Journal of Computational Physics 248 (2013) 87–108

Contents lists available at SciVerse ScienceDirect

Journal of Computational Physics

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

Reweighted ℓ_1 minimization method for stochastic elliptic differential equations

Xiu Yang, George Em Karniadakis *

Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

ARTICLE INFO

Article history:
Received 1 June 2012
Received in revised form 28 March 2013
Accepted 1 April 2013
Available online 18 April 2013

Keywords:
Compressive sensing
Generalized polynomial chaos
Chebyshev probability measure
High-dimensions

ABSTRACT

We consider elliptic stochastic partial differential equations (SPDEs) with random coefficients and solve them by expanding the solution using generalized polynomial chaos (gPC). Under some mild conditions on the coefficients, the solution is "sparse" in the random space, i.e., only a small number of gPC basis makes considerable contribution to the solution. To exploit this sparsity, we employ reweighted ℓ_1 minimization to recover the coefficients of the gPC expansion. We also combine this method with random sampling points based on the Chebyshev probability measure to further increase the accuracy of the recovery of the gPC coefficients. We first present a one-dimensional test to demonstrate the main idea, and then we consider 14 and 40 dimensional elliptic SPDEs to demonstrate the significant improvement of this method over the standard ℓ_1 minimization method. For moderately high dimensional ($\sim$10) problems, the combination of Chebyshev measure with reweighted ℓ_1 minimization performs well while for higher dimensional problems, reweighted ℓ_1 only is sufficient. The proposed approach is especially suitable for problems for which the deterministic solver is very expensive since it *reuses* the sampling results and exploits all the information available from limited sources.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Due to the great interest in uncertainty quantification (UQ) for computational engineering applications in the past decade, several new methods have been proposed for the numerical solution of stochastic partial differential equations (SPDEs). For example, *generalized polynomial chaos* (gPC), [1,2] and its extensions, e.g., *multi-element generalized polynomial chaos* (ME-gPC) [3,4] have been successfully applied to a stochastic flow and other problems, where the number of uncertain parameters is not too large. Alternatively, *probabilistic collocation method* (PCM) based on sparse grid integration/interpolation [5–7] is also very popular due to its simplicity. Different adaptivity strategies [8–10] have further improved the efficiency of these types of methods. Another way of implementing adaptivity based on the analysis of variance (ANOVA) has also been very effective, e.g., [11–14], and can potentially be used in high-dimensions if sufficient sparsity in the representation of the solution exists.

In this paper we consider problems which exhibit sparsity, i.e., the quantity of interest is "sparse" in random space, and hence the solution can be accurately represented with only a few terms when linearly expanded into a stochastic, e.g., gPC, basis. This is similar to adaptive gPC approach [15]. It is also similar to dimension reduction methods, where the aim is to retain only the most important dimensions, hence the solution can be approximated economically without significantly

* Corresponding author. Tel.: +1 401 863 1217.
E-mail address: george_karniadakis@brown.edu (G.E. Karniadakis).

0021-9991/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcp.2013.04.004

sacrificing accuracy [9,10,13,16]. For many high-dimensional problems, careful analysis has shown that the number of basis functions with large coefficients is small relative to the cardinality of the full basis. For example, it has been shown in [17,18] that under some mild conditions, solutions to *elliptic* SPDEs with high-dimensional random coefficients admit sparse representations with respect to gPC basis; see also [19,20].

Doostan and Owhadi [21] proposed a method for gPC expansion of sparse solutions to SPDEs based on compressive sampling techniques. This method is non-adapted, provably convergent and well suited to problems with high-dimensional random inputs. They applied it to an elliptic equation with random coefficients and demonstrated good results. Their approach provides a more flexible method than the sparse grid method in that the number of the sampling points at a different level of the sparse grid method is fixed once the dimension is decided while number of the sampling points of this method is arbitrary (but of course generally speaking, more samples will lead to more accurate results). For example, in one of our numerical examples in this paper, where $d = 40$, i.e., 40 random variables are considered, level 1 sparse grid method requires 81 samples while level 2 sparse grid method needs 3281 samples. In dynamic data-driven applications, e.g., weather forecast, petroleum engineering, ocean modelings, the system can be extremely complicated and the simulation is very costly, and only a small number, e.g., $\mathcal{O}(100)$, of simulations can be afforded. Therefore, it is impossible to obtain more accurate results than level 1 sparse grid or even the level 1 sparse grid may not be applicable when the dimension is really high. Hence, the compressive sampling technique provides a feasible approach for such type of problems provided that based on physical knowledge, mathematical analysis or experience, the solution is sparse in random space. Our main improvement of the method proposed by Doostan and Owhadi is that by employing the reweighted procedure, which is popular in compressive sensing [22–24], we can greatly enhance the accuracy of the approximated solution. Furthermore, we combine the technique of selecting sampling points based on Chebyshev measure [25] to enhance the performance of the results.

