

Effectively Subsampled Quadratures for Least Squares Polynomial Approximations

Pranay Seshadri ^{*}, Akil Narayan [†], Sankaran Mahadevan [‡]

May 3, 2017

Abstract

This paper proposes a new deterministic sampling strategy for constructing polynomial chaos approximations for expensive physics simulation models. The proposed approach, *effectively subsampled quadratures* involves sparsely subsampling an existing tensor grid using QR column pivoting. For polynomial interpolation using hyperbolic or total order sets, we then solve the following *square* least squares problem. For polynomial approximation, we use a column pruning heuristic that removes columns based on the highest total orders and then solves the *tall* least squares problem. While we provide bounds on the condition number of such tall submatrices, it is difficult to ascertain how column pruning effects solution accuracy as this is problem specific. We conclude with numerical experiments on an analytical function and a model piston problem that show the efficacy of our approach compared with randomized subsampling. We also show an example where this method fails.

1 Introduction & motivation

Polynomial chaos is a powerful tool for uncertainty quantification that has seen widespread use in numerous disciplines [35, 7, 33, 42]. It approximates a model’s response with respect to uncertainty in the input parameters using orthogonal polynomials. One important challenge in computing these polynomial approximations lies in determining their coefficients. Standard approaches to compute these coefficients include tensor grids as in Figure 1(a) and sparse grids in (b). In this paper we present a deterministic strategy to construct stable polynomial approximations by *subsampling points* from a tensor grid, as in Figure 1(c).

The motivation for moving away from tensor product spaces is that the number of points grow exponentially with dimension. While sparse grids [18, 8, 12] are one way forward, they still restrict users to very specific index sets—even with their various growth rules. Least squares based methods on the other hand offer more flexibility. The idea of employing least squares techniques in a polynomial approximation context is not a new idea, but recent theoretical results [43] motivate the idea of randomized or *subsampled* quadratures. In a nutshell the idea is to evaluate the simulation at only a small subset of points from a suitable tensorial grid, and to subsequently construct an approximation using either compressed sensing [14, 22, 30, 38] or least squares. The goal is to subsample in a way such that the subsample-constructed surrogate approximates the surrogate that would be obtained by using all the points from the tensor grid, by minimizing the mean-square error in the coefficients. After subsampling, the approximation scheme we focus on is a least-squares-type approach; subsampling approaches for least-squares have received little attention [43] compared to compressive sampling.

Despite ease of implementation, discrete least-squares has known stability issues. For instance, when using abscissae associated with Newton-Cotes quadrature (equidistant point sets), it is relatively unstable even for an

^{*}Research Associate, Department of Engineering, University of Cambridge, U.K., ps583@cam.ac.uk

[†]Assistant Professor, Department of Mathematics, University of Utah, Salt Lake City, U.S.A.

[‡]John R. Murray Sr. Professor, Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, U.S.A.

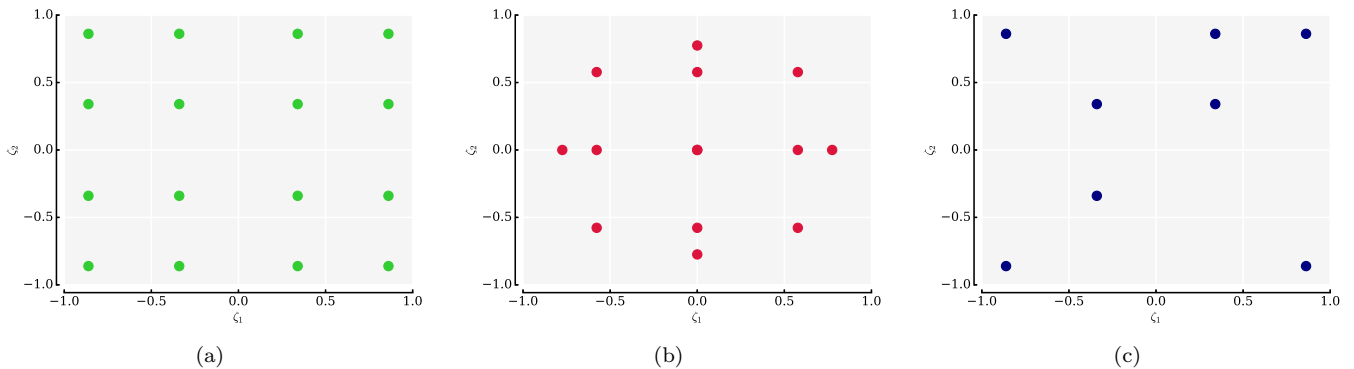


Figure 1: Sample stencils for (a) Tensor grid; (b) Sparse grid and (c) Effectively subsampled quadratures, with a maximum univariate degree of 4 using Gauss-Legendre quadrature points.

infinitely smooth noiseless function [31]. Recent theoretical work [28, 27, 43, 11, 29] has centered around determining the stability conditions for least squares, when using *independent and identically distributed* (iid) or random sampling. In Cohen et al. [11] the authors analyze univariate polynomials and observe for approximating an m -th order polynomial approximately cm^2 points are required when sampling randomly from a uniform distribution on the inputs, or approximately $cm\log(m)$ points when sampling from a Chebyshev distribution, for some constant c . Theoretical extensions to multivariate polynomial spaces can be found in Chkifa et al. [10], and with applications to multivariate orthogonal polynomial expansions in tensor-product and total-order index spaces in [43].

In this paper we introduce a new approach for deterministically subsampling quadratures in the context of least squares approximations. Given a polynomial subspace and a tensor product quadrature grid, the central idea is to select a number of subsamples from the grid equal to the dimension of the subspace using a QR factorization column pivoting strategy. Further pruning of the polynomial subspace is performed via heuristics. We remark here that good performance of our method depends on the pruning strategy adopted, and in general it is difficult to develop rigorous bounds.

Details of the approach are in section 2 with a discussion in section 3. This is followed by numerical examples in sections 4 and 5. All our results, and code to produce them, can be found at: www.effective-quadratures.org/papers. These codes use routines from our effective-quadratures toolkit [34].

1.1 Preliminaries & notation

Let $\zeta = (\zeta^{(1)}, \dots, \zeta^{(d)})$ be a d -dimensional vector of mutually independent random variables with joint probability density ρ and marginal densities ρ_i related by $\rho(\zeta) = \prod_{i=1}^d \rho_i(\zeta^{(i)})$ defined on \mathbb{R}^d .

1.2 On polynomials

Let $\{\psi_j^{(i)}\}_{j=0}^{\infty}$ denote a family of polynomials L^2 -orthogonal on \mathbb{R} when weighted by the density ρ_i :

$$\int_{\mathbb{R}} \psi_i^{(k)}(s) \psi_j^{(k)}(s) \rho_k(s) ds = \mathbb{E} \left[\psi_i^{(k)}(\zeta^{(k)}) \psi_j^{(k)}(\zeta^{(k)}) \right] = \delta_{i,j}, \quad (1)$$

where $\delta_{i,j}$ is the Kronecker delta. Existence of such a family is ensured under mild assumptions on ρ_i [17]; the ρ_i -weighted L^2 -completeness of the polynomial family can be established under some additional technical assumptions [26, 15].

A multivariate polynomial $\psi_{\mathbf{j}} : \mathbb{R}^d \rightarrow \mathbb{R}$ can be defined as a product of univariate polynomials,

$$\psi_{\mathbf{j}}(\zeta) = \prod_{k=1}^d \psi_{j_k}^{(k)}(\zeta^{(k)}), \quad \mathbf{j} = (j_1, \dots, j_d) \in \mathbb{N}_0^d, \quad (2)$$

where \mathbf{j} is a multi-index that denotes the order (degree) of $\psi_{\mathbf{j}}$ and its composite univariate polynomials $\psi_{j_k}^{(k)}$. The family $\{\psi_{\mathbf{j}}\}_{\mathbf{j} \in \mathbb{N}_0^d}$ defined in this way is mutually orthogonal in L^2 weighted by ρ .

For computational purposes, we require a finite number of polynomials $\psi_{\mathbf{j}}$. This finite set is chosen by restricting the multi-index \mathbf{j} to lie in a finite multi-index set \mathcal{J} . There are four well-known multi-index sets \mathcal{J} that have proven fruitful for parametric approximation: tensor product index sets, total order index sets, hyperbolic cross spaces [40, 8] and hyperbolic index sets [2]. Each of these index sets in d dimensions is well-defined given a fixed $k \in \mathbb{N}_0$, which indicates the maximum polynomial degree associated to these sets.

Isotropic tensor product index sets consist of multi-indices satisfying $\max_k j_k \leq k$, and have a cardinality (number of elements) equal to $(k+1)^d$. Total order index sets contain multi-indices satisfying $\sum_{i=1}^d j_i \leq k$. Loosely speaking, total order indices disregard some higher order interactions between dimensions present in tensorized spaces, and a total order index set \mathcal{J} has cardinality

$$|\mathcal{J}| = \binom{k+d}{k}. \quad (3)$$

Hyperbolic cross sets contain indices governed by the rule $\prod_{i=1}^d (j_i + 1) \leq k + 1$, and prune even more tensorial interaction terms than total order index spaces. The cardinality of this last index set is approximately $(k+1)(1 + \log(k+1))^{d-1}$ [43]. Finally, a hyperbolic index set contain indices that satisfy

$$\left(\sum_{i=1}^d j_i^q \right)^{1/q} \leq k, \quad (4)$$

where q is a user-defined constant that can be varied from 0.2 to 1.0. When $q = 1$ the hyperbolic index space is equivalent to a total order index space, while for values less than unity higher-order interactions terms are eliminated [2]. In the numerical examples in this paper we will form approximations from the hyperbolic index space (not the hyperbolic cross set).

