Journal of Computational Physics 229 (2010) 4664–4682

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

A method for stochastic constrained optimization using derivative-free surrogate pattern search and collocation

Sethuraman Sankaran [a], Charles Audet [b], Alison L. Marsden [a,*]

[a] Department of Mechanical and Aerospace Engineering, EBU II 569, University of California San Diego, La Jolla, CA 92093-0411, USA
[b] GERAD and Département de mathématiques et de génie industriel, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal (Québec), Canada H3C 3A7

ARTICLE INFO

Article history:
Received 24 February 2009
Received in revised form 1 March 2010
Accepted 2 March 2010
Available online 15 March 2010

Keywords:
Stochastic optimization
Uncertainty quantification
Surrogate Management Framework (SMF)
Derivative-free optimization
Mesh Adaptive Direct Search (MADS)
Probabilistic constraints

ABSTRACT

Recent advances in coupling novel optimization methods to large-scale computing problems have opened the door to tackling a diverse set of physically realistic engineering design problems. A large computational overhead is associated with computing the cost function for most practical problems involving complex physical phenomena. Such problems are also plagued with uncertainties in a diverse set of parameters. We present a novel stochastic derivative-free optimization approach for tackling such problems. Our method extends the previously developed surrogate management framework (SMF) to allow for uncertainties in both simulation parameters and design variables. The stochastic collocation scheme is employed for stochastic variables whereas Kriging based surrogate functions are employed for the cost function. This approach is tested on four numerical optimization problems and is shown to have significant improvement in efficiency over traditional Monte-Carlo schemes. Problems with multiple probabilistic constraints are also discussed.

1. Introduction

The literature is replete with optimization techniques, both gradient-based and derivative-free, that are tailored for a specific set of problems. However, many of these optimization techniques are not applicable for problems involving complex physics due to numerical challenges, noise, and/or lack of gradient information. In large-scale simulation-based design problems, it is not uncommon for the computation of the cost function (which involves running a simulation of the physical process) to take hours or even several days.

The focus of this work is to develop a computational optimization technique that can be seamlessly applied to stochastic physical problems for which evaluating cost functions involves significant computational expense. In particular, we are interested in complex systems that require large-scale computation of partial differential equations (PDEs) during the design process. Examples of such problems include fluid flow applications (groundwater flow in porous media [1], flow over aerodynamic structures [2]), solid mechanics (robust design of forming processes [1]), biological applications (design of surgical procedures [3]), atomistic simulations, and fluid-structure interaction (FSI) problems. Surrogate models have also been used to successfully accelerate evolutionary algorithms [4]. In this paper we present an expanded version of the derivative-free surrogate management framework (SMF) that incorporates stochastic capabilities for robust design.

* Corresponding author. Tel.: +1 858 822 3744.
E-mail address: amarsden@ucsd.edu (A.L. Marsden).

0021-9991/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2010.03.005

1.1. Stochastic PDEs

In many complex systems, it is essential to account for uncertainties and quantify the propagation of uncertainties from system inputs to outputs. Uncertainties can be present in different forms such as geometry, material properties, input parameters or boundary conditions. A number of numerical schemes are available for quantifying uncertainties. One of the first established methods was the Monte-Carlo (MC) sampling technique [5]. MC's popularity originated from its non-intrusive nature, and dimension-independent convergence properties. However, MC is plagued with extremely slow convergence rates, and it is usually impractical for complex PDE optimization problems due to computational expense.

In this work, we use the concept of stochastic spaces for representing and quantifying random variables [6]. This concept has been used successfully in many engineering approaches over the past two decades. Similar to space and time being seen as dimensions, randomness is seen as an additional dimension. Initially, the Generalized Polynomial Chaos (GPCE) approach (spectral expansion of variables in stochastic space) was used successfully to tackle problems involving uncertainties [7,8]. However, one disadvantage of this technique is the need to solve a coupled set of resultant matrix equations in the random domain. The non-intrusive spectral projection method generates approximations by sampling the polynomials in GPCE using latin hypercube sampling technique. Recently, a stochastic collocation scheme was developed in which simulations are performed at specific collocation points in the stochastic space [9–11]. This technique combines the exponential convergence rates of the GPCE scheme with the decoupled nature of Monte-Carlo techniques. It is non-intrusive and can be used with legacy codes or in situations where source code is not available. Another similar technique is the probabilistic collocation (PC) method [12], where Lagrange interpolating polynomials are employed. Collocation points and weights are computed based on the input probability density function (PDF) and Gauss-quadrature rules are computed through the Golub–Welsch algorithm. While the simplest means for extension to multiple stochastic dimensions is the tensor-product rule, sparse grids can also be used with PC techniques. As we deal with arbitrary continuous PDFs and are interested in higher order statistical moments, techniques such as ANOVA (analysis of variance) and DoE (design of experiments) are not applicable here. Therefore, in the present work we employ sparse grid stochastic collocation with the Smolyak algorithm.

1.2. Stochastic optimization

In developing an optimization method suitable for PDE problems with uncertainty, there are several important considerations that must be addressed. First, for time-dependent PDE problems, obtaining gradient information can pose formidable challenges. Specialized numerical methods such as differentiable mesh movement schemes have been used for computing gradients in flow problems. However, the complexity of geometries and unstructured meshes common in such problems makes it difficult to evaluate gradients without significant noise (error). Although there has been impressive success with adjoint methods, they present challenges in unsteady flow problems because of the need to store large time histories. Finite difference techniques are often prohibitively expensive and subject to non-trivial numerical noise. Automatic differentiation schemes pose difficulties with shape optimization problems. Techniques for computing sensitivities using complex variables [13] are also infeasible since a number of sensitivity equations need to be solved, and they are not applicable for shape sensitivities. As a result, we devise an algorithm based on a derivative-free optimization technique.

In coupling a stochastic method to optimization, efficiency is a key consideration. The advent of efficient collocation methods has led to the recent development of stochastic gradient-based tools for performing optimization [1,14,15] that significantly improve efficiency compared to Monte-Carlo methods. Techniques for performing stochastic inverse problems [16] have also been recently developed. However, despite this progress, there remains a large class of problems with expensive simulation-based cost functions for which gradient-based methods are impractical.

Keeping the above issues in mind, we present a derivative-free optimization method that is well suited for computationally expensive stochastic problems. Our technique couples surrogate pattern search techniques with a stochastic collocation scheme, and does not require any gradient information. In this work, we present a coupled framework employing the surrogate management framework (SMF) [17], mesh adaptive direct search methods (MADS) [18] and the stochastic collocation technique [9]. The stochastic SMF method relies on convergence theory of derivative-free pattern search methods for optimization and the stochastic collocation theory for convergence of the cost function. Deterministic SMF has been used successfully in unsteady fluid mechanics problems [2,3], helicopter rotor blade design [17], and in multi-objective liquid-rocket injector design [19].

Table 1
Nomenclature for the optimization framework.

Term	Description
design variable	unknown entities that are to be computed through optimization
parameter	given values that are used as inputs to the PDE
stochastic or uncertain	prefix to denote that an entity is not known with certainty
inputs	entities necessary to solve the PDE
outputs	post-processed entities obtained from the solution of the PDE

Some of the nomenclature used in this paper is described in Table 1. For example, some design variables for beam design problems are geometrical variables (width, height, thickness) and choice of material. Typical simulation input parameters for such problems are forces and moments acting on the beam, and material properties such as Young's modulus. Bold mathematical symbols refer to vectors throughout. In this paper, we will analyze two situations: (Case 1) the parameters are stochastic but design variables are deterministic or (Case 2) both, the parameters and design variables needs to be computed. In problems with insufficient knowledge about the parameter, techniques such as the Maximum Entropy scheme [20–22] can be used to yield the PDF of parameters. The paper is organized as follows. In Section 2, we discuss some mathematical formalism and background of the techniques developed. In Section 3, we describe the stochastic-surrogate based optimization scheme. In Section 4, we provide implementation details of the scheme, in Section 5, we show some convergence proofs, and follow it up with some numerical examples in Section 6.

2. Mathematical background

We employ a stochastic collocation scheme that uses Lagrange polynomials for interpolation in the stochastic space. We first detail the formalism for the stochastic collocation scheme and then provide details of the optimization algorithm.

2.1. Stochastic collocation scheme

A probability space is specified as $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is the sample space of elementary events, $\mathcal{F}$ is the σ-algebra $\mathcal{F} \subset 2^{\Omega}$ and $\mathcal{P}$ is a probability measure. The σ-algebra consists of all allowed permutations of the basic outcomes in Ω [23].

A discrete random variable is defined as one that maps each point on the sample space to a scalar probability measure that takes values between 0 and 1. A continuous random variable is defined as one that maps each infinitesimal region on the sample space to a scalar probability measure. In short, a real-valued random variable X can be written as $X : (\Omega, \mathcal{F}, \mathcal{P}) \mapsto \mathbb{R}$. For notational convenience, a random variable will be denoted as $X(\Omega) \mapsto \mathbb{R}$ where Ω denotes $(\Omega, \mathcal{F}, \mathcal{P})$.

In this paper, instead of dealing with PDFs directly, we use the concept of stochastic spaces in our numerical solutions. Stochastic space is the space of $\boldsymbol{\xi} = [\xi^1, \xi^2, \ldots, \xi^N]$ where ξ^i represents either uniform or normally distributed random variables. Any construct on the stochastic space has an unique PDF associated with it. A finite dimensional random support space is described by its truncated descriptor (random vector) $\boldsymbol{\xi}$ so that,

$$\boldsymbol{\xi} = [\xi^1, \xi^2, \ldots, \xi^N] : \Omega \to \mathbb{R}^N, \tag{1}$$

where the dimensionality N of the stochastic support space is problem dependent. We denote the approximations to a function g by $g(\mathbf{x}, \mathbf{t}, \boldsymbol{\xi})$.

In the collocation scheme, the stochastic space is approximated using mutually orthogonal interpolating functions. To represent a function at any point in the stochastic space, the function $g(\mathbf{x}, \mathbf{t}, \boldsymbol{\xi})$ is written as $g(\mathbf{x}, \mathbf{t}, \boldsymbol{\xi}) = \sum_{i=1}^{M} g(\mathbf{x}, \mathbf{t}, \boldsymbol{\xi}_i)\mathcal{L}_i(\boldsymbol{\xi})$ where $\mathcal{L}_i(\cdot)$ are the orthogonal interpolating polynomials and M is the number of collocation points.

2.1.1. Construction of interpolating polynomials for one-dimensional random variable:

We denote the interpolant to a function f as $\mathcal{I}f$. The interpolant constructed using the function values at n distinct points is denoted as $\mathcal{I}^n f$. The interpolant, $\mathcal{I}^n f$ is chosen to ensure that the error induced by utilizing this interpolant goes to 0 as the number of collocation points increases. For a Chebychev node based interpolating polynomial, $\|f - \mathcal{I}\backslash f\|_{\infty} \leqslant Cn^{-k}log(n)$ where n denotes the number of collocation points, $f \in C^k$, that is, the kth derivative of f is continuous. In general, for one-dimensional functions, Gauss points and Chebechev points have the least interpolation error [24]. The results of simulations always contain noise (due to numerical convergence) and this can limit the order of interpolating polynomial. Higher order polynomials can oscillate if this error is significant.