This paper is organized as follows. In Section 2 we briefly review the main theorems of compressive sensing method as well as the reweighted procedure and the sampling strategy based on Chebyshev probability measure. In Section 3 we describe the set up of the problem of which numerical tests are performed. In Section 4, we present the results of numerical tests of 1D, 14D and 40D (in random space) linear elliptic SPDEs.

2. Brief review of ℓ_1 minimization

2.1. Basic concepts

Consider an underdetermined linear system of equations $\boldsymbol{\Psi c} = \boldsymbol{u}$, where $\boldsymbol{\Psi}$ (the "measurement matrix") is an $m \times N$ matrix with $m < N$ (usually $m \ll N$), and $\boldsymbol{c}$ and $\boldsymbol{u}$ (the "observation") are vectors of length N and m, respectively. In order to find a "sparse" solution, we consider the following optimization problem:

$$(P_{h,\epsilon}) : \min_{c} \|\boldsymbol{c}\|_h \quad \text{subject to} \quad \|\boldsymbol{\Psi c} - \boldsymbol{u}\|_2 \leqslant \epsilon, \tag{2.1}$$

where we introduce the tolerance ϵ to make $(P_{h,\epsilon})$ more general since in data-driven systems there may be noise in the observations. The sparsity of the solution $\boldsymbol{c}$ is best described by the ℓ_0 "norm" (number of nonzero entries of $\boldsymbol{c}$):

$$\|\boldsymbol{c}\|_0 \overset{\text{def}}{=} |\{i : c_i \neq 0\}|,$$

as smaller ℓ_0 "norm" indicates more sparse $\boldsymbol{c}$. In order to avoid the NP-hard problem of exhaustive search to solve $(P_{0,\epsilon})$, a type of greedy algorithm called *orthogonal matching pursuit* (OMP) can be employed [26]. However, this single-term-at-a-time strategy can fail badly, i.e., there are explicit examples (see [27–29]) where a simple k-term representation is possible, but this approach yields an n-term (i.e., dense) representation. Alternatively, one can relax the ℓ_0 "norm" to ℓ_1 norm, therefore replace $(P_{0,\epsilon})$ with a convex optimization problem $(P_{1,\epsilon})$, which is the ℓ_1 minimization method in compressive sensing. If the solution is "nearly sparse" instead of "sparse", i.e., the zero entries of $\boldsymbol{c}$ are replaced by relatively small numbers which will only change $\boldsymbol{\Psi c}$ a little, we can still use $(P_{h,\epsilon})$ by considering this small change as "noise" in the observation. In the problems we consider in this paper, the exact solutions are always *nearly sparse*. Next, we recall two fundamental concepts in the ℓ_1 minimization.

Definition 2.1 (*Mutual coherences [26,30]*). The mutual coherence $\mu(\boldsymbol{\Psi})$ of a matrix $\boldsymbol{\Psi} \in \mathbb{R}^{m\times N}$ is the maximum of absolute normalized inner-products of its columns. Let Ψ_j and Ψ_k be two columns of $\boldsymbol{\Psi}$. Then,

$$\mu(\boldsymbol{\Psi}) \overset{\text{def}}{=} \max_{1\leqslant j,k\leqslant N j\neq k} \frac{|\Psi_j^T \Psi_k|}{\|\Psi_j\|_2 \|\Psi_k\|_2}. \tag{2.2}$$

In other words, the mutual coherence measures how close $\boldsymbol{\Psi}$ is to orthogonal matrix. It is clear that for a general matrix A,

$$0 \leqslant \mu(A) \leqslant 1.$$

For instance, if A is unitary matrix, then $\mu(A) = 0$. However, since we always consider the case of $m < N$ in compressive sensing, $\mu(\boldsymbol{\Psi})$ is strictly positive. It is understood that a measurement matrix with smaller mutual coherence can better recover a sparse solution by compressive sensing techniques, e.g., Lemma 3.4 of [21].

