Stochastic Finite Element Methods for Transport Equations

by

Regina C. Almeida
A Class of Discontinuous Petrov-Galerkin Methods. Part II: Optimal Test Functions
by
L. Demkowicz and J. Gopalakrishnan
Stochastic Finite Element Methods
for Transport Equations

Regina C. Almeida *

In association with the J. T. Oden Faculty Fellowship Research Program

Institute for Computational Engineering and Sciences
University of Texas at Austin

Department of Computational Mechanics
Laboratório Nacional de Computação Científica - LNCC/MCT

June, 2009

Abstract

This text reviews some finite element methods for solving stochastic PDEs for transport problems, focusing on stochastic collocation methods. The main issues we will consider in this report are: Lagrange polynomials interpolation; isotropic Smolyak sparse grids; choice of cubature rules; error estimators and anisotropic grids.

*rcca@ices.utexas.edu; rcca@lncc.br
1 Problem definition

The focus of this work is to investigate the effects of input uncertainties in transport problems defined in a $d$-dimensional bounded domain $D \subset \mathbb{R}^d$ with a Lipschitz boundary $\partial D$ where Dirichlet boundary conditions are applied. Consider a complete probability space $(\Omega, U, \mathbb{P})$, where $\Omega$ is the set of all possible outcomes $w$, $U$ is the $\sigma$-algebra of subsets of $\Omega$ and $\mathbb{P} : U \to [0, 1]$ is the probability measure.

The transient transport of a stochastic scalar process in the presence of a divergence-free stochastic velocity field $\beta(x, w)$ in a medium with random non-negative diffusion coefficient $\epsilon(x, w)$ under a local reaction $r(x, w)$ can be defined as:

Find $\phi(x, t, w) : (D \times (0, T] \times \Omega) \to \mathbb{R}$ such that:

$$
\partial_t \phi + \beta \cdot \nabla \phi - \nabla \cdot \epsilon \nabla \phi + r \phi = f, \quad (x, t, w) \in (D \times (0, T] \times \Omega); (1)
\phi(x, t, w) = g(x, t, w), \quad (x, t, w) \in (\partial D \times (0, T] \times \Omega);
\phi(x, 0, w) = \phi_0(x, w), \quad (x, w) \in (D \times \Omega).
$$

$f(x, t, w) : (D \times (0, T] \times \Omega) \to \mathbb{R}$ is a source term and $g(x, t, w) : (\partial D \times (0, T] \times \Omega) \to \mathbb{R}$ is the stochastic Dirichlet boundary condition.

**Remark 1** It is assumed that the weak form for (1) complies with the requirements of the Lax-Milgram theorem. The following assumptions must be accomplished: $\exists r_0 > 0$ such that $-\frac{1}{2} \nabla \cdot \beta(x, w) + r(x, w) \geq r_0$; $\epsilon(x, w)$ is uniformly bounded and coercive. Thus, the Lax-Milgram theorem ensures existence and uniqueness of the solution.
Remark 2 In the following, we only consider stochastic processes which can be modeled as second order random parameters. This implies that the randomness is viewed as an additional dimension and is modeled with a Hilbert space \( L_2^p (\Omega; W(D)) \) of \( L^2 \) functions with respect to \( P \), where \( W(D) \) is a Banach space of functions \( z : D \to \mathbb{R} \).

\[
L_2^p (\Omega; W(D)) = \left\{ z(\mathbf{x}, w) : \Omega \to W(D) \mid z \text{ is strongly measurable and} \right. \\
\int_{\Omega} z(\mathbf{x}, w) z(\mathbf{x}, w) \, dP(\Omega) < \infty \left. \right\},
\]

with the corresponding norm

\[
\|z(\mathbf{x}, w)\|_0 = \left\{ \int_{\Omega} z(\mathbf{x}, w) z(\mathbf{x}, w) \, dP(\Omega) \right\}^{1/2}.
\]

Thus for two random variables \( z_1(\mathbf{x}, w) \) and \( z_2(\mathbf{x}, w) \), one has \( <z_1, z_2>=\int_{\Omega} z_1 z_2 dP(w) \). We also consider finite dimensional subspaces of \( L_2^p (\Omega) \), yielding the characterization of the random inputs into \( N \) random variables \( Y_i(w), i = 1, \ldots, N \).

The solution of (1) can be described by the random variables \( Y_i(w), i = 1, \ldots, N \). We assume that \( Y_i(w) \) are independent random variables with pdf \( \rho_i : \Gamma^i \to \mathbb{R} \) with bounded ranges \( \Gamma^i \). The joint probability density is then given by

\[
\rho(Y) = \prod_{i=1}^{N} \rho_i(Y_i), \quad \forall Y \in \Gamma, \quad \text{where} \quad \Gamma = \prod_{i=1}^{N} \Gamma^i \subset \mathbb{R}^N. \quad (2)
\]

Since \( N \) and \( d \) are the dimensionality of the truncated random space \( \Gamma \) and the physical space \( D \), respectively, the equation (1) is written in \((N+d)\)
\[ \partial_t \phi + \beta \cdot \nabla \phi - \nabla \cdot \epsilon \nabla \phi + r \phi = f, \quad (x, t, Y) \in (D \times (0, T] \times \Gamma); \tag{3} \]
\[ \phi (x, t, Y) = g (x, t, Y), \quad (x, t, Y) \in (\partial D \times (0, T] \times \Gamma); \]
\[ \phi (x, 0, Y) = \phi_0 (x, Y), \quad (x, Y) \in (D \times \Gamma). \]

For the sake of simplicity, let us first consider the stochastic steady state diffusion problem:

Find \( \phi (x, w) : (D, \Gamma) \to \mathbb{R} \) such that:

\[ L (\phi; x, Y) = - \nabla \cdot (\epsilon (x, Y) \nabla \phi (x, Y)) = f (x, Y), \quad (x, Y) \in (D \times \Gamma); \]
\[ \phi (x, Y) = 0, \quad (x, Y) \in (\partial D \times \Gamma). \tag{4} \]

A weak formulation for (4) reads: find \( \phi (x, Y) \in Z \subset L_2^p (\Gamma) \) such that

\[ \int_{\Gamma} L (\phi; x, Y) v (Y) \rho (Y) dY = \int_{\Gamma} f (x, Y) v (Y) \rho (Y) dY, \quad \forall v (Y) \in Z; x \in D. \tag{5} \]

Alternatively, let \( \phi (x, Y) \in X \times Z \), where \( X \) is an appropriate Hilbert space \( (X = H^1_0 (D)) \). The Galerkin weak formulation in the physical and parameter spaces reads: find \( \phi (x, Y) \in X \times Z \) such that

\[ \int_{\Gamma} \int_{D} L (\phi; x, Y) v (x, Y) \rho (Y) dx dY = \int_{\Gamma} \int_{D} f (x, Y) v (x, Y) \rho (Y) dx dY, \quad \forall v (x, Y) \in X \times Z. \tag{6} \]

**Remark 3** If \( E [\cdot] \) denotes the expected value, the previous weak formulation may be written as

\[ E \left[ \int_{D} L (\phi; x, Y) v (x, Y) dx - \int_{D} f (x, Y) v (x, Y) dx \right] = 0.\]
Remark 4 An auxiliary joint probability density function $\hat{\rho}(\textbf{Y})$ of $N$ independent random variables may be defined when $Y_i(w)$ are not independent random variables. The case when $\rho/\hat{\rho}$ is a smooth function was analyzed in [1].

In the following we review some finite element methods for solving stochastic transport problems. We focus on collocation methods. Also, an overview of past and recent developments in the stochastic finite element area, indicating future directions and addressing some open issues to be considered by the engineering community in the future can be found in [22].

2 The Stochastic Galerkin Methods

The spatial discretization is performed here by a standard finite element discretization in the usual manner by choosing a $J$-dimensional space $X_J \subset X$. Let $\{\chi_j(\textbf{x})\}_{j=1}^{J}$ denote a basis for $X_J$. Functions of the parameters may be discretized in the same way. The stochastic Galerkin methods are methods for which discretization with respect to the parameter space is also effected by using a Galerkin approach, i.e.,

- a $M$-dimensional space $Z_M \subset Z$ is chosen;
- $\{\Phi_m(\textbf{Y})\}_{m=1}^{M}$ is a (stochastic) basis for $Z_M$.

As approximations are defined in $X_J \otimes Z_M$, we seek an approximate solution of the stochastic partial differential equation (SPDE) of the form

$$\phi_{JM} = \sum_{j=1}^{J} \sum_{m=1}^{M} c_{jm} \chi_j(\textbf{x}) \Phi_m(\textbf{Y}) \in X_J \times Z_M.$$
In particular, we may use Lagrange interpolants for building the parameter basis. Let us denote by \( y = (Y_1, \ldots, Y_N) \) any point in the random space \( \Gamma \subset \mathbb{R}^N \). Given a set of points \( \Theta = \{ y_i \}_{i=1}^{M_{LI}} \), the Lagrange interpolating polynomials for these points are such that

\[
L_k(y_l) = \delta_{kl} \quad \text{for all} \quad k, l \in \{1, \ldots, M_{LI}\}.
\]

A \( M_{LI} \)-dimensional space \( Z_{M_{LI}} \subset Z \) may be built such that \( Z_{M_{LI}} = \text{span} \{ L_m(Y) \}_{m=1}^{M_{LI}} \) and the solution of the SPDE takes the form

\[
\phi_{JM_{LI}} = \sum_{j=1}^{J} \sum_{m=1}^{M_{LI}} c_{jm} \chi_j(x) L_m(Y) \in X_J \times Z_{M_{LI}}.
\]

The integration on the parameter space may be performed by using a cubature rule defined by \( \{ \hat{y}_r \}_{r=1}^{R} \) points and respective weights \( \{ W_i \}_{i=1}^{M_{LI}} \) yielding

\[
\sum_{r=1}^{R} W_r \rho(\hat{y}_r) L_{m'}(\hat{y}_r) \int_D L \left( \sum_{j=1}^{J} \sum_{m=1}^{M_{LI}} c_{jm} \chi_j(x) L_m(\hat{y}_r), \hat{y}_r \right) \chi_{j'}(x) \, dx
\]

\[=
\sum_{r=1}^{R} W_r \rho(\hat{y}_r) L_{m'}(\hat{y}_r) \int_D f(x, \hat{y}_r) \chi_{j'}(x) \, dx,
\quad (7)
\]

\[\forall j' \in \{1, \ldots, J\}; m' \in \{1, \ldots, M_{LI}\}.
\]

If the interpolating points \( \{ y_i \}_{i=1}^{M_{LI}} \) are chosen to be the same as the cubature points \( \{ \hat{y}_r \}_{r=1}^{R} \) we have

\[
L_m(\hat{y}_r) = \delta_{mr}, \quad \forall r, m \in \{1, \ldots, R = M_{LI}\}.
\]

As a result, the discretized SPDE reduces to

\[
\int_D L \left( \sum_{j=1}^{J} c_{jm} \chi_j(x), \hat{y}_r \right) \chi_{j'}(x) \, dx = \int_D f(x, \hat{y}_r) \chi_{j'}(x) \, dx,
\quad (8)
\]

\[\forall j' \in \{1, \ldots, J\}; r \in \{1, \ldots, R = M_{LI}\}.
\]
yielding a total uncoupling in parameter space. This is the stochastic collocation method.

