent that

$$\begin{aligned} \|\xi^{r(X)} - \xi^{l(X)}\|_{\max} &= \Big\| \mathrm{diag}\Big((U_D)_{*(n-\rho+1:n)} \Big(\frac{1}{\mathrm{Trace}(\bar{\Sigma}_\rho^2)} \bar{\Sigma}_\rho^2 - \frac{1}{\rho} I_\rho\Big) (U_D)_{*(n-\rho+1:n)}^T\Big) \Big\|_{\max} \\ &\le \Big\| (U_D)_{*(n-\rho+1:n)} \Big(\frac{1}{\mathrm{Trace}(\bar{\Sigma}_\rho^2)} \bar{\Sigma}_\rho^2 - \frac{1}{\rho} I_\rho\Big) (U_D)_{*(n-\rho+1:n)}^T \Big\| \\ &= \Big\| \frac{1}{\mathrm{Trace}(\bar{\Sigma}_\rho^2)} \bar{\Sigma}_\rho^2 - \frac{1}{\rho} I_\rho \Big\| \\ &= \Big(\frac{\lambda_{n-\rho+1}(\Sigma_D)^2}{\sum_{i=n-\rho+1}^n \lambda_i(\Sigma_D)^2} - \frac{1}{\rho}\Big) \vee \Big(\frac{1}{\rho} - \frac{\lambda_n(\Sigma_D)^2}{\sum_{i=n-\rho+1}^n \lambda_i(\Sigma_D)^2}\Big). \end{aligned} \tag{43}$$

Besides, taking the 2-norm gives

$$\begin{aligned} &\|\xi^{r(X)} - \xi^{l(X)}\| = \|\xi^{r(X)} - \xi^{l(X)}\|_F \\ &\le \Big\| (U_D)_{*(n-\rho+1:n)} \Big(\frac{1}{\mathrm{Trace}(\bar{\Sigma}_\rho^2)} \bar{\Sigma}_\rho^2 - \frac{1}{\rho} I_\rho\Big) (U_D)_{*(n-\rho+1:n)}^T \Big\|_F \\ &\le \Big\| \frac{1}{\mathrm{Trace}(\bar{\Sigma}_\rho^2)} \bar{\Sigma}_\rho^2 - \frac{1}{\rho} I_\rho \Big\|_F = \sqrt{\sum_{i=n-\rho+1}^n \Big(\frac{\lambda_i(\Sigma_D)^2}{\sum_{j=n-\rho+1}^n \lambda_j(\Sigma_D)^2} - \frac{1}{n}\Big)^2}. \end{aligned}$$

□

Remark 3.7. 1. If we define $\zeta_j := \frac{\lambda_j(\Sigma_D)^2}{\sum_{i=n-\rho+1}^n \lambda_i(\Sigma_D)^2}$, then the upper bound of $\|\xi^{r(X)} - \xi^{l(X)}\|$ as shown in Lemma 3.6 characterises the discrepancy between ζ and the uniform probability. Meanwhile, the upper bound of $\|\xi^{r(X)} - \xi^{l(X)}\|_{\max}$ measures the largest deviation of ζ from the uniform probability.

2. Under the condition

$$\frac{\lambda_{n-\rho+1}(\Sigma_D)^2}{\sum_{i=n-\rho+1}^n \lambda_i(\Sigma_D)^2} > \frac{2}{\rho}, \tag{44}$$

the upper bound for $\|\xi^{r(X)} - \xi^{l(X)}\|_{\max}$ in (38) is $\frac{\lambda_{n-\rho+1}(\Sigma_D)^2}{\sum_{i=n-\rho+1}^n \lambda_i(\Sigma_D)^2} - \frac{1}{\rho}$. This suggests that there may exist a ρ such that (36) holds. To see this, if we define

$$F(\rho) = \frac{\lambda_{n-\rho+1}(\Sigma_D)^2}{\sum_{i=n-\rho+1}^n \lambda_i(\Sigma_D)^2} - \frac{1}{\rho},$$

and find ρ such that $F(\rho) \ll F(n)$, we may then expect that

$$\|\xi^{r(X)} - \xi^{l(X)}\|_{\max} < \|\xi^{r(Y)} - \xi^{l(Y)}\|_{\max},$$

where $F(n)$ is obviously the upper bound for $\|\xi^{r(Y)} - \xi^{l(Y)}\|_{\max}$. Besides, we can also find a range of ρ such that $F(\rho)$ is non-decreasing with respect to ρ and ρ satisfies (44). Within this range, as the upper bound is increasing with ρ, we may expect the measure $\|\xi^{r(X)} - \xi^{l(X)}\|_{\max}$ to increase with ρ as well. Conversely, as ρ reduces therein, $\xi^{r(X)}$ may converge to $\xi^{l(X)}$.

Theorem 3.8. *In the homogeneous model, i.e., $Z = zI$ with $z > 0$ (37) holds if ρ satisfies either*

$$\frac{kd+n}{\sum_{i=1}^n \lambda_i(\Sigma_D)^2} \leq z \leq \frac{1}{\lambda_{n-\rho+1}(\Sigma_D)^2}, \tag{45}$$

or

$$\left(\frac{n}{\sum_{i=1}^n \lambda_i(\Sigma_D)^2}\right) \vee \left(\frac{2}{\lambda_{n-\rho+1}(\Sigma_D)^2}\right) \leq z \leq \frac{kd-n}{kd\lambda_{n-\rho+1}(\Sigma_D)^2 - \sum_{i=1}^n \lambda_i(\Sigma_D)^2}. \tag{46}$$

Proof. From (41) and (42) we have that

$$\begin{aligned}
&\|l(X) - r(X)\|_{\max} = \left\|\operatorname{diag}\Big((U_D)_{*(n-\rho+1:n)}(z\bar{\Sigma}_\rho^2 - I_\rho)(U_D)_{*(n-\rho+1:n)}^T\Big)\right\|_{\max} \\
&\leq \left\|(U_D)_{*(n-\rho+1:n)}(z\bar{\Sigma}_\rho^2 - I_\rho)(U_D)_{*(n-\rho+1:n)}^T\right\| \\
&\leq |1 - z\lambda_{n-\rho+1}(\Sigma_D)^2| \vee |1 - z\lambda_n(\Sigma_D)^2|,
\end{aligned} \tag{47}$$

while on the other hand, from (40) we can deduce that

$$
\begin{aligned}
\|\ell(Y) - r(Y)\|_{\max} &= \left\| \mathrm{diag}\Big(U_D(z\Sigma_D^2 - I_n)U_D^T\Big)\right\|_{\max} \\
&= \left\|\mathrm{diag}\left(U_D \begin{pmatrix} I_{(n-\rho)} - z\bar{\Sigma}_D^2 & 0 \\ 0 & I_\rho - z\bar{\Sigma}_\rho^2 \end{pmatrix} U_D^T\right)\right\|_{\max} \\
&\geq \frac{1}{kd}\left|\mathrm{Trace}\left(U_D \begin{pmatrix} I_{(n-\rho)} - z\bar{\Sigma}_D^2 & 0 \\ 0 & I_\rho - z\bar{\Sigma}_\rho^2 \end{pmatrix} U_D^T\right)\right| \\
&= \frac{1}{kd}\left|n - z\sum_{i=1}^{n} \lambda_i(\Sigma_D)^2\right| = \frac{1}{kd}\left(z\sum_{i=1}^{n}\lambda_i(\Sigma_D)^2 - n\right),
\end{aligned}
\tag{48}
$$

where the last equality can be deduced from conditions (45) or (46). Under condition (45), the upper bound for $\|l(X) - r(X)\|_{\max}$ is 1, while the lower bound for $\|l(Y) - r(Y)\|_{\max}$ is at least 1, hence (37) holds under condition (45). Alternatively, under condition (46), the upper bound for $\|l(X) - r(X)\|_{\max}$ becomes $z\lambda_{n-\rho+1}(\Sigma_D)^2 - 1$, while the lower bound for $\|l(Y) - r(Y)\|_{\max}$ is no smaller than $z\lambda_{n-\rho+1}(\Sigma_D)^2 - 1$, thus (37) holds under the condition (46). □

Reasoning similarly to (48) for $\|l(X) - r(X)\|_{\max}$ shows that both the upper and lower bounds are increasing with increasing ρ, as the following corollary concludes.