2.1.2. Extension to multi-dimensional randomness – the Smolyak algorithm:

A trivial means to extend the above-mentioned scheme for multi-dimensions is by constructing a simple tensor product space. However, this is subject to the curse of dimensionality as the number of collocation points grows exponentially. As a result, we employ computationally efficient schemes of approximating the stochastic space. The Smolyak algorithm is a method for computing the collocation points.

The Smolyak algorithm reduces the number of collocation points necessary for interpolation in multi-dimensional random space. This has been explained in [10] in which the sparse grid interpolant is given as:

$$A_{s,N}(f) = \sum_{s-N+1 \leqslant |\mathbf{i}| \leqslant s} (-1)^{s-|\mathbf{i}|} \cdot \binom{N-1}{s-\mathbf{i}} \cdot (\mathcal{I}^{i_1} \otimes \cdots \otimes \mathcal{I}^{i_N})(f), \tag{2}$$

with $A_{N-1,N} = 0$, $\mathbf{i} = (i_1, \ldots, i_N)$ and $|\mathbf{i}| = i_1 + \ldots + i_N$. Denoting $\Delta^i = \mathcal{I}^i - \mathcal{I}^{i-1}$ as the incremental interpolant, the Smolyak interpolation can be written as:

$$\mathcal{A}_{s,N}(f) = \sum_{|\mathbf{i}| \leqslant s} (\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N})(f) = \mathcal{A}_{s-1,N}(f) + \sum_{|\mathbf{i}| = s} (\Delta^{i_1} \otimes \cdots \otimes \Delta^{i_N})(f), \tag{3}$$

which has the error bounded as:

$$\|f - A_{s,N}(f)\| = \mathcal{O}(M^{-2}|log_2 M|^{3(N-1)}), \tag{4}$$

where M is the number of interpolation points.

Further details of the algorithm are given in [24–26]. Algorithms for integration based on sparse grids are provided in [27]. Collocation points constructed using Smolyak's algorithm in a two-dimensional stochastic space are shown in Fig. 1.

3. Optimization framework

3.1. The surrogate management framework for deterministic problems

In this section, we review some mathematical preliminaries required for the surrogate management framework (SMF) for optimization of deterministic functions. This technique has been discussed in [28] but we summarize important aspects for the sake of completeness. This technique was originally introduced by Booker et al. [17] and belongs to a general class of pattern search optimization methods. There are a number of techniques that rely on approximating the cost function including neural networks, response surface techniques, and polynomial-based statistical regression. An advantage of the SMF optimization scheme is that it has a well-established convergence theory and uses a surrogate function that could, if required, be tailored according to the problem.

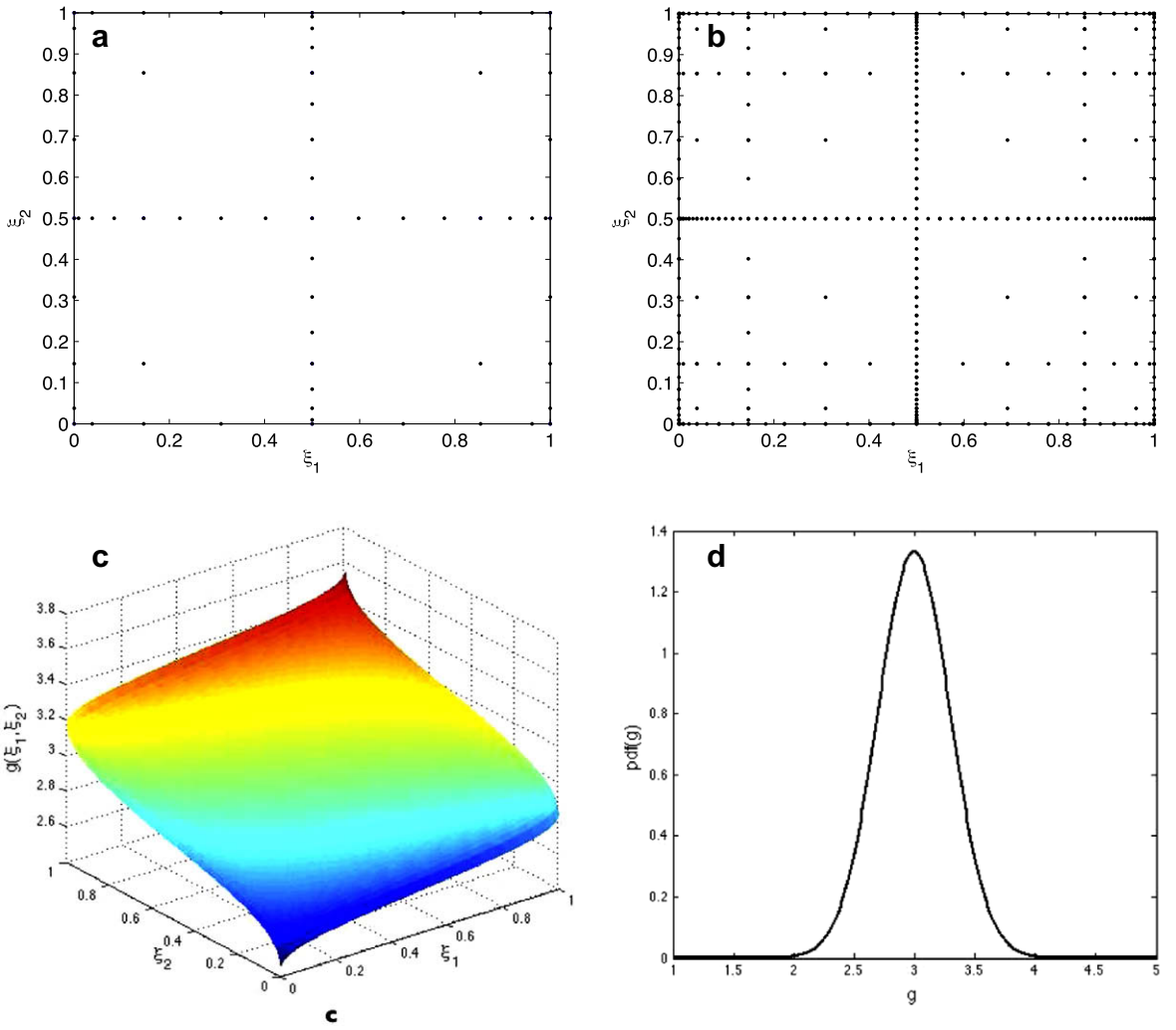

Fig. 1. The figure shows (a) stochastic collocation points in a two-dimensional stochastic space using a level-4 depth of interpolation and (b) level-6 depth of interpolation. (c) Shows the function g(·) which is a sum of two Gaussian distributions, $\mathcal{N}(2, 0.2)$ and $\mathcal{N}(1, 0.1)$. The PDF of g is shown in (d).

In this work, we use Kriging surrogate functions, which were originally employed in the field of geostatistics, to interpolate the cost function using measurements at specific points. It is to be noted that the stochastic model associated with Kriging is not related to the input uncertainties. Kriging functions ensure smoothness in interpolation and do not have sharp variations that are prevalent in polynomial interpolation. The Kriging function has enjoyed wide popularity in diverse engineering fields. The unknown function is written as a linear combination of its value at specific points and is computed by minimizing the variance of the predicted error. The cost function is mathematically described using Kriging functions [29] by matching its values at select locations. Combined with the stochastic collocation approach that interpolates the cost function in the stochastic domain using Lagrange polynomials, we represent the function in its entirety using its values at specific points.

The SMF algorithm is based on two fundamental tasks - SEARCH and POLL. The SEARCH step uses the surrogate to identify one or more points that are likely to improve the cost function. The POLL step consists of a local exploration in the space of d.v.s, and provides the convergence framework for the algorithm. The SMF algorithm constructs an initial surrogate. We employ the latin hypercube sampling (LHS) technique that distributes the initial set of d.v.s so that they maximally avoid each other [30]. While any other problem specific sampling technique can be substituted, we choose LHS since it provides a good distribution over the space of design variables. As we show later, this does not affect the convergence properties of our optimization algorithm. The SEARCH step optimizes the surrogate to predict the location of one or more minimizing points. The true cost function is then evaluated at these points. For the range of surrogate points used for examples in this paper, the cost of searching the surrogate is negligible compared to the cost of running a simulation. The SEARCH may also include additional points as desired (such as points at maximum mean squared error for model improvement) as long as they belong to a discretization of the space of variables called the mesh. The mesh is refined as the algorithm evolves. The SEARCH is considered successful if an improved cost function value is found, at which point the surrogate is updated, and another SEARCH step is performed. The SEARCH is considered unsuccessful if it fails to find an improved point, and a POLL step is performed. The basis for convergence of the SMF algorithm is given by the POLL step. In the POLL step, the function is evaluated at a set of mesh points that form a positive spanning set of directions in order to evaluate whether the current best point is a mesh local optimizer. A set of at least $n+1$ POLL points are required to generate a positive basis, where n is the number of optimization parameters. In this work, the POLL set is chosen using the mesh adaptive direct search (MADS) [18] algorithm.

We present a detailed convergence analysis in Section 5. This uses the results of some convergence proofs given in [18]. Our proof is also influenced by earlier work on pattern search convergence theory by Audet and Dennis [31,32] and Torczon [33]. We refer the interested reader to [3,28] for details on the deterministic optimization technique.

For constrained optimization problems, several approaches are available. With the extreme barrier approach [18], the cost function values at infeasible points are set to infinity. In other approaches, a constraint violation function, H is defined where $H \leqslant 0$ represents the feasible region while $H > 0$ represents the infeasible region. In the filter approach [28,32], an infeasible point x_0 is considered filtered, or dominated, if there is an infeasible point x for which $H(x) \leqslant H(x_0)$ and $J(x) \leqslant J(x_0)$, where J represents the cost function. A filter is defined as the finite set of non-dominated infeasible points found so far. At any iteration of the algorithm, the best feasible point and the least infeasible point (infeasible point with least constraint violation function value) are identified. In this technique, the conditions for the SEARCH and POLL steps to be considered successful rely on the filter, rather than the cost function value alone (algorithmic and implementation details are found in [28]). In the recent progressive barrier approach [34], any trial point whose constraint violation function exceeds a threshold are deemed infeasible. The threshold value is reduced as the algorithm evolves. In this paper, we use the filter approach, but the algorithm is flexible enough to incorporate any of the above methods without restriction.