3 The Generalized Polynomial Chaos (gPC) \[2, 34, 33]\n
It is worth mentioning a particular case of Stochastic Galerkin Methods yielded by choosing orthogonal polynomials as a basis for the parameter space. To this end, let us define one-dimensional subspaces for \( L^2_{\rho_i}(\Gamma^i) \equiv L^2_{\rho_i}(\Gamma^i) \) as

\[
Z_i^{d,i} \equiv \{ v : \Gamma^i \rightarrow \mathbb{R}, v \in \text{span} \{ \varphi_m(Y_i) \}_{m=1}^{m_i} \} , \quad i = 1, \ldots, N,
\]

where \( \varphi_{d,i}(Y_i) \equiv \{ \varphi_m(Y_i) \}_{m=1}^{m_i} \) is a set of orthogonal polynomials satisfying the orthogonality condition

\[
\int_{\Gamma^i} \varphi_m(Y_i) \varphi_n(Y_i) \rho_i(Y_i) dY_i = h_m^2 \delta_{mn} = \left\{ \int_{\Gamma^i} [\varphi_m(Y_i)]^2 \rho_i(Y_i) dY_i \right\} \delta_{mn},
\]

where \( \delta_{mn} \) is the Kronecker delta function and \( h_m^2 \) is a normalization factor (with proper scaling, the basis can be normalized and \( h_m^2 = 1 \)).

Remark 5 The pdf \( \rho_i(Y_i) \) defines the type of orthogonal polynomials \( \{ \varphi_m(Y_i) \} \).
For example, for Gaussian distributed random variables, its pdf defines Hermite polynomials and this is the classical polynomial chaos method. If \( Y_i \) is a uniformly distributed random variable in a bounded interval, its pdf is a constant and define Legendre polynomials. See \[34\] for more details.
The corresponding multivariate \((N\text{-variate})\) orthogonal polynomial space of total degree at most \(K\) in \(\Gamma\) is then defined as

\[
Z_N^P \equiv \bigotimes_{|d| \leq P} Z_i^{d_i},
\]

where the tensor product is over all possible combinations of the multi-index \(d = (d_1, \ldots, d_N) \in \mathbb{N}_+^N\) satisfying \(|d| = \sum_{i=1}^N d_i \leq K\) (complete polynomials). Let \(\{\Phi_m(Y)\}\) be the \(N\)-variate orthogonal polynomials from \(Z_N^K\). They are built as products of a sequence of univariate polynomials in each direction of \(Y_i, i = 1, \ldots, N\), as

\[
\Phi_m(Y) = \varphi_{d_1}(Y_1) \cdots \varphi_{d_N}(Y_N), \quad d_1 + \ldots + d_N \leq K
\]

and the number of basis functions is

\[
\dim(Z_N^K) = \binom{N + K}{N} = \frac{(N + K)!}{N!K!}.
\]

**Remark 6** With this construction we have

\[
E[\Phi_k(Y) \Phi_j(Y)] = \int_{\Gamma} \Phi_k(Y) \Phi_j(Y) \rho(Y) dY = \delta_{kj},
\]

\(\forall 1 \leq k, j \leq \dim(Z_N^K)\).

This is the reason why \(Y_i(w)\) must be independent random variables, i.e. \(\rho(Y) = \prod_{i=1}^N \rho_i(Y_i)\).

**Remark 7** Another way to construct a subspace for \(L_2^p(\Gamma)\) is based on full tensor products of the one-dimensional polynomial spaces with fixed highest polynomial orders in all directions (dimensions). In this case, \(W_N^K \equiv \bigotimes_{i} Z_i^{d_i}\) such that \(\max_i d_i = K\), yielding \(\dim(W_N^K) = (K + 1)^N\). \(W_N^K\) and \(Z_N^K\) are denoted by full and complete tensor polynomial spaces.
Remark 8 The gPC presents high resolution and fast convergence for low to moderate $N$. For large value of $N >> 1$, the growing of basis functions reduces its efficiency. In this case, anisotropic full tensor grids have been developed in the sake of avoiding the curse of dimensionality [2].

4 The stochastic collocation method [25, 19, 33, 1, 32]

It is an attempt to combine the strength of the high resolution of gPC and the ease of (non-intrusive) Monte Carlo methods. As mentioned before, it is based on (Lagrange) polynomials interpolations in the multidimensional random space and choosing the interpolation points equal to the cubature points used to integrate on the parameter space.

Let us denote by $y = (Y_1, \ldots, Y_N)$ any point in the random space $\Gamma \subset \mathbb{R}^N$, by $\Pi^N$ the space of all $N$-variate polynomials with real coefficients, and by $\Pi^N_K$ the subspace of polynomials of total degree at most $K$ (note that $Z^N_K$ is a subspace of $\Pi^N_K$).

Given a set of distinct collocation points $\Theta = \{y_i\}_{i=1}^M$, the Lagrange interpolating polynomials for these points are such that

$$L_k(y_l) = \delta_{kl} \quad \text{for all } k, l \in \{1, \ldots, M\}.$$ 

The Lagrange interpolation of a smooth function $F : \mathbb{R}^N \rightarrow \mathbb{R}$ reads: find $I(F) \in V_I \subset \Pi_N$ such that $I(F)(y_i) = F(y_i), \forall i = 1, \ldots, M$. The interpolation polynomial can be expressed by using Lagrange interpolation
polynomials, i.e.,
\[ \mathcal{I}(\mathcal{F})(\mathbf{y}) = \sum_{k=1}^{M} \mathcal{F}(\mathbf{y}_k) L_k(\mathbf{y}), \]
where \( L_k(\mathbf{y}) \in V_I \).

**Remark 9** The Lebesgue theorem states that
\[
\| \mathcal{F}(\mathbf{y}) - \mathcal{F}^*(\mathbf{y}) \|_\infty \leq \| \mathcal{F}(\mathbf{y}) - \mathcal{I}(\mathcal{F})(\mathbf{y}) \|_\infty \leq (1 + \Lambda) \| \mathcal{F}(\mathbf{y}) - \mathcal{F}^*(\mathbf{y}) \|_\infty,
\]
(10)
where \( \mathcal{F}^*(\mathbf{y}) \) is the best approximating polynomial and \( \Lambda \) is the Lebesgue constant. The estimation of \( \Lambda \) depends on the particular choice of the interpolation set \( \Theta \).

By denoting
\[
\hat{\phi}(\mathbf{y}) \equiv \mathcal{I}(\phi)(\mathbf{y}) = \sum_{k=1}^{M} \phi(\mathbf{y}_k) L_k(\mathbf{y}),
\]
(11)
the collocation procedure to solve the stochastic elliptic problem (4) requires
\[
R\left( \hat{\phi}(\mathbf{y}) \right)|_{\mathbf{y}_k} = 0, \quad \forall k = 1, \ldots, M,
\]
(12)
where \( R(\phi) = \mathcal{L}(\phi) - f \) is the residual of (4). By using (11) in (12) we obtain
\[
\mathcal{L}(\phi; \mathbf{x}, \mathbf{y}_k) = f(\mathbf{x}, \mathbf{y}_k), \quad \mathbf{x} \in D; \\
\phi(\mathbf{x}, \mathbf{y}_k) = 0, \quad \mathbf{x} \in \partial D.
\]
(13)
Hence, the stochastic collocation method is equivalent to solving \( M \) deterministic problem (13) at each point \( \mathbf{y}_k, k = 1, \ldots, M, \) in a given set \( \Theta \). The deterministic problem (13) is solved here by using standard finite element
methods. We define the finite dimension space $X_h$ of dimension $J$ such that $X_h \subset X = H^1_0(D)$. We assume that $X_h$ contains continuous piecewise polynomials defined on a regular triangulation $\tau_h$ that have a maximum characteristic length $h > 0$. For a given function $\varphi \in H^2(D) \cap H^1_0(D)$, the well known approximation property holds for piecewise linear finite element subspaces:

$$\min_{\phi \in X_h} \| \varphi - \phi \|_{H^1_0(D)} \leq c \ h \ \| \varphi \|_{H^2(D)},$$

(14)

where $c$ is independent of $h$. Moreover, for all $\tilde{\phi} \in H^1_0(D)$, we also assume that there exists a finite element operator $P_h : X \to X_h$ with the optimality condition

$$\| \tilde{\phi} - P_h \tilde{\phi} \|_{H^1_0(D)} \leq C \ \min_{\phi \in X_h} \| \tilde{\phi} - \phi \|_{H^1_0(D)},$$

(15)

where $C$ is independent of $h$. Hence, the stochastic collocation method requires the evaluation of the approximate values $P_h \phi(x, y_k) = \phi_h(x, y_k)$ on a suitable set of points $\Theta$. Some choices for $\Theta$ will be discussed in the following. To simplify notation, we will adopt $\phi_h(x, y_k) \equiv \phi(x, y_k)$ henceforth.

Once the numerical solutions of (13) are obtained at all collocation points $y_k$, the statistics of the random solution can be evaluated as

$$E \left[ \hat{\phi}(x) \right] = \sum_{k=1}^{M} \phi(x, y_k) \int \Gamma L_k(y) \rho(y) \, dy,$$

requiring the explicit knowledge of each $L_k(y)$. To avoid this, we may choose $\Theta$ to be the same as a cubature point set, yielding

$$E \left[ \hat{\phi}(x) \right] = \sum_{k=1}^{M} \phi(x, y_k) \, w_k,$$

where $\{w_k\}$ are the weights of the cubature rule.
Remark 10  The stochastic collocation method can easily deal either with unbounded random variables or a diffusivity coefficient with unbounded second moment (see [1] and Appendix 1 for more details).

Remark 11  Both the stochastic collocation method and the gPC were extended to multi-element approach in [10] and [27], respectively. Given a stochastic input with respectively probability measure, this approach basically consists in decomposing the random space into smaller elements. In each element, a new random variable with respect to a conditional probability density function is defined. Subsequently, either the collocation or gPC methods are implemented element by element. The convergence rate were verified numerically in a variety of examples, showing that this approach is especially efficient for solving problems related to long-term integration and stochastic discontinuities.

5 Choices of collocation points

The collocation points set should be chosen with the fewest possible number of points under a prescribed accuracy requirement for computing the multivariate integral (7). The bounded random space is a $N$–hypercube such that $\Gamma = [-1, 1]^N$. We then consider the integration of functions $f$ from a function class $\mathcal{F}$ over the unit hypercube $\Gamma$,

$$\mathcal{I}^NF := \int_\Gamma f(x,y)\,dy \approx \mathcal{Q}_N^N = \sum_{j=1}^M f(x,y_j)\,w_j.$$ (16)

Random variables $Y^i$ with bounded support $[a^i, b^i]$ can always be mapped to $[-1, 1]$, and any domain $\Gamma^* = \prod_{i=1}^N [a^i, b^i]$ to $[-1, 1]^N$.  

12
The cubature formulas for integrating (16) are obtained by a sequence of formulas of level \( l \in \mathbb{N}_+ \), \( M = n^N_l \) and \( n^N_l < n^N_{l+1} \), using the weights \( w_j \) and abscissas \( \Theta^N_l = \{y_j\}^{n^N_l}_{j=1} \). The cubature formulas are nested if \( \Theta^N_l \subset \Theta^N_{l+1} \).