Corollary 3.9. *If there exists an integer $q \in [n]$ such that*

$$
z\lambda_{n-q+1}(\Sigma_D)^2 > 2 \quad \textit{and} \quad z \sum_{i=n-q+1}^{n} \lambda_i(\Sigma_D)^2 > q,
$$

then for any integer $\rho \in [n]$ such that $q \leq \rho$, we have that

$$
\frac{1}{kd}\left(z \sum_{i=n-\rho+1}^{n} \lambda_i(\Sigma_D)^2 - \rho\right) \leq \|l(X) - r(X)\|_{\max} \leq \lambda_{n-\rho+1}(\Sigma_D)^2 - 1.
\tag{49}
$$

Remark 3.10. As both bounds in (49) increase with increasing ρ this implies that $\|l(X) - r(X)\|_{\max}$ is non-decreasing within the range $[q, n]$, which in turn implies (37).

4 General case: inhomogeneous model

In the general case, the parameter vector p and the element volumes ω will be variable, and thus our aim is to show that the effect of the projection as demonstrated above for the homogeneous case is sustained.

Lemma 4.1. *In the inhomogeneous model, we have*

$$\|\xi^{r(X)} - \xi^{l(X)}\|_{\max} \le \big(\max_i \pi_i(\rho) - \frac{1}{\rho}\big) \vee \big(\frac{1}{\rho} - \min_i \pi_i(\rho)\big), \tag{50}$$

and

$$\|\xi^{r(X)} - \xi^{l(X)}\| \le \sqrt{\sum_{i=1}^{\rho} \pi_i(\rho)^2 - \frac{1}{\rho}}, \tag{51}$$

where $\pi_i(\rho) := \frac{\lambda_i(\Sigma_X^2)}{\|X\|_F^2}$ *for* $i \in [\rho]$.

Remark 4.2.

1. As in the homogeneous case (see Lemma 3.6), the upper bound of $\|\xi^{l(X)} - \xi^{r(X)}\|$ in (51) also characterises the discrepancy between the probability $\pi(\rho)$ and the uniform probability. Meanwhile, the upper bound of $\|\xi^{l(X)} - \xi^{r(X)}\|_{\max}$ in (50) is measured by the largest deviation of probability $\pi(\rho)$ from the uniform probability.

2. For the term $\max_i \pi_i(\rho) - \frac{1}{\rho}$ in the upper bound of $\|\xi^{l(X)} - \xi^{r(X)}\|_{\max}$, we may set as

$$\bar{F}(\rho) \doteq \max_i \pi_i(\rho) - \frac{1}{\rho} = \frac{\lambda_1(\Sigma_X^2)}{\sum_{i=1}^{\rho} \lambda_i(\Sigma_X^2)} - \frac{1}{\rho}. \tag{52}$$

   Assume that we find a range of ρ such that $\bar{F}(\rho) \ge \frac{1}{\rho}$. This range will reveal the valid ρ where the upper bound of $\|\xi^{l(X)} - \xi^{r(X)}\|_{\max}$ in (50) is indeed $\bar{F}(\rho)$. In general, large integers for ρ can be included. Then within such a range find a sub-range of ρ such that $\bar{F}(\rho)$ is non-decreasing. Within this sub-range, we thus expect $\|\xi^{l(X)} - \xi^{r(X)}\|_{\max}$ to be increasing with ρ. This argument is consistent with the numerical results presented in figure 2.

Corollary 4.3. *For an arbitrary* Y *of size* $kd \times n$ *and rank* n*, we have*

$$\|\xi^{r(Y)} - \xi^{l(Y)}\|_{\max} \le \Big(\frac{\|Y\|^2}{\|Y\|_F^2} - \frac{1}{\rho}\Big) \vee \Big(\frac{1}{\rho} - \frac{\lambda_n(\Sigma_Y)^2}{\|Y\|_F^2}\Big), \tag{53}$$

and

$$\|\xi^{r(Y)} - \xi^{l(Y)}\| \le \sqrt{\sum_{i=1}^{\rho} \Big(\frac{\lambda_i(\Sigma_Y)^2}{\sum_{j=1}^{n} \lambda_j(\Sigma_Y)^2} - \frac{1}{\rho}\Big)^2}. \tag{54}$$

This corollary is a consequence of Lemma 4.1 with $\Psi = I$. The result shows that the difference of choosing between the two sampling probabilities is mainly determined by the dispersion in the singular values of Y.

Theorem 4.4. *In the inhomogeneous model,* (37) *holds if ρ satisfies*

$$2 \leq \lambda_{n-\rho+1}(\Sigma_Y)^2 \tag{55}$$

and if the whole system satisfies

$$kd - n > kd\lambda_1(\Sigma_Y)^2 - \sum_{i=1}^{n} \lambda_i(\Sigma_Y)^2. \tag{56}$$

Remark 4.5. Let us consider the term $\|l(X) - r(X)\|_{\max}$ alone. Pick $q \in [n]$ as the integer that satisfies

$$2 \leq \lambda_{n-q+1}(\Sigma_Y)^2 \quad \text{and} \quad \sum_{i=n-q+1}^{n} \lambda_i(\Sigma_Y)^2 \geq 1.$$

Then for any $q \leq \rho \leq n$, from inequality (65) we have that the upper bound for $\|l(X) - r(X)\|_{\max}$ in (63) is

$$\|l(X) - r(X)\|_{\max} \leq \lambda_1(\Sigma_X)^2 - 1 \leq \lambda_1(\Sigma_Y)^2 - 1.$$

Besides, similar to Corollary 3.9, together with inequality (65) we have a lower bound as

$$\frac{1}{kd}\Big(\sum_{i=n-\rho+1}^{n} \lambda_i(\Sigma_Y)^2 - 1\Big) \leq \frac{1}{m}\Big(\sum_{i=1}^{\rho} \lambda_i(\Sigma_X)^2 - 1\Big) \leq \|\ell(X) - r(X)\|_{\max}.$$

Note however that unlike the homogeneous case (see Remark 3.10), the upper bound $\lambda_1(\Sigma_X)^2 - 1$ or $\lambda_1(\Sigma_Y)^2 - 1$ does not have an obvious increasing trend, as the lower bound is increasing in ρ. This may also imply that $\|l(X) - r(X)\|_{\max}$ has a roughly nondecreasing trend within the range $\rho \in [q, n]$.

4.1 Error Analysis

Our approach for simulating a projected solution the FEM system contends with three main challenges. With reference to the schematic in figure 4.1, there is an *approximation error* component associated with restricting the solution to lie in the subspace $\mathcal{S}_\rho$, where this error can be further decomposed into two parts, one termed as the *projection error*, given by $\|u - \Pi u\|$, measuring the distance between the exact solution u and its projection onto the subspace Πu; and a second component, the *subspace approximation error*, given by $\|\Pi u - \Psi r\|$, measuring the distance between the projection of the true solution u to the best approximation of the regression point r from (23) within $\mathcal{S}_\rho$. It is easy to show that the distance between Πu and Ψr can be bounded in terms of approximation error.