3.2. Stochastic optimization scheme

The general constrained stochastic optimization problem is formulated as follows,

$$\begin{aligned} &\underset{z}{\text{minimize}}\, J(\mathbf{z}, \mathbf{y}(\xi)), \\ &\text{subject to } \mathbf{z} \in \mathcal{Q}, \end{aligned} \tag{5}$$

where $J : \mathbb{R}^{d_v} \times \Omega^{d_v} \to \mathbb{R}$ is the cost function, and $\mathbf{z}$ is the vector of design variables (d.v.s) in the domain $\mathcal{Q}$ defined as follows:

1. If the d.v.s are deterministic, then

$$\mathcal{Q} = \{\mathbf{z} \in \mathbb{R}^{d_v} | \mathbf{l} \leqslant \mathbf{z} \leqslant \mathbf{u}, \mathbb{M}(g(\mathbf{z}, \mathbf{y}(\xi))) \in \mathcal{S}\}.$$

2. If the d.v.s are stochastic, then

$$\mathcal{Q} = \{\mathbf{z} \in \mathbb{R}^{d_v} \times \Omega^{d_v} | \mathbf{l} \leqslant \mathbf{z} \leqslant \mathbf{u}, \mathbb{M}(g(\mathbf{z}(\xi^{\mathbf{z}}), \mathbf{y}(\xi^{\mathbf{y}}))) \in \mathcal{S}\}.$$

In the above, $\mathbf{y}$ represents stochastic parameters. $\xi^{\mathbf{z}}$ and $\xi^{\mathbf{y}}$ represent the corresponding discretized stochastic spaces and $\xi = \xi^{\mathbf{z}} \times \xi^{\mathbf{y}}$. The functional $\mathbb{M}(\cdot)$ represents moments or any integral over the probability space and can be over any non-linear function g, and $\mathcal{S} \subset \Omega$. $\mathbf{l}$ and $\mathbf{u}$ denote the upper and lower bounds on the problem d.v.s.

A deterministic black box takes in deterministic inputs and returns deterministic outputs. We define a stochastic black box as taking an input PDF and generating the deterministic cost function J (which is usually composed of moments or statistic such as confidence intervals of the output σ of the embedded deterministic black box). The deterministic black box might consist of legacy or binary codes, and we create a wrapper around this to generate the stochastic black box as shown in Fig. 2. Note that σ can represent more than one output. In practice, to obtain the output of the stochastic black box, multiple runs of the deterministic PDE solver are usually required.

We first describe the algorithm for a problem with no constraints and deterministic d.v.s and then generalize to include the constraints and stochastic d.v.s.

3.3. Stochastic optimization algorithm

Deterministic d.v.s

The procedure described below is very similar to the deterministic scheme described in [28] since only parameters are considered stochastic. The critical difference occurs in step (b). In the algorithm, the mesh defines a discretization of the space of d.v.s. These d.v.s are forced to lie on the nodes of the mesh, which is gradually refined.

1. SEARCH
   (a) Identify a finite set T_k of trial points on the mesh.
   (b) Evaluate $J(\mathbf{x}_{search}, \mathbf{y}(\xi)) = J(\mathbf{x}_{search}, \sum_{i=1}^{N} \mathbf{y}(\xi_i)\mathcal{L}_i(\xi))$ for all trial points $\mathbf{x}_{search} \in T_k$.
   (c) If $J(\mathbf{x}_{search}, \mathbf{y}(\xi)) < J(\mathbf{x}_k, \mathbf{y}(\xi))$ for any trial point $\mathbf{x}_{search}$ in T_k, then a lower cost function value has been found, and the SEARCH is *successful*. Set $\mathbf{x}_{k+1} = \mathbf{x}_{search}$, increment k and go back to (a).
   (d) Else, if no trial point in T_k improves the cost function, the SEARCH is *unsuccessful*. Go to the POLL step.
2. POLL
   (a) Choose a set of positive spanning directions, and form the POLL set X_k as the set of mesh points adjacent to $\mathbf{x}_k$.
   (b) If $J(\mathbf{x}_{poll}, \mathbf{y}(\xi)) < J(\mathbf{x}_k, \mathbf{y}(\xi))$ for any point $\mathbf{x}_{poll} \in X_k(\xi)$, then a lower cost function value has been found and the POLL is *successful*. Set $\mathbf{x}_{k+1} = \mathbf{x}_{poll}$, increment k and go to the SEARCH step.
   (c) Else, if no point in X_k improves the cost function, the POLL is unsuccessful.
      (i) If convergence criteria are satisfied, a converged solution has been found. STOP.
      (ii) Else, refine the mesh, increment k and go to the SEARCH.

This algorithm is similar to the deterministic optimization scheme due to the non-intrusive nature of the stochastic collocation algorithm. The stochastic interpolates are built in the second step of the algorithm.

Stochastic d.v.s

In the case where the stochastic collocation problem has uncertain d.v.s, the stochastic space is defined as $\Omega = \Omega(\mathbf{z}) \times \Omega(\mathbf{y})$ where $\Omega(\mathbf{y})$ is a known stochastic space based on stochastic parameters and $\Omega(\mathbf{z})$ is an unknown stochastic space to be determined. Following [14], we can assume without any loss of generality that $\xi^{\mathbf{z}}$s are defined on the support for the standard uniform variable. Any arbitrary distribution can be defined in this space. The stochastic optimization algorithm involves the following changes:

- Replace T_k by $T_k(\xi^{\mathbf{z}})$. Step 1(a) is replaced by: Identify a finite set $T_k(\xi_i^{\mathbf{z}})$ of trial points on the mesh for all collocation points i of the parameters $\mathbf{z}$.
- Replace $\mathbf{x}_{search}$ by $\mathbf{x}_{search}(\xi^{\mathbf{z}})$, and $\mathbf{y}(\xi)$ by $\mathbf{y}(\xi^{\mathbf{y}})$.
- Replace $\mathbf{x}_{poll}$ by $\mathbf{x}_{poll}(\xi^{\mathbf{z}})$, and X_k by $X_k(\xi^{\mathbf{z}})$.

In this algorithm, each of the collocation points in the d.v. space is associated with a mesh. Since the cost function is dependent on the values of parameter and d.v. at each collocation point, it is important to store this vector at each iteration. Further, if the cost function involves, for example, only expectation of some function, then there will be multiple PDF of d.v.s which will give us the same optimal cost function. A schematic of the stochastic optimization algorithm is shown in Fig. 3.

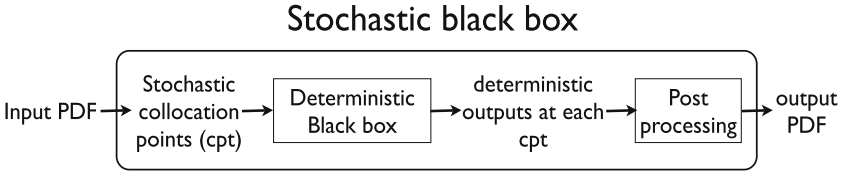

Fig. 2. The figure shows a schematic of a stochastic black box.

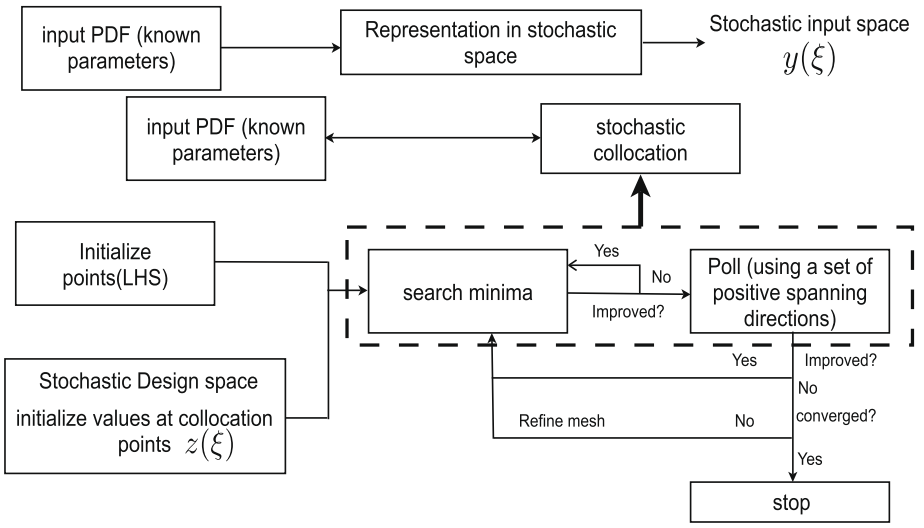

Fig. 3. The figure shows a schematic of the SMF optimization algorithm employing the stochastic collocation technique.

3.4. Hierarchical stochastic optimization algorithm

A hierarchical technique can be employed for performing stochastic optimization. This algorithm cannot be used for the problems in which the collocation points define the space of d.v.s. The optimization algorithm is initially run with a coarse depth of interpolation, d_l and the following steps are employed.

1. Set depth of interpolation, $d = d_l$, usually a coarse value.
2. Perform the optimization steps given in Section 3.1. Terminate if the optimal solution point of the last two iterations is the same.
3. Increment the depth of interpolation, $d = d + 1$. Set the initial trial points to be the filter points (non-dominated points) of the previous optimization run. Modify the domain of the `SEARCH` step based on extreme Filter points. Go to step 2. (The filter points can be replaced with trial points that have the least cost function value for unconstrained optimization problems.)

This algorithm will perform well if the lower depths of interpolation give an approximate picture of the stochastic variation of the cost function. In the worst case, it will reduce to the algorithm given in Section 3.1 (as exemplified in one of the example problems). This happens when the stochastic space construct of the cost function is not smooth and will need a very accurate construct to compute either the cost function or constraints.

4. Implementation

The main advantage of the stochastic collocation technique is its non-intrusive nature, i.e., it can be seamlessly wrapped over existing deterministic algorithms. Hence, the interpolation in the stochastic space is built independently on top of the Kriging interpolation of the cost function. The code was implemented in MATLAB with the following steps:

1. The collocation points are first evaluated based on the stochastic dimensions as well as the depth of interpolation. The sparse grid toolkit *spinterp* [26] is employed for this purpose.
2. If moments are to be computed, the weights corresponding to each collocation point is computed by integrating the Lagrange functions over the stochastic space
3. The SMF optimization algorithm makes repeated calls to the cost function to update the surrogate function. The DACE Matlab Toolbox [35] is used for Kriging. The cost function now employs the stochastic collocation points as well as corresponding weights for its evaluation. The rest of the details of the SMF algorithm remain the same.

When coupled with the SMF framework, the total number of evaluations of the function is significantly reduced compared to traditional techniques like Monte-Carlo. The simulation is performed at collocation points and the sparse grid toolkit is used to interpolate and construct the stochastic space. The PDF is computed by sampling the stochastic space using MC (which takes negligible computational expense compared to the cost of running a simulation) and binning (constructing a histogram). The cost function, J and constraints are evaluated from this PDF.

5. Convergence analysis

In this section, we discuss details of the convergence of the stochastic algorithm. We split the discussion into three parts. First we show that the interpolated cost function (using collocation in the stochastic domain and Kriging in the deterministic parameter space) converges to its true value under a specific set of conditions. Second, we show that the unconstrained optimization algorithm generates a stationary point using the generalized pattern search (GPS) technique, and then extend it to the SMF algorithm using MADS. Finally, we discuss convergence of the constrained optimization algorithm using the filter approach.

5.1. Convergence of the interpolated function

The analysis relies on the definition of stochastic sensitivities.