The error \( E^N_l = I^N f - Q^N_l \) must be measured in appropriate norms and may be obtained by assuming certain regularity requirements. Notice that, as the interpolation points are chosen to be the same of the cubature points for the collocation method, we may approach this problem from the interpolation point of view.

Conventional cubature formulas are often limited by the computing cost, and its respective lower bounds, that grows exponentially with the dimension of the problem (curse of dimension) [13]. In the following, we review some algorithms. Special attention is given to Smolyak’s sparse grids, which can overcome the curse of dimension to a certain extent.

### 5.1 Tensor product of one-dimensional sets

In this approach, multivariate quadrature formulas are constructed using tensor products of suitable one-dimensional formulas. Let \( l \in \mathbb{N}_+ \) and \( \Theta^l_1 \equiv \{y^l_1, \ldots, y^l_{m_l}\} \subset [-1, 1] \) so that \( n^l_1 = m_l \). Consider a sequence of one-dimensional quadrature formulas for a univariate function \( f(x, y) \) defined by

\[
Q^l_1 f = \sum_{k=1}^{m_l} f(x, y^l_k) \ w^l_k.
\]  

(17)

Define now \( Q^0_1 = 0 \) and the differential (incremental) quadrature formula by

\[
\Delta^l_1 f = (Q^l_k - Q^l_{k-1}) \ f.
\]  

(18)
Then
\[ Q_l f = \sum_{k \leq l} \Delta_k f = Q_{l-1} f + \Delta_l f, \]
meaning that to extend the integration from level \( l - 1 \) to \( l \), one has only to evaluate the differential formula. By choosing appropriate nested points for the one dimensional formula, like Chebyshev and Gauss-Lobatto points, one only has to evaluate the function at the grid points that are unique to level \( l \), defined as
\[ \Theta_{1}^{l \Delta} = \Theta_{1}^{l} \setminus \Theta_{1}^{l-1}. \]

**Example:** For formulas based on the extrema of the Chebyshev polynomials (Clenshaw-Curtis formulas \([26]\)), the quadrature points are non-equidistant and given by
\[ y_1^l = 0; \quad m_1 = 1 \quad \text{and} \quad y_j^l = -\cos \frac{\pi (j-1)}{m_l-1}, j = 1, \ldots, m_l; \quad m_l \geq 2. \]

In order to obtain nested points such that \( \Theta_{1}^{l-1} \subset \Theta_{1}^{l} \), one chooses
\[ m_1 = 1 \quad \text{and} \quad m_l = 2^{l-1} + 1 \quad \text{for} \quad l \geq 2. \]

The polynomial degree of exactness is \( K = m_l - 1 \) and for functions \( f \) belonging to
\[ \mathcal{F}^r := \left\{ f : [-1, 1] \to \mathbb{R} \mid \left\| \frac{\partial^\beta \varphi}{\partial y^\beta} \right\|_\infty < \infty, \ \beta \in \mathbb{N}_0, \ \beta \leq r \right\}, \]
the approximation error is \([13]\)
\[ |E_l^1| = O \left( 2^{-l r} \right) = O \left( (n_1^1)^{-r} \right). \]

Figure 1 shows a sequence of mesh points for the one-dimensional Clenshaw-Curtis formulas. In this figure, \( q = l - 1 \) indicates the interpolation level.
Figure 1: Smolyak 1D grid for $0 \leq q \leq 4$.

For the multivariate case $N > 1$, a multi-index vector $l = (l_1, \ldots, l_N) \in \mathbb{N}_+^N$ is defined, where $l_k$ represents the level of interpolation along the $k-$th direction. For each $f(x, y)$, the full tensor product cubature formula is given by

$$I_N f \approx Q_l^N f = (Q_{l_1}^1 \otimes \ldots \otimes Q_{l_N}^1) f = \sum_{j=1}^N \sum_{k_j=1}^{l_j} (\Delta_{k_1}^1 \otimes \ldots \otimes \Delta_{k_N}^1) f,$$

which requires $n_l^N = m_{l_1} \ldots m_{l_N}$ nodal points (or function evaluations). For example, if $m_{l_1} = \ldots = m_{l_N} = m$, the total number of points is $n_l^N = m^N$. This number grows quickly in high dimensions (curse of dimension).

**Example:**[5] By the full tensor product approach, a 2D rule is made by taking the product between two 1D rules. The number of points in a product
grid is the product of the sizes of the 1D rules. The resulting rule captures monomials up to \((y_1)^{K_1} (y_2)^{K_2}\), where \(K_1\) and \(K_2\) are the individual accuracies in each direction and \(K_1 + K_2\) is the monomial degree. The Pascal triangle for the product of two 1D rules of degree 4 is shown below. One may see that a rule of degree 4 must capture the monomials \((y_1)^4\), \((y_1)^3 (y_2)^1\), \((y_1)^2 (y_2)^2\), \((y_1)^1 (y_2)^3\), \((y_2)^4\), and all smaller ones (above the line). The monomials under the line are also captured although not aimed by the 4 degree rule.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>((y_1)^2)</td>
<td>((y_1)^2 (y_2))</td>
<td>((y_1) (y_2)^2)</td>
<td>((y_2)^2)</td>
<td>((y_1)^2 (y_2)^2)</td>
<td>((y_1) (y_2)^3)</td>
<td>((y_2)^3)</td>
</tr>
<tr>
<td>3</td>
<td>((y_1)^3)</td>
<td>((y_2))</td>
<td>((y_1)^3 (y_2))</td>
<td>((y_1)^2 (y_2)^2)</td>
<td>((y_1) (y_2)^3)</td>
<td>((y_2)^3)</td>
<td>((y_1)^2 (y_2)^3)</td>
<td>((y_1) (y_2)^4)</td>
<td>((y_2)^4)</td>
</tr>
<tr>
<td>4</td>
<td>((y_1)^4)</td>
<td>((y_2))</td>
<td>((y_1)^3 (y_2))</td>
<td>((y_1)^2 (y_2)^2)</td>
<td>((y_1) (y_2)^3)</td>
<td>((y_2)^3)</td>
<td>((y_1)^2 (y_2)^3)</td>
<td>((y_1) (y_2)^4)</td>
<td>((y_2)^4)</td>
</tr>
</tbody>
</table>

The previous example indicates that multivariate quadrature formulas with same accuracy may be built by using some suitable combination of the tensor products points. This is the role of sparse grids.

### 5.2 Sparse grids

Smolyak (1963) developed special combinations of tensor product formulas \(A_{q,N}(f)\) which lead to almost (up to logarithmic factors) optimal error rates, as soon as the one-dimensional formulas possess such property. They are based on the following assumptions:

\[\text{This technique is also known under many different names such as Boolean method, discrete blending method, hyperbolic cross points or sparse grid methods.}\]
• only products with relatively small number of points are used;

• the interpolation property for \( N = 1 \) is preserved for \( N > 1 \).

These requirements reduce substantially the number of points when \( N \geq 10 \). Defining \( q \in \mathbb{N} \) as the formula level and \( \mathbf{l} = (l_1, \ldots, l_N) \in \mathbb{N}_+^N \), the Smolyak’s formulas are given by

\[
\mathcal{A}_{q,N}(f) = \sum_{|\mathbf{l}| \leq q} (\Delta^1_{l_1} \times \cdots \times \Delta^1_{l_N})(f) = \mathcal{A}_{q-1,N}(f) + \sum_{|\mathbf{l}| = q} (\Delta^1_{l_1} \times \cdots \times \Delta^1_{l_N})(f),
\]

where \( |\mathbf{l}| = l_1 + \cdots + l_N \), \( \mathcal{A}_{N-1,N} = 0 \) and \( K = q - N \) is the order of interpolation. Alternatively, Smolyak’s formula can be written in terms of \( Q^1_l \) in the following way

\[
\mathcal{A}_{q,N}(f) = \sum_{q-N+1 \leq |\mathbf{l}| \leq q} (-1)^{q-|\mathbf{l}|} \binom{N - 1}{q - |\mathbf{l}|} (Q^1_{l_1} \times \cdots \times Q^1_{l_N})(f)
\]
or

\[
\mathcal{A}_{q,N}(f) = \sum_{p=0}^{N-1} (-1)^p \binom{N - 1}{p} \sum_{|\mathbf{l}| = q-p} (Q^1_{l_1} \times \cdots \times Q^1_{l_N})(f).
\]

Thus, the Smolyak algorithm builds the interpolation function by adding a combination of one-dimensional functions of level \( l_k, k = 1, \ldots, N \), with the constraint that the total sum (\( |\mathbf{l}| \)) across all dimensions is between \( K+1 \) and \( K + N \) (or \( q - N + 1 \) and \( q \)). The multivariate index \( \mathbf{l} \) may also be defined by the index set

\[
X(q,N) := \left\{ \mathbf{l} \in \mathbb{N}_+^N, \ \mathbf{l} \geq \mathbf{1} : \sum_{k=1}^N l_k - 1 \leq q \right\}.
\]
To compute $A_{q,N}(f)$ it is only necessary to know the function values at the sparse grid

$$H_{q,N} = \bigcup_{q-N+1 \leq |l| \leq q} (\Theta_{l_1}^1 \times \ldots \times \Theta_{l_N}^1),$$

where $\Theta_{l_i}^1 = \{y_{l_i}^1, \ldots, y_{m_i}^1\} \subset [-1,1]$ denotes the set of points used by $Q_{l_i}^1$.

As mentioned before, by choosing appropriate one-dimensional formulas, one can ensure that the sets of points $H_{q,N}$ are nested ($H_{q,N} \subset H_{q+1,N}$).

Thus, to extend the order of interpolation from $K-1$ to $K$ in $N$ dimensions, one only needs to evaluate the function at the differential nodes $\Delta H_{q,N}$ given by

$$\Delta H_{q,N} = \bigcup_{|l|=q} (\Theta_{l_1}^{1:\Delta} \times \ldots \times \Theta_{l_N}^{1:\Delta}),$$

where $\Theta_{l}^{1:\Delta}$ is defined in (19).

**Example:** The Smolyak’s sparse grid for $N=2$ and $q=4$ ($K=2$) is constructed by the formula:

$$A_{4,2} = \sum_{3 \leq |l| \leq 4} (-1)^{4-|l|} \begin{pmatrix} 1 \\ 4 - |l| \end{pmatrix} (Q_{l_1}^1 \otimes Q_{l_2}^1).$$

This means that the grid is obtained by assembling 9 nested grids under the following combination:

$$A_{4,2} = - Q_1^1 \otimes Q_1^1 - Q_3^1 \otimes Q_2^1 - Q_2^1 \otimes Q_3^1 - Q_1^1 \otimes Q_1^1 + Q_2^1 \otimes Q_1^1 + Q_1^1 \otimes Q_2^1 + Q_3^1 \otimes Q_1^1 + Q_2^1 \otimes Q_1^1 + Q_1^1 \otimes Q_3^1.$$

The final grid with 65 points, shown in figure 2, is generated by adding each partial grid points, getting rid of repeated points,

$$65 \text{ points} \quad \Rightarrow \quad (9 \times 1) + (5 \times 3) + (3 \times 5) + (1 \times 9) + (17 \times 1) + (9 \times 3) + (5 \times 5) + (3 \times 9) + (1 \times 17).$$
Regarding collocation methods, it was suggested in [33] to use Smolyak’s formulas that are based on polynomial interpolation at the extrema of the Chebyshev polynomials although other rules may also be used. In this case, $\mathcal{A}_{q,N}$ is exact for all polynomials in $\Pi^N_K$. According to [21], the total number of points required in this case satisfies the following bound:

$$ N \left(2^{K+N} - 1\right) \leq n_K^N \leq (2eN)^{K+N} \min \{K + N + 1, 2eN\}. \quad (24) $$

If $N \gg 1$, then

$$ n_K^N = \dim (\mathcal{A}_{q,N}) \sim \frac{2^K}{K!} N^K, \quad K \text{ fixed.} $$

Figure 2: Smolyak 2D grid for $q = 4$. 
This number of nodes is larger than the number of expansion terms of a stochastic Galerkin method in the complete polynomial space $Z_N^P$ (9). For large $N$, the complete polynomial space uses about $N^K/K!$ points. Thus, the sparse grid uses about $2^K$ times as many function values [33]. However, as $N$ grows ($N \geq 10$), the number of nodes for the Smolyak sparse grid is significantly less than that required by the full tensor product. Table 1 shows the evolution of the number of points as a function of the dimension of the probabilistic space and the cubature level.