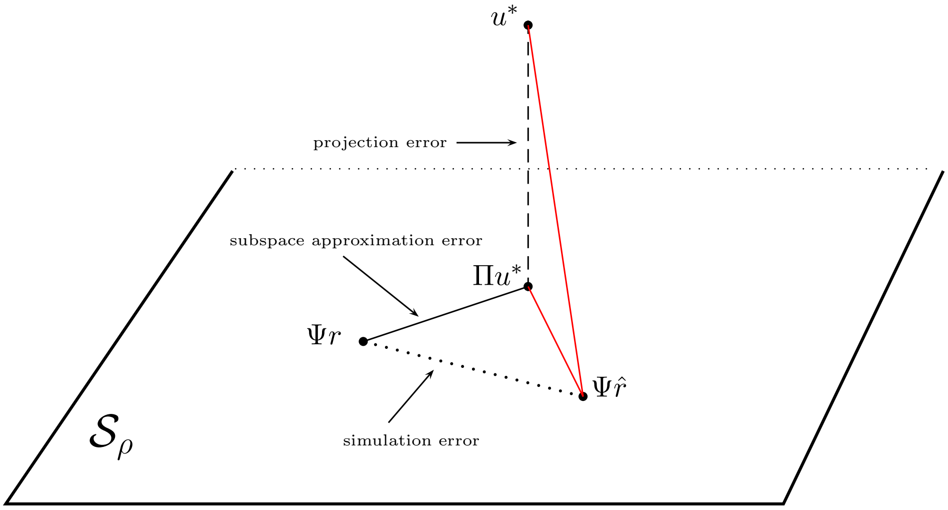

Figure 1: A geometric interpretation of the error components imparted in the sketched solution $\Psi\hat{r}$. Starting from the high-dimensional, 'exact' FEM solution $u^* = A^{-1}b$ we project orthogonally onto the subspace $\mathcal{S}_\rho$ arriving at Πu^* while incurring some projection error. The projected problem then leads to a low-dimensional solution $\Psi r = \Psi G^{-1}\Psi^T b$ that in turn incurs a subspace approximation error due to the condition of the projected matrix G, and ultimately Ψr is approximated via its sketched version $\Psi\hat{r} = \Psi\hat{G}^{-1}\Psi^T b$ that includes simulation error due to the variance in the estimated $\hat{G}$.

Proposition 4.6. *Considering the regression problem* (23) *and recalling that $G = \Psi^T A \Psi$, then we have*

$$\|\Pi u - \Psi r\| \leq \frac{\lambda_{\max}(A)}{\lambda_{\min}(G)}\|u - \Pi u\|.$$

Proof. From the expression for r in formula (24), we directly have that

$$\|\Psi r - \Pi u\| = \|\Psi(\Psi^T A \Psi)^{-1}\Psi^T A(I - \Pi)u\| \leq \frac{\lambda_{\max}(A)}{\lambda_{\min}(G)}\|u - \Pi u\|.$$

□

The *simulation error* associated with replacing Ψr by the sketching-based approximation $\Psi\hat{r}$ from (27), is given by $\|\Psi r - \Psi \hat{r}\|$. The simulation error can be bounded in terms of $\|\Psi r\|$.

Proposition 4.7. *Assume problem settings as discussed in sections 2.3 and 3, and consider the sketched system in* (27)*. Then any $\epsilon, \delta \in (0, 1)$ and c chosen as*

$$(1 + \frac{1}{\epsilon})^2 \frac{\big((\sum_{\ell=1}^{kd} z_\ell \|(D\Psi)_{(\ell)*}\|)^2 - \|G\|_F^2\big)}{\lambda_{\min}(G)^2\, \delta} \leq c \tag{57}$$

satisfy

$$\|\Psi r - \Psi\hat{r}\| \leq \epsilon \|\Psi r\| \tag{58}$$

with probability $1 - \delta$.

An application of Proposition 4.6 and Proposition 4.7 leads to the following main result.

Theorem 4.8. *Assume the settings as in Proposition 4.7 with ϵ, δ and c. Then*

$$\|\Psi\hat{r} - u\| \leq \frac{\lambda_{\max}(A)}{\lambda_{\min}(G)}\|u - \Pi u\| + \epsilon\|\Psi r\|, \tag{59}$$

with probability $1 - \delta$.

Remark 4.9. Due to the Cauchy interlacing theorem [Woo14], we have

$$\lambda_{\min}(A) \leq \lambda_{\min}(\Psi^T A \Psi) \leq \lambda_\rho(A).$$

Thus the best to be expected from (59) is

$$\|\Psi\hat{r} - u\| \leq \kappa_\rho(A)\|u - \Pi u\| + \epsilon\|\Psi r\|,$$

where $\kappa_\rho(A) := \frac{\|A\|_2}{\lambda_\rho(A)}$ with c chosen to as in Proposition 4.7. On the other hand the worst case is

$$\|\Psi\hat{r} - u\| \leq \kappa(A)\|u - \Pi u\| + \epsilon\|\Psi r\|.$$

test	ρ	c	time	ratio	$\frac{\Vert\Pi u-u\Vert}{\Vert u\Vert}$	$\frac{\Vert\hat{G}-G\Vert_F}{\Vert G\Vert_F}$	$\kappa(G)$	$\frac{\Vert\hat{r}-r\Vert}{\Vert r\Vert}$	$\frac{\Vert\Psi\hat{r}-u\Vert}{\Vert u\Vert}$
A	100	5000	658	0.0087	0.0420	0.1312	16.3	0.0796	0.0914
B	50	5000	609	0.0087	0.0675	0.1309	10.9	0.0783	0.0913
C	50	10000	396	0.0087	0.0675	0.0924	10.2	0.0624	0.0992
D	50	10000	448	0.0087	0.0662	0.0923	10.9	0.0613	0.0942
E	50	50000	495	0.0806	0.0675	0.0292	10.5	0.0193	0.0861
F	50	100000	574	0.1496	0.0675	0.0207	10.8	0.0137	0.0854

Table 1: The table above summarises the findings of our simulation tests on the Dirichlet problem. ρ is the number of basis functions spanning the projection subspace, c the number of samples used in the sketch, time in seconds is the time taken for 1000 sketched problem evaluations, and ratio is the percentage of the rows of X utilised in the sketch. The remaining quantities are relative values for the subspace approximation, sketching and overall solution errors and the condition number of G averaged over 1000 FEM solutions. In all tests the parameter vectors were drawn from the uniform distribution $\mathcal{U}[10^{-1}, 10^{2}]$ apart from test D where p was sampled from $\exp(-\mathcal{U}[10^{-4}, 1])$. Characteristic to these tests are the relative low overall error levels, due to the suppressed projection and subspace approximation errors in conjunction to the small condition number of the projected matrix.

5 Numerical experiments

To verify the performance of our algorithm and to test the derived error bounds we report on a number of numerical experiments based on the Dirichlet and Neumann problems (1)-(2). For these we consider Ω to be a spherical domain of unit radius centred at the origin and discretised into $k = 190955$ unstructured linear tetrahedral elements. The model comprises $n + n_{\partial} = 34049$ nodes of which $n_{\partial} = 4217$ are on the boundary. In the tests discussed below we run a sequence of 1000 FEM problems where p is chosen from a random distribution, and present our findings on average over the samples. For the subspace projection, a basis Ψ consisting of singular vectors of $\Delta = D^T D$ was used throughout. Ahead of the tests we compute and store the mesh-dependent sparse gradients matrix D modified to conform to the imposed boundary conditions, and the tall matrix $D\Psi$. Effectively, given p one readily forms the diagonal Z and thereafter the solution is computed direclty as `u=A\b` once the stiffness matrix `A=D'ZD` is assembled. Our code was implemented in Matlab R2018b and executed on a workstation equipped with two 14-core Intel Xeon dual processors, running Linux NixOS with 384GB RAM.