1.3 On quadrature rules

We assume existence of a quadrature rule $\{(\zeta_i, \omega_i^2)\}_{i=1}^m$, with non-negative weights ω_i^2 , such that

$$\sum_{i=1}^m \omega_i^2 \phi_{\mathbf{j}}(\zeta_i) \phi_{\mathbf{l}}(\zeta_i) = \int_{\mathbb{R}^d} \phi_{\mathbf{j}}(\zeta) \phi_{\mathbf{l}}(\zeta) \rho(\zeta) dx \zeta = \delta_{\mathbf{l}, \mathbf{j}}, \quad \mathbf{j}, \mathbf{l} \in \mathcal{J}. \quad (5)$$

Thus the choice of quadrature rule is intimately tied to the choice of \mathcal{J} . Ideally, the cardinality m of this quadrature rule should be as small as possible. In one dimension, we can achieve equality in the above expression with $m = |\mathcal{J}|$ for essentially any density ρ by using a *Gaussian* quadrature rule [16]. However, rules of this optimal (smallest) cardinality are not known in the multivariate setting for general ρ and \mathcal{J} . The construction of quasi-optimal multivariate quadrature rules, even over canonical domains, is a challenging computational problem [39] and no universal solutions are currently known.

On tensor-product domains (as is assumed in this paper) one quadrature rule satisfying (5) can be constructed by tensorizing univariate rules. For example, (5) holds if the univariate rules are Gaussian quadrature rules with sufficiently high accuracy. Since the cardinality of the resulting tensorial rule grows exponentially with dimension d , evaluating the model over a full tensor-product quadrature rule quickly becomes infeasible. In this paper, we will

use a tensorial Gauss quadrature rule with high enough accuracy to ensure (5), but in principle it is also reasonable to apply our approach when the first equality in (5) is only approximate.

This motivates the goal of this paper: prune a full tensorized quadrature rule $\{(\zeta_i, \omega_i^2)\}_{i=1}^m$ via a subsampling procedure so that the subsampled grid has an acceptable cardinality while hopefully maintaining the accuracy properties associated to an approximation using the index set \mathcal{J} .

1.4 On matrices

Matrices in this paper are denoted by upper case bold letters, while vectors are denoted by lower case bold letters. For a matrix, $\mathbf{D} \in \mathbb{R}^{m \times n}$ with $m \geq n$, singular values are defined by $\sigma_i(\mathbf{D})$, with $i = 1, \dots, n$. Unless explicitly stated otherwise, the singular values are arranged in descending order, i.e. $\sigma_1(\mathbf{D}) \geq \dots \geq \sigma_n(\mathbf{D})$. The ℓ^2 condition number of \mathbf{D} is denoted $\kappa(\mathbf{D})$ and is the ratio of the largest singular value to the smallest. The singular values of \mathbf{D} coincide with those of \mathbf{D}^T .

2 Effectively subsampled quadratures

In this section we describe polynomial least squares via *effectively subsampled quadratures*. To aid our discussion, the overall strategy is captured in Figure 2; we describe the details below.

2.1 Setting up the \mathbf{A} matrix

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix formed by evaluating n multivariate orthonormal polynomials (constructed as described in Section 1.2) at a tensor grid formed of m quadrature points (described in Section 1.3). Individual entries of \mathbf{A} are given by

$$\mathbf{A}(i, \mathbf{j}) = \omega_i \psi(\zeta_i), \quad \mathbf{j} \in \mathcal{J}, \quad i = 1, \dots, m. \quad (6)$$

where in this paper we implicitly assume that the index set \mathcal{J} is a hyperbolic¹ or total order index set with the condition that $|\mathcal{J}| = n \ll m$. We assume the quadrature rule is accurate enough so that (5) holds. We consider \mathbf{A} a matrix by assuming a linear ordering of the elements in \mathcal{J} ; the particular ordering chosen may be arbitrary in the context of this paper.

We define $\mathbf{b} \in \mathbb{R}^m$ to be the vector of weighted model evaluations at the m quadrature points, where individual entries of \mathbf{b} are given by

$$\mathbf{b}(i) = \omega_i f(\zeta_i), \quad (7)$$

where $f(\cdot)$ represents the quantity of interest from our computational model; we seek to approximate this with $g(s)$

$$f(s) \approx g(s) = \sum_{i \in \mathcal{J}} x_i \psi_i(s). \quad (8)$$

This is equivalent to solving the least squares problem $\mathbf{A}\mathbf{x} = \mathbf{b}$ for the coefficients $\mathbf{x} \in \mathbb{R}^n$. This requires our model to be evaluated at each tensor grid quadrature point. A non-deterministic (i.e., randomized) approach to reduce the cost associated with this least squares problem is to randomly subsample tensor grid quadrature points as outlined in [43]. This strategy requires a reduction in the number of basis terms in \mathbf{A} to promote stability. We make specific comments regarding the randomized strategy in the numerical studies section of this paper. To contrast, our approach in this paper is deterministic and rooted in a heuristic that is tailored for least squares problems.

¹Either a hyperbolic cross space or a hyperbolic space

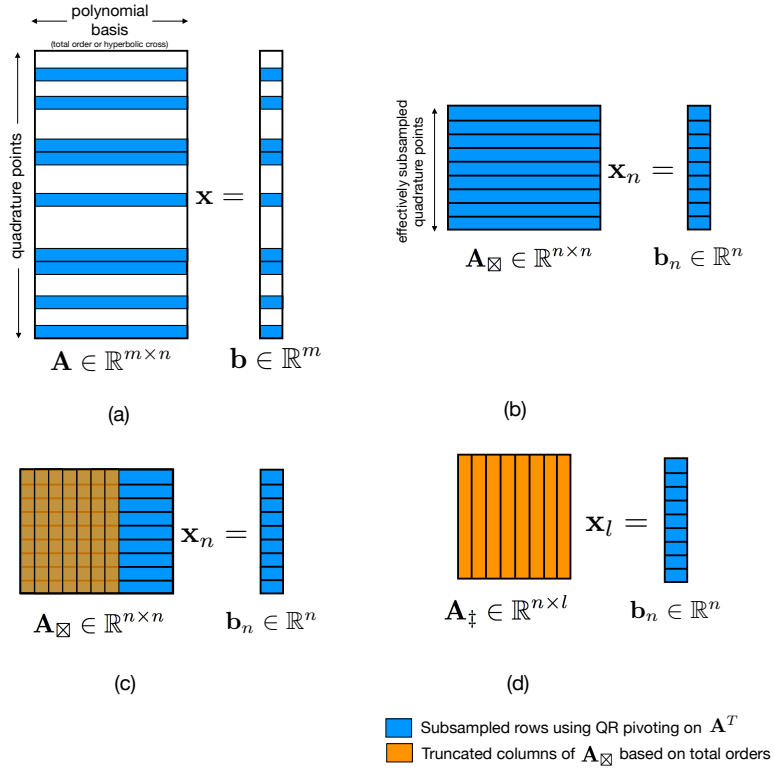


Figure 2: Step-by-step outline of the effectively subsampled quadrature technique for computing polynomial least squares approximations: (a) Setting up the \mathbf{A} matrix; (b) QR factorization with column pivoting on \mathbf{A}^T ; (c) Column pruning; (d) Solving the least squares problem.

2.2 QR factorization with column pivoting on \mathbf{A}^T

To reduce the cost associated with evaluating our model at each point in a tensor grid, we utilize QR column pivoting—a well-known heuristic for solving rank deficient least squares problems. QR column pivoting works by (i) determining the numerical rank $r < n$ of an m -by- n matrix, and (ii) permuting columns of the matrix such that the first r columns are linearly independent [23]. Here, we apply this heuristic for subselecting rows of \mathbf{A} . Let the QR column pivoting factorization of \mathbf{A}^T be given by

$$\mathbf{A}^T \mathbf{P} = \mathbf{Q} \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_2 \end{pmatrix} \quad (9)$$

where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is orthogonal, $\mathbf{R}_1 \in \mathbb{R}^{n \times n}$ is nonsingular² and upper triangular with positive diagonal elements, and $\mathbf{R}_2 \in \mathbb{R}^{n \times (m-n)}$. The matrix \mathbf{P} is a *permutation matrix* that permutes columns of \mathbf{A}^T such that the diagonal entries of \mathbf{R}_1 are greedily maximized. It should be noted that the factorization in (9) is not necessarily unique as there may be numerous ways to select n linearly independent columns of \mathbf{A} [1]. To determine precisely which rows of \mathbf{A} to subselect (equivalent to determining which points to subsample) we define a vector $\boldsymbol{\pi}$ that converts the pivots encoded in the matrix \mathbf{P} to the specific rows in \mathbf{A} by

$$\boldsymbol{\pi} = \mathbf{P}^T \mathbf{u}, \quad (10)$$

where $\mathbf{u} = (1, 2, \dots, m)^T$ is a vector of integers from 1 to m . The vector $\boldsymbol{\pi}$ contains ordered rows of \mathbf{A} that are subselected via the QR factorization. For clarity let $\boldsymbol{\pi}_n = \boldsymbol{\pi}(1:n)$ be the first n entries of $\boldsymbol{\pi}$, and the operator $\mathcal{G}_{\boldsymbol{\pi}_n}$ that selects the rows indexed by $\boldsymbol{\pi}_n$. We define

$$\mathbf{A}_{\boxtimes} = \mathcal{G}_{\boldsymbol{\pi}_n}(\mathbf{A}) \quad (11)$$

where $\mathbf{A}_{\boxtimes} \in \mathbb{R}^{n \times n}$ (see Figure 2(b)). For ease in notation we define column-wise vectors of \mathbf{A}^T using \mathbf{c} as follows:

$$\mathbf{A}^T = \begin{pmatrix} - & \mathbf{a}_1^T & - \\ & \vdots & \\ - & \mathbf{a}_n^T & - \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{c}_1 & \dots & \mathbf{c}_n \\ | & & | \end{pmatrix}. \quad (12)$$

The QR pivoting algorithm used in this study—shown in Algorithm 1—is based on the work of Dax [13] and uses modified Gram-Schmidt QR factorization. As shown in lines 14-15 of the algorithm, the vector \mathbf{c}_j is orthogonalized by iteratively projecting it on to the subspace orthogonal to $\text{span}(\mathbf{g}_1, \dots, \mathbf{g}_{k-1})$, where $\mathbf{g}_k = \mathbf{c}_k / \|\mathbf{c}_k\|_2$ [23]. Householder QR pivoting (see page 227 of [20]) may also be used instead. Both algorithms require an initial computation of the column norms followed by subsequent updates after iterative orthogonalization. As Björck [1] notes, once the initial column norms have been computed, they may be updated using the identity

$$\begin{aligned} \|\mathbf{c}_j^{(k+1)}\|_2 &= \sqrt{\|\mathbf{a}_k^{(k)}\|_2^2 - (\mathbf{g}_k^T \mathbf{c}_j^{(k)})^2} \\ &= \|\mathbf{c}_k^{(k)}\|_2 \left(1 - \left(\frac{\mathbf{g}_k^T \mathbf{c}_j^{(k)}}{\|\mathbf{c}_k^{(k)}\|_2} \right)^2 \right)^{1/2}, \end{aligned} \quad (13)$$

instead of directly computing the column norms. The above identity is applied in line 16 of Algorithm 1. While (13) reduces the overhead of pivoting from $\mathcal{O}(mn^2)$ flops to $\mathcal{O}(mn)$ flops [20], there are some salient computational issues that need to be considered when using such an updating rule. (See page 6 of [13].) A small but notable difference in our implementation of the QR with column pivoting algorithm is that our for-loop terminates at $\mathbf{k}=\mathbf{n}$, which is sufficient for computing the n pivots we need. A python implementation of the above algorithm is included in our `effective-quadratures` toolkit, and is utilized in all the numerical studies in this paper. For QR factorizations with column pivoting, there are comprehensive LAPACK routines, such as DGEQP3 [32].