Table 1: Number of nodes for the Smolyak sparse grid.

<table>
<thead>
<tr>
<th>$q \backslash N$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>13</td>
<td>25</td>
<td>41</td>
<td>61</td>
<td>85</td>
<td>113</td>
<td>145</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>29</td>
<td>69</td>
<td>137</td>
<td>241</td>
<td>389</td>
<td>589</td>
<td>849</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>65</td>
<td>177</td>
<td>401</td>
<td>801</td>
<td>1457</td>
<td>2465</td>
<td>3937</td>
</tr>
</tbody>
</table>

**Remark 12** Bathelmann et al. [3] have proved that the multivariate Smolyak’s formula is actually interpolatory whenever nested points are used and the one-dimensional has this property.

**Remark 13** The formula $A_{q,N}$ is exact on $E_{K,N} = \sum_{|u| = K+N} \left( \Pi_{m_1-1}^{1} \otimes \ldots \otimes \Pi_{m_N-1}^{1} \right)$.

**Remark 14** Since the number of points in each direction is equal, all the stochastic dimensions are treated equally (isotropic grid). However, many
practical problems exhibit highly anisotropic behavior, whose specific structure is usually not known a priori. Some procedures to develop adaptive sparse grid are to assess the stochastic dimensions according to some error estimation. This subject will be addressed later.

5.3 Other cubature rules

Multivariate integration is still an active research field and there are many choices of one-dimensional formulas. In [26], the Gauss and Clenshaw-Curtis quadrature formulas are compared with respect to convergence. That work does not address multiple dimensions and it essentially shows that the two formulas are about equally accurate for most integrands, although a factor-of-2 advantage in efficiency for finite number of points is proved for Gauss quadrature. In [21], the analysis of Smolyak’s sparse grid stochastic collocation method is presented when the sparse tensor product is built using either Clenshaw-Curtis or Gaussian abscissas. Both cubature rules presented similar convergence rates, with the Clenshaw-Curtis rate slightly better. On the other hand, Xiu and Hesthaven [33] showed that Stroud’s cubature of degrees 2 and 3 are highly efficient in high-dimensional spaces for relatively low-order accuracy requirement in the solution statistic due to the small number of function evaluations. The Stroud’s cubature (1963) was developed to interpolate harmonic functions by generalizing spherical product formulas based on Gauss-Lobatto and Gauss-Radau points. It can be summarized as in the following.

Let $B_{N+1}$ denote the unit ball in Euclidean $(N + 1)$-space with boundary $S_N$. If $f$ is a continuous real-valued function on $S_N$, the $f$ has a unique
harmonic extension on $B_{N+1}$. Stroud developed interpolation formulas for the value of the extension at an arbitrary point in the interior of $B_{N+1}$. This is done by means of a finite number of values of $f$ at $S_N$ in such a way that the interpolation is exact for all harmonic polynomials belonging to $H_K(B_{N+1})$, where $H_K(B_{N+1})$ is the set of all polynomials of degree $\leq K$ satisfying the homogeneous elliptic problem defined in $B_{N+1}\setminus S_N$ with Dirichlet boundary conditions. For $N$-hypercube spaces $[-1,1]^N$, Stroud constructed a set of cubature points with $(N+1)$-point, that is accurate for multiple integrals of polynomials of degree 2, and a set of degree 3 with $2N$-point with unity weight function. They are constructed as:

- **degree 2 (Stroud-2):** set of points $\{Y_k\}_{k=0}^N$ such that
  $$Y_k^{2r-1} = \sqrt{\frac{2}{3}} \cos \frac{2rk\pi}{N+1}, \quad Y_k^{2r} = \sqrt{\frac{2}{3}} \sin \frac{2rk\pi}{N+1}, \quad r = 1, 2, \ldots, [N/2],$$
  where $[N/2]$ is the greatest integer not exceeding $N/2$. If $N$ is odd $Y_k^N = (-1)^k / \sqrt{3}$;

- **degree 3 (Stroud-3):** set of points $\{Y_k\}_{k=1}^{2N}$ such that
  $$Y_k^{2r-1} = \sqrt{\frac{2}{3}} \cos \frac{(2r-1)k\pi}{N}, \quad Y_k^{2r} = \sqrt{\frac{2}{3}} \sin \frac{(2r-1)k\pi}{N}, \quad r = 1, 2, \ldots, [N/2],$$
  and $Y_k^N = (-1)^k / \sqrt{3}$ if $N$ is odd.

Figures 3(a-b) show the sparse grid for a hypercube space $[-1,1]^2$ with Stroud’s cubature rule of degree 2 and 3, respectively. Figure 4 shows the sparse grid for a hypercube space $[-1,1]^3$. 

22
Figure 3: 2D Stroud grids.

Figure 4: 3D Stroud grid.
Remark 15 Gauss formulas are, in general, not nested. However, there exist nested Gauss-Patterson formulas starting with the 3-point formula. Although being highly accurate, the number of points can be twice as large as Clenshaw-Curtis formulas [13].

Remark 16 To have an overview of all known cubature rules for the hypercube as well as the hypersphere, the simplex and the entire space see [6] (www.cs.kuleuven.ac.be/~nines/ecf/) and the references there in.

Remark 17 One may find at John Burkardt’s web page a variety of FORTRAN90 codes for building sparse grids based on many open and closed rules. For more details, see people.sc.fsu.edu/~burkardt/f_src/f_src.html.

6 Quantities of interest

As mentioned before, once the numerical solutions of (13) are obtained at all collocation points \( y_k \), the mean of the random solution can be evaluated by using the same cubature rule as

\[
E \left[ \hat{\phi}(x) \right] = \langle \phi(x, y) \rangle = \int_\Gamma \phi(x, y) \rho(y) \, dy = \sum_{k=1}^{n_N^N} \phi(x, y_k) \int_\Gamma L_k(y) \rho(y) \, dy \approx \sum_{k=1}^{n_N^N} \phi(x, y_k) w_k.
\]

Similarly, its variance can be determined by

\[
Var [\phi(x, y)] = \langle \phi(x, y) \phi(x, y) \rangle - \langle \phi(x, y) \rangle < \phi(x, y) >.
\]
The errors related to these statistical moments may also be easily evaluated. For example, the errors in mean ($e_\mu$) and in variance ($e_{\sigma^2}$) are defined as

\[
e_\mu = \langle \phi(x, y) \rangle - \langle \phi_e(x, y) \rangle;
\]

\[
e_{\sigma^2} = \left( \langle \phi(x, y) \phi(x, y) \rangle - \langle \phi(x, y) \rangle \langle \phi(x, y) \rangle \right) - \left( \langle \phi_e(x, y) \phi_e(x, y) \rangle - \langle \phi_e(x, y) \rangle \langle \phi_e(x, y) \rangle \right),
\]

where $\phi_e(x, y)$ stands for a reference (exact) solution. These errors are usually measured by using the $L^2(D)$ and $L^\infty(D)$ -norms.

We may also be interested in assessing some functional of the numerical solution of (4), a specific quantity of interest. The stochastic collocation approximation of this quantity of interest is then simply defined by

\[
\int_\Gamma G(\phi(x, y)) \rho(y) \, dy \approx \sum_{k=1}^{n_N^\gamma} G(\phi(x, y_k)) \rho(y_k) w_k.
\]

The previous expressions show that it is not necessary to explicitly obtain the Lagrange interpolating polynomials to determine the approximation of a quantity of interest. The computational cost of obtaining a quantity of interested is then dictated by how many times the deterministic problem (13) is solved, i.e., the number of collocation points.

### 7 Error bounds

In this section, we review some results from the literature associated to the Smolyak approximation error

\[
\| P_h \phi - A_{q,N} (P_h \phi) \|_{L^2_\mu(\Omega; W(D))},
\]
mainly for the Clenshaw-Curtis formulas. The error estimates essentially depend on the solution regularity, which also depends on the regularity of the random input.

Barthelmann et al. presented in [3] the analysis of the approximation error of the sparse grid method. As indicated in (21), it is valid for functions with bounded derivatives. For $N > 1$, they considered the class of functions $\mathcal{F}_N^r$ with bounded mixed derivatives of order $r$,

$$
\mathcal{F}_N^r := \left\{ f : [-1,1]^N \rightarrow \mathbb{R} \mid D^\beta f \in C^0 \left([-1,1]^N\right), \beta \in \mathbb{N}_0^N, \beta_i \leq r \ \forall i \right\},
$$

where

$$
D^\beta f = \frac{\partial^{\vert \beta \vert} \phi}{\partial y_1^{\beta_1} \ldots \partial y_N^{\beta_N}}, \quad \vert \beta \vert = \sum_{i=1}^N \beta_i
$$

with the associate norm

$$
\|f\| = \max \left\{ \|D^\beta f\|_{\infty} \mid \beta \in \mathbb{N}_0^N, \beta_i \leq r \right\}.
$$

According to [3], one always gets at least

$$
\|P_h \phi - A_{K,N} (P_h \phi)\|_{\infty} \leq c_{N,r} \left( n_K^N \right)^{-r} \left( \log n_K^N \right)^{(r+1)(N-1)} \quad (25)
$$

where $c_{N,r}$ denotes a constant that depends only on $N$ and $r$. The bound in (25) holds only if one starts with univariate formulas that yield the optimal order $\left( n_K^N \right)^{-r}$. This is the case when the sparse grid is based on the Clenshaw-Curtis, Gauss-Patterson and Gauss-Legendre formulas.

The interpolation error can be used to develop adaptive schemes during the cubature process. Nested cubature formulas are particularly useful since they allow the error evaluation based on the difference of two subsequent formulas [14, 11]. This issue will be covered in the next section.
Nobile et al. have analyzed in [21] the Smolyak’s approximation error based on the regularity of the function with respect to the random variables, which is obviously strongly related to the allowed regularity of the input data. More precisely, the basic assumption is that \( \phi(\mathbf{x}, y) \) admits an analytic extension with respect to each of the random variables, which can be stated by the following:

Let define \( \tilde{y}^j \) as an arbitrary element of \( \tilde{\Gamma}^j = \prod_{i \neq j}^N \Gamma^i \). For each \( y^j \in \Gamma^j \), there exists \( \tau^j > 0 \) such that the function \( \phi(\mathbf{x}, y^j, \tilde{y}^j) \) as a function of \( y^j \), \( \phi : \Gamma^j \rightarrow C^0(\tilde{\Gamma}^j, D) \) admits an analytic extension \( \phi(\mathbf{x}, z, \tilde{y}^j), z \in \mathbb{C} \), in the region of the complex plane \( \Sigma(\Gamma^j; \tau^j) \equiv \{ z \in \mathbb{C}, \text{dist} (z, \Gamma^j) \leq \tau^j \} \).