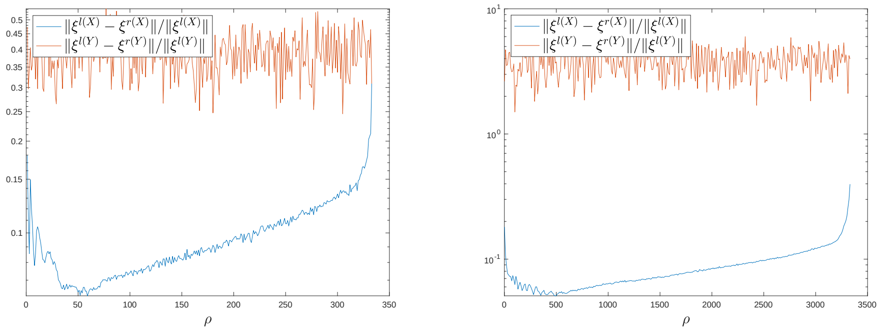

Figure 2: Numerical investigation of the relative discrepancy between the utilised and optimal sampling distributions for the projected $G = X^T X$ (blue) and the original $A = Y^T Y$ (red) matrix products for varying ρ on two coarser meshes of the domain with $n = 334$ (left) and $n = 3335$ (right) degrees of freedom respectively. The optimal sampling distributions $\xi^{l(X)}$ and $\xi^{l(Y)}$ are taken to be those based on the statistical leverage scores as in [DM10] of X ans Y respectively, while the $\xi^{r(X)}$ and $\xi^{r(Y)}$ are those implemented in our algorithm and are based on the Euclidean norm of the matrix rows. Note that the plots are averaged over 5 (X, Y) pairs involving randomly drawn vectors p from the uniform distribution $\mathcal{U}[10^{-2}, 1]$. The graphs show explicitly that sketching the projected product $X^T X$ for $\rho \ll n$ with $\xi^{r(X)}$ is near optimal, as well as illustrating the range of ρ values where the discrepancy between the two distributions exhibits a monotonic behaviour.

5.1 The Dirichlet problem

We first address the Dirichlet problem with a uniform boundary condition $u = 0$ on $\partial\Omega$ yielding a FEM system with $n = 29832$ degrees of freedom, one for each interior node in the mesh. The forcing term is taken to be a piecewise constant approximation of the function

$$f(x, y, z) = \begin{cases} 5 & \text{if } \sqrt{(x + \frac{1}{2})^2 + y^2 + z^2} \leq 0.3, \\ 0 & \text{otherwise,} \end{cases}$$

on the elements. Two matrices Ψ were computed using the last $\rho = 50$ and $\rho = 100$ singular vectors of Δ for the needs of the tests referred to as A, B, C, D, E and F in table 5.1. In each of these tests we solve for both the exact u^* and sketched solutions $\Psi\hat{r}$ for each of the 1000 parameter vectors and record their corresponding timings, the later of which includes forming the sampling distribution, taking c iid samples, sketching the matrix G and solving the projected problem for $\hat{r}$. The particular settings for these tests and the average values of the errors obtained are tabulated in table 5.1. From this it appears that $\rho = 50$ yields a projection error of about 6% despite that p varies over four orders of magnitude, and that the overall relative error is bounded below 10% in all tests. As anticipated, with the sampling budget increasing from $c = 5000$ to $c = 100000$ the simulation error $\|G - \hat{G}\|_F / \|G\|_F$ drops from 13% to about 2%, even though no more than 8% of the rows of X are sampled in the sketching process, which indicates that the sampling is highly inhomogeneous. Finally, the times for 1000 sketched solutions were found to be in the range 500 - 600 s, yielding an average of about 0.55 s per FEM problem, which compares favourably to the recorded 3.1 s taken on average for each high-dimensional solution. Critical to this desirable performance is the small condition number $\kappa(G) \approx 10$ which implies that $\lambda_{\min}(G)$ is bounded away from zero, in agreement to the bound in Theorem 4.8. A further insight into the dependence of the solution errors on the parameter vector can be obtained by the histograms in figure 3 illustrating the variation of the projection, sketching, subspace approximation and total errors, along with the condition number of G across the range of the simulated problems in test C, where p was sampled from the uniform distribution $\mathcal{U}[10^{-1}, 10^2]$.

5.2 The Neumann problem

For the Neumann problem we consider $f = 0$ in the interior of the domain and the condition

$$g^{(N)}(x, y, z) = \begin{cases} 1 & \text{if } \sqrt{x^2 + (y - 1)^2 + z^2} \leq 0.4 \\ 0 & \text{otherwise} \end{cases}, \quad (x, y, z) \in \partial\Omega$$

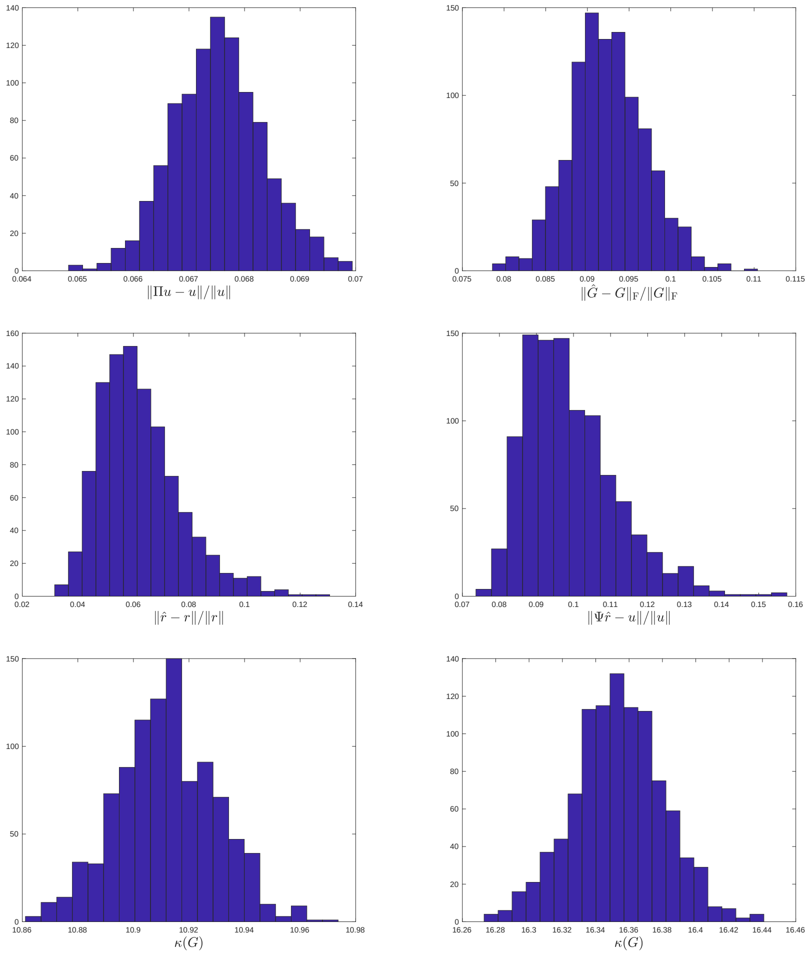

Figure 3: Histograms depicting the relative variation of the projection, subspace approximation, simulation and total solution errors for the 1000 simulations in test C where $c = 10000$, $\rho = 50$ and p was drawn from the uniform distribution $\mathcal{U}[10^{-1}, 10^{2}]$. The figures at the top row show that the projection and sketching errors are symmetrically concentrated around some small values without any outliers, while those in the second row for the subspace approximation and overall errors appear to be somewhat skewed towards zero. This desirable behaviour can be explained via the condition number of the projected matrix G that controls the overall error amplification, as shown at the bottom left figure. For comparison, we plot to its right the respective histogram for $\rho = 100$ indicating that G remains a well-conditioned matrix for these choices of ρ.