Definition 5.1. Stochastic sensitivities are defined as the change in output when the PDF of the d.v.s is perturbed. The sensitivity field (directional derivative) $\Theta_\sigma(\mathbf{z}(\xi^{\mathbf{z}}),\mathbf{y}(\xi^{\mathbf{y}})) \equiv D_{\Delta\mathbf{z}(\xi^{\mathbf{z}})}\sigma(\mathbf{z}(\xi^{\mathbf{z}}),\mathbf{y}(\xi^{\mathbf{y}}))$ is defined to be the linear part in the Taylor expansion of the process $\sigma(\mathbf{z}(\xi^{\mathbf{z}}),\mathbf{y}(\xi^{\mathbf{y}}))$ i.e.

$$\sigma(\mathbf{z}(\xi^{\mathbf{z}}),\mathbf{y}(\xi^{\mathbf{y}});\Delta\mathbf{z}(\xi^{\mathbf{z}})) = \sigma(\mathbf{z}(\xi^{\mathbf{z}}),\mathbf{y}(\xi^{\mathbf{y}})) + D_{\Delta\mathbf{z}(\xi^{\mathbf{z}})}\sigma(\mathbf{z}(\xi^{\mathbf{z}}),\mathbf{y}(\xi^{\mathbf{y}})) + \mathcal{O}(\|\Delta\mathbf{z}\|^2). \tag{6}$$

For convenience, the sensitivities are denoted as $\Theta\sigma_{\Delta\mathbf{z}}$.

We show below that sensitivities can be defined using perturbations of stochastic collocation coefficients (which preserves normality [1]).

The following observations will be useful for deriving our main results. The cost function J is related to the input parameters through the output of the deterministic black box, σ, and therefore:

$$\frac{\partial J}{\partial \mathbf{z}(\xi_i)} = \frac{\partial J}{\partial \sigma(\xi_i)}\frac{\partial \sigma(\xi_i)}{\partial \mathbf{z}(\xi_i)}. \tag{7}$$

Furthermore, we can write the stochastic sensitivities for J provided $\Theta J_{\Delta\mathbf{z}} \in \mathcal{C}^0$ for piecewise linear interpolates, and $\Theta J_{\Delta\mathbf{z}} \in \mathcal{C}^{N_z}$ for Lagrange polynomials.

$$\Theta J_{\Delta\mathbf{z}} = \sum_{i=1}^{N_z}\frac{\partial J}{\partial \mathbf{z}(\xi_i)}\mathcal{L}_i(\xi) = \sum_{i=1}^{N_z}\frac{\partial J}{\partial \sigma(\xi_i)}\frac{\partial \sigma(\xi_i)}{\partial \mathbf{z}(\xi_i)}\mathcal{L}_i(\xi). \tag{8}$$

This will be used for making conclusions on $\Theta J_{\Delta\mathbf{z}}$ based on results for $\frac{\partial\sigma(\xi_i)}{\partial\mathbf{z}(\xi_i)}$.

The norm of a stochastic function, $f(.,\xi)$, is defined as $g(\cdot) \equiv \|f(.,\xi)\|_s = \int f(.,\xi)p(\xi)d\xi$.

The variance of a random variable satisfies $var(x) = \mathbf{E}(x^2) - (\mathbf{E}(x))^2 \geqslant 0$. If $var(x) = \mathbf{E}(x) = 0$ for a variable random x, then $P(x \neq 0) = 0$. This will be useful to show that infimum of the iterates generated by our algorithm attains a point-distribution centered at 0.

With this background, we now show convergence properties on the optimization algorithm.

Theorem 5.1. *Suppose that (a) $J \in \mathcal{C}^0$ for piecewise linear stochastic interpolates and (b) $J \in \mathcal{C}^{N_y+N_z}$ for Lagrange interpolates. Then as the number of Kriging samples, N_y and N_z goes to infinity, the interpolated cost function $\widehat{J}$ converges to its true value*

$$\|\widehat{J} - J\| \to 0.$$

Proof. Based on the construction of interpolates, the function $\widehat{J}$ is given by

$$\widehat{J}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{N_s} J(\mathbf{z}^{\mathbf{i}}(\xi^{\mathbf{z}}),\mathbf{y}(\xi^{\mathbf{y}}))s^i(\mathbf{x}(\xi^{\mathbf{z}})),$$

where N_s represents the number of Kriging samples (which we use for interpolation of the cost function in parameter space) and s^i represent the Kriging weights corresponding to an arbitrary parameter $\mathbf{x}$. Employing stochastic interpolation of the cost function using the collocation approach, we have:

$$\widehat{J}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{N_s}\sum_{j=1}^{N_z}\sum_{k=1}^{N_y} J(\mathbf{z}^{\mathbf{i}}(\xi_j^{\mathbf{z}}),\mathbf{y}(\xi_k^{\mathbf{y}}))s^i(\mathbf{x}(\xi^{\mathbf{z}}))\mathcal{L}_j(\xi)\mathcal{L}_k(\xi).$$

Using the triangle inequality, $\|a+b\| \leqslant \|a\| + \|b\|$, we have

$$\|J(\mathbf{x},\mathbf{y}) - \widehat{J}(\mathbf{x},\mathbf{y})\| \leqslant \|J(\mathbf{x},\mathbf{y}) - J^*\| + \|J^* - \widehat{J}(\mathbf{x},\mathbf{y})\|, \tag{9}$$

where J^* is chosen as $J^* = \sum_{j=1}^{N_z}\sum_{k=1}^{N_y} J(\mathbf{x}(\xi_j^{\mathbf{z}}),\mathbf{y}(\xi_k^{\mathbf{y}}))\mathcal{L}_j(\xi)\mathcal{L}_k(\xi)$. In a similar fashion, the spaces $\mathcal{L}_j$ and $\mathcal{L}_k$ can be split and the error is bounded by the additive summation of errors as:

$$\|J - J^*\| \leqslant CN_z^{-2}|log_2 N_z|^{3(d_z-1)} + CN_y^{-2}|log_2 N_y|^{3(d_y-1)}, \tag{10}$$

where d_y and d_z represent the stochastic dimensions of the known variables and unknown parameters respectively. As N_z and N_y tend to ∞, $\|J - J^*\| \to 0$. Eq. 10 follows from [24]. The second term is given by

$$\|J^* - \widehat{J}\| \leqslant \sum_{j=1}^{N_z}\sum_{k=1}^{N_y}\|R_{j,k}\| \int \mathcal{L}_j(\xi)\mathcal{L}_k(\xi)p(\xi)d\xi, \tag{11}$$

where $R_{j,k} = J(\mathbf{x}(\xi_j^{\mathbf{z}}), \mathbf{y}(\xi_k^{\mathbf{y}})) - \sum_{i=1}^{N_s} J(\mathbf{z}^{\mathbf{i}}(\xi_j^{\mathbf{z}}), \mathbf{y}(\xi_k^{\mathbf{y}}))s^i(\mathbf{x}(\xi^{\mathbf{z}}))$.

By virtue of Kriging interpolation, as the number of Kriging samples N_s increases, $\|R_{j,k}\| \to 0$.

Using Eqs. (9)–(11), we have:

$$\|J - \widehat{J}\| \to 0. \qquad \square$$

We now show convergence of the stochastic optimization procedure. We first show convergence on GPS in Section 5.2, and then extend to the SMF procedure using MADS in Section 5.3. Before that, we note some properties of the sensitivities and moments of σ.

Note 5.2. If $\Theta\sigma_{\Delta z} \in C^n$, then $\Theta\sigma_{\Delta z}^m \in \mathcal{C}^{n+m}$, $m = 1, 2$.

In the equation above, $\sigma_{\Delta_z}^m$ denotes the mth moment of the stochastic sensitivities. Essentially, continuity properties on the function in the stochastic space imposes sufficient continuity properties on the probability density function itself. This is because arbitrary moments are at least $\mathcal{C}^0$ continuous.

5.2. Stochastic unconstrained optimization with pattern search

Initially, we assume that $\sigma(\xi_i) \in \mathcal{C}^1$ as was assumed by Torczon [33]. We later lift some of the constraints on σ.

Theorem 5.3. *Let the stochastic sensitivities $\Theta\sigma_{\Delta\mathbf{z}}$ be defined on the neighborhood of a compact set $L(\mathbf{z}_0)$. Then for the sequence of iterates that the GPS algorithm produces $x_k(\xi)$, we have*

$$\liminf_{k\to\infty} \|\Theta\sigma_{\Delta\mathbf{z}}^m\| = 0 \quad m = 1, 2$$

where the superscript m denotes the moment of stochastic sensitivities.

Proof. We represent the moments of the stochastic sensitivities using the collocation approach [1] as:

$$\Theta\sigma_{\Delta\mathbf{z}}^m = \sum_{i=1}^{N_z} \Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^m \mathcal{L}_i(\xi).$$

where $\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}$ are the deterministic sensitivities computed by perturbing the parameters at the collocation points [1]. Using the triangle inequality and since $w_i \geqslant 0 \forall i$,

$$\|\Theta\sigma_{\Delta\mathbf{z}}^m\| \leqslant \sum_{i=1}^{N_z} \|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^m\| \int \mathcal{L}_i(\xi)p(\xi)d(\xi) = \sum_{i=1}^{N_z} w_i \|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^m\|. \tag{12}$$

Using Theorem 3.5 in [33], we have $\liminf_{k\to\infty}\|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^m\| = 0$. This theorem holds here since $\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}$ is deterministic and the update of parameters at each collocation point coincides with the update technique of the GPS algorithm. Let the iterate at which $\|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^2\| = 0$ be k'. At this iterate, owing to the semi-positiveness of the variance of σ, we have:

$$\|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^2\| - \left(\|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^1\|\right)^2 \geqslant 0 \quad \text{and therefore } \|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^1\| = 0. \tag{13}$$

Hence, the infimum for both the moments occur together. Also since the norm is always positive, we have

$$0 \leqslant \liminf_{k\to\infty} \|\Theta\sigma_{\Delta\mathbf{z}}^m\| \leqslant \sum_{i=1}^{N_z} w_i \left(\liminf_{k\to\infty} \|\Theta\sigma_{\Delta\mathbf{z}(\xi_i)}^m\|\right) = 0.$$

Hence, we have $\liminf_{k\to\infty}\|\Theta\sigma_{\Delta\mathbf{z}}^m\| = 0$. $\square$

Corollary 5.4. *The stochastic sensitivity field satisfies:*

$$\liminf_{k\to\infty} \|\Theta\sigma_{\Delta\mathbf{z}}\| \equiv 0.$$

Proof. The proof directly follows from Theorem 5.3. Since the mean and the variance of the stochastic sensitivities are zero, the stochastic sensitivity field is identically zero. □

Theorem 5.5. *Let the stochastic sensitivities $\Theta\sigma_{\Delta\mathbf{z}}$ be defined on the neighborhood of a compact set $L(\mathbf{z}_0)$. Assume the columns of the generating matrix [33] are bounded in norm for each collocation point ξ_i, (i.e.) $\lim_{k\to\infty}\Delta_k(\xi_i) = 0$ and that the GPS method enforces the Strong hypotheses on exploratory moves. Then for the sequence of iterates that the algorithm produces $x_k(\xi)$, we have*

$$\lim_{k\to\infty} \|\Theta\sigma^m_{\Delta\mathbf{z}}\| = 0 \quad m = 1, 2.$$

Proof. Using Theorem 3.6 in [33], we have $\lim_{k\to\infty}\|\Theta\sigma^m_{\Delta\mathbf{z}(\xi_i)}\| = 0$. Since the norm is always positive and using Eq. 12, we have:

$$0 \leqslant \lim_{k\to\infty} \|\Theta\sigma^m_{\Delta\mathbf{z}}\| \leqslant \sum_{i=1}^{N_z} w_i \left(\lim_{k\to\infty} \|\Theta\sigma^m_{\Delta\mathbf{z}(\xi_i)}\| \right) = 0. \tag{14}$$

Hence, we have $\lim_{k\to\infty}\|\Theta\sigma^m_{\Delta\mathbf{z}}\| = 0$. The interpretation of Torczon's convergence proof for the stochastic algorithm is that η can be chosen differently at different collocation points, (i.e.) $\eta = \eta(\xi)$. Another interpretation is by choosing ϵ in Torczon's paper as $\epsilon = \epsilon_{min}(\xi)$ over all points in the stochastic space. □

Corollary 5.6. *The stochastic sensitivity field satisfies:*

$$\lim_{k\to\infty} \|\Theta\sigma_{\Delta\mathbf{z}}\| \equiv 0.$$

Proof. The proof directly follows from Theorem 5.5. Since the mean and variance of the stochastic sensitivities are zero, the stochastic sensitivity field is identically zero. □

In Theorems 5.3 and 5.5, we showed convergence on the first two moments of the stochastic sensitivities using convergence proofs published for the deterministic problem given in [33].