Moreover, \( \forall z \in \Sigma(\Gamma^j; \tau^j) \),

\[
\| \phi(z) \|_{C^0(\tilde{\Gamma}^j, D)} < \lambda, \]

with \( \lambda \) a constant independent of \( j \).

\( \tau^j \) is related to the boundness of the \( r \)-th derivative of the input data. For the problem (4), once the following bound for the diffusion coefficient holds

\[
\left\| \frac{1}{\epsilon(\mathbf{x}, \mathbf{Y})} \frac{\partial^r \epsilon(\mathbf{x}, \mathbf{Y})}{\partial y^j r} \right\|_{L^\infty(D)} \leq \gamma_j^r \gamma^r, \]

then

\[
0 < \tau^j < \frac{1}{\gamma_j^r \gamma^r}, \]

and a similar condition holds for the source term (see [1]).

If the dependence on each random variable is approximated with polynomials, the best approximation error decays exponentially fast with respect to the polynomial degree. As shown in [1], for an univariate function
\( \nu : \Gamma^1 \rightarrow X \) which admits an analytic extension in the region of the complex plane \( \Sigma (\Gamma^1; \tau^1) \equiv \{ z \in \mathbb{C}, \text{dist}(z, \Gamma^1) \leq \tau^1 \} \), for some \( \tau^1 > 0 \), this error is bounded as

\[
E_{\alpha_k}^i \equiv \min_{\phi \in \Pi_{\alpha_k}^n} \| \nu - \phi \|_{C^0(\Gamma^1, D)} \leq \frac{2}{e^{2g} - 1} e^{-2 \alpha_k^i} g \max_{z \in \Sigma(\Gamma^1; \tau^1)} \| \nu(z) \|_{X(D)},
\]

with

\[
0 < g < \frac{1}{2} \ln \left( a + \sqrt{1 + a^2} \right), \quad \text{and} \quad a = \frac{2\tau^1}{|\Gamma^1|}.
\]

For \( N > 1 \), \( \tau_j \) depends on the direction \( j \) and so does the coefficient \( g \), denoted henceforth by \( g_j \). The exponential convergence rate is higher for bigger \( g_j \) and the worst case occurs for the smallest \( g_j \). This is the key issue in the development of the anisotropic sparse grid proposed in [20]. The interpolation estimates for the Clenshaw-Curtis abscissas is the following:

\[
\| P_h \phi - A_{q,N}(P_h \phi) \|_\infty \leq \hat{c}_{N, \sigma} \left( n_q^N \right)^{-\mu_1},
\]

where

\[
\mu_1 = \frac{g}{1 + \log(2N)},
\]

and

\[
g = \frac{1}{2} \min_{1 \leq k \leq N} g_k
\]

\[
0 < g_k < \log \left( a + \sqrt{1 + a^2} \right), \quad a = \frac{2\tau^1}{|\Gamma^1|}.
\]

The algebraic convergence stated in (27) holds when \( q \leq \frac{N}{\log(2)} \). Otherwise, subexponential convergence is obtained (see [21]).

Using the idea of non-isotropic formulas [13], [29, 20] defined an \( N \)-dimensional weight vector \( \alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{R}_+^N \) such that the Smolyak
anisotropic is

\[ A_{q,N}^\alpha (f) = \sum_{l \in X_{\alpha}(q,N)} (\Delta_{l_1}^1 \otimes \ldots \otimes \Delta_{l_N}^1)(f) \] (30)

where the index set is defined by

\[ X_{\alpha}(q,N) := \left\{ l \in \mathbb{N}_+^N, l \geq 1 : \sum_{k=1}^N (l_k - 1) \alpha_k \leq \alpha q \right\} \] (31)

and \( \alpha = \min_{1 \leq k \leq N} \alpha_k \). They assumed that the weights \( \alpha_k \) are the rate of exponential convergence rate in each direction:

\[ \alpha_k = g(k), k = 1, \ldots, N. \]

A priori selection can be performed for some applications by first estimating \( \tau^k \) and then taking

\[ g(k) = \frac{1}{2} \ln \left( \frac{1}{a + \sqrt{1 + a^2}} \right), \quad \text{and} \quad a = \frac{2\tau^k}{|\Gamma^k|}. \]

Alternatively, [20] proposed the following more conservative relation

\[ g(k) = \frac{1}{2} \ln \left( \frac{2\tau^k}{|\Gamma^k|} + 1 \right). \]

The maximum index vector used to build the Smolyak grid, denoted by \( l(\alpha, q) \), is obtained by

\[ l_k(\alpha, q) = \max_{l \in X_{\alpha}(q,N)} \{l_k\} = 1 + \text{int} \left( \frac{\alpha q}{\alpha_k} \right), \quad \text{where} \quad \alpha_k = g(k). \] (32)

If the analyticity region increases as \( k \) increases, and so does \( g(k) \), the second term in (32) decreases and such anisotropic behavior is incorporated into the Smolyak grid introducing lesser and lesser points for growing \( k \). The following examples clarifies this issue.
Example: The isotropic Smolyak’s sparse grid for $N = 2$ and $q = 4$ ($K = 2$) built in section 6.2 is now reconstructed by considering admissible indices such that $\alpha_1(l_1 - 1) + \alpha_2(l_2 - 1) \leq 4$ with $\alpha = (1, 2)^t$. The grid is now obtained by assembling 5 nested grids under the following combination:

$$
A_{4,2}^\alpha = -Q_3^1 \otimes Q_1^1 - Q_1^1 \otimes Q_2^1 + Q_1^1 \otimes Q_1^1 + Q_3^1 \otimes Q_2^1 + Q_1^1 \otimes Q_3^1.
$$

(33)

The final grid with 29 points, shown if figure 5, is obtained by adding each partial grid, getting rid of repeated points. The partial grids are:

$$29 \text{ points} \leftarrow (5 \times 1) + (1 \times 3) + (17 \times 1) + (5 \times 3) + (1 \times 5).$$

Example: Consider a Smolyak approximation level $q = 5$ for two problems with $N = 8$ and respective $g(k)$ (known a priori). With respect to the direction $k$ of the probability space, problems 1 and 2 present anisotropic and isotropic behaviors, respectively, as shown in tables 2 and 3 (extracted from [20]). The maximum indices values for each direction are also indicated in this table.

Table 2: Problem 1.

<table>
<thead>
<tr>
<th></th>
<th>$g(1)$</th>
<th>$g(2) = g(3)$</th>
<th>$g(4) = g(5)$</th>
<th>$g(6) = g(7)$</th>
<th>$g(8)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_1$</td>
<td>0.10</td>
<td>0.09</td>
<td>0.21</td>
<td>0.62</td>
<td>1.55</td>
</tr>
<tr>
<td>$l_2 = l_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$l_4 = l_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$l_6 = l_7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$l_8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$5$</td>
<td>$6$</td>
<td>$3$</td>
<td>$1$</td>
<td>$1$</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5: Anisotropic grid for \( q = 4 \).

The grids obtained by truncating the stochastic dimension at \( N = 3 \) for both problems are shown in figure 6. The anisotropic grid has 109 and the isotropic one has 249 points. The un-weighted index set yields a grid with higher number of points (441) and is shown in figure 7.

Defining \( G(N) = \sum_{k=1}^{N} g(k) \), the algebraic convergence when \( (q \leq \frac{G(N)}{g \log(2)}) \) is given by:

\[
\| P_h \phi - A_{q,N} (P_h \phi) \|_\infty \leq \tilde{c}_{N,g} \left( \frac{n_q}{2} \right)^{-\mu_2},
\]

(34)

where

\[
\mu_2 = \frac{g \left( \log(2) e - 1/2 \right)}{\log(2) + \sum_{n=1}^{N} g/g(k)}. \tag{35}
\]

When a priori selection is not possible, [29, 20] also provided a way to
Table 3: Problem 2.

<table>
<thead>
<tr>
<th></th>
<th>g(1)</th>
<th>g(2) = g(3)</th>
<th>g(4) = g(5)</th>
<th>g(6) = g(7)</th>
<th>g(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.39</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>l_1</td>
<td>l_2 = l_3</td>
<td>l_4 = l_5</td>
<td>l_6 = l_7</td>
<td>l_8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

a posteriori estimate $g(k)$, denoted by $\overline{g}(k)$, by approximating the dependence of the solution on the $k^{th}$ parameter $y^k$ by polynomial interpolation using either Clenshaw-Curtis or Gaussian abscissas while freezing the other parameters $\tilde{y}^k$ [20]. From (26), the interpolation error is expected to decay as

$$\varepsilon_k \approx d_k e^{-2m_k\overline{g}(k)}, \quad (36)$$

where $m_k$ is the number of collocation points in the $k^{th}$ direction. By comparing two consecutive interpolants with $m_k$ and $m_k + 1$ collocation points, $\overline{g}(k)$ is obtained by determining the slope of the plot $\log_{10}(\varepsilon_k)$ versus the number of collocation points.

**Remark 18** The results presented in [29, 20] show quite different values for $g(k)$ obtained using a priori and a posteriori selections. This may be due to the fact that, in the a posteriori selection, $d_k$ (36) does not consider any dependence on $g(k)$ as predicted by (26).
Figure 6: Anisotropic grids obtained using the a priori estimator proposed in [20].

8 Generalized sparse grid

As pointed out by [13], the anisotropic formula (30) is, in fact, a special case of a wider class of cubature formulas which can be generated by the following generalization of Smolyak’s construction. A general index set $I_l$ is constructed by all $l \in \mathbb{N}_+^N$ such that

$$l - e_j \in I_l \quad \text{for} \quad 1 \leq j \leq N, \ l_j > 1,$$

(37)

with $e_j$ denoting the $j$th unit vector. Thus, for every index $l$, the general index set contains all indices which have smaller entries than $l$ in at least one dimension. We may notice that the criterion (37) still ensures the telescopic property of the differential interpolant (18) [14]. By defining the character-
istic function $\chi^l$ of $I_l$ as

$$
\chi^l(I) = \begin{cases} 
1 & \text{if } l \in I_l; \\
0 & \text{otherwise},
\end{cases}
$$

the generalized sparse grid is defined as

$$
A^l(f) = \sum_{l \in I_l} c^l(l) \left( Q_{i_1}^l \otimes \ldots \otimes Q_{i_N}^l \right)(f),
$$

where

$$
c^l(l) = \sum_{z=(0,\ldots,0)}^{(1,\ldots,1)} (-1)^{|z|} \chi^l(l + z). \quad \text{(38)}
$$

**Example:** For the index set that yields the grid (33), only some indices contribute to the final anisotropic grid. They are the ones whose the sum over $\chi^l(I + z)$ on the four index $l + z^i$, $i = 1, \cdots, 4$, is non-zero. The vectors $z^i$ are all vectors in the colexical order from the null vector until the base $1^N$. In this case we have $z^1 = (0,0)^t$, $z^2 = (1,0)^t$, $z^3 = (0,1)^t$ and $z^4 = (1,1)^t$.  