²By assuming (5), \mathbf{A} must be of full rank, equal to $\min\{m, n\} = n$. Thus \mathbf{R}_1 also has rank n .

Algorithm 1 QR with column pivoting on $\mathbf{A}^T = (\mathbf{c}_1, \dots, \mathbf{c}_m) \in \mathbb{R}^{n \times m}$ where $m > n$

```

1: procedure QR COLUMN PIVOTING( $\mathbf{A}^T$ )
2:   Compute and store column norms of  $\mathbf{A}^T$  in colnorms
3:   Declare permutation vector  $\pi=1:m$ 
4:   for  $k=1:n$  do
5:     Store  $\max(\text{colnorms}(k:n))$  index as jmax
6:     Set  $\text{jmax} = \text{jmax} + (k-1)$ 
7:     if  $k$  is not equal to jmax then
8:       Swap  $\mathbf{c}_k$  with  $\mathbf{c}_{\text{jmax}}$ 
9:       Swap  $\text{colnorms}(k)$  with  $\text{colnorms}(\text{jmax})$ 
10:      Swap  $\pi(k)$  with  $\pi(\text{jmax})$ 
11:    end if
12:    if  $k$  is not equal to  $n$  then
13:      for  $j=k+1:n$  do
14:         $\mathbf{g} = \mathbf{c}_k / \|\mathbf{c}_k\|_2$ 
15:         $\mathbf{c}_j = \mathbf{c}_j - \mathbf{g}^T \mathbf{c}_j \mathbf{g}$ 
16:        Update  $\text{colnorms}(j)$ 
17:      end for
18:    end if
19:    if  $k$  is not equal to 1 then
20:      for  $i=1:k-1$  do
21:         $\mathbf{h} = \mathbf{c}_i / \|\mathbf{c}_i\|_2$ 
22:         $\mathbf{c}_k = \mathbf{c}_k - \mathbf{h}^T \mathbf{c}_k \mathbf{h}$ 
23:      end for
24:    end if
25:  end for
26:  return  $\pi$ 
27: end procedure

```

▷ Swaping
 ▷ Orthogonalization
 ▷ Reorthogonalization

2.2.1 Relation to rank revealing QR factorizations

The factorization in (9) is called a *rank revealing QR (RRQR)* factorization if it satisfies the property that $\sigma_{\min}(\mathbf{R}_1) \geq \sigma_k(\mathbf{A}^T)/p(n, m)$, where $p(n, m)$ is a function bounded by a lower order polynomial in n and m . The Businger and Golub Householder QR with pivoting algorithm [20], the Chandrasekaran and Ipsen algorithm [9] and that of Golub, Klema and Stewart [19] fall into this category. Bounds on all n singular values of \mathbf{R}_1 —and not just the minimum—can be obtained when using algorithms that yield a *strong RRQR* factorization; a term coined by Gu and Eisenstat [21]. The latter also provide an efficient algorithm for computing a strong RRQR. Broadbent et al. [6] follow this work and proved that a strong RRQR applied to matrices with more columns than rows—as in the case for our QR factorization with \mathbf{A}^T —yields the following identity

$$\sigma_i(\mathbf{R}_1) \geq \frac{\sigma_i(\mathbf{A})}{\sqrt{1 + \delta^2 n(m - n)}} \quad \text{for } 1 \leq i \leq n, \quad (14)$$

for some constant $\delta > 1$. As the singular values of \mathbf{R}_1 are equivalent to those of \mathbf{A}_{\boxtimes} , this implies that singular values $\sigma_i(\mathbf{A}_{\boxtimes})$ lie between $\sigma_i(\mathbf{A})$ and the right-hand-side of (14). We briefly analyze the stability of solving a linear system involving the rank- n matrix \mathbf{A}_{\boxtimes} by bounding its condition number relative to the condition number.

Lemma 1 *With δ the parameter in the inequality (14), and $\kappa(\cdot)$ the 2-norm condition number of a matrix, then*

$$\kappa(\mathbf{A}_{\boxtimes}) \leq \kappa(\mathbf{A}) \sqrt{1 + \delta^2 n(m - n)} \quad (15)$$

Proof 1 *We first note that*

$$\kappa(\mathbf{A}) := \frac{\sigma_1(\mathbf{A})}{\sigma_n(\mathbf{A})}. \quad (16a)$$

By (14), we have

$$\frac{1}{\sigma_n(\mathbf{A}_{\boxtimes})} \leq \frac{\sqrt{1 + \delta^2 n(m - n)}}{\sigma_n(\mathbf{A})} \quad (16b)$$

Finally, since \mathbf{A}_{\boxtimes} is precisely a submatrix of \mathbf{A} (see (11)), then the singular values of \mathbf{A} and \mathbf{A}_{\boxtimes} interlace [41], in particular,

$$\sigma_1(\mathbf{A}_{\boxtimes}) \leq \sigma_1(\mathbf{A}). \quad (16c)$$

Combining the three relations (16) with $\kappa(\mathbf{A}_{\boxtimes}) := \sigma_1(\mathbf{A}_{\boxtimes}) / \sigma_n(\mathbf{A}_{\boxtimes})$ proves (15).

2.2.2 Relation to subset selection

Like QR with column pivoting, *subset selection* is an alternative heuristic that aims to produce a well-conditioned submatrix with linearly independent columns. In practice, subset selection can produce a submatrix with smaller condition number than that provided by QR with column pivoting [20]. The algorithm has two key steps that can be adapted to determine which rows of \mathbf{A} to subselect for \mathbf{A}_{\boxtimes} . The first step involves computing the singular value decomposition of \mathbf{A}^T . The next step requires QR column pivoting to be applied to a matrix formed by the transpose of the first n right-singular vectors of \mathbf{A} ,

$$\mathbf{V}(:, 1 : n)^T \mathbf{P} = \mathbf{Q}\mathbf{R}, \quad (17)$$

where as before the columns of \mathbf{P} encode the permutations. Equations (10) and (11) can subsequently be used to determine \mathbf{A}_{\boxtimes} . One of main computational bottlenecks with *subset selection* is the aggregated cost of performing both an SVD—costing $\mathcal{O}(m^2 n^3)$ flops—with a QR column pivoting routine [20].

2.3 Column pruning

In (7) we defined elements of the vector \mathbf{b} to be the weighted model evaluations at all m quadrature points. In practice we only require the model evaluations at the n quadrature points identified by $\boldsymbol{\pi}_n$. We define the *effectively subsampled quadrature* points and weights to be

$$\zeta_{e,j} = \zeta_{\boldsymbol{\pi}(j)}, \quad \omega_{e,j} = \omega_{\boldsymbol{\pi}(j)} \quad (18)$$

respectively, for $j = 1, \dots, n$, where $\boldsymbol{\pi}$ is the QR permutation vector defined in (11). Thus the vector of weighted model evaluations at these points is given by

$$\mathbf{b}_n(j) = \omega_{e,j} f(\zeta_{e,j}), \quad j = 1, \dots, n. \quad (19)$$

Assembling the *square* linear system of equations yields

$$\underset{\mathbf{x}_n}{\text{minimize}} \|\mathbf{A}_{\boxtimes} \mathbf{x}_n - \mathbf{b}_n\|_2. \quad (20)$$

The subscript n in \mathbf{x}_n simply denotes the number of coefficient terms—equivalent to the cardinality of the polynomial basis defined by \mathcal{J} —that are to be solved for. While (20) can be solved to yield accurate coefficient estimates for smooth functions, it is generally ill-advised, as we wish to *approximate* rather than *interpolate*. Consequently, we prune down the number of columns of \mathbf{A}_{\boxtimes} from n to l . It is difficult to offer heuristics for column pruning as this will no doubt vary across applications, as it is dependent on which polynomial basis terms can be truncated without significant loss of accuracy in approximating f . Our experience (from the examples later in this paper) suggests that better results are obtained by eliminating columns with the highest total degrees first, and our results using $n/l \in [1, 1.5]$ show promise.

The matrices \mathbf{A} and \mathbf{A}_{\boxtimes} are formed from the polynomial basis defined by \mathcal{J} . The above procedure prunes elements from \mathcal{J} . We define the index set that results from this pruning as \mathcal{I} . I.e., \mathcal{I} is defined by

$$|\mathcal{I}| = l \quad \mathcal{I} \subseteq \mathcal{J}, \quad \mathbf{k} \in \mathcal{J} \setminus \mathcal{I} \implies \sum_{i=1}^d k_d \geq \sum_{i=1}^d j_d \quad \text{for all } \mathbf{j} \in \mathcal{I}. \quad (21)$$

This does not uniquely define \mathcal{I} as there is usually not a unique element of \mathcal{J} with highest total order. In this paper we perform the following methodology for pruning a single element from \mathcal{J} : We specify an (arbitrary) ordering of elements in \mathcal{J} , and based on this ordering prune the first $\mathbf{k} \in \mathcal{J}$ achieving the maximum total order.