Definition 2.2 (*Restricted isometries [31]*). Let $\boldsymbol{\Psi}$ be a matrix with a finite collection of vectors $(\Psi_j)_{j\in J} \in \mathbb{R}^m$ as columns, where J is a set of indices. For every integer $1 \leqslant s \leqslant |J|$, we define the s-restricted isometry constant δ_s to be the smallest quantity such that $\boldsymbol{\Psi}_T$ obeys

$$(1-\delta_s)\|\boldsymbol{c}\|_2^2 \leqslant \|\boldsymbol{\Psi}_T\boldsymbol{c}\|_2^2 \leqslant (1+\delta_s)\|\boldsymbol{c}\|_2^2 \tag{2.3}$$

for all subsets $T \subset J$ of cardinality at most s, and all real coefficients $(c_j)_{j\in T}$. Here $\boldsymbol{\Psi}_T$ is the submatrix of $\boldsymbol{\Psi}$ with column indices $j \in T$ so that

$$\boldsymbol{\Psi}_T\boldsymbol{c} = \sum_{j\in T} c_j\Psi_j.$$

In other words, for any *s-sparse* vector $\boldsymbol{c}$ (the number of non-zero entries is at most s), δ_s measures whether $\boldsymbol{\Psi}_T$ behaves like an orthonormal matrix. Informally, the matrix $\boldsymbol{\Psi}$ is said to have the restricted isometry property (RIP) if δ_s is small for s reasonably large compared to m.

Note that both the concepts of mutual coherence and restricted isometry describe the same property of the measurement matrix, i.e., roughly speaking, how far sparse subsets of the columns of it are from being an isometry. This is a key idea in compressive sensing, and to construct a measurement matrix with small mutual coherence or small restricted isometry constant is crucial. In this paper we mainly concentrate on the latter when we refer to theoretical analysis. A well known theorem is the following:

Theorem 2.3 (*Sparse recovery for RIP-matrices [32]*). *Assume that the restricted isometry constant of $\boldsymbol{\Psi}$ satisfies $\delta_{2s} < \sqrt{2}-1$. Let $\boldsymbol{c}$ be an arbitrary signal with noisy measurements $\boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{c} + \boldsymbol{e}$, where $\|\boldsymbol{e}\|_2 < \epsilon$. Then the approximation $\hat{\boldsymbol{c}}$ to $\boldsymbol{c}$ from ℓ_1 minimization $(P_{1,\epsilon})$ satisfies*

$$\|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_2 \leqslant C_1\epsilon + C_2\frac{\sigma_s(c)_1}{\sqrt{s}}, \tag{2.4}$$

where

$$C_1 = \frac{2\alpha}{1-\rho}, \quad C_2 = \frac{2(1+\rho)}{1-\rho}, \quad \rho = \frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}, \quad \alpha = \frac{2\sqrt{1+\delta_{2s}}}{\sqrt{1-\delta_{2s}}}, \quad \sigma_s(\boldsymbol{c})_p = \inf_{z:\|\boldsymbol{z}\|_0\leqslant s}\|\boldsymbol{c}-\boldsymbol{z}\|_p.$$

Remark 2.4. Notice that in Theorem 2.3 the bound for δ_{2s} is $\sqrt{2}-1$. There are sharper estimates for this bound, e.g. [33,34]. Since we are not concentrating on ℓ_1 minimization itself, we quote the original estimate by Candès [32].

Remark 2.5. Instead of Eq. (2.4), there is another form of estimate [21,26,30], which relates the error with the tolerance ϵ and the number of basis N and provides the probability of solutions satisfying the error bound.

In classical compressive sensing method, several choices of the measurement matrix are employed, e.g., Gaussian random matrix, Bernoulli random matrix, etc. For SPDEs, we can expand the solution with gPC basis:

$$u(\boldsymbol{x};\xi) = \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{x})\psi_{\boldsymbol{\alpha}}(\xi), \tag{2.5}$$

where $\boldsymbol{x}$ is the variable in the physical space and ξ is the random vector, $\psi_{\boldsymbol{\alpha}}$ are gPC basis functions, and $\boldsymbol{\alpha}$ are the indices. For each fixed $\boldsymbol{x}, u$ is a linear combination of the gPC basis, hence when different sampling points $\xi_1, \xi_2, \ldots,$ are employed, Eq. (2.5) can be cast into a linear system:

$$\boldsymbol{\Psi}\boldsymbol{c} = \boldsymbol{u}, \tag{2.6}$$

where each entry of the measurement matrix is $\Psi_{i,j} = \psi_{\boldsymbol{\alpha}_j}(\xi_i)$ with ξ_i being sampling points in random space, and each entry of vector $\boldsymbol{u}$ being $u_i = u(\boldsymbol{x};\xi_i)$, i.e., the output of the deterministic solver with sampling point ξ_i. We call these u_i samples of the solution. In [21] the authors expanded the solutions with Legendre polynomial and used Monte Carlo points based on uniform distribution. According to the law of large numbers and the orthogonality of the Legendre polynomials, the mutual coherence converges to zero for asymptotically large random sample sizes m. Also, with this choice of sampling points, the compressive sensing method can be considered as a post processing of Monte Carlo method and the benefit is that we can arbitrarily increase the number of sampling points without wasting any realizations we already obtained. We note that the nested sparse grid method can also reuse the sampling points of the lower levels while the number of additional samples from one level to the next level is fixed.