Theorem 5.7. *The stochastic sensitivity of the cost function, $\Theta J_{\Delta z}$ is identically zero for a sufficiently large k.*

$$\lim_{k\to\infty} \|\Theta J_{\Delta z}\| \equiv 0. \tag{15}$$

Proof. The chain rule for the cost function ensures that:

$$\Theta J_{\Delta z} = \sum_i \frac{\partial J}{\partial \sigma(\xi_i)} \frac{\partial \sigma(\xi_i)}{\partial z(\xi_i)} \mathcal{L}_i(\xi). \tag{16}$$

From Corollary 5.6, the stochastic sensitivities of σ are zero, hence $\frac{\partial\sigma(\xi_i)}{\partial z(\xi_i)} = 0$. Therefore, we can conclude that $\Theta J_{\Delta z} = 0$. □

The convergence analysis presented here is restricted to unconstrained GPS algorithms. We have assumed that $\sigma(\xi_i) \in \mathcal{C}^1$. We will first extend the proof to MADS polling for all Lipschitz functions, and then discuss optimization with constraints.

5.3. Stochastic unconstrained optimization with MADS polling

Audet and Dennis [31] showed that some restrictions on the GPS techniques can be removed. Some of their important results for $\hat{x}$, the limit point of a refining subsequence using the GPS algorithm are:

- If σ is Lipschitz near $\hat{x}$, then Clarke's generalized derivatives satisfy

  $$\sigma^\circ(\hat{x}, d) \equiv \limsup_{y\to\hat{x}, t\downarrow 0} \frac{\sigma(y + td) - \sigma(y)}{t} \geqslant 0$$

  for a subset of directions $d \in D$ that form a positively spanning basis.
- If σ is regular at $\hat{x}$, then the directional derivatives satisfy

  $$\sigma'(\hat{x}, d) \equiv \lim_{t\downarrow 0} \frac{\sigma(\hat{x} + td) - \sigma(y)}{t} \geqslant 0$$

  for a subset of directions $d \in D$ that form a positively spanning basis.
- If σ is strictly differentiable at $\hat{x}$, then $\nabla\sigma(\hat{x}) = 0$.

By virtue of the definition of the generalized derivatives, and the proof of Theorem 3.7 in [31], we can show that at a specific collocation point:

$$\sigma^{\circ}(\tilde{x}(\xi_i); d(\xi_i)) \geqslant \limsup_{k \in K} \frac{\sigma(x_k(\xi_i) + \Delta_k d(\xi_i)) - \sigma(x_k(\xi_i))}{\Delta_k} \quad \forall i.$$

The main difference between GPS and MADS is that we choose a dense set of searching directions in MADS (called polling). In this case, the set D approaches a dense set as the mesh is refined [18]. Hence, with the MADS algorithm, $D \equiv \mathbb{R}^n$. A detailed proof, showing that the MADS algorithm produces a constrained Clarke stationary point for Lipschitz functions (which becomes a stationary point for strictly differentiable functions) is given in [18]. Following these proofs, we outline a proof for stochastic MADS polling that is analogous to the proof for stochastic GPS. We first expand the first two moments of σ and its sensitivities using Lagrange interpolates as:

$$\sigma^m_{\Delta \mathbf{z}} = \sum_{i=1}^{N_z} \sigma^m_{\Delta \mathbf{z}(\xi_i)} \mathcal{L}_i(\xi) \quad m = 1, 2.$$

For the stochastic SMF algorithm, the deterministic proofs can be applied directly to the moments at stochastic collocation points, $\sigma^m_{\Delta \mathbf{z}(\xi_i)}$. The deterministic proofs can be extended to their stochastic counterparts in the spirit of the proofs for Theorems 5.1 through 5.5. By showing that any arbitrary moment converges, we conclude that the PDF converges. Hence, the SMF stochastic optimization technique with MADS converges to a stochastic stationary point for strictly differentiable functions (a stochastic Clarke stationary point for Lipschitz functions).

5.4. Stochastic constrained optimization with MADS polling

Different techniques are available for implementing constraints with the GPS techniques which include – (a) extreme barrier approach where the cost function value is set to ∞ outside the feasible region, (b) the filter method [32] which accepts a trial point if it either improves feasibility or cost function value, and (c) a progressive barrier approach where a constraint barrier is constantly evolved [34]. Convergence proofs for GPS with the extreme barrier approach for linear and bound constraints were given by Lewis and Torczon [36,37] for continuously differentiable functions. Audet and Dennis [32] propose the filter method for general constraints and proved convergence with the GPS technique. Their main convergence result shows that for any Lipschitz σ, $\sigma^{\circ}(\tilde{x}(\xi_i); d(\xi_i)) \geqslant 0$ for all refining d generated by the GPS algorithm and such that $d \in T(\omega, \tilde{x})$, the tangent cone in the feasible domain at the limit point $\tilde{x}$.

The progressive barrier method for imposing constraints with MADS polling has shown to possess stronger convergence properties [34]. However, no convergence proof has been published yet for the MADS technique with the filter approach. An attempt for such a proof is beyond the scope of this article. We feel that the choice of filter approach is justified based on some of the results (filter plots) obtained in this paper. In the future, we also plan to implement the progressive barrier approach [38] which possesses a good convergence theory.

6. Numerical examples

The first two numerical examples are reliability-based problems in mechanics that are motivated by the examples treated in [15].

6.1. Problem 1: stochastic optimization with probabilistic constraints – robust beam design

In this example, we explore the efficiency of the technique for a four stochastic dimensional problem. In addition, probabilistic inequality constraints are imposed on the problem.

Problem statement

This problem is based on a robust design problems involving beams. The objective is to minimize the cost associated with construction of the beam while certain failure-based criterion must be adhered to for desirable performance. In addition, there are diverse sources of uncertainties such as material uncertainties (Young's modulus, yield stress etc.) or uncertainties in boundary conditions (loads, displacements). The problem is posed mathematically as :

$$\begin{aligned}
&\underset{w,h}{\text{minimize}} \ \ wh \\
&\text{such that} \ \int \mathbf{I}(\sigma(\boldsymbol{\xi}) - \sigma_c) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \leqslant p_f \\
&\int \mathbf{I}(d(\boldsymbol{\xi}) - d_c) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \leqslant q_f \\
&1 \leqslant w \leqslant 4 \\
&1 \leqslant h \leqslant 4,
\end{aligned}$$

where $\mathbf{I}(\cdot)$ represents the indicator function and takes the value 1 when its argument is greater than 0 and is 0 otherwise. The maximum stress in the beam, σ, is given by $\sigma = \frac{600}{wh^2}F_y + \frac{600}{w^2h}F_x,\ \sigma_c = R,\ d = \frac{4L^3}{Ewh}\sqrt{\left(\frac{F_y}{h^2}\right)^2 + \left(\frac{F_x}{w^2}\right)^2}$ and $d_c = D_0$. F_x and F_y represent the x and y direction forces, σ_c represents a critical value of stress based failure which is a material property, d represents the displacement, L is a characteristic length, E represents Young's modulus, and d_c represents critical displacement. w and h are the width and height respectively (and the d.v.s), and the objective is to minimize weight of the beam. The uncertain or stochastic variables are $R = \mathcal{N}(40000, 2000),\ E = \mathcal{N}(2.9e7, 1.45e6),\ F_x = \mathcal{N}(500, 100)$ and $F_y = \mathcal{N}(1000, 100)$. Deterministic parameters include $L = 100$ in and $D_0 = 2.2535$ in. $\mathcal{N}(\mu, \sigma^2)$ represents a Gaussian distribution with mean μ and variance σ^2.p_f and q_f are set to 10^{-5}.

To solve this problem, a four stochastic dimensional sparse grid was constructed first. This was constructed using the Smolyak algorithm on a $[0, 1]^4$ hypercube. Using the inverse-cdf transformation of the normal distribution, this was converted into $\mathcal{R}^4$ space. Different levels of interpolants were tested to check how the stochastic SMF optimization algorithm performed. The parameters of the SMF algorithm were fixed through all the cases for the stochastic interpolation. The termination criterion for optimization is when the MADS mesh refinement parameter reaches a critical value of $\delta = 1/32$. The constraints are incorporated into the optimization algorithm using a filter method that has been described in [28]. Denoting, $c_1 = \left(\int \mathbf{I}(\sigma(\xi) - \sigma_c)p(\xi)d\xi - p_f\right)$ and $c_2 = \left(\int \mathbf{I}(d(\xi) - d_c)p(\xi)d\xi - q_f\right)$, we define the constraint violation function H as:

$$H = \mathbf{I}(c_1)|c_1| + \mathbf{I}(c_2)|c_2|.$$

H quantifies how much the constraints are violated and is non-zero only if one of the constraints is violated. Any number of constraints can be encompassed within H by summation. We enforce strong constraint conditions which means H must equal zero for any acceptable solution.

The constraints are evaluated using the stochastic collocation technique as follows:

$$\int \mathbf{I}(\sigma(\xi) - \sigma_c)p(\xi)d\xi = \int \mathbf{I}\left(\sum_i \sigma(\xi_i)\mathcal{L}_i(\xi) - \sigma_c\right)p(\xi)d\xi,$$

$$\int \mathbf{I}(d(\xi) - d_c)p(\xi)d\xi = \int \mathbf{I}\left(\sum_i d(\xi_i)\mathcal{L}(\xi) - d_c\right)p(\xi)d\xi.$$

The solution vectors and the cost function associated with them are shown in Table 2.

As observed in the table, we note that it is not mathematically possible to distinguish between two d.v. vectors that give the optimal cost function while satisfying the constraints. Convergence of the cost function is achieved with increasing depth of interpolation as shown in Fig. 4. The reason for the slow convergence is because tails of the PDF are used to compute the constraints and an accuracy greater than 3σ is required to capture the constraint violation. Fig. 5 shows the PDF of the value of constraint violation plotted on its sample space. A better approach will be to adapt the collocation points according to the function which has been attempted in [39]. A plot of the objective and the constraint violation function through the algorithm is depicted in Fig. 6. A satisfactory solution is obtained with depth of interpolation 3, and a modest further improvement is made in level 7. The optimal d.v. computed $w = 2.38, h = 3.36$ using the final interpolation depth 7 is indeed a feasible point.

6.2. Problem 2: stochastic constraint optimization

In this example, we explore how input PDFs that are non-normal/uniform can be captured in the optimization procedure. This problem has the same physical motivation as the previous problem. We include non-normal PDFs here.