---

Figure 7: Isotropic sparse grid for $q = 4$ and $N = 3$. 
Table 4 illustrates the computation of the classifier (38) for this example. The final grid is then the one indicated in (33) obtained by the partial grids indicated in boldface.

Table 4:

| $I_l$ |  
|---|---|---|---|---|---|---|---|---|
| $l_1$ | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 1 |
| $l_2$ | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 3 |
| $l_1 + z_1^1$ | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 1 |
| $l_2 + z_2^1$ | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 3 |
| $\chi^l (l + z^1)$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $l_1 + z_1^2$ | 2 | 3 | 4 | 5 | 6 | 2 | 3 | 4 | 2 |
| $l_2 + z_2^2$ | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 3 |
| $\chi^l (l + z^2)$ | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| $l_1 + z_1^3$ | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 1 |
| $l_2 + z_2^3$ | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 |
| $\chi^l (l + z^3)$ | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| $l_1 + z_1^4$ | 2 | 3 | 4 | 5 | 6 | 2 | 3 | 4 | 2 |
| $l_2 + z_2^4$ | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 |
| $\chi^l (l + z^4)$ | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $c_{I_l}(l)$ | 0 | 0 | -1 | 0 | 1 | -1 | 0 | 1 | 1 |

The main goal in the adaptive process is to choose the index set $I_l$ so as to substantially reduce the computational cost in high dimensional problems.

The dimension-adaptive cubature developed in [14] increases accuracy by
increasing the number of indices following the equation (22). Beginning with the one element index set, indices are added successively such that [11]:

- the resulting index set remains admissible;
- a large reduction in the interpolation error is achieved.

The differential interpolation operator (18) is used as the error indicator $\hat{\varepsilon}_l$,

$$\hat{\varepsilon}_l := \Delta_l f = \sum_{l \in I_l} \left( \Delta^1_{l_1} \otimes \ldots \otimes \Delta^1_{l_N} \right) (f),$$

which amounts to evaluate the difference between the computed value of the function at the newly added index and the interpolated value of the function using all the old indices. A detailed description of the basic algorithm is made in [14]. With the help of the basic dimension adaptive Algorithm 1, we repeat here the general definitions for clarity. The algorithm is constructed by splitting the current index set $I_l$ into two disjoint sets, the active index set $A$ and the old index set $O$.

The adaptive procedure always starts with the smallest index (step 1), builds the initial index sets and computes the index error (steps 2-5). If the global error estimate falls below a given threshold the computation is stopped and the computed integral value is returned. The global error can be estimated as $\eta = \sum_{l \in I_l} \hat{\varepsilon}_l$. The index with the largest error indicator is selected from $A$ and is added to the old indices set (steps 7-9). Once an index is added, its forward neighborhood is scanned for new admissible indices and their error indicators are computed (steps 11-19). The forward neighborhood of an index $l$ is the $N$ indices $l + e_j$, $1 \leq j \leq N$. The admissible indices are
put into the active index set (step 14) and the corresponding values of the
differential integral are added to the current cubature result (step 16) and the
corresponding values of the error indicators are added to the current global
error estimate (step 17). After this scan, the index in \( A \) with the current
largest error is selected, and the procedure goes on.

**Example:** A detailed two-dimensional example of the dimensional adap-
tive procedure is illustrated in the following.

Consider that the indices \((3, 1), (2, 2)\) and \((4, 1)\) form a sequence of the se-
lected indices from \( A \) (largest error). At the \( m^{th} \) step of the algorithm, the old
and active index sets are \( O = \{(1, 1), (2, 1), (1, 2)\} \) and \( A = \{(3, 1), (2, 2), (1, 3)\} \).
The index \((3, 1)\) is moved from \( A \) to \( O \) and its forward neighbors are \((4, 1), (3, 2)\).
Only the index \((4, 1)\) is admissible, which is moved to \( A \). Thus, at the \((m + 1)^{th}\)
step, the old and active index sets become \( O = \{(1, 1), (2, 1), (1, 2), (3, 1)\} \)
and \( A = \{(2, 2), (1, 3), (4, 1)\} \). The process continues so that the complete in-
der set at the \((m + 3)^{th}\) step is \( I_{m} = \{(1, 1), (2, 1), (1, 2), (3, 1), (2, 2), (1, 3),
(4, 1), (3, 2), (5, 1)\} \).

This dimension-adaptive sparse grid methodology is proved to be useful
in many areas such as data mining [4], fuzzy arithmetic [16] and machine
learning [12]. For choosing the collocation points of stochastic collocation
method, it was used in [11] for solving stochastic natural convection problems.
The latter work showed that the dimension-adaptive scheme provides at least
an order of magnitude improvement in the number of function evaluations

\[m + 3\]
Algorithm 1 Dimension adaptive cubature [14]: integrate $f$.

1: $i := (1, \cdots, 1)$
2: $O := \emptyset$
3: $A := \{i\}$
4: $r := \Delta_i f$
5: $\eta := \hat{\epsilon}_i$
6: while $\eta > TOL$ do
7: Select $i$ from $A$ with largest $\hat{\epsilon}_i$
8: $A := A \backslash \{i\}$
9: $O := O \cup \{i\}$
10: $\eta := \eta - \hat{\epsilon}_i$
11: for $k : 1, \ldots, N$ do
12: $j := i + e_k$
13: if $j - e_n \in O$ $\forall n = 1, \ldots, N$ then
14: $A := A \cup \{j\}$
15: $s := \Delta_j f$
16: $r := r + s$
17: $\eta := \eta + \hat{\epsilon}_j$
18: end if
19: end for
20: end while
required to construct comparable interpolation functions as compared to the conventional sparse grid method [11].

A slight change in Algorithm 1 is also proposed in [14] to incorporate the work involved for the computation of the differential integral. A generalized error indicator is defined as

\[ \hat{\varepsilon}_l := \max \left\{ w \frac{\Delta_l f}{\Delta_1 f}, (1 - w) \frac{n_1}{n_k} \right\}, \]

where \( n_1 = 1 \) and it is assumed that \( \Delta_1 f \neq 0 \). This value can be replaced by a suitable normalizing constant. While classical sparse grid is obtained by setting \( w = 0 \), values of \( w \) between 0 and 1 are supposed to balance the computational work with the error estimate.

9 Numerical examples

9.1 Steady state problems

The sparse grid collocation method presented in the previous sections are used in a variety of stochastic problems. Different cubature rules are used as well dimension-adaptive techniques (a priori and a posteriori) in order to evaluate the main features of these procedures.

9.1.1 1D diffusion problem

Consider the following 1D problem \( (d = 1) \) and \( N > 1 \) random dimensions

\[
\frac{d}{dx} \left[ \epsilon(x, y) \frac{d\phi(x, y)}{dx} \right] = 0, \quad (x, y) \in (0, 1) \times \Gamma,
\]

(39)
with boundary conditions

\[ \phi(0, y) = 0, \quad \phi(1, y) = 1, \quad y \in \Gamma. \]

The random diffusivity has the form

\[ \epsilon(x, y) = 1 + \sigma \sum_{k=1}^{N} \frac{1}{k^2 \pi^2} \cos(2\pi k x) y^k, \quad (40) \]

where \( y^k \in [-1, 1]^N, \) \( k = 1, \ldots, N, \) are independent uniformly distributed random variables\(^3\). This series converge as \( N \to \infty \) and

\[ E[\epsilon(x, y)] = 1, \quad 1 - \frac{\sigma}{6} < \epsilon(x, y) < 1 + \frac{\sigma}{6}. \]

The deterministic problem defined on the physical space is solved by using the Galerkin method with 200 linear elements. We first choose \( N = 2 \) and \( q = 10 \), yielding 7169 collocation points \((M = 7169)\). For \( \sigma = 0.1 \), figure 8 shows the mean solution and the variance of (39). The \( L^2(D) \) approximation error to the mean of the random solution for different interpolation levels and number of collocation points is shown in figure 9, in which both Clenshaw-Curtis and Stroud cubature rules are used. The exact solution is the Smolyak isotropic grid \( A_{10,2} \). They present good convergence rates, although the Stroud rule yields greater error due to the fewer knots. We also assess the error for \( 1 \leq N \leq 8 \) with increasing \( q \) \((0 \leq q \leq 4)\) as shown in figure 10. The error is evaluated by selecting an exact solution obtained with \( A_{5,N} \), which corresponds to one level higher than the used more refined grid. We clearly note that there is a limit under which the level and the dimension

\(^3\)This form is similar to those obtained from a Karhunen-Loève (K-L) expansion, with eigenvalues delaying as \( 1/k^4 \).
Figure 8: 1D diffusion problem: mean and variance of (39) with $N = 2$ and $q = 10$. 
of the cubature rule can be increased separately to improve accuracy. Having
in mind that the increase of one dimension of the cubature formula increases
the number of nodes in about one order for the corresponding cubature level,
both the level and the dimension of the cubature formula must be adapted
in order to gain accuracy.

Remark 19 As the representation (40) may come from a K-L expansion, the
value of $N$ is primarily governed by the accuracy of the eigen-pairs in rep-resenting the covariance function of the diffusive random process. As pointed
out in [15, 7], smooth covariance models exhibit faster convergence than models with less smooth covariance function. Moreover, more terms are needed
to represent the random process for a given accuracy in using numerical K-L
expansion rather than analytical one [15].

To estimate the $L^2(D)$ computational error to the mean in the level $q$,
$\|E[\varepsilon^q]\|$, we set $\|E[\varepsilon^q]\| \approx \|E[A_{q,N} (P_h \phi_N) - A_{q+1,N} (P_h \phi_N)]\|$. Figure 11
shows the behavior of $\|E[\varepsilon^q]\|$ for $1 \leq N \leq 8$ with increasing $q$ (up to
$q \leq 4$) so that $\|E[\varepsilon^q]\| \leq 10^{-12}$, in log-log scale. As expected [21], the
error decreases sub-exponentially as $q$ increases and the convergence rate is
dimension dependent. To determine the a priori error estimator we need to
consider the following bound

$$
\left\| \frac{1}{\epsilon(x,Y)} \frac{\partial^r \epsilon(x,Y)}{\partial y^r} \right\|_{L^\infty(D)} \leq \begin{cases} 
\frac{\sigma}{k^{2\pi^2}} & \text{for } r = 1 \\
0 & \text{for } r > 1
\end{cases},
$$

and then we may use

$$
\tau^k = \frac{k^2 \pi^2}{2\sigma}.
$$
Figure 9: 1D diffusion problem: $L^2$-error to the mean for Clenshaw-Curtis points with $1 \leq q \leq 4$ with and for Stroud points with degree 2 and 3.
Figure 10: 1D diffusion problem: $L^2$-error to the mean for $2 \leq N \leq 8$ and $0 \leq q \leq 4$ versus the number of collocation points in log-log scale.