Let $\mathbf{A}_{\dagger} \in \mathbb{R}^{n \times l}$ be the submatrix of \mathbf{A}_{\boxtimes} associated with the pruned set, \mathcal{I} , i.e.,

$$\mathbf{A}_{\dagger}(i, \mathbf{j}) = \omega_{e,i} \psi_{\mathbf{j}}(\zeta_{e,i}), \quad \mathbf{j} \in \mathcal{I}, \quad i = 1, \dots, n. \quad (22)$$

Regardless of how pruning is performed, the following result holds:

Corollary 1 *With δ as in Lemma 1, let \mathbf{A}_{\dagger} be the column-pruned version of \mathbf{A}_{\boxtimes} . Then*

$$\kappa(\mathbf{A}_{\dagger}) \leq \kappa(\mathbf{A}_{\boxtimes}) \leq \kappa(\mathbf{A}_n) \sqrt{1 + \delta^2 n(m - n)} \quad (23)$$

Proof 2 *From the interlacing property of singular values we have*

$$\sigma_1(\mathbf{A}_{\boxtimes}) \geq \sigma_1(\mathbf{A}_{\dagger}) \quad \text{and} \quad \sigma_l(\mathbf{A}_{\dagger}) \geq \sigma_n(\mathbf{A}_{\boxtimes}), \quad (24)$$

where the singular values are ordered such that $\sigma_1(\cdot) \geq \sigma_2(\cdot) \geq \dots \geq \sigma_n(\cdot)$. Then

$$\frac{\sigma_1(\mathbf{A}_{\dagger})}{\sigma_l(\mathbf{A}_{\dagger})} \leq \frac{\sigma_1(\mathbf{A}_{\boxtimes})}{\sigma_n(\mathbf{A}_{\boxtimes})} \implies \kappa(\mathbf{A}_{\dagger}) \leq \kappa(\mathbf{A}_{\boxtimes}). \quad (25)$$

The second inequality in (23) is an application of Lemma 1.

This implies that reducing the number of columns—in particular for our case of eliminating the basis terms with the highest total degrees—will not cause an increase in the condition number of \mathbf{A}_{\dagger} compared to \mathbf{A}_{\boxtimes} . Thus compared to solving a least-squares problem with the best rank- n approximation to \mathbf{A} , we suffer a penalty of the order $\delta\sqrt{mn}$ (when $m \gg n$).

2.4 Solving the least squares problem

We now reach the main objective of this paper, which is to solve the least squares problem given by

$$\underset{\mathbf{x}_l}{\text{minimize}} \|\mathbf{A}_{\dagger} \mathbf{x}_l - \mathbf{b}_n\|_2, \quad (26)$$

for the coefficients $\mathbf{x}_l \in \mathbb{R}^l$. There are two salient points we wish to emphasize upon when solving (26). The first pertains to preconditioning. We impose a unit length diagonal column scaling, a preconditioner that is frequently selected when solving least squares problems [23]. We define the preconditioner (a nonsingular matrix) $\mathbf{S} \in \mathbb{R}^{n \times n}$ as

$$\mathbf{S} = \begin{pmatrix} \|\mathbf{a}_{\dagger 1}\|_2 & & \\ & \ddots & \\ & & \|\mathbf{a}_{\dagger n}\|_2 \end{pmatrix} \quad (27)$$

where vectors $\{\mathbf{a}_{\dagger 1}, \dots, \mathbf{a}_{\dagger n}\}$ are the columns of \mathbf{A}_{\dagger} . This yields the modified least squares problem

$$\underset{\mathbf{z}_l}{\text{minimize}} \|\mathbf{A}_{\dagger} \mathbf{S}^{-1} \mathbf{z}_l - \mathbf{b}_n\|_2 \quad \text{with} \quad \mathbf{S} \mathbf{x}_l = \mathbf{z}_l. \quad (28)$$

This brings us to a well-known observation about solving least-squares problems: Solving the normal equations, e.g., computing the inverse of $\mathbf{S}^{-T} \mathbf{A}_{\dagger}^T \mathbf{A}_{\dagger} \mathbf{S}^{-1}$ or even forming an associated Cholesky factorization, is relatively unstable. Our results reported here solve the system (28) via the more stable *QR* factorization approach (see Chapter 4 of [23]).

2.5 A step-by-step algorithm

We summarize this section with an algorithm incorporating the previously defined computational techniques given a function $f = f(\boldsymbol{\zeta})$ defined on $\mathcal{R} = \text{supp}(\rho)$ with joint probability density function $\rho(\boldsymbol{\zeta})$.

1. **Polynomial basis selection:** Let a joint density $\rho(\boldsymbol{\zeta})$ be given. Select a hyperbolic or total order index set \mathcal{J} , with a cardinality $|\mathcal{J}| = n$.
2. **Initial grid selection:** Choose a tensorized quadrature rule consisting of m points and weights $\{\zeta_j, \omega_j\}_{j=1}^m$ such that the quadrature rule satisfies (5). This defines the matrix \mathbf{A} .
3. **Subselecting points from initial grid:** Compute the QR with pivoting factorization $\mathbf{A}^T \mathbf{P} = \mathbf{Q} \mathbf{R}$ and select the first n entries encoded in \mathbf{P} to compute the matrix \mathbf{A}_{\boxtimes} . Evaluate the model at the quadrature points corresponding to these n pivots to compute \mathbf{b}_n .
4. **Column pruning:** For noisy f , prune down the number of columns by eliminating the columns of \mathbf{A}_{\boxtimes} that correspond to the highest total orders. Store the remaining columns in a new matrix, \mathbf{A}_{\dagger} . As a heuristic, we recommend pruning down by ratios of 1.25 and 1.50.
5. **Least squares:** Solve the least squares problem with \mathbf{A}_{\dagger} and \mathbf{b}_n .

3 Discussion & heuristics

The strategy described is similar to the procedure of constructing *approximate Fekete points* [4]; this latter approach has been used in the context of finding near-optimal interpolation points in multidimensional space. An alternative way to think about the strategy above is that for a given design matrix \mathbf{A} , and a maximum number of permitted model evaluations n , we are extracting a set of at most n sample points. Our algorithm in this context offers a deterministic recipe for subsampling a tensor grid. The notable difference between our algorithm and that produced

by approximate Fekete points is that we introduce (square-root) quadrature weights ω_i in the definition of \mathbf{A} . The algorithm as we have described it can produce accurate low-rank approximations to the matrix \mathbf{A} [24].

We note that we have described this quadrature subselection strategy as an attempt to construct a well-conditioned design matrix. Remarkably, if $\omega_i \equiv 1$, then the limiting behavior of the points selected via this algorithm is known. Consider the univariate case on a bounded set but with large m and n . In this case it is known that as m and n tend to infinity appropriately, the QR selection strategy chooses points that distribute according to the Chebyshev (arcsine) measure [3]. In addition, this property holds in the multidimensional setting on a hypercube (the set formed from the Cartesian product of univariate bounded intervals). If we use the QR strategy to select points from a sufficiently dense grid on a hypercube, these points distribute according to the product Chebyshev measure on the hypercube. Precise conditions on m , n , the type of grid, and the type of convergence is given in [3].

3.1 Memory requirements

One of the disadvantages of QR with column pivoting is that to pivot the column with the largest norm, we must compute the norms of all the remaining columns. In the next stage of the for-loop, we then need to downdate these norms. This is where Identity (13) is useful. Occasionally, a cancellation occurs in (13) requiring knowledge of all the column entries to recompute its norm. Thus, it is not possible to carry out QR with column pivoting by storing 2-3 rows or columns at a time; access to the full matrix is required. This implies that the cost of our technique scales exponentially with the dimension of f 's inputs—owing to QR column pivoting. This is one drawback of the current approach. This scaling is however independent from the number of evaluations of f that we require.

While writing the entire matrix \mathbf{A} on the disk and then extracting rows and columns as required is one option, it is far from elegant. One possible path forward lies in randomized QR column pivoting techniques [5]. These techniques are promising because they restrict the size of \mathbf{A} ; however, a detailed investigation of this procedure is outside the scope of this manuscript.

4 An Analytical Example

In this simple analytical example we set

$$f(\zeta) = \exp(\zeta^{(1)} + \zeta^{(2)}) \quad (29)$$

defined over $\mathcal{R} = [-1, 1]^2$ with $\rho(\zeta)$ the uniform density. We wish to approximate f using a basis of Legendre orthonormal polynomials with *effectively subsampled quadratures*. We subsample an isotropic tensor grid formed from a 21-point Gauss-Legendre stencil in each dimension. We let \mathcal{J} be a total order basis for our least squares computations. In this example, we compare our method with the randomized sampling approach of [43] and investigate the sensitivity of the procedure to column pruning.

Figure 3(a) plots the approximation errors in the polynomial coefficients using both the randomized and effectively subsampled methods. Here values on the x -axis represent the maximum degree defining the total-order index set \mathcal{J} . The y -axis defines the coefficient error on a base-10 logarithmic scale. For each x -axis value, this error is computed using

$$\epsilon = \|\mathbf{x}_{\otimes, \mathcal{J}} - \mathbf{x}_n\|_2 \quad (30)$$

where \mathbf{x}_{\otimes} are the coefficients estimated from a 21-point tensor grid quadrature rule. The subscript \mathcal{J} in $\mathbf{x}_{\otimes, \mathcal{J}}$ denotes the coefficient values only associated with the multi-indices in the total order index set \mathcal{J} . As mentioned earlier, the coefficients \mathbf{x}_n (see Figure 2) are obtained by solving the least squares problem with the polynomial basis \mathcal{J} . The coefficient errors in Figure 3(a) correspond to the errors associated with solving the least squares problem on the square matrix \mathbf{A}_{\otimes} . The green line in the figure shows the coefficient error resulting from using *effectively subsampled quadratures*, while the yellow shaded regions denote the minimum and maximum values of ϵ obtained using randomized subsampling—with 20 repetitions of the experiment. The red line represents the mean result from those 20 trials.