Problem Statement: The problem is posed as follows :

$$\begin{aligned}
&\underset{b,h}{\text{minimize}}\ \ bh\\
&\text{such that}\ \int \mathbf{I}(g(\xi) - g_c)p(\xi)d\xi \leqslant g_f\\
&5 \leqslant b \leqslant 15\\
&15 \leqslant h \leqslant 25,
\end{aligned}$$

where $\mathbf{I}(\cdot)$ represents the indicator function and takes the value 1 when its argument is greater than 0, and is 0 otherwise. The function g is given by $g = 1 - \frac{4M}{bh^2Y} - \frac{P^2}{b^2h^2Y^2}$ and $g_c = 0$. The uncertain or stochastic variables are: the load $P = \mathcal{N}(500, 100)$, the external moment $M = \mathcal{N}(2000, 400)$ and Young's modulus $Y = logn(5, 0.5)$. The design variables are the width b and the height h, and the objective is to minimize weight. g_f is set to 10^{-5}.

Collocation points are constructed in a 3D stochastic space using Smolyak's algorithm. Increasing levels of interpolants were tested to assess the performance of the stochastic optimization algorithm. This test problem reveals the performance

Table 2
A tabulation of the optimal solutions computed at different depths of sparse grid interpolation for Problem 1. We note that there are multiple parameter vectors providing the same cost function, hence the solution vectors are different though the cost is identical and constraints are satisfied. "Hierch" denotes the hierarchical optimization algorithm. The total number of evaluations is high since we need an exact representation of the entire PDF with four stochastic dimensions (to capture the probabilistic constraints accurately). With this technique, even level 2 interpolation gives a solution close to the optimal.

Depth of interp.	No. colloc. pts	Best feas. pt.	Opt. cost fn	Tot. fn. eval.
2	41	(2.2,3.85)	8.47	2419
3	137	(2.35,3.42)	8.03	8905
5	1105	(2.46,3.26)	8.03	72930
7	7537	(2.38,3.36)	8.01	550201
Hierch	NA	(2.38,3.36)	8.01	359336
MC	100000	(2.38,3.36)	8.01	8400000

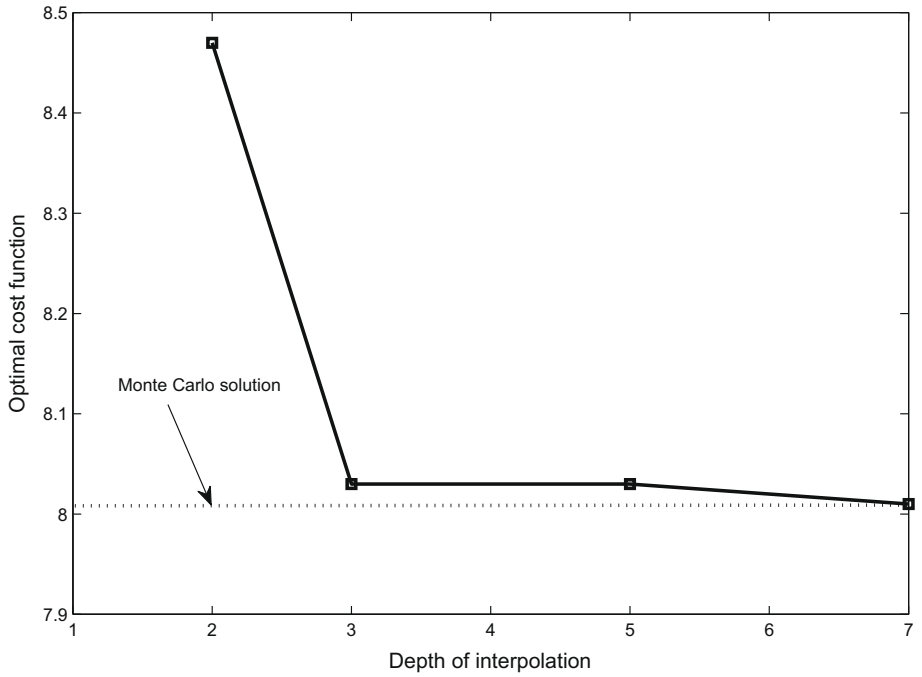

Fig. 4. Convergence of best feasible solutions for different depths of interpolation for Problem 1 with multiple probabilistic constraints. The optimal solution computed using the Monte-Carlo technique is shown with a dashed line.

of our technique when the initial surrogates are a poor approximation due to extremely sharp variations in the function. Similar to the previous example problem, we define a constraint function H as:

$$H = \mathbf{I}(c_3)|c_3|,$$

where $c_3 = \left(\int \mathbf{I}(g(\xi) - g_c)p(\xi)d\xi - g_f\right)$. H quantifies the magnitude of constraint violation. The constraint is evaluated using stochastic collocation technique in a similar fashion to the previous problem.

$$\int \mathbf{I}(g(\xi) - g_c)p(\xi)d\xi = \int \mathbf{I}\left(\sum_i g(\xi_i)\mathcal{L}_i(\xi) - g_c\right)p(\xi)d\xi.$$

The stochastic SMF technique with filtering was employed to compute the optimal solution as shown in Table 3. Due to the lognormal nature of the distribution of Y, the constraints cannot be captured using small to moderate depths of interpolation. Hence, levels 2, 4 and 6 showed the absence of any feasible region owing to poor interpolation of the function. We observe that as we further refine the stochastic space, we are able to capture the optimal solution. In fact, at level 10, the optimal point violated the constraint by less than 0.3% and the optimal value is shown to be less than that computed using Monte-Carlo technique. A plot of the constraint function violation versus the cost function is depicted in Fig. 7.

The number of evaluation points in Tables 1 and 2 are very large and it may not be feasible to perform as many function evaluations in practical engineering problems. Techniques such as adaptive collocation [39] can be used to improve the efficiency, but the reduction in number of collocation points is problem specific. Hence, depending on the problem, this might not represent an accurate picture of the stochastic SMF algorithm. We show here that we can build accuracy by increasing

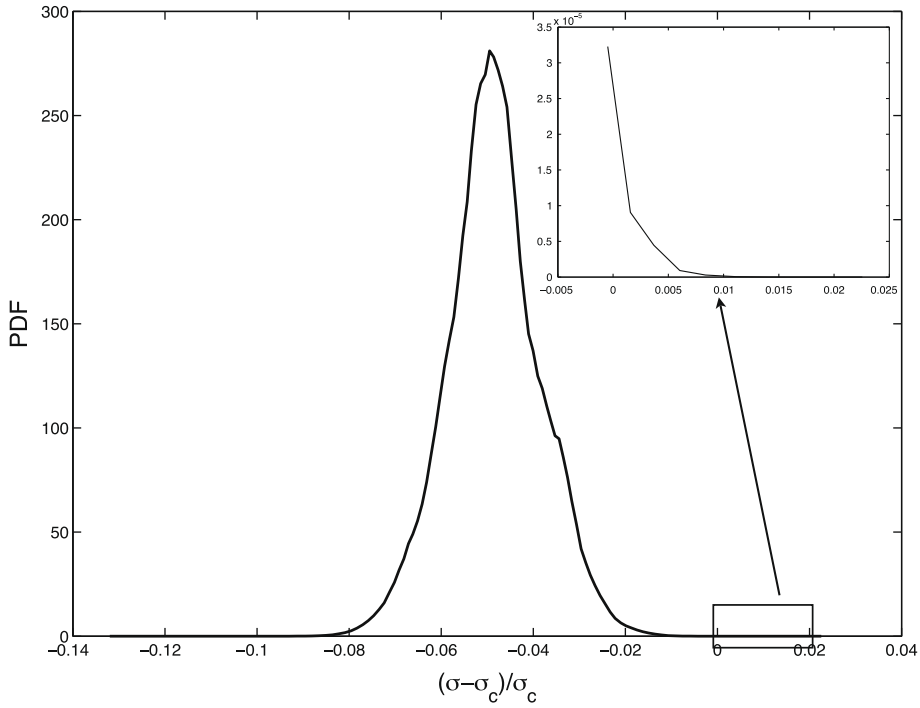

Fig. 5. The figure shows PDF of how much the stresses exceed their critical values normalized with σ_c. The PDF of the sample space where constraints are violated is shown in the inset.

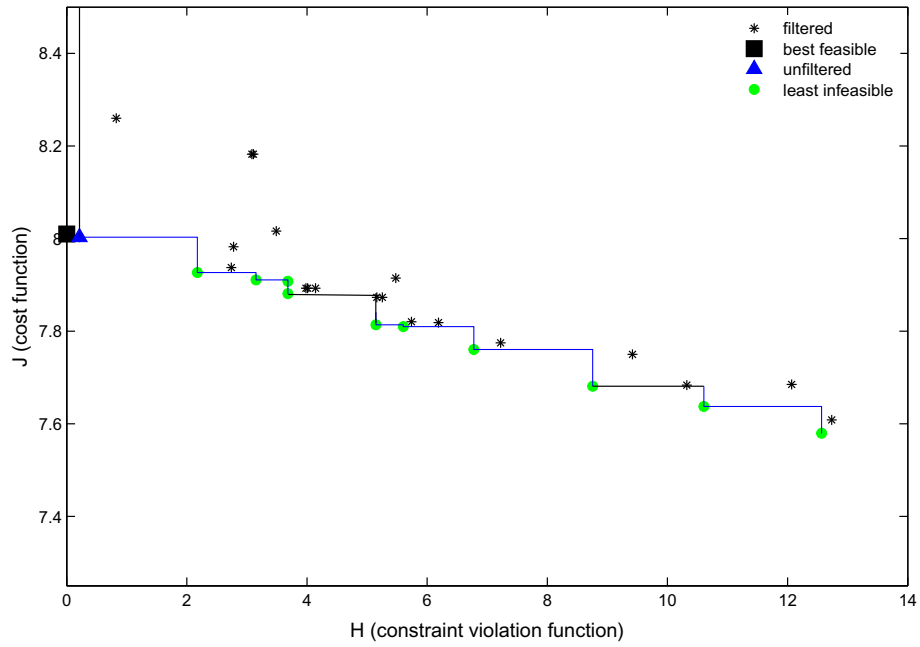

Fig. 6. A plot of the cost function versus the constraint violation function for Problem 1.

the depth of stochastic interpolation. One advantage of the filter method is that the user can specify a tolerance for the constraint function and at the cost of accuracy, stop the simulations after a reasonable improvement in design solution is achieved. Since these were benchmark problems, we enforce the constraint conditions to match exactly. In practical engineering design problems, the goal often lies in a significant improvement of the cost function rather than computing the exact optima, and the number of evaluations will be significantly less in practice.

Table 3
Optimal solutions at different depths of sparse grid interpolation for Problem 2 are shown. For depths of interpolation less than 10, no feasible points were found due to poor interpolation at certain regions in the stochastic space. Finally, at the depth of interpolation 10, optimal solutions are captured.