Figure 11: 1D diffusion problem: $\|E[\varepsilon^q]\|$ for $1 \leq N \leq 8$ and $0 \leq q \leq 4$ versus the number of collocation points in log-log scale.
Since $|\Gamma^k| = 2$, the convergence rates for each direction is approximated by

$$g(k) = \frac{1}{2} \ln \left( \frac{k^2 \pi^2}{2\sigma} + 1 \right).$$

For $\sigma = 0.1$, those values are shown in Table 5, which also depicts the maximum indices $\bar{l}_k(\alpha, q)$ for $0 \leq q \leq \bar{q} = 4$. Because all directions yield almost equal weighting, except for the first one, their indices are pretty much the same. The $L^2(D)$ approximation error to the mean of the random solution is obtained by truncating the probability space at $N = 4$ and $N = 8$ and computing solutions up to $\bar{q} = 4$, as shown in Figure 12. In this case, the exact solution is an enriched isotropic Smolyak solution obtained by using the Smolyak’s index set $X(5, 9)$. Note that, although the anisotropic behavior of this problem is mild, there still some gain in using the anisotropic sparse grid.

<table>
<thead>
<tr>
<th>$g(k)$</th>
<th>$g(1)$</th>
<th>$g(2)$</th>
<th>$g(3)$</th>
<th>$g(4)$</th>
<th>$g(5)$</th>
<th>$g(6)$</th>
<th>$g(7)$</th>
<th>$g(8)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_k$</td>
<td>1.96</td>
<td>2.65</td>
<td>3.05</td>
<td>3.34</td>
<td>3.56</td>
<td>3.74</td>
<td>3.90</td>
<td>4.03</td>
</tr>
<tr>
<td>$q$</td>
<td>$\bar{l}_1(\alpha, q)$</td>
<td>$\bar{l}_2(\alpha, q)$</td>
<td>$\bar{l}_3(\alpha, q)$</td>
<td>$\bar{l}_4(\alpha, q)$</td>
<td>$\bar{l}_5(\alpha, q)$</td>
<td>$\bar{l}_6(\alpha, q)$</td>
<td>$\bar{l}_7(\alpha, q)$</td>
<td>$\bar{l}_8(\alpha, q)$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

According to the methodology proposed in [20], the a posteriori error estimator may be obtained by using (36). For each stochastic direction $k$, we evaluate here the error in the $k$-th direction between two consecutive solutions.
Figure 12: 1D diffusion problem: $L^2$-error to the mean for $N = 4$ and $N = 8$, with $0 \leq q \leq 4$ versus the number of collocation points. Comparison between isotropic and anisotropic grids.
obtained using either consecutive number of points (as proposed in [20]) or consecutive interpolation orders, using indices \( l = (1, \ldots, 1, l_k, 1, \ldots, 1) \) and \( \tilde{l} = (1, \ldots, 1, l_k + 1, 1, \ldots, 1) \). The error is the \( L_2 (D) \) approximation error in the expected value, \( \|E [\varepsilon_k]\|_{L_2(D)} \), so that

\[
\log_{10} \left( \|E [\varepsilon_k]\|_{L_2(D)} \right) \approx \log_{10} (d_k) - 2m_k \log_{10}(e) \overline{g} (k).
\]

Thus, denoting by \( a_k \) the slope of the linear least square approximation to fit \( \log_{10} \left( \|E [\varepsilon_k]\|_{L_2(D)} \right) \) versus \( m_k \), the a posteriori error estimator is

\[
\overline{g} (k) = a / \lfloor 2 \log_{10}(e) \rfloor.
\]

This procedure yields the computational results presented in figure 13, in which the linear least square approximations is also presented. Table 6 shows the computed \( \overline{g}(k) \) as well as the maximum indices \( \overline{l}_k (\alpha, q) \). They are similar to the ones obtained with the a priori error estimator, which can be seen in figure 14. This figure shows the \( L_2 (D) \) approximation error to the mean for \( N = 4 \) and \( N = 8 \) using anisotropic grids, built a priori and a posteriori. Here, the exact solution is again an enriched isotropic Smolyak solution obtained with \( X(5,9) \). Although there are some slight differences among the grids obtained using the a priori and a posteriori error estimators, the a posteriori error estimator was able to capture the dimension-dependent convergence rate behavior. The a posteriori approach of using consecutive interpolation orders yielded a grid with fewer points (177) than that using consecutive number of points.
Figure 13: 1D diffusion problem: Linear least square aproximation to fit \( \log_{10} \left( \| E[\varepsilon_k] \|_{L^2(D)} \right) \) versus \( m_k \).

### 9.1.2 1D advection-diffusion problem

Consider now the following transport problem and \( N > 1 \) random dimensions

\[
-\epsilon \frac{d^2 \phi(x, y)}{dx^2} + \beta(x, y) \frac{d\phi(x, y)}{dx} = 0, \quad (x, y) \in (0, 1) \times \Gamma, \tag{41}
\]

with boundary conditions

\[
\phi(0, y) = 0, \quad \phi(1, y) = 1, \quad y \in \Gamma.
\]

The random velocity field has the form

\[
\beta(x, y) = 10^5 \left( 1 + \sum_{k=1}^{N} \frac{1}{k^4 \pi^2} y^k \right), \tag{42}
\]

where \( y^k \in [-1, 1]^N, k = 1, \ldots, N, \) are independent uniformly distributed random variables. The diffusion coefficient is constant and set equal to \( 10^3 \).
Figure 14: 1D diffusion problem: $L^2$-error to the mean for $N = 4$ and $N = 8$, with $0 \leq q \leq 4$ versus the number of collocation points. Comparison between the grids obtained using the a priori and a posteriori error estimators.
Table 6: Convergence rates - a posteriori error estimator.

<table>
<thead>
<tr>
<th>( g(k) \rightarrow g(1) )</th>
<th>( g(2) )</th>
<th>( g(3) )</th>
<th>( g(4) )</th>
<th>( g(5) )</th>
<th>( g(6) )</th>
<th>( g(7) )</th>
<th>( g(8) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consecutive interpolation orders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n_k \rightarrow 3.88 )</td>
<td>( 5.57 )</td>
<td>( 5.48 )</td>
<td>( 6.74 )</td>
<td>( 7.25 )</td>
<td>( 7.67 )</td>
<td>( 8.03 )</td>
<td>( 8.03 )</td>
</tr>
<tr>
<td>( q )</td>
<td>( \Gamma_1(\mathbf{x}, q) )</td>
<td>( \Gamma_2(\mathbf{x}, q) )</td>
<td>( \Gamma_3(\mathbf{x}, q) )</td>
<td>( \Gamma_4(\mathbf{x}, q) )</td>
<td>( \Gamma_5(\mathbf{x}, q) )</td>
<td>( \Gamma_6(\mathbf{x}, q) )</td>
<td>( \Gamma_7(\mathbf{x}, q) )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Consecutive number of points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n_k \rightarrow 4.55 )</td>
<td>( 5.96 )</td>
<td>( 6.11 )</td>
<td>( 6.30 )</td>
<td>( 7.76 )</td>
<td>( 8.18 )</td>
<td>( 8.46 )</td>
<td>( 8.78 )</td>
</tr>
<tr>
<td>( q )</td>
<td>( \Gamma_1(\mathbf{x}, q) )</td>
<td>( \Gamma_2(\mathbf{x}, q) )</td>
<td>( \Gamma_3(\mathbf{x}, q) )</td>
<td>( \Gamma_4(\mathbf{x}, q) )</td>
<td>( \Gamma_5(\mathbf{x}, q) )</td>
<td>( \Gamma_6(\mathbf{x}, q) )</td>
<td>( \Gamma_7(\mathbf{x}, q) )</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

As the Peclet number is small, the deterministic problem defined on the physical space is also solved by using the Galerkin method with 200 linear elements. Figure 15 compares the \( L^2 \)-error to the mean using isotropic and anisotropic grids. The anisotropic grids are the same used in the previous example and they yield some computational gain as compared with the isotropic grids.

9.1.3 2D advection dominated advection-diffusion problem

We consider now the problem

\[-\varepsilon \Delta \phi (x, z, y) + \mathbf{\beta} (x, z, y) \cdot \nabla \phi (x, z, y) = 0, \quad (x, z, y) \in (0, 1)^2 \times \Gamma\]
Figure 15: 1D advection-diffusion problem: $L^2$-error to the mean for $N = 4$ with $0 \leq q \leq 4$ versus the number of collocation points. Comparison between the isotropic and anisotropic (a priori) grids.
with homogeneous boundary conditions everywhere except on the bottom part of the domain in which it is assumed
\[
\phi(x, 0, y) = \begin{cases} 
0 & \text{if } x < 0.3 \\
1 & \text{if } x \geq 0.3, \quad y \in \Gamma.
\end{cases}
\]

The random velocity field \( \beta(x, z, y) = (\beta_1, \beta_2) \) has the form
\[
\beta_1 = 0.4 + \sigma \sum_{k=1}^{N} \frac{1}{k^2 \pi^2} y^k; \quad \beta_2 = 1 + \sigma \sum_{k=1}^{N} \frac{1}{k^2 \pi^2} y^k.
\]

where \( y^k \in [-1, 1]^N, k = 1, \ldots, N, \) are independent uniformly distributed random variables. The diffusion coefficient is constant and set equal to \( 10^{-9} \).

The inflow discontinuity propagates towards the interior of the domain and gives rise of an internal layer and an external layer appears at \((1, z, y)\) and at the top right part of the domain. The deterministic problem is solved using piecewise linear triangular elements and, due to the advection dominated behavior, the SUPG method is used to avoid global oscillations although spurious modes appear in the neighborhood of the layers. The finite element mesh has 40 divisions in each direction of the physical domain, yielding 3200 elements. Figure 16 shows the mean and the variance obtained with an isotropic \( A_{4}A_{4} \) Smolyak grid. As expected, we may observe the presence of spurious modes in the vicinity of sharp gradients in the mean. However, no oscillations are observed in the variance. The \( L^2(D) \) computational error to the mean, \( \| E[\varepsilon^q] \| \), does not decreases monotonically as \( q \) increases. The solution does not have the regularity requirements to allow the use neither of the a priori nor a posteriori error estimators.
Figure 16: 2D advection dominated advection-diffusion problem: mean and variance of (39) with \( N = 4 \) and \( q = 4 \).
9.2 Transient problems

In the following, we investigate the accuracy of the stochastic collocation method for transient advection-diffusion problems.

We consider both one and two-dimensional cases in which the stochastic velocity field is perturbed by a Gaussian variable with zero mean and unit variance. The collocation points are the roots of the Hermite polynomials and the Gauss-Hermite quadrature rule is used. Alternatively, any other cubature rule may be employed by using a mapping between the stochastic space and an artificial “standard” stochastic space whose properties are known. This approach, introduced in [18, 19], is presented in the Appendix 1 and allows using all the sparse grid ingredients shown previously, although it may eventually lead to a decrease in the convergence rate.