Table 1: Input parameters and ranges for the piston problem

Input parameters	Range	Description
M	[30, 60]	Piston weight (kg)
S	[0.005, 0.0020]	Piston surface area (m^2)
V_0	[0.002, 0.010]	Initial gas volume (m^3)
k	[1000, 5000]	Spring coefficient (N/m)
P_0	[90000, 110000]	Atmospheric pressure (N/m^2)
T_a	[290, 296]	Ambient temperature (K)
T_0	[340, 360]	Filling gas temperature (K)

In Figures 3(b-d), for a fixed maximum degree k , we prune down the number of columns in \mathbf{A}_{\boxtimes} from $n = |\mathcal{J}|$ to l to yield a smaller index set \mathcal{I} and matrix \mathbf{A}_{\ddagger} . This *column pruning* starts with those columns that have the highest total orders. The errors we report here are given by

$$\epsilon = \|\mathbf{x}_{\otimes, \mathcal{I}} - \mathbf{x}_l\|_2 \quad (31)$$

where $\mathbf{x}_{\otimes, \mathcal{I}}$ are only the coefficients \mathbf{x}_{\otimes} that have multi-indices in \mathcal{I} , and where \mathbf{x}_l are the coefficients obtained via least squares on the matrix \mathbf{A}_{\ddagger} . In Figures 3(b-d) we plot ϵ values for varying total orders with n/l ratios of 1.15, 1.25 and 1.5. Condition numbers associated with the matrices \mathbf{A}_{\boxtimes} and \mathbf{A}_{\ddagger} in Figure 3 are shown in Figure 4.

First let us consider the results when $n/l=1.0$ as in Figure 3(a) and Figure 4(a). Random draws of rows for low maximum degrees—from $k = 2$ to $k = 8$ —leads to, on average, matrices that are nearly singular. This results in coefficient errors that extremely high as illustrated in Figure 3(a). By pruning down the number of columns, this effect can be reduced as shown in Figures 3(b-d). This also results in better coefficient error estimates, however this comes at the cost of requiring more model evaluations for a given number of coefficients to be estimated. In contrast, our effectively subsampled procedure offers reduced error estimates even at a n/l ratio of 1.0; condition numbers for all k never exceed 10.

5 Piston model problem

In this example, we apply our method on a non-linear model of the cycle time of a piston given in [25]. The piston cycle time C is expressed as

$$C = 2\pi \sqrt{\frac{M}{k + S^2 \frac{P_0 V_0 T_a}{T_0 V^2}}}, \quad (32)$$

with

$$V = \frac{S}{2k} \left(\sqrt{A^2 + 4k \frac{P_0 V_0}{T_0} T_a} - A \right) \quad \text{and} \quad A = P_0 S + 19.62M - \frac{kV_0}{S} \quad (33)$$

which depends on the seven inputs given in Table 1. These inputs are uniformly distributed over their respective ranges. Our task here is to contrast *effective quadrature subsampling* with *randomized quadrature subsampling* on a total order basis.

To begin, we compute the full tensor grid solution using 5 points in each direction, yielding a total of $5^7 = 78,125$ function evaluations. We then use these coefficients for computing the errors from both effectively subsampled and randomly subsampled approaches—i.e., ϵ computed as per (30) and (31). We also use these coefficients to compute the *full tensor grid* Sobol' indices shown Figure 5.

We run numerical experiments using total order basis with maximum degrees of $k = 2, 3$ and 4. Table 2 shows the number of basis terms at each of these k values for three different n/l ratios. Both randomly subsampled and

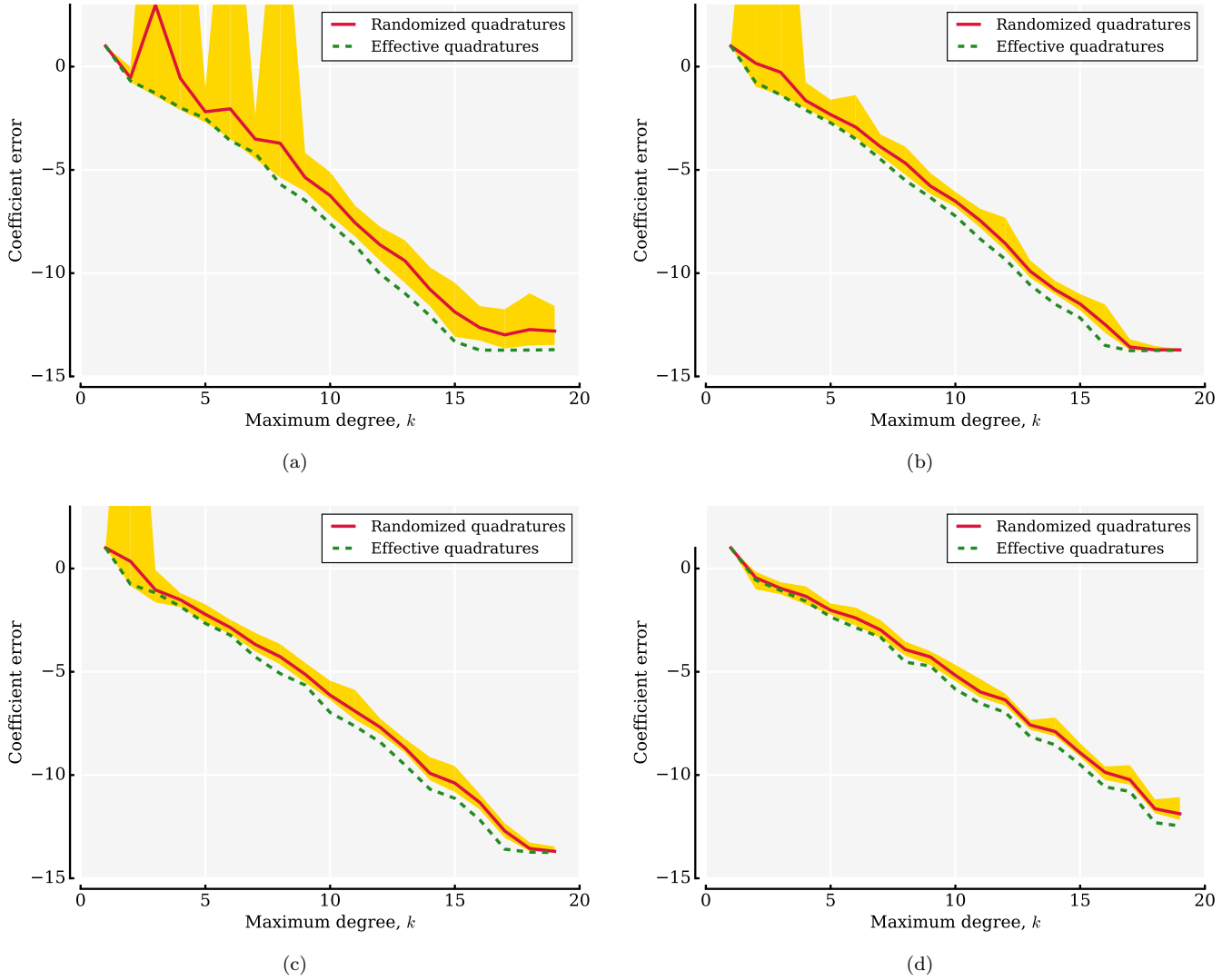


Figure 3: Approximation errors in polynomial coefficients—computed using (30) and (31)—plotted on a base-10 logarithmic scale for the bivariate function, $f(\zeta) = \exp(\zeta^{(1)} + \zeta^{(2)})$ defined on the $[-1, 1]$ hypercube, using Legendre orthonormal polynomials. Polynomial approximations are constructed via least squares using randomized quadratures—repeated 20 times—and effectively subsampled quadratures. Results are plotted for n/l ratios of (a) 1.0; (b) 1.15; (c) 1.25; (d) 1.5.