Depth of interp.	No.colloc. pts	Best feas. pt.	Opt. cost fn	Tot. fn. eval.
2	25	–	–	1250
4	177	–	–	10089
6	1073	–	–	59015
10	32001	(5,19.5)	97.5	1856058
Hierch	NA	(5,19.5)	97.5	1760055
MC	500000	(5,19.75)	98.75	25000000

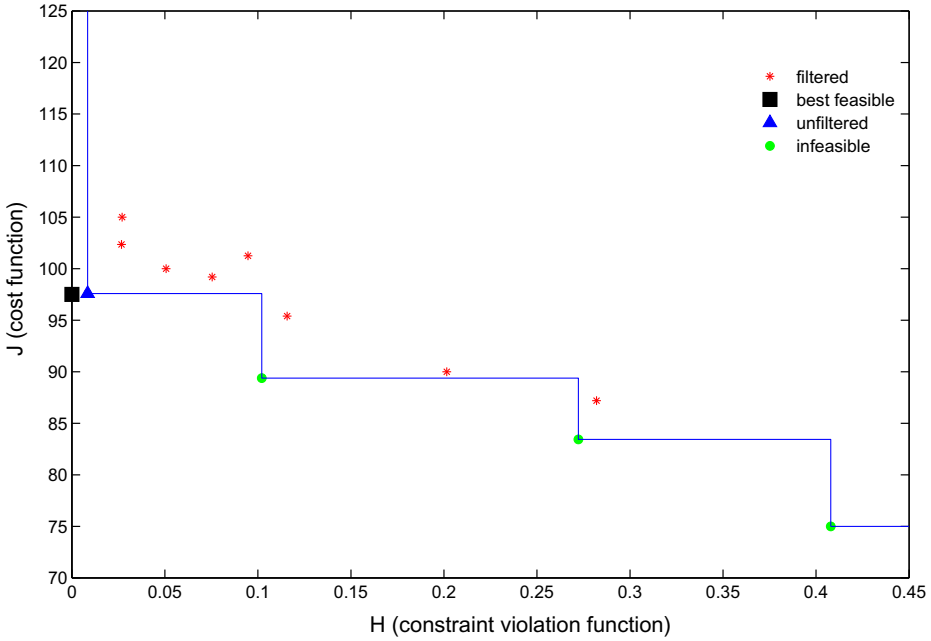

Fig. 7. A plot of the cost function versus the constraint violation function for Problem 2.

6.3. Problem 3: optimization with stochastic d.v.s

In this problem, we present an example in which the d.v.s themselves can be PDFs.

Problem statement

The governing DE for heat conduction equation in any crust can be expressed as:

$$\frac{d}{dz}\left(K\frac{dT}{dz}\right) = -A(z) = -A_0 exp\left(-\frac{z}{D}\right), \tag{17}$$

where K represents thermal conductivity, T represents temperature, z represents the depth, D represents the characteristic depth and $A(z)$ is an exponential radiogenic heat source. The solution to the one-dimensional heat conduction problem is given analytically by [40]:

$$T = T_0 + \frac{Q}{K}z + \frac{A_0 D^2}{K}\left(1 - exp\left(-\frac{z}{D}\right) - \frac{z}{D}exp\left(-\frac{L}{D}\right)\right). \tag{18}$$

In the stochastic inverse heat conduction problem, a PDF of temperature is provided and the PDF of heat flux is to be computed. The problem we solve is representative of a stochastic inverse thermal problem [1] in which we typically employ data (such as Temperature) to reconstruct some parameter (such as heat flux at the boundary). We also show that multimodal PDFs can be accurately captured by this technique by analyzing two scenarios - an unimodal heat flux and a bimodal heat flux.

The cost function is given by $J = \int (T - T_{\text{measured}})^2 p(\xi) d\xi$. The d.v.s are the heat flux at specific collocation points. We assume that no information other that the solution to the direct problem (computing temperature given a heat flux) is available. The unknown parameter space is modeled using a single stochastic variable (it has been shown in [10] that even if the actual stochastic dimension of the heat flux is more than one, its PDF can always be captured using one stochastic variable). Of course, as the PDF becomes complex, more collocation points are needed which means the stochastic problem will reduce to a high dimensional deterministic optimization problem. The deterministic parameters used in this problem are $T_0 = 0°\text{C},\ K = 2.5\ \text{W m}^{-1}\ \text{C}^{-1},\ A_0 = 20\ \mu\text{W m}^{-3},\ D = 10\ \text{km},\ L = 35\ \text{km}$ and z at which temperature is measured is 20 km.

The recomputed heat flux for the unimodal case is shown in Fig. 8. Due to the unimodal nature of the heat flux, even the level-1 sparse grid interpolant with three points is able to approximately capture the PDF and with nine stochastic points (at level 3), the PDF of heat flux is captured almost exactly.

The results for the bimodal case (Fig. 9) are in stark contrast to case 1. In this case, the level-2 sparse grid interpolant (five points) is not able to capture the heat flux. We notice that only with a interpolation level of around 8 (257 points), are we able to accurately capture the actual PDF. This confirms that the required depth of interpolation depends on the underlying PDF to be constructed. Yet, with sufficient computational resources, the stochastic SMF optimization technique can capture arbitrary PDFs having multiple modes.

6.4. Problem 4: concentration reconstruction

Our final example is a porous media problem. The temporal evolution of solute concentration is specified. The governing equations are given by:

$$\frac{\partial c}{\partial t} = -M(t)\left[v_\psi \frac{\partial}{\partial x} c(x,t) - D_\psi \frac{\partial^2}{\partial x^2} c(x,t)\right], \tag{19}$$

where M accounts for unknown heterogeneities below the level of measurement resolution and c represents resident concentration. The transport velocity, v_ψ is the unknown variable that is to be computed and due to the multiscale nature of the problem, the dispersion coefficient, D_ψ is uncertain. We assume here that D_ψ follows the lognormal distribution with mean -4 and standard deviation 0.2. The reader is referred to [41] and [42] for further details. The CTRW toolbox [43] for contaminant transport in porous media is used as a black box simulator for the problem. The following transform is used to convert stochastic collocation points from [0,1] support space (ξ space) to the lognormal support space (η space):

$$\eta = e^{\left(erf^{-1}(2\xi-1)\sqrt{2}\sigma+\mu\right)}.$$

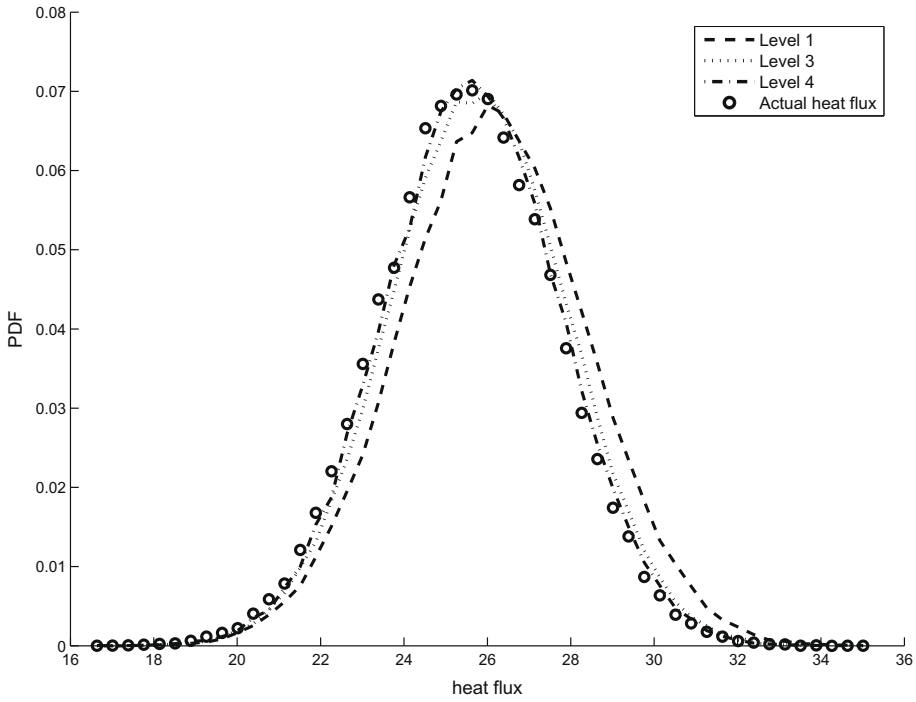

Fig. 8. Comparison of optimal solution from SMF to actual heat flux. The convergence of PDF of heat flux with increasing number of collocation points for unimodal case is shown.

The concentration measurements are given temporally as $c_{\text{meas}}(1,t)$. The task of reconstructing ν_ψ is posed as an optimization problem and the cost function is given by $J = (\mathbf{E}(c(1,t)) - c_{\text{meas}}(1,t))^2$ where $c(1,t)$ is computed using the CTRW toolbox

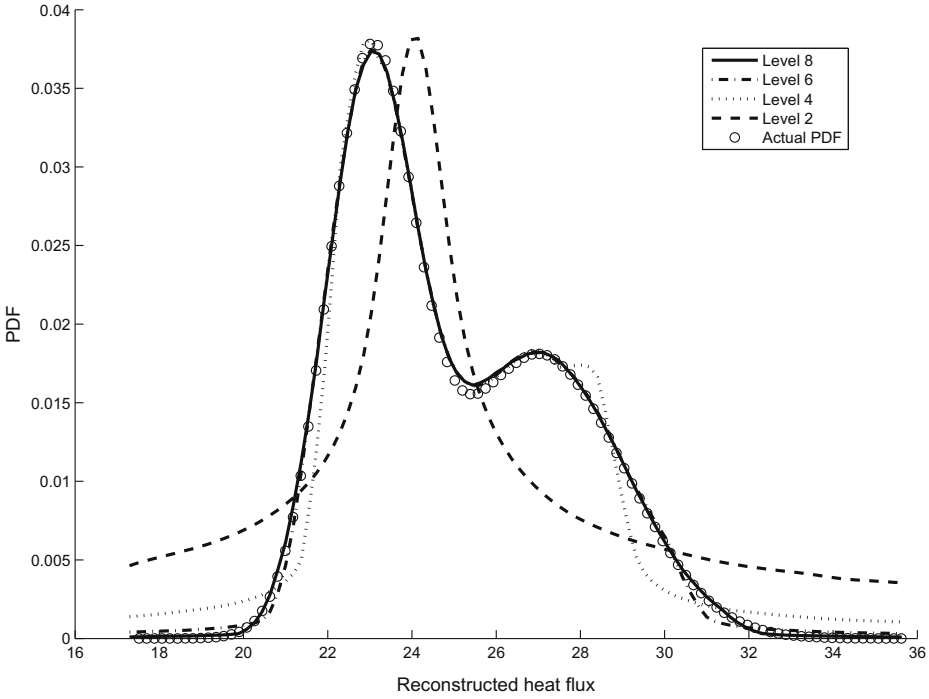

Fig. 9. Convergence of PDF of heat flux with different levels of interpolation for the bimodal case.