9.2.1 1D transient advection-diffusion problem

Let us consider now the problem

\[
\frac{\partial \phi (t, x, y)}{\partial t} - \epsilon \frac{\partial^2 \phi (t, x, y)}{\partial x^2} + \beta (x, y) \frac{\partial \phi (t, x, y)}{\partial x} = 0,
\]

\[(t, x, y) \in (0, T] \times (0, 2) \times \Gamma, \quad \text{(44)}\]

with homogeneous Dirichlet boundary conditions and the initial condition given by

\[
\phi (0, x) = \begin{cases} 
1, & \text{if } x \in [0.2, 0.7]; \\
0, & \text{otherwise.}
\end{cases}
\]

The random velocity field has the form

\[
\beta (x, y) = \bar{\beta} + \sigma y,
\]

\[(45)\]
where $\beta = 1$, $y$ is a Gaussian distributed random variable with zero mean and unit variance and $\sigma$ is a parameter to control the amplitude of the stochastic perturbation. The diffusion coefficient is constant and set equal to $10^{-2}$. The deterministic problem defined on the physical space is solved by using the SUPG method with 400 linear elements. The resulting system of differential algebraic equations is solved using the Crank-Nicolson method and the timestep is set equal to 0.005. These discrete choices are such that the deterministic solution of (44) coincides with the analytical one at $T = 0.8$. The analytical solution far from $t = 0$ is given by

$$
\phi(t,x) = \frac{1}{2} \left[ \text{erf} \left( \frac{x - \beta t - 0.2}{\sqrt{4\epsilon t}} \right) - \text{erf} \left( \frac{x - \beta t - 0.7}{\sqrt{4\epsilon t}} \right) \right].
$$

Figure 17 shows the numerical expected solution and respective variance for two different values of $\sigma$ in comparison with the exact deterministic solution at $T = 0.8$ using seven collocation points. As expected, the largest $\sigma$ yields a higher variance but we may also notice a larger spread of the mean solution in the physical space with a smaller maximum value. [8] blames the uncertainty for this behavior, which is supposed to play an extra diffusion role in the system. This conclusion can be misleading, since the mean solution in this case fails to have any relevant meaning. This is clearly shown by the plot of the solutions at the collocation points shown in Figure 18 in which no smoothing effect is observed. A better understanding of the system can be provided by a plot of the stochastic solution using error bars, with a length of two standard deviations around the mean, shown in Figure 19. In this case, if the output solution follows the normal distribution, about 90% of all possible solutions are in the shaded region.

We also solved this problem using the Monte-Carlo method. The sampling
Figure 17: 1D transient advection-diffusion problem: mean and variance obtained by using 7 Gauss-Hermite collocation points.
points are chosen in a naive way, using routines from the Numerical Recipes (gardev and ran1) (http://sci.ui.ac.ir/~sjalali/nrf/index.html). In Figure 20 we compare the expected value and the variance obtained using the stochastic collocation (SC) method (7 Gauss-Hermite points) with the Monte Carlo (MC) method (100 sample points), for $\sigma = 0.05$. The histogram of the sample points distribution is shown in Figure 21. Although quite sparse, the MC stochastic solution has similar mean of that one obtained using the stochastic collocation method. A better shot of the system is depicted in Figure 22.
Figure 19: 1D transient advection-diffusion problem: error bars of the numerical stochastic solution.
Figure 20: 1D transient advection-diffusion problem: mean and variance obtained by using the Monte Carlo Method with 100 sampling points ($\sigma = 0.05$).
Figure 21: 1D transient advection-diffusion problem: histogram of the sample points distribution of the Monte Carlo method.

Figure 22: 1D transient advection-diffusion problem: error bars of the numerical Monte Carlo solution ($\sigma = 0.05$).
9.2.2 2D transient advection-diffusion problem

We consider now the problem
\[
\frac{\partial \phi (t, x, z, y)}{\partial t} - \varepsilon \Delta \phi (t, x, z, y) + \beta (x, z, y) \cdot \nabla \phi (t, x, z, y) = 0,
\]
\[(t, x, z, y) \in (0, T] \times (-1, 1)^2 \times \Gamma \tag{46} \]
with homogeneous boundary conditions everywhere. The initial condition is a Gaussian cone given by
\[
\phi (0, x, z) = \exp \left[ -\frac{(x - x_0)^2 + (z - z_0)^2}{2 \lambda^2} \right],
\]
where \((x_0, z_0) = (-0.5, 0.0)\) is the initial center of the Gaussian cone and \(\lambda = 1/8\). The random velocity field yields a circular motion and has the form
\[
\beta (x, z, y) = (z + \sigma y, -x - \sigma y), \tag{47}
\]
where \(y\) is a Gaussian distributed random variable with zero mean and unit variance and \(\sigma = 0.05\). The diffusion coefficient is constant and set equal to \(10^{-5}\) so that the transport of the Gaussian cone occurs mainly due to the advection. In the associate deterministic problem, the Gaussian cone travels along the circular movement almost without changing its shape. The uncertainty in the velocity field perturbs the circular movement and yields the following analytical stochastic solution
\[
\phi (t, x, z, y) = \frac{\lambda^2}{\lambda^2 + 2 \varepsilon t} \exp \left[ -\frac{(\overline{x} + \overline{z})^2}{2 (\lambda^2 + 2 \varepsilon t)} \right],
\]
where
\[
\overline{x} = x + \sigma y - (x_0 + \sigma y) \cos t - (z_0 + \sigma y) \sin t; \nonumber
\]
\[
\overline{z} = z + \sigma y + (x_0 + \sigma y) \sin t - (z_0 + \sigma y) \cos t. \nonumber
\]
For further reference, the contours for the exact expectation is shown in Figure 23.

The deterministic problem defined on the physical space is solved by using the Galerkin method with 5,000 linear triangular elements. The resulting system of differential algebraic equations is solved using the Crank-Nicolson method and the timestep is set equal to $\pi/200$. These discrete choices are such that the deterministic solution of (46) coincides with its analytical at $T = 2\pi$.

The contours of the numerical expectation and the variance are shown in Figures 24 and 25 for $t = \pi/2, \pi, 3\pi/2$ and $2\pi$. Seven Gauss-Hermite collocation points were used. The numerical expectation agrees very well with the exact one (see figures 24 and 23). We may see that the mean solution becomes asymmetric with the time and returns to the symmetric Gaussian cone shape after one complete cycle ($t = 2\pi$).

A similar problem was studied in [8] using polynomial chaos and a semi-Lagrangian method in the finite element framework. It was also studied in [28] by employing the generalized polynomial chaos and the spectral/hp element method. They both blamed the uncertainty in the velocity field for that asymmetric behavior of the mean solution. They claimed that the presence of the random perturbation in the transport velocity introduces extra diffusion in the mean solution. Although the presence of the velocity field perturbation may be seen as yielding an extra diffusion in the mean solution, such smoothness can again be misleading. It is important to distinguish between the solution of the stochastic problem and the mean of all possible solutions, which is rather a smoothness. Figure 26 depicts the evolution
Figure 23: 2D transient advection-diffusion problem: exact expected solution ($\sigma = 0.05$).

of the peak solution along with the deterministic numerical solution. The stochastic solution is plotted using a linear interpolation of the mean and of the standard deviation around the mean, so that about 50% of all possible solutions lie between the green and blue lines.
Figure 24: 2D transient advection-diffusion problem: mean solution obtained by using the Stochastic Collocation Method with 7 Gauss-Hermite points ($\sigma = 0.05$) at $t = \pi/2, \pi, 3\pi/2$ and $2\pi$. 
Figure 25: 2D transient advection-diffusion problem: variance obtained by using the Stochastic Collocation Method with 7 Gauss-Hermite points ($\sigma = 0.05$) at $t = \pi/2, \pi, 3\pi/2$ and $2\pi$. 
Figure 26: 2D transient advection-diffusion problem: error bars of the Stochastic Collocation Method with 7 Gauss-Hermite points ($\sigma = 0.05$).
10 Summary

In this report we examine the stochastic collocation (SC) method for solving transport problems. As pointed out in [24], the stochastic collocation method is a generalization of the stochastic Galerkin approach and has the advantage of being non-intrusive and easily deal with nonlinear systems, nonindependent and unbounded random variables [1]. Some limitations have been reported lately associated with long-term integration and convergence rate. These issues can be overcome at some extend by using the multi-element [10] and the pseudo-spectral [31, 17, 30] approaches.

The choice of collocation points (or the cubature rule) is a key issue in SC methods, especially for high dimensions of the probability space. In this case, sparse grids, being subsets of the full tensor product grids, play important role in efficiency since they reduce the number of collocation points. Under the assumption that the solution depends analytically on each random variable, algebraic convergence with respect to the total number of collocation points in a sparse grid is proved in [21], using either Clenshaw-Curtis or Gaussian abscissas. The exponent of the algebraic convergence depends on the regularity assumption and the number of input random variables. Error estimates by using anisotropic sparse grid were derived in [20], overcoming the latter dependence. That work also developed both a priori and a posteriori error indicators to drive the construction of the anisotropic grid. Due to the regularity assumption, they both work well for diffusion dominated problems.

From the numerical experiments conducted in section 9, it is important to keep in mind that “the solution of a stochastic problem, being a proba-
bility distribution, is never one image, but a set of images, samples of the posterior probability density” ([23], page 47). Tarantola also adds that “the common practice of plotting the mean image should be abandoned, even if it is accompanied by some analysis of error and resolution”. The mean solution may lead to an error in understanding the behavior of the system in the presence of uncertainties. This issue is clear in the transient numerical experiments conducted in section 9.2.

Finally, a recent work compared the performance of the non-intrusive polynomial chaos (PC) and stochastic collocation methods by using a number of simple algebraic test problems. “Performance between these methods is shown to be very similar and both demonstrate impressive efficiency relative to Monte Carlo sampling methods. When a difference is observed between traditional PC and SC, SC has been the consistent winner, typically manifesting in the reduction of the required integration by one order or level” [9].

Acknowledgment

This work has been partially supported by the Brazilian government, through Agency CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), under grant # 0858/08-0.
References


**Appendix 1**

The stochastic collocation method introduced in [18, 19] is based on a transformation between the stochastic space and an artificial space whose properties are known. In the following, we describe this mapping for an univariate case, although the procedure is automatically extended for multivariate cases.
Let $\xi$ denote a point in the artificial stochastic domain, where $\xi \in [-1, 1]$ is a uniformly distributed random variable. The stochastic collocation method uses collocation points $\xi_i, i = 1, \ldots, M$, where $M$ is the total number of quadrature points. According to the used rule, these have associate quadrature weights $w_i$.

The probability density function (PDF) of the random variable $y, \rho(y)$, is the basis for the transformation by means of its cumulative distribution function (CDF), whose range is $[0, 1]$. The CDF is denoted by $F(y)$. If $\rho(y)$ is nonzero in the interior of the stochastic domain, $F(y)$ is strictly monotonic so that the transformation is bijective. A transformation of this kind is illustrated in Figure 27 when $y$ is a Gaussian random variable with mean zero and unit variance. $F(y)$ may be conveniently mapped to the standard domain $[-1, 1]$ by defining

$$F(y) = 2F(y) - 1.$$ 

Thus, $\xi$ is a point in the standard domain $[-1, 1]$ of the inverse of $F$, ie,

$$\xi = F(y) = 2F(y) - 1 \quad \text{and} \quad y = F^{-1}(\xi). \quad (48)$$

These formulas associate the random variable to each collocation point so that there are $y_i = F^{-1}(\xi_i), i = 1, \ldots, M$. Thus, the numerical solutions of the deterministic problem (13) are obtained at all collocation points $y_i$. Once they are obtained, the mean, for example, is simply determined by

$$E[\hat{\phi}(x)] = \sum_{k=1}^{M} \phi(x, y_k) w_k.$$ 

This procedure is specially attractive for general PDFs due to its simplicity, and standard interpolation methods may be used for the mapping (48).
Figure 27: Use of the CFD to transform the stochastic space into the artificial stochastic space.

It is also shown in [19] that, for one stochastic variable, the stochastic collocation method applied to a nonlinear 1D nozzle problem yields significant reduction in cost as compared to using polynomial chaos method.