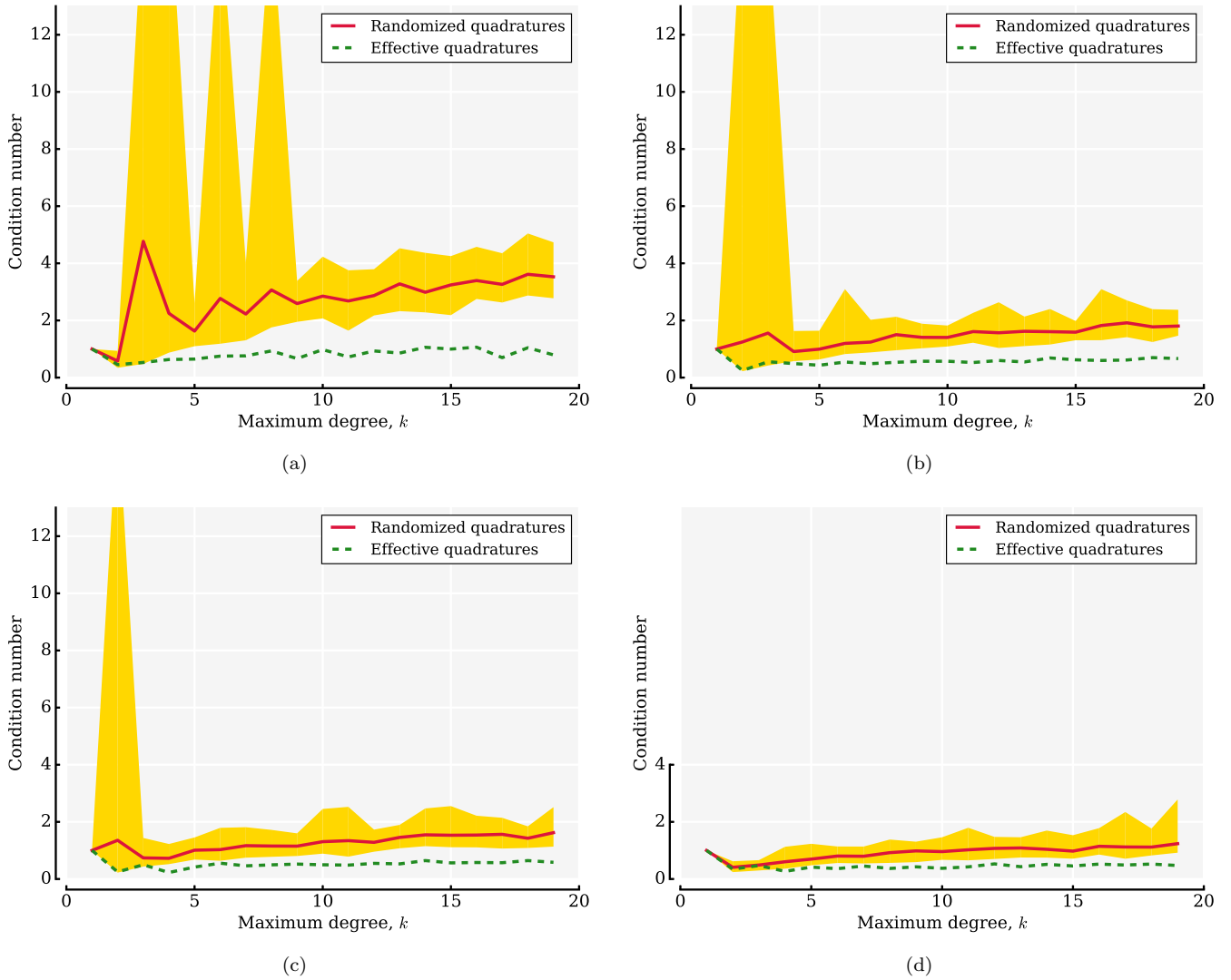


Figure 4: Condition numbers of \mathbf{A}_{\boxtimes} and \mathbf{A}_{\ddagger} , plotted on a base-10 logarithmic scale for the bivariate function, $f(\zeta) = \exp(\zeta^{(1)} + \zeta^{(2)})$ defined on the $[-1, 1]$ hypercube, using Legendre orthonormal polynomials. Polynomial approximations are constructed via least squares using randomized quadratures—repeated 20 times—and effectively subsampled quadratures. Results are plotted for n/l ratios of (a) 1.0; (b) 1.15; (c) 1.25; (d) 1.5.

effectively subsampled condition numbers and errors for these numerical experiments are reported in Tables 3-5. The randomized experiments were repeated 20 times and their minimum, maximum and mean values are reported.

Table 2: Number of tensor grid points and cardinality of total order basis used for the piston problem numerical experiments

k	Points	Cardinality		
		$n/l = 1.0$	$n/l = 1.15$	$n/l = 1.25$
2	$3^7 = 2,187$	36	31	29
3	$4^7 = 16,384$	120	104	96
4	$5^7 = 78,125$	330	287	264

Table 3: Comparison of error ϵ in the coefficients and the condition number κ of \mathbf{A}_{\boxtimes} for randomized and effective quadratures for $n/l=1.00$. Reported randomized results are the outcome of 20 repetitions. Here k indicates the maximum degree of the 7D total order polynomial.

k			Randomized				Effective	
	$\min(\epsilon)$	$\max(\epsilon)$	$\mu(\epsilon)$	$\min(\kappa)$	$\max(\kappa)$	$\mu(\kappa)$	ϵ	κ
2	0.051	3.061	0.448	48.244	1559.1	301.56	0.0252	5.185
3	0.066	3.083	0.671	115.35	12422.1	2118.2	0.01463	15.098
4	0.066	0.901	0.199	595.4	32782.6	2989.0	0.0597	791.98

Table 4: Comparison of error ϵ in the coefficients and the condition number κ of \mathbf{A}_{\ddagger} for randomized and effective quadratures for $n/l=1.15$. Reported randomized results are the outcome of 20 repetitions. Here k indicates the maximum degree of the 7D total order polynomial.

k			Randomized				Effective	
	$\min(\epsilon)$	$\max(\epsilon)$	$\mu(\epsilon)$	$\min(\kappa)$	$\max(\kappa)$	$\mu(\kappa)$	ϵ	κ
2	0.0367	0.252	0.0779	15.02	60.83	23.41	0.0375	3.912
3	0.0284	0.051	0.0403	22.92	36.98	29.60	0.0196	6.821
4	0.0151	0.023	0.0194	30.275	40.97	35.317	0.0159	33.652

Table 5: Comparison of error ϵ in the coefficients and the condition number κ of \mathbf{A}_{\ddagger} for randomized and effective quadratures for $n/l=1.25$. Reported randomized results are the outcome of 20 repetitions. Here k indicates the maximum degree of the 7D total order polynomial.

k			Randomized				Effective	
	$\min(\epsilon)$	$\max(\epsilon)$	$\mu(\epsilon)$	$\min(\kappa)$	$\max(\kappa)$	$\mu(\kappa)$	ϵ	κ
2	0.0440	0.129	0.0719	9.809	27.043	16.008	0.0362	4.9126
3	0.0244	0.0469	0.0329	14.98	23.61	19.63	0.0186	5.7832
4	0.0145	0.0217	0.0168	18.72	25.164	21.95	0.0163	20.507

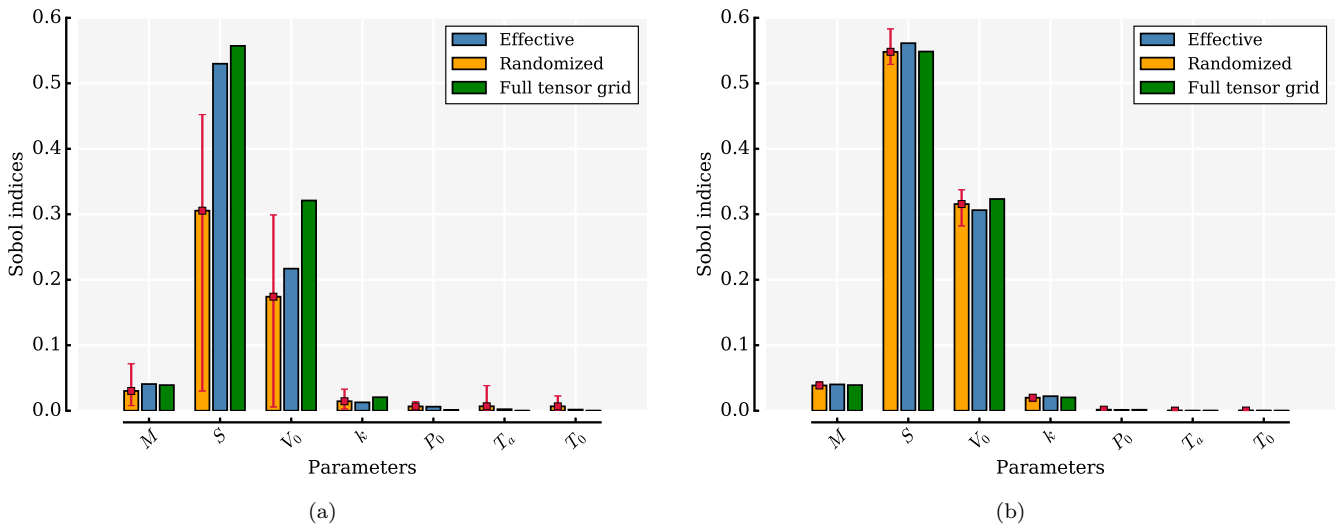


Figure 5: First order Sobol' indices for the seven parameters with $k = 4$ for (a) Total order basis with $n/l = 1.00$; (b) Total order basis with $n/l = 1.25$. The error bars indicate the minimum and maximum Sobol' indices from the mean (red dot) obtained from 20 repetitions of the randomized approach.

For all k across all n/l ratios, the effectively subsampled approach yields lower coefficient errors and lower condition numbers compared to the averaged randomized quadrature results. In particular, when $n/l = 1.0$ the ϵ values from effectively subsampled quadratures are 1-3 orders of magnitude below those of the corresponding randomized values. Randomized subsampling in general yields lower coefficient errors at n/l ratios greater than one. For a fixed k , as the n/l ratio is increased, both condition numbers and errors are found to drop for both approaches.

For completeness, we compare the Sobol' indices of both the random and effectively subsampled quadratures approaches in Figure 5. These indices are computed using the coefficient estimates obtained from the randomized, effectively subsampled and full tensor grid quadrature rule—using the approach in Sudret [37], which is detailed in the Appendix of this paper. These results highlight the point that even with an $n/l = 1.0$, effectively subsampled quadratures does yield results that are comparable to those obtained from the full tensor grid.

6 A problem where effectively subsampled quadratures fails

In this section we present an example that illustrates a limitation of our method. Consider the function

$$f(\zeta) = \frac{1}{(1 + 50(\zeta_1 - 0.9)^2 + 50(\zeta_2 + 0.9)^2)} \quad (34)$$

defined over $\mathcal{R} = [-1, 1]^2$. The contours of this function are shown in Figure 6(a) and its bi-variate Legendre polynomial approximation—obtained by evaluating the function at a tensor grid with 10 points in each direction and computing the coefficients with the corresponding tensor product integration rule—is shown in Figure 6(b). The function is relatively flat throughout most of its domain, but it exhibits a steep variation in the lower right hand corner around $(-0.9, -0.9)$.

In Figure 7 we plot the effectively subsampled quadrature approximation with different hyperbolic basis index sets that all have a maximum order of 9 in each direction. Figure 7(a-b) shows the results for a q factor of 0.3; (c-d) for a q factor of 0.5 and (e-f) for a q factor of 1.0 which is equivalent to a total order basis (For the definition of q see (4)). The tensor grid points in Figures (a,c,e) are given by red circular markers, while the effectively subsampled points—used for generating the polynomial approximation contours—are shown as green “x” markers.