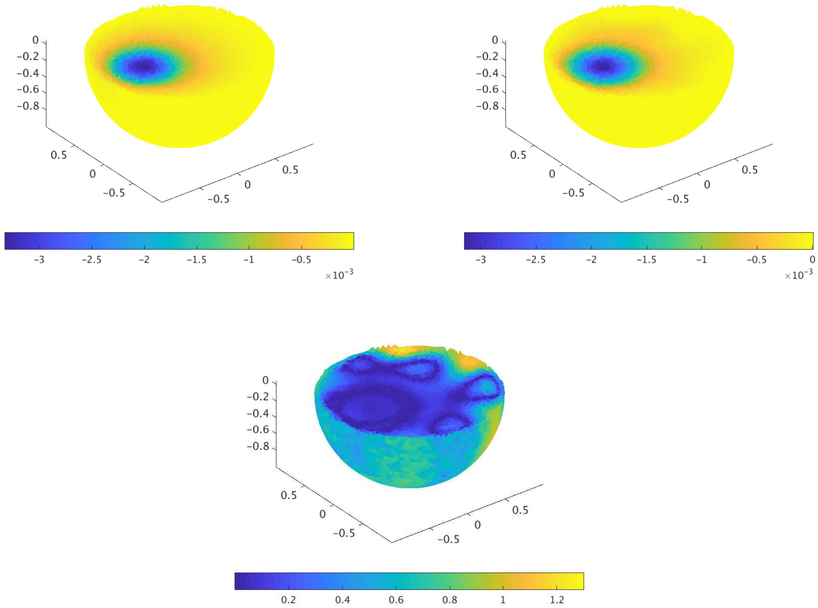

Figure 4: At the top row, an extract of the exact u^* (left) and sketched $\Psi\hat{r}$ (right) solution of the Dirichlet problem at one instance of p in test E. The recorded times were 0.4 s and 3.1 s respectively, and the percentage relative error mapped to this section of the domain is shown below.

test	ρ	c	time	ratio	$\frac{\|\Pi u - u\|}{\|u\|}$	$\frac{\|\hat{G}-G\|_F}{\|G\|_F}$	$\kappa(G)$	$\frac{\|\hat{r}-r\|}{\|r\|}$	$\frac{\|\Psi\hat{r}-u\|}{\|u\|}$
A	100	5000	600	0.0087	0.0040	0.2079	1728	0.4946	0.4418
B	100	50000	776	0.0814	0.0039	0.0649	1743	0.1107	0.1365
C	50	100000	605	0.1539	0.0053	0.0293	1153	0.0873	0.1294
D	50	100000	562	0.1574	0.0053	0.0293	1062	0.0792	0.1204
E	50	500000	1897	0.5496	0.0053	0.0131	1130	0.0375	0.1126
F	50	500000	1133	0.5085	0.0053	0.0131	1055	0.0383	0.1223

Table 2: The table above summarises the findings of our simulation tests on the Neumann problem. ρ is the number of basis functions spanning the projection subspace, c is the number of samples used in the sketching, time is the duration in seconds taken for a 1000 sketched problem evaluations and ratio is the percentage of the rows of X utilised in the sketch. The remaining quantities are relative errors for the projection, simulation, subspace approximation, overall solution error, and the condition of the projected matrix G averaged over 1000 problem solutions. The parameter vectors were drawn from the uniform distribution $\mathcal{U}[10^{-1}, 10^2]$, apart from tests D and E where $\exp(-\mathcal{U}[10^{-4}, 1])$ was invoked.

at the boundary. Similarly to the Dirichlet case we set to investigate the performance of our algorithm in approximating u^* on a series of tests whose results are tabulated in table 5.2. To aid the comparison with the Dirichlet setting the same mesh is used, although in the Neumann problem u^* has $n + n_\partial - 1 = 34048$ degrees of freedom, incorporating all nodes of the mesh apart from one whose value is fixed in order to enforce uniqueness [ESW14]. Overall, the error measures recorded show that despite the small projection error, the total errors observed are substantially larger compared to those at the Dirichlet tests. Increasing the sampling budget to 500000, sampling 54% of the rows of X, suppresses the total error to around 12% with a linear reduction in the sketching error in $\hat{G}$. However the relative regression solution error, and thus the total solution error are significantly larger than those observed in the Dirichlet tests. Noticeably, the condition number of G is also about two orders of magnitude larger by comparison, which explains the error amplification. To see this, from (57) and Theorem 4.8 notice that the total error is bounded by $\|(I - \Pi)u\|$ and $\|\Psi r\|$ where the latter can be further developed following the last line of (24) as

$$\begin{aligned}\|\Psi r\| &\leq \|\Psi\Psi^T u\| + \|\Psi(\Psi^T A \Psi)^{-1}\Psi^T A(I - \Pi)u\| \\ &\leq \|\Pi u\| + \big(\kappa(G) + \frac{\lambda_{\max}(A) - \lambda_{\max}(G)}{\lambda_{\min}(G)}\big)\|(I - \Pi)u\|,\end{aligned}$$

therefore when $\kappa(G)$ is large, $\|\Psi r\|$ and in turn the total error increases. The large values of $\kappa(G)$ are also confirmed in the associated histogram plots for test F in figure 5.2. In terms of the computational times, the sketched approach maintains its advantage against the deterministic solution since for $c = 100000$ we approximate a solution with about 12% error in about 0.6 s while the corresponding u^* takes 4.2 s.

6 Conclusions

We developed a fast, randomised implementation of the finite element method for solving elliptic partial differential equations, suited in particular to high-dimensional models in the many-query context. Using a low-dimensional subspace projection, we reduce the dimensionality of the problem and cast the resulting projected stiffness matrix as a matrix product that we approximate through a randomised sketch. We analyse the errors imparted to the solution and provide bounds for the overall error with respect to the choice of subspace and the spectra of the stiffness matrices involved. Moreover, invoking the projection onto a low-dimensional subspace effectively allows to approximate the sampling distribution based on the statistical leverage scores in a computationally efficient way. Our numerical results are in alignment with the derived approximation bounds, while overall, our algorithm yields substantial computational savings for a moderate solution error.