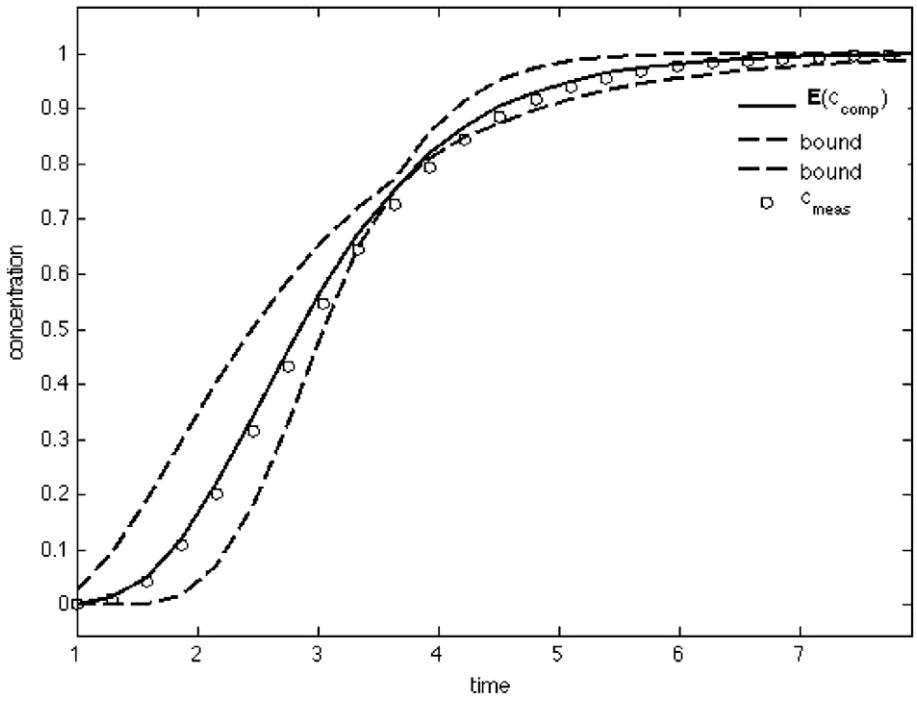

Fig. 10. The figure shows a comparison between the mean of the reconstructed concentration and the original concentration profiles. Also, the bounds on the reconstructed concentration profiles are shown which originates due to uncertainty in the dispersion coefficient. c_{comp} represents the recomputed concentrations.

by solving Eq. 19 using the Laplace transform method, and $\mathbf{E}(\cdot)$ represents the expectation operator. The stochastic SMF optimization technique was employed to solve this problem. A depth of interpolation of 5 with 33 collocation points was found to be sufficient (Fig. 10).

Since the dispersion coefficient is not known with certainty, only a range of possible concentration profiles can be computed using the stochastic optimization technique. However, since the target data was the mean concentration, the expected value of the predicted concentrations matches the measured concentrations. This problem illustrates how an arbitrary black box PDE-based solver can be seamlessly integrated with our algorithm.

7. Summary and future work

In this work, we developed a stochastic optimization toolbox to efficiently perform optimization in the presence of uncertainties and in the absence of gradient information. A general algorithm was developed to perform optimization with both stochastic parameters and stochastic design variables. We extended the convergence proof of the GPS technique by showing convergence of mean and variances of the cost function and its gradient, and extending it to PDFs. In addition to the cost function, the constraints can also be probabilistic. We successfully tested the method on two problems in solid mechanics with reliability constraints. We then computed the PDF of heat fluxes based on a PDF of temperature measurements for a stochastic inverse heat conduction problem. The technique can be used with a deterministic black box solver and this ability was showcased using an example involving contaminant flow in porous media.

With this technique, we are able to get a computational gain of the order of 15 to 20 for problems with strict probabilistic constraints, where tails need to be resolved to more than three standard deviations. However, for unconstrained problems with smooth PDFs, a computational gain of more than 2 orders of magnitude is achieved over the MC technique. We can further improve this by employing an adaptive collocation algorithm [39]. The number of d.v.s that can be optimized is limited since the number of function evaluations can become unwieldy if we consider hundreds of d.v.s. One way to deal with this is to construct a coarse stochastic response surface and identify the most important d.v.s before performing optimization.

In future work, we aim to make the optimization process more robust by improving the `POLL` step. The technique presented here is flexible by design. It can be easily extended to incorporate recent advances in both polling [38] as well as implementing constraints [34]. In a recently developed technique called OrthoMADS, the directions of polling are chosen in such a way that they are orthogonal to other polling directions. This ensures a more uniform distribution of polling directions compared to the original MADS technique. We also plan to extend the algorithm to include the progressive barrier approach for implementing constraints. Further, we plan to use an adaptive technique for selecting the stochastic collocation points that will generate a denser array of points close to areas of steep gradients.

Acknowledgments

This work was supported by a Burroughs Wellcome Fund Career Award at the Scientific Interface. We would like to thank Prof. John Dennis Jr. for his insightful comments and suggestions on this work.

References

[1] S. Sankaran, Stochastic optimization using a sparse grid collocation scheme, Probabilistic Engineering Mechanics 24 (3) (2009) 382–396.
[2] A.L. Marsden, M. Wang, J.E. Dennis Jr., P. Moin, Trailing-edge noise reduction using derivative-free optimization and large-eddy simulation, Journal of Fluid Mechanics 572 (2007) 13–36.
[3] A.L. Marsden, J.A. Feinstein, C.A. Taylor, A computational framework for derivative-free optimization of cardiovascular geometries, Computer Methods in Applied Mechanics and Engineering 197 (21–24) (2008) 1890–1905.
[4] D. Buche, N.N. Schraudolph, P. Koumoutsakos, Accelerating evolutionary algorithms with Gaussian process fitness function models, IEEE Transactions on Systems, Man and Cybernetics, Part C 35 (2) (2005) 183–194.
[5] N. Metropolis, S. Ulam, The Monte Carlo method, Journal of the American Statistical Association 44 (247) (1949) 335–341.
[6] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, New York, 1991.
[7] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM Journal of Scientific Computing 24 (2002) 619–644.
[8] D. Xiu, G.E. Karniadakis, Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, Computer Methods in Applied Mechanics and Engineering 191 (2002) 4927–4948.
[9] D. Xiu, J.S. Hesthaven, High order collocation methods for the differential equation with random inputs, SIAM Journal of Scientific Computing 27 (2005) 1118–1139.
[10] B. Ganapathysubramaniam, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, Journal of Computational Physics 225 (2007) 652–685.
[11] I. Babuska, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM Journal of Numerical Analysis 45 (3) (2007) 1005–1034.
[12] G.J.A. Loeven, J.A.S. Witteveen, H. Bijl, Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties, AIAA paper 2007-317, Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit, Reno (NV), United States, 2007.
[13] J.R.R.A. Martins, I.M. Kroo, J.J. Alonso, A method for sensitivity analysis using complex variables, 38th AIAA Aerospace Sciences Meeting and Exhibit, AIAA Paper Reno (NV), 2000-0689, 2000.
[14] N. Zabaras, B. Ganapathysubramanian, A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach, Journal of Computational Physics 227 (2008) 4697–4735.
[15] M.S. Eldred, C.G. Webster, P.G. Constantine, Design under uncertainty employing stochastic expansion methods, American Journal of Aeronautics and Astronautics (2008) 1–22.

[16] B. Faverjon, R. Ghanem, Stochastic inversion in acoustic scattering, Journal of Acoustic Society of America 119 (6) (2006) 3577–3588.
[17] A.J. Booker, J.E. Dennis Jr., P.D. Frank, D.B. Serafini, V. Torczon, M.W. Trosset, A rigorous framework for optimization of expensive functions by surrogates, Structural Optimization 17 (1) (1999) 1–13.
[18] C. Audet, J.E. Dennis Jr., Mesh adaptive direct search algorithms for constrained optimization, SIAM Journal on Optimization 17 (1) (2006) 2–11.
[19] N.V. Queipo, R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, P. Kevin Tucker, Surrogate-based analysis and optimization, Progress in Aerospace Sciences 41 (1) (2005) 1–28.
[20] A.L. Berger, V.J.D. Pietra, S.A.D. Pietra, A maximum entropy approach to natural language processing, Computational Linguistics 22 (1) (1996) 39–71.
[21] S.C. Zhu, Y. Wu, D. Mumford, Filters, random fields and maximum entropy (FRAME): towards a unified theory for texture modeling, International Journal of Computer Vision 27 (2) (2004) 107–126.
[22] S. Sankaran, N. Zabaras, Computing property variability of polycrystals induced by grain size and orientation uncertainties, Acta Materialia 55 (7) (2007) 2279–2290.
[23] M. Loève, Probability Theory, fourth ed., Springer-Verlag, 1977.
[24] A. Klimke, Uncertainty Modeling using Fuzzy Arithmetic and Sparse Grids, Ph.D. Thesis, Universitt Stuttgart, Shaker Verlag, Aachen, 2006.
[25] A. Klimke, B. Wohlmuth, Algorithm 847: spinterp: Piecewise Multilinear Hierarchical Sparse Grid Interpolation in MATLAB, ACM Transactions on Mathematical Software 31 (2005).
[26] A. Klimke, Sparse Grid Interpolation Toolbox Users Guide, IANS report 2006/001, University of Stuttgart, 2006.
[27] T. Gerstner, M. Griebel, Numerical integration using sparse grids, Numerical Algorithms 18 (1998) 209–232.
[28] A.L. Marsden, M. Wang, J.E. Dennis Jr., P. Moin, Optimal aeroacoustic shape design using the surrogate management framework, optimization and engineering, Special Issue: Surrogate Optimization 5 (2004) 235–262.
[29] G. Matheron, Principles of Geostatistics, Economic Geology 58 (1963) 1246–1266.
[30] M.D. McKay, W.J. Conover, R.J. Beckman, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (1979) 239–245.
[31] C. Audet, J.E. Dennis Jr., Analysis of generalized pattern searches, SIAM Journal on Optimization 13 (3) (2003) 889–903.
[32] C. Audet, J.E. Dennis Jr., A pattern search filter method for nonlinear programming without derivatives, SIAM Journal on Optimization 14 (4) (2004) 980–1010.
[33] V. Torczon, On the convergence of pattern search algorithms, SIAM J. Optimization 7 (1997) 1–25.
[34] C. Audet, J.E. Dennis Jr., A progressive barrier approach to derivative-free nonlinear programming, SIAM Journal on Optimization 20 (1) (2009) 445–472.
[35] <http://www2.imm.dtu.dk/~hbn/dace/>.
[36] R.M. Lewis, V. Torczon, Pattern search algorithms for bound constrained minimization, SIAM Journal on Optimization 9 (1999) 1082–1099.
[37] R.M. Lewis, V. Torczon, Pattern search methods for linearly constrained minimization, SIAM Journal on Optimization 10 (2000) 917–941.
[38] M.A. Abramson, C. Audet, J.E. Dennis Jr., S. Le Digabel, OrthoMADS: a deterministic MADS instance with orthogonal directions, SIAM Journal on Optimization 20 (2) (2009) 948–966.
[39] S. Sankaran, A.L. Marsden, A stochastic collocation method for uncertainty quantification in cardiovascular simulations, Journal of Biomechanical Engineering, submitted for review.
[40] K. Srivastava, R.N. Singh, A stochastic model to quantify the steady-state crystal geotherms subject to uncertainties in thermal conductivity, Geophysical Journal International 138 (1999) 895–899.
[41] A. Cortis, Y. Chen, H. Scher, B. Berkowitz, Quantitative characterization of pore-scale disorder effects on transport in homogeneous granular media, Physical Review E 70 (4) (2004) 1–8.
[42] A. Cortis, C. Gallo, H. Scher, B. Berkowitz, Numerical simulation of non-Fickian transport in geological formations with multiple-scale heterogeneities, Water Resources Research 40 (4) (2004) 1–16.
[43] <http://www.weizmann.ac.il/ESERold/People/Brian/CTRW/software.html>.