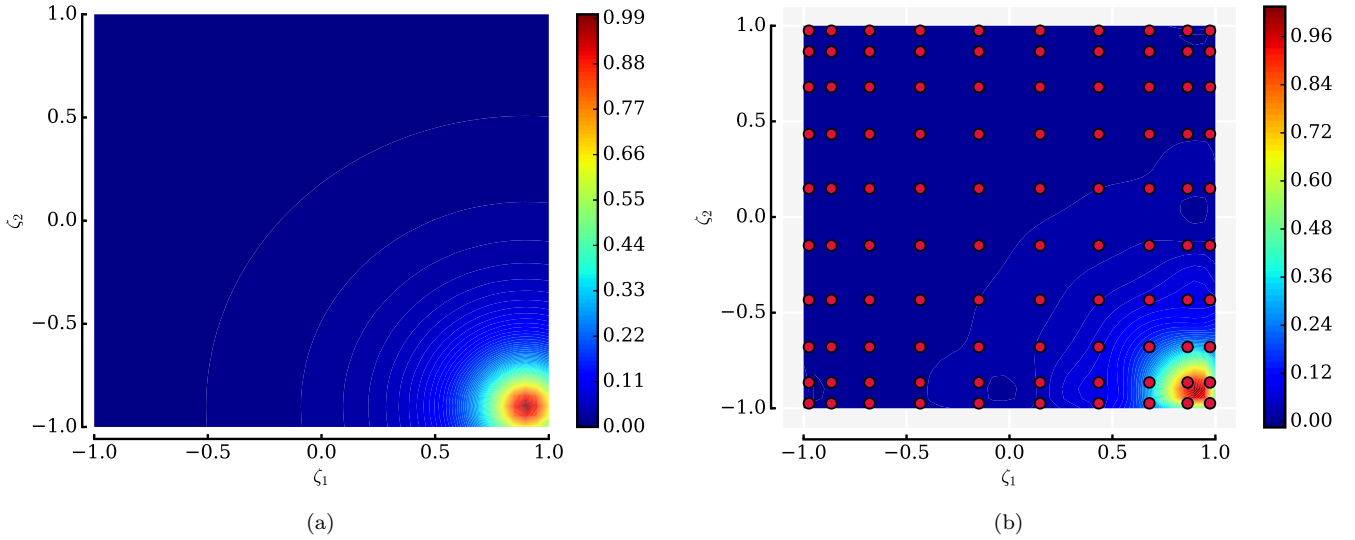


Figure 6: Contour plots for (34) with the real function in (a) and its tensor grid polynomial approximant in (b).

It is clear that in Figure 7(a) and to a certain extent in (c), the subsampled points do not lie in the vicinity of $(-0.9, 0.9)$, hence the polynomial approximation does a rather poor job of approximating the overall response of the function. This approximation is however improved in Figure 7(e-f) when using a total order basis, as the subsampled points do lie around $(-0.9, 0.9)$. This illustrates a limitation of effectively subsampled quadratures—i.e., they are based solely on $\rho(\zeta)$, the choice of the tensor grid and basis—and are invariant to the function response.

7 Conclusion

In this paper, we proposed a new sampling strategy for generating polynomial least squares approximations, titled *effectively subsampled quadratures*. This technique uses a QR column pivoting heuristic for subsampling points from a tensor grid. Numerical results showed the advantages of this technique compared with *randomized quadrature* subsampling. In future work we will investigate the incorporation of gradients.

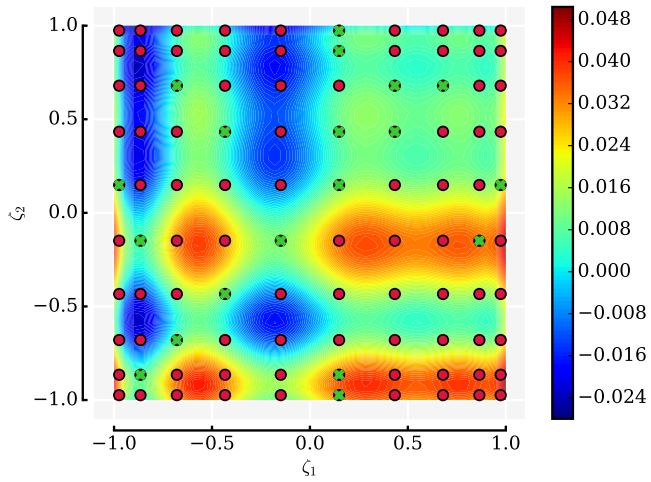
8 Acknowledgements

The first author would like to thank Paul Constantine for valuable discussions on polynomial least squares, Alireza Doostan for insightful exchanges on compressed sensing strategies with pseudospectral approximations and Tiziano Ghisu for his feedback on the randomized results. Thanks are also due to Gregorio Quintana-Ortí and Mario Arioli for their insights into QR with column pivoting. The authors also thank the reviewers for their suggestions and comments, which improved the overall quality of this manuscript.

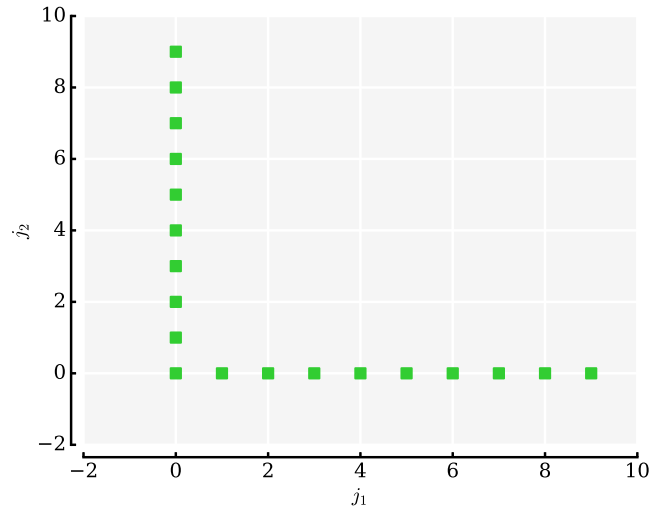
Appendix: Computing Sobol' indices

Consider the polynomial approximation of $f(s)$ provided earlier (see (8))

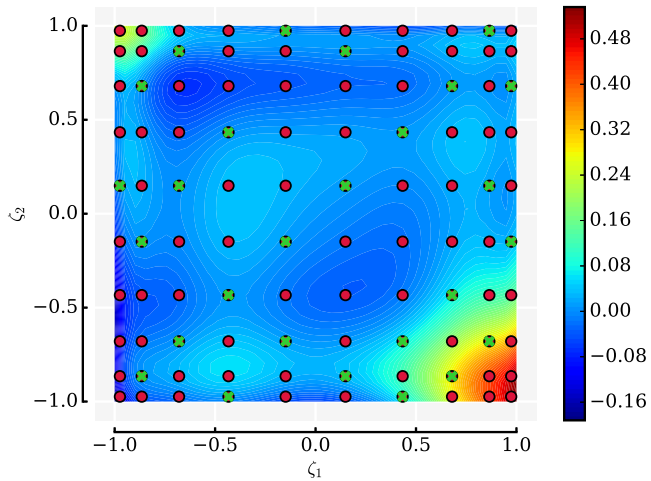
$$f(s) \approx g(s) = \sum_{j \in \mathcal{J}} x_j \psi_j(s). \quad (35)$$



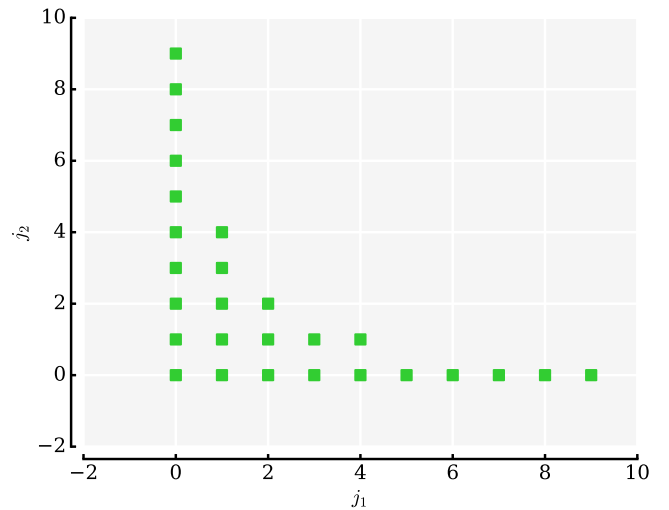
(a)



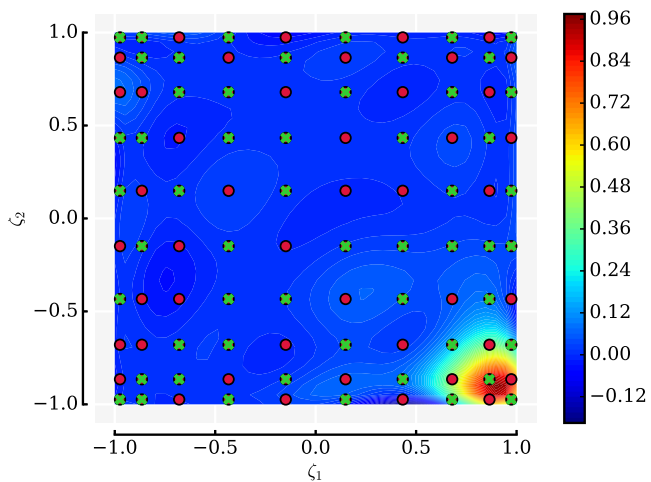
(b)



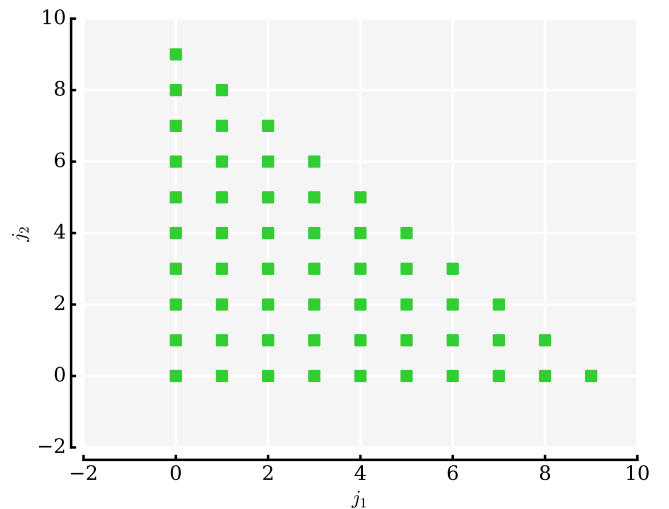
(c)



(d)



(e)



(f)

Figure 7: Effectively subsampled quadratures approximations of (34) with the chosen hyperbolic basis with: (a-b) $q = 0.3$; (c-d) $q = 0.5$; (e-f) $q = 1.0$.