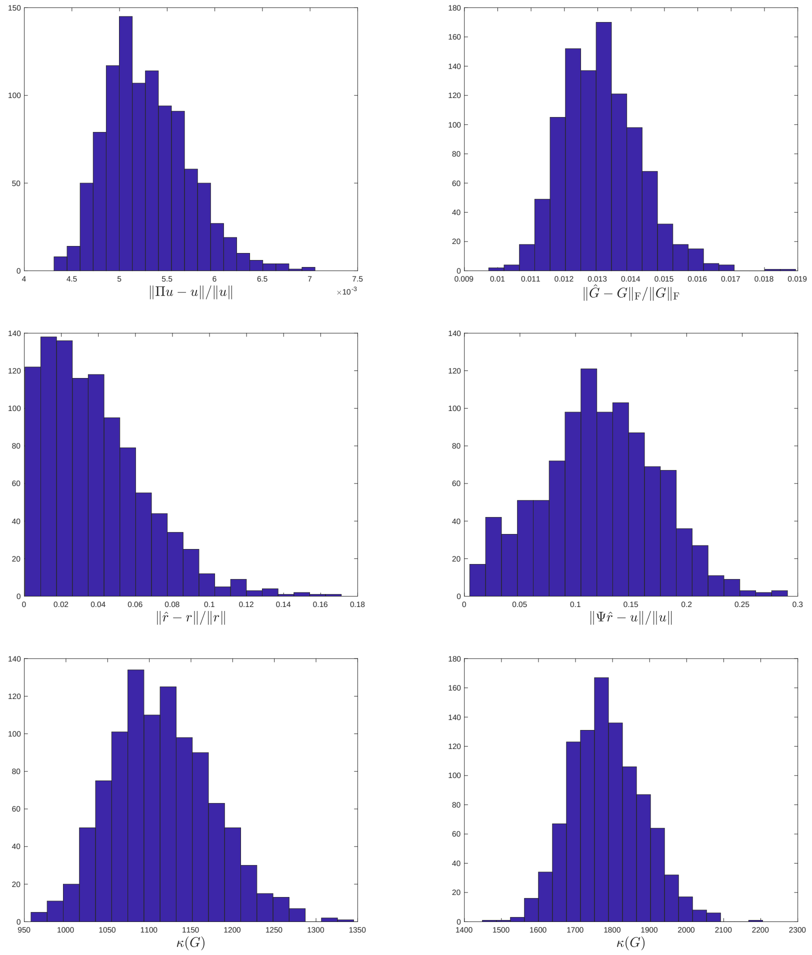

Figure 5: Histograms of the various error components affecting the sketched solution of the Neumann problem in 1000 simulations of test F, where $p \sim \mathcal{U}[10^{-1}, 10^{2}]$, $c = 500000$ and $\rho = 50$. Notice that the total error spans over a larger range of values as a affected by the relatively large condition number of the G matrices, as shown at the bottom left figure. For comparison we plot to its right the respective histogram for $\rho = 100$ showing that $\kappa(G)$ remains significantly higher compared to the Dirichlet case and that the situation worsens as ρ increases.

Acknowledgements

NP and YW are grateful to EPSRC for funding this work through the project EP/R041431/1, titled 'Randomness: a resource for real-time analytics'. NP acknowledges additional support from the Alan Turing Institute.

Appendix

Proof of Proposition 3.1. The claim can be shown through the linearity property of expectation and its definition as follows

$$\begin{aligned}\mathbb{E}_\xi[\hat{G}] &= \mathbb{E}_\xi\Big[\frac{1}{c}\sum_{t=1}^{c}\frac{1}{\xi_{r_t}}z_{r_t}(D\Psi)^T_{(r_t)*}(D\Psi)_{(r_t)*}\Big] \\ &= \sum_{t=1}^{c}\sum_{\ell=1}^{kd}\xi_\ell\frac{z_\ell}{c\xi_\ell}(D\Psi)^T_{(\ell)*}(D\Psi)_{(\ell)*} \\ &= \sum_{\ell=1}^{kd}z_\ell(D\Psi)^T_{(\ell)*}(D\Psi)_{(\ell)*} = G.\end{aligned}$$

□

Proof of Corollary 3.2. Recall that SS^T is a $kd \times kd$ diagonal matrix, whose ℓth index has a value of the product of the cardinality of index ℓ in the sample set $\{r_1, \ldots, r_t\}$ denoted by c_ℓ with $\frac{1}{c\xi_\ell}$. From this it remains to show that each entry has expectation 1. For an index ℓ, the sampling procedure can be treated as a sequence of c binomial trials each with success probability ξ_ℓ. Therefore we have

$$\mathbb{E}_\xi[(SS^T)_{\ell\ell}] = \mathbb{E}\big[\frac{c_\ell}{c\xi_\ell}\big] = \frac{1}{c\xi_\ell}\mathbb{E}[c_\ell] = \frac{1}{c\xi_\ell}c\xi_\ell = 1,$$

where the penultimate equality is by virtue of the properties of binomial random variables. □

Proof of Proposition 3.3. First we note that for all $1 \leq i, j \leq \rho$ we have

$$\hat{G}_{ij} = \sum_{t=1}^{c}(G_t)_{ij},$$

where

$$(G_t)_{ij} = \frac{1}{c\xi_{r_t}}z_{r_t}(D\Psi)_{r_t i}(D\Psi)_{r_t j}$$

A direct consequence of $\mathbb{E}_\xi[\hat{G}] = G$ is that $\mathbb{E}_\xi[\hat{G}_{ij}] = G_{ij}$, and $\mathbb{E}_\xi[(G_t)_{ij}] = \frac{1}{c}G_{ij}$. We then have that

$$\mathrm{Var}_\xi[\hat{G}_{ij}] = \sum_{t=1}^{c}\mathrm{Var}_\xi[(G_t)_{ij}] = \sum_{t=1}^{c}\Big(\mathbb{E}[(G_t)^2_{ij}] - \mathbb{E}[(G_t)_{ij}]^2\Big)$$

from where we get

$$\mathrm{Var}_\xi[\hat{G}_{ij}] = \frac{1}{c}\Big(\sum_{\ell=1}^{kd} \frac{1}{\xi_\ell} z_\ell^2 (D\Psi)_{\ell i}^2 (D\Psi)_{\ell j}^2 - G_{ij}^2\Big) \quad 1 \le i,j \le \rho.$$

To bound the simulation-induced error we have

$$\mathbb{E}_\xi[\|G-\hat{G}\|_F^2] = \sum_{i=1}^{\rho}\sum_{j=1}^{\rho} \mathbb{E}_\xi[(G-\hat{G})_{ij}^2] = \sum_{i=1}^{\rho}\sum_{j=1}^{\rho} \mathrm{Var}_\xi[\hat{G}_{ij}].$$

In fixing the matrix indices we have

$$\begin{aligned}
\sum_{i=1}^{\rho}\sum_{j=1}^{\rho} \mathrm{Var}_\xi[\hat{G}_{ij}] &= \sum_{i=1}^{\rho}\sum_{j=1}^{\rho} \frac{1}{c}\Big(\sum_{\ell=1}^{kd} \frac{z_\ell^2}{\xi_\ell} (D\Psi)_{\ell i}^2 (D\Psi)_{\ell j}^2 - G_{ij}^2\Big)\\
&= \frac{1}{c}\sum_{\ell=1}^{kd} \frac{z_\ell^2}{\xi_\ell} \sum_{i=1}^{\rho}\sum_{j=1}^{\rho} (D\Psi)_{\ell i}^2 (D\Psi)_{\ell j}^2 - \frac{1}{c}\sum_{i=1}^{\rho}\sum_{j=1}^{\rho} G_{ij}^2\\
&= \frac{1}{c}\sum_{\ell=1}^{kd} \frac{z_\ell^2}{\xi_\ell} \sum_{i=1}^{\rho} (D\Psi)_{\ell i}^2 \sum_{j=1}^{\rho} (D\Psi)_{\ell j}^2 - \frac{1}{c}\|G\|_F^2\\
&= \frac{1}{c}\Big(\sum_{\ell=1}^{kd} \frac{z_\ell^2}{\xi_\ell} \|(D\Psi)_{(\ell)*}\|^4 - \|G\|_F^2\Big) \qquad (60)\\
&\le \frac{1}{c}\Big(\sum_{\ell=1}^{kd} \frac{z_\ell^2}{\xi_\ell} \|D_{(\ell)*}\|^4 - \|G\|_F^2\Big) \qquad (61)
\end{aligned}$$

where the first inequality holds true since $\mathrm{Var}[X] \le \mathbb{E}[X^2]$ for any real-valued random variable X, and the last from $\Psi^T\Psi = I$. In order to optimise the choice of ξ in reducing the simulation error we invoke the Lagrangian function based on (60)

$$\mathcal{L}(\xi;\lambda) = \sum_{\ell=1}^{kd} \frac{z_\ell^2}{\xi_\ell} \|(D\Psi)_{(\ell)*}\|^4 + \lambda\Big(\sum_{\ell=1}^{kd} \xi_\ell - 1\Big).$$

for which the method of Lagrange multipliers returns

$$\xi_\ell = \frac{z_\ell \|(D\Psi)_{(\ell)*}\|^2}{\sum_{\ell=1}^{kd} z_\ell \|(D\Psi)_{(\ell)*}\|^2}, \quad 1 \le \ell \le kd.$$