By using the orthogonality properties of the polynomial basis, the mean μ and variance σ^2 of $g(s)$ can be expressed solely using the expansion's coefficients (see [36] page 210)

$$\mu = x_{\mathbf{0}}, \quad \sigma^2 = \sum_{j \in \mathcal{J}, j \neq \mathbf{0}} x_j^2. \quad (36)$$

Recall that Sobol' indices represent a fraction of the total variance that is attributed to each input variable (the first order Sobol' indices) or combinations thereof (higher order Sobol' indices). Let $\mathcal{J}_{\mathbf{s}}$ be the set of multi-indices that depend only on the subset of variables $\mathbf{s} = \{j_1, \dots, j_s\}$, i.e.,

$$\mathcal{J}_{\mathbf{s}} = \{\mathbf{j} \in \mathbb{N}^d : l \in \mathbf{s} \Leftrightarrow j_l \neq 0\}. \quad (37)$$

The first order partial variances σ_i^2 are then obtained by summing up the square of the coefficients in $\mathcal{J}_{\mathbf{s}}$

$$\sigma_i^2 = \sum_{j \in \mathcal{J}_i} x_j^2, \quad \mathcal{J}_i = \{\mathbf{j} \in \mathbb{N}^d : j_i > 0\}, \quad (38)$$

and the higher order variances $\sigma_{\{j_1, \dots, j_s\}}^2$ can be written as

$$\sigma_{\mathbf{s}}^2 = \sum_{j \in \mathcal{J}_{\mathbf{s}}} x_j^2, \quad \mathcal{J}_{\{j_1, \dots, j_s\}} = \{\mathbf{j} \in \mathbb{N}^d : l \in \mathbf{s} \Leftrightarrow j_l > 0\}. \quad (39)$$

The first and higher order Sobol indices are then given by

$$S_i = \frac{\sigma_i^2}{\sigma^2} \quad \text{and} \quad S_{\mathbf{s}} = \frac{\sigma_{\mathbf{s}}^2}{\sigma^2} \quad (40)$$

respectively (for further details see [37]).

References

- [1] BJÖRCK, Å. *Numerical methods in matrix computations*. Springer, 2015.
- [2] BLATMAN, G., AND SUDRET, B. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics* 230, 6 (2011), 2345–2367.
- [3] BOS, L., CALVI, J.-P., LEVENBERG, N., SOMMARIVA, A., AND VIANELLO, M. Geometric weakly admissible meshes, discrete least squares approximations and approximate Fekete points. *Mathematics of Computation* 80, 275 (Jan. 2011), 1623–1638.
- [4] BOS, L., DE MARCHI, S., SOMMARIVA, A., AND VIANELLO, M. Computing Multivariate Fekete and Leja Points by Numerical Linear Algebra. *SIAM Journal on Numerical Analysis* 48, 5 (2010), 1984.
- [5] BOUTSIDIS, C., MAHONEY, M. W., AND DRINEAS, P. An improved approximation algorithm for the column subset selection problem. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms* (2009), Society for Industrial and Applied Mathematics, pp. 968–977.
- [6] BROADBENT, M. E., BROWN, M., AND PENNER, K. Subset selection algorithms: Randomized vs. deterministic. *SIAM Undergraduate Research Online* 3 (2010).
- [7] BRUNO, L., CANUTO, C., AND FRANSOS, D. Stochastic aerodynamics and aeroelasticity of a flat plate via generalised polynomial chaos. *Journal of Fluids and Structures* 25, 7 (2009), 1158 – 1176.

- [8] BUNGARTZ, H. J., AND GRIEBEL, M. Sparse grids. *Acta numerica* 13 (2004), 147–269.
- [9] CHANDRASEKARAN, S., AND IPSEN, I. C. F. On rank-revealing factorisations. *SIAM Journal on Matrix Analysis and Applications* 15, 2 (1994), 592–622.
- [10] CHKIFA, A., COHEN, A., MIGLIORATI, G., NOBILE, F., AND TEMPONE, R. Discrete least squares polynomial approximation with random evaluations- application to parametric and stochastic elliptic pdes. *ESAIM: Mathematical Modelling and Numerical Analysis* (2015).
- [11] COHEN, A., DAVENPORT, M. A., AND LEVIATAN, D. On the stability and accuracy of least squares approximations. *Foundations of computational mathematics* 13, 5 (2013), 819–834.
- [12] CONSTANTINE, P. G., ELDRED, M. S., AND PHIPPS, E. T. Sparse pseudospectral approximation method. *Computer Methods in Applied Mechanics and Engineering* 229 (2012), 1–12.
- [13] DAX, A. A modified gram–schmidt algorithm with iterative orthogonalization and column pivoting. *Linear algebra and its applications* 310, 1 (2000), 25–42.
- [14] DOOSTAN, A., AND OWHADI, H. A non-adapted sparse approximation of pdes with stochastic inputs. *Journal of Computational Physics* 230, 8 (2011), 3015–3034.
- [15] ERNST, O. G., MUGLER, A., STARKLOFF, H., AND ULLMANN, E. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis* 46, 2 (2012), 317–339.
- [16] FREUD, G. *Orthogonal Polynomials*. Elsevier, 1971.
- [17] GAUTSCHI, W. *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press, USA, 2004.
- [18] GERSTNER, T., AND GRIEBEL, M. Numerical integration using sparse grids. *Numerical algorithms* 18, 3-4 (1998), 209–232.
- [19] GOLUB, G. H., KLEMA, V., AND STEWART, G. W. Rank degeneracy and least squares problems. Tech. rep., DTIC Document, 1976.
- [20] GOLUB, G. H., AND VAN LOAN, C. F. *Matrix computations*, vol. 3. JHU Press, 2012.
- [21] GU, M., AND EISENSTAT, S. C. Efficient algorithms for computing a strong rank-revealing qr factorization. *SIAM Journal on Scientific Computing* 17, 4 (1996), 848–869.
- [22] HAMPTON, J., AND DOOSTAN, A. Compressive sampling of polynomial chaos expansions: convergence analysis and sampling strategies. *Journal of Computational Physics* 280 (2015), 363–386.
- [23] HANSEN, P. C., PEREYRA, V., AND SCHERER, G. *Least squares data fitting with applications*. JHU Press, 2012.
- [24] HARBRECHT, H., PETERS, M., AND SCHNEIDER, R. On the low-rank approximation by the pivoted Cholesky decomposition. *Applied Numerical Mathematics* 62, 4 (Apr. 2012), 428–440.
- [25] KENETT, R., ZACKS, S., AND AMBERTI, D. *Modern Industrial Statistics: with applications in R, MINITAB and JMP*. John Wiley & Sons, 2013.
- [26] LUBINSKY, D. A survey of weighted polynomial approximation with exponential weights. *Surveys in Approximation Theory* 3 (2007), 1–105.

- [27] MIGLIORATI, G., NOBILE, F., VON SCHWERIN, E., AND TEMPONE, R. Approximation of quantities of interest in stochastic pdes by the random discrete l2 projection on polynomial spaces. *SIAM Journal on Scientific Computing* 35, 3 (2013), A1440–A1460.
- [28] MIGLIORATI, G., NOBILE, F., VON SCHWERIN, E., AND TEMPONE, R. Analysis of discrete l2 projection on polynomial spaces with random evaluations. *Foundations of Computational Mathematics* 14, 3 (2014), 419–456.
- [29] NARAYAN, A., JAKEMAN, J. D., AND ZHOU, T. A christoffel function weighted least squares algorithm for collocation approximations. *To appear: Mathematics of Computation* (2014). arXiv:1412.4305 [math].
- [30] PENG, J., HAMPTON, J., AND DOOSTAN, A. A weighted l_1 -minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics* 267 (2014), 92–111.
- [31] PLATTE, R., TREFETHEN, L. N., AND KUIJLAARS, A. Impossibility of fast stable approximation of analytic functions from equispaced samples. *SIAM Review* 53, 2 (2011), 308–318.
- [32] QUINTANA-ORTÍ, G., SUN, X., AND BISCHOF, C. H. A blas-3 version of the qr factorization with column pivoting. *SIAM Journal on Scientific Computing* 19, 5 (1998), 1486–1494.
- [33] SANDU, A., SANDU, C., AND AHMADIAN, M. Modeling multibody systems with uncertainties. part i: Theoretical and computational aspects. *Multibody System Dynamics* 15, 4 (2006), 369–391.
- [34] SESHADRI, P., AND PARKS, G. Effective-quadratures (EQ): Polynomials for computational engineering studies. *The Journal of Open Source Software*, 2(11). DOI: 10.21105/joss.00166.
- [35] SESHADRI, P., PARKS, G. T., AND SHAHPAR, S. Leakage uncertainties in compressors: The case of rotor 37. *Journal of Propulsion and Power* 31, 1 (2014), 456–466.
- [36] SMITH, R. C. *Uncertainty quantification: theory, implementation, and applications*, vol. 12. Siam, 2013.
- [37] SUDRET, B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety* 93, 7 (2008), 964–979. Bayesian Networks in Dependability.
- [38] TANG, G., AND IACCARINO, G. Subsampled gauss quadrature nodes for estimating polynomial chaos expansions. *SIAM/ASA Journal on Uncertainty Quantification* 2, 1 (2014), 423–443.
- [39] TAYLOR, M. A., WINGATE, B. A., AND BOS, L. P. A cardinal function algorithm for computing multivariate quadrature points. *SIAM Journal on Numerical Analysis* 45, 1 (2007), 193–205.
- [40] TEMPLJAKOV, V. N. Approximation of periodic functions of several variables with bounded mixed difference. *Mathematics of the USSR-Sbornik* 41, 1 (1982), 53.
- [41] THOMPSON, R. C. Principal submatrices IX: Interlacing inequalities for singular values of submatrices. *Linear Algebra and its Applications* 5, 1 (Jan. 1972), 1–12.
- [42] ZHANG, Y., AND SAHINIDIS, N. V. Uncertainty quantification in co2 sequestration using surrogate models from polynomial chaos expansion. *Industrial & Engineering Chemistry Research* 52, 9 (2012), 3121–3132.
- [43] ZHOU, T., NARAYAN, A., AND XIU, D. Weighted discrete least-squares polynomial approximation using randomized quadratures. *Journal of Computational Physics* 298 (2015), 787–800.