Plugging in the optimal expression of ξ into simulation error expression yields

$$\begin{aligned}
\mathbb{E}_\xi[\|G-\hat{G}\|_F^2] &\le \frac{1}{c}\sum_{\ell=1}^{kd} z_\ell \|(D\Psi)_{(\ell)*}\|^2 \sum_{\ell=1}^{kd} z_\ell \|(D\Psi)_{(\ell)*}\|^2 - \frac{1}{c}\|G\|_F^2\\
&\le \frac{1}{c}\Big(\big(\sum_{\ell=1}^{kd} z_\ell \|(D\Psi)_{(\ell)*}\|\big)^2 - \|G\|_F^2\Big).
\end{aligned}$$

□

Proof of Proposition 3.4. Applying the induced norm to the expression of $\hat{G}$ in (29) yields

$$\begin{aligned}\|\hat{G}\| &\le \frac{1}{c}\sum_{t=1}^{c}\frac{z_{r_t}}{\xi_{r_t}}\big\|(D\Psi)^T_{(r_t)*}(D\Psi)_{(r_t)*}\big\| = \frac{1}{c}\sum_{t=1}^{c}\frac{z_{r_t}}{\xi_{r_t}}\big\|(D\Psi)_{(r_t)*}\big\|^2\\ &= \frac{1}{c}\sum_{t=1}^{c}\sum_{\ell=1}^{kd} z_\ell\big\|(D\Psi)_{(r_\ell)*}\big\|^2 = \sum_{\ell=1}^{kd} z_\ell\big\|(D\Psi)_{(r_\ell)*}\big\|^2 \le d\,p_\Omega\|D\|^2.\end{aligned}$$

where $p_\Omega = \sum_{\ell=1}^{k} p_\ell|\Omega_\ell|$. □

Proof of Proposition 3.5. Applying directly the eigenvalue perturbation result from [Mey13], we immediately have

$$\sigma_i(\hat{G}) \ge \sigma_i(G) + \lambda_{\min}(\hat{G}-G), \quad \text{for} \quad i = 1,\dots,\rho \tag{62}$$

where $\lambda_{\min}$ represents the minimum eigenvalue of a matrix. Note for the symmetric matrix $\hat{G}-G$, $|\lambda_{\min}(\hat{G}-G)| \le \|\hat{G}-G\|$. Markov's inequality leads to

$$\mathbb{P}_\xi\big(\|\hat{G}-G\| \le \gamma\sigma_{\min}(G)\big) \ge 1 - \min\Big\{1, \frac{\mathbb{E}_\xi[\|\hat{G}-G\|_F]}{\gamma\sigma_{\min}(G)}\Big\}.$$

Thus with the above indicated probability, we have

$$|\lambda_{\min}(\hat{G}-G)| \le \gamma\sigma_{\min}(G),$$

which implies that $\lambda_{\min}(\hat{G}-G) \ge -\gamma\sigma_{\min}(G)$. Substituting back into (62) yields the final assertion. □

Proof of Lemma 4.1. With the definitions as before, i.e, $Y = Z^{\frac{1}{2}}D$, $X = Y\Psi$ and $\Psi^T\Psi = I$, let $X = U_X\Sigma_X V_X^T$, where $U_X \in \mathbb{R}^{kd\times\rho}$, $\Sigma_X \in \mathbb{R}^{\rho\times\rho}$, $V_X \in \mathbb{R}^{\rho\times\rho}$, and $\beta = \|X\|_F^{-2}$. Then

$$\|\xi^{l(X)} - \xi^{r(X)}\| = \Big\|\frac{1}{\rho}l_X - \beta r_X\Big\| = \Big\|\text{diag}\Big\{U_X\Big(\frac{1}{\rho}I - \beta\Sigma_X^2\Big)U_X^T\Big\}\Big\|,$$

where $\|\cdot\|$ can now be taken as an arbitrary vector norm to be determined. Taking the 2-norm gives

$$\begin{aligned}\|\xi^{l(X)} - \xi^{r(X)}\| &= \Big\|\text{diag}\Big\{U_X\Big(\frac{1}{\rho}I - \beta\Sigma_X^2\Big)U_X^T\Big\}\Big\|_F\\ &\le \Big\|U_X\Big(\frac{1}{\rho}I - \beta\Sigma_X^2\Big)U_X^T\Big\|_F \le \|U_X\|^2\Big\|\frac{1}{\rho}I - \frac{\Sigma_X^2}{\|X\|_F^2}\Big\|_F\\ &\le \sqrt{\sum_{i=1}^{\rho}\big(\frac{1}{\rho} - \pi_i(\rho)\big)^2} = \sqrt{\sum_{i=1}^{\rho}\pi_i(\rho)^2 - \frac{1}{\rho}}.\end{aligned}$$

On the other hand, taking the max-norm yields

$$\begin{aligned}
&\|\xi^{l(X)} - \xi^{r(X)}\|_{\max} = \Big\|\mathrm{diag}\Big\{U_X\Big(\frac{1}{\rho}I - \beta\Sigma_X^2\Big)U_X^T\Big\}\Big\|_{\max} \\
&\le \Big\|U_X\Big(\frac{1}{\rho}I - \beta\Sigma_X^2\Big)U_X^T\Big\| \le \|U_X\|^2\Big\|\frac{1}{\rho}I - \frac{\Sigma_X^2}{\|X\|_F^2}\Big\| \\
&\le \big(\max_i \pi_i(\rho) - \frac{1}{\rho}\big) \vee \big(\frac{1}{\rho} - \min_i \pi_i(\rho)\big).
\end{aligned}$$

□

Proof of Theorem 4.4. We have that

$$\begin{aligned}
\|l(X) - r(X)\|_{\max} &\le \Big\|\mathrm{diag}\Big\{U_X\Big(I_\rho - \Sigma_X^2\Big)U_X^T\Big\}\Big\| \\
&= \Big\|\mathrm{diag}\Big\{U_X\Big(I_\rho - \Sigma_X^2\Big)U_X^T\Big\}\Big\| \\
&\le |1 - \lambda_1(\Sigma_X)^2| \vee |1 - \lambda_\rho(\Sigma_X)^2|.
\end{aligned} \tag{63}$$

On the other hand,

$$\begin{aligned}
&\|l(Y) - r(Y)\|_{\max} = \Big\|\mathrm{diag}\Big\{U_Y\Big(I_n - \Sigma_Y^2\Big)U_Y^T\Big\}\Big\| \\
&\ge \frac{1}{kd}\Big|\mathrm{Trace}\Big(U_Y\Big(I_n - \Sigma_Y^2\Big)U_Y^T\Big)\Big| = \frac{1}{kd}\Big|n - \sum_{i=1}^n \lambda_i(\Sigma_Y)^2\Big|.
\end{aligned} \tag{64}$$

Besides, lemma 4.4 in [Vol04] suggests that

$$\lambda_{n-\rho+i}(\Sigma_Y^2) \le \lambda_i(\Sigma_X^2) \le \lambda_i(\Sigma_Y^2) \tag{65}$$

for $i \in [\rho]$. Then under the condition (55), the upper bound of (47) is

$$|1 - \lambda_1(\Sigma_X)^2| \vee |1 - \lambda_\rho(\Sigma_X)^2| = \lambda_1(\Sigma_X)^2 - 1 \le \lambda_1(\Sigma_Y)^2 - 1.$$

Condition (56) suggests that

$$\frac{1}{kd}\Big(\sum_{i=1}^n \lambda_i(\Sigma_Y)^2 - n\Big) \ge \lambda_1(\Sigma_Y)^2 - 1,$$

which also implies that the lower bound of (64) is $\frac{1}{kd}\Big(\sum_{i=1}^n \lambda_i(\Sigma_Y)^2 - n\Big)$.
□

Proof of Proposition 4.7. We have the normal equations

$$X^TXr = X^TZ^{-\frac{1}{2}}(D^T)^\dagger b = \Psi^T b,$$

and

$$X^TSS^TX\hat{r} = \Psi^T b.$$

Subtracting the latter equation from the first one gives

$$X^T S S^T X(r - \hat{r}) = -X^T (I - SS^T) X r.$$

Taking 2-norm yields

$$\lambda_{\min}(\hat{G}) \|r - \hat{r}\| \leq \|G - \hat{G}\| \|r\|.$$

Assume that matrix $\hat{G}$ is invertible. For the estimation of $\lambda_{\min}(\hat{G})$, which is exactly $\lambda_\rho(\hat{G})$, defining $\gamma \doteq \frac{\epsilon}{\epsilon+1}$ and following similar arguments as in the proof of Proposition 3.5 gives

$$\lambda_\rho(\hat{G}) \geq \lambda_\rho(G) + \lambda_\rho(\hat{G} - G) \geq (1 - \gamma)\lambda_\rho(G)$$

and

$$\|\hat{G} - G\| \leq \gamma \lambda_\rho(G)$$

with probability at least $1 - \min\left\{1, \frac{\mathbb{E}_\xi[\|G - \hat{G}\|_F]}{\gamma \lambda_\rho(G)}\right\}$. Besides, based on the assumptions of c and δ, we have through Proposition 3.3 that

$$\frac{\mathbb{E}_\xi[\|G - \hat{G}\|_F]}{\gamma \lambda_\rho(G)} \leq \frac{\sqrt{\mathbb{E}_\xi[\|G - \hat{G}\|_F^2]}}{\gamma \lambda_\rho(G)} \leq \frac{\sqrt{(\sum_{\ell=1}^{kd} z_\ell \|D_{(\ell)*}\|_2)^2 - \|G\|_F^2}}{\sqrt{c}\gamma \lambda_\rho(G)} \leq \delta.$$

Thus the probability above can be lower-bounded by $1 - \delta$. In summary, these estimations lead to

$$\|\Psi r - \Psi \hat{r}\| \leq \|r - \hat{r}\| \leq \frac{\gamma}{1 - \gamma} \|r\| \leq \epsilon \|\Psi r\|$$

with probability $1 - \delta$ for any $\epsilon, \delta \in (0, 1)$ with c chosen to be

$$(1 + \frac{1}{\epsilon})^2 \frac{\left((\sum_{\ell=1}^{kd} z_\ell \|(D\Psi)_{(\ell)*}\|)^2 - \|G\|_F^2\right)}{\delta \lambda_{\min}(G)^2} \leq c.$$

□

References

[AT11] Haim Avron and Sivan Toledo. *Effective Stiffness: Generalizing Effective Resistance Sampling to Finite Element Matrices.* ArXiv, oct 2011.

[BJMS15] Alexandros Beskos, Ajay Jasra, Ege A. Muzaffer, and Andrew M. Stuart. Sequential Monte Carlo methods for Bayesian elliptic inverse problems. *Statistics and Computing*, 25(4):727–737, jul 2015.

[BOCW17] Peter Benner, Mario Ohlberger, Albert Cohen, and Karen Willcox, editors. *Model Reduction and Approximation.* Society for Industrial and Applied Mathematics, Philadelphia, PA, jul 2017.

[BY09] Dimitri P Bertsekas and Huizhen Yu. Journal of Computational and Applied Projected equation methods for approximate solution of large linear systems. *Journal of Computational and Applied Mathematics*, 227(1):27–50, 2009.

[CDSS18] Daniela Calvetti, Matthew Dunlop, Erkki Somersalo, and Andrew Stuart. Iterative updating of model error for Bayesian inversion. *Inverse Problems*, 34(2):025008, feb 2018.

[CKM$^+$14] Michael B. Cohen, Rasmus Kyng, Gary L. Miller, Jakub W. Pachocki, Richard Peng, Anup B. Rao, and Shen Chen Xu. *Solving SDD linear systems in nearly m log 1/2 n time.* ACM Press, New York, New York, USA, 2014.

[DB09] P. Drineas and C. Boutsidis. Random projections for the nonnegative least-squares problem. *Linear Algebra and Its Applications*, 431(5-7):760–771, 2009.

[DM10] Petros Drineas and Michael W. Mahoney. *Effective Resistances, Statistical Leverage, and Applications to Linear Equation Solving.* ArXiv, may 2010.

[DMMS11] Petros Drineas, Michael W Mahoney, S. Muthukrishnan, and Tamás Sarlós. Faster least squares approximation. *Numerische Mathematik*, 117(2):219–249, feb 2011.

[ESW14] Howard Elman, David Silvester, and Andy Wathen. *Finite Elements and Fast Iterative Solvers.* Oxford University Press, 2nd edition, 2014.

[GR15] Robert M. Gower and Peter Richtárik. Randomized Iterative Methods for Linear Systems. *SIAM Journal on Matrix Analysis and Applications*, 36(4):1660–1690, jan 2015.

[GR16] Robert M. Gower and Peter Richtárik. *Linearly Convergent Randomized Iterative Methods for Computing the Pseudoinverse.* ArXiv, dec 2016.

[HMT11] Nathan Halko, P. G. Martinsson, and Joel A. Tropp. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Review*, 53(2):217–288, jan 2011.

[KHX14] Lennard Kamenski, Weizhang Huang, and Hongguo Xu. Conditioning of finite element equations with arbitrary anisotropic meshes. *Mathematics of Computation*, 83(289):2187–2211, mar 2014.

[KL07] C. Robert Kirby and Anders Logg. Efficient Compilation of a Class of Variational Forms. *ACM Transactions on Mathematical Software*, 33(3):025008, aug 2007.

[LPS14] Gabriel J. Lord, Catherine E. Powell, and Tony Shardlow. *An introduction to computational stochastic PDEs*. Cambridge University Press, 2014.

[Mey13] C Meyer. Matrix analysis and applied linear algebra. *Choice Reviews Online*, 38(09):38–5065–38–5065, 2013.

[PW14] Mert Pilanci and Martin J. Wainwright. Iterative Hessian sketch: Fast and accurate solution approximation for constrained least-squares. *Journal of Machine Learning Research*, 17:1–38, 2014.

[PWB12] Nick Polydorides, Mengdi Wang, and Dimitri P. Bertsekas. A Quasi Monte Carlo Method for Large-Scale Inverse Problems. In H Woźniakowski, editor, *Springer Proceedings in Mathematics and Statistics*, volume 23, pages 623–637. Monte Carlo and Quasi-Monte Carlo Methods 2010. Springer Proceedings in Mathematics & Statistics, 23, Springer, 2012.

[Saa03] Yousef Saad. *Iterative Methods for Sparse Linear Systems, Second Edition*. 2003.

[ST06] Daniel A. Spielman and Shang-Hua Teng. Nearly-Linear Time Algorithms for Preconditioning and Solving Symmetric, Diagonally Dominant Linear Systems. *SIAM J. Matrix Anal*, 35(3):835–885, 2006.

[Vol04] Stefan Volkwein. Condition number of the stiffness matrix arising in POD Galerkin schemes for dynamical systems. In *PAMM*, volume 4, pages 39–42. Math. Mech. 4, dec 2004.

[Woo14] David P. Woodruff. Computational Advertising: Techniques for Targeting Relevant Ads. *Foundations and Trends® in Theoretical Computer Science*, 10(1-2):1–157, 2014.

[YB10] Huizhen Yu and Dimitri P. Bertsekas. Error Bounds for Approximations from Projected Linear Equations. *Mathematics of Operations Research*, 35(2):306–329, may 2010.