2.2. Reweighted ℓ_1 minimization

Candès et al. [22] proposed the reweighted ℓ_1 minimization, which employed the weighted norm and iterations to enhance the sparsity of the solution. This algorithm consists of the following four steps:

1. Set the iteration count l to zero and $w_i^{(0)} = 1,\ i = 1,\ldots,N$.
2. Solve the weighted ℓ_1 minimization problem

$$\boldsymbol{c}^{(l)} = \arg\min \|W^{(l)}\boldsymbol{c}\|_1 \quad \text{subject to} \quad \|\Psi\boldsymbol{c} - \boldsymbol{u}\|_2 \leqslant \epsilon,$$

where W is a diagonal matrix with $W_{jj}^{(l)} = w_j^{(l)}$.
3. Update the weights: for each $i = 1, 2, \ldots, N$,

$$w_i^{(l+1)} = \frac{1}{|c_i^{(l)}| + \tau}.$$

4. Terminate upon convergence or when l reaches a specified maximum number of iterations $l_{\max}$. Otherwise, increment l and go to step 2.

Remark 2.6. According to [22], in step 3, the parameter τ is introduced to provide stability and to ensure that a zero-valued component in $\boldsymbol{c}^{(l)}$ does not strictly prohibit a nonzero estimate at the next step. It should be set slightly smaller than the expected nonzero magnitudes of $\boldsymbol{c}_0$, which is a sparse vector approximating $\boldsymbol{c}$. Empirically, τ is set around 0.1–0.001. In step 4, the convergence of $\boldsymbol{c}$ can be tested by comparing the difference between $\boldsymbol{c}^{(l)}$ and $\boldsymbol{c}^{(l+1)}$. Usually, $l_{\max}$ is set so that it controls the number of iterations.

From the description of this algorithm, we notice that it repeats the ℓ_1 minimization for different weighted ℓ_1 norm in different steps. Hence, when $l_{\max}$ is fixed, we actually need $l_{\max}$ additional optimization solves compared to the regular ℓ_1 minimization. A series of tests in [22–24] demonstrated remarkable performance and broad applicability of this algorithm in the areas of sparse signal recovery, statistical estimation, error correction and image processing with not only sparse but also nearly-sparse representations. Moreover, these tests showed that usually three or four iterations are enough to obtain good performance. Notice that in this algorithm there is no requirement or modification for the measurement matrix, therefore it can be applied to any linear system including (2.6). Needell [35] provided an analytical result of the improvement in the error bound by using the reweighted ℓ_1 minimization over the ℓ_1 minimization under some conditions:

Theorem 2.7 [35]. *Assume that ψ satisfies the RIP condition with $\delta_{2s} < \sqrt{2} - 1$. Let $\boldsymbol{c}$ be an arbitrary vector with noisy measurements $\boldsymbol{y} = \Psi\boldsymbol{c} + \boldsymbol{e}$, where $\|\boldsymbol{e}\|_2 < \epsilon$. Assume that the smallest nonzero coordinate η of $\boldsymbol{c}_s$ satisfies $\eta > \frac{4\alpha\epsilon_0}{1-\rho}$, where $\epsilon_0 = 1.2(\|\boldsymbol{c} - \boldsymbol{c}_s\|_2 + \frac{1}{\sqrt{s}}\sigma_s(\boldsymbol{c})_1 + \epsilon$ and $\|\boldsymbol{c}_s - \boldsymbol{c}\|_1 = \sigma_s(\boldsymbol{c})_1$. Then the limiting approximation from reweighted ℓ_1 minimization satisfies*

$$\|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_2 \leqslant \frac{4.1\alpha}{1+\rho}\left(\epsilon + \frac{\sigma_{s/2}(\boldsymbol{c})_1}{\sqrt{s}}\right), \tag{2.7}$$

and

$$\|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_2 \leqslant \frac{2.4\alpha}{1+\rho}\left(\epsilon + \frac{\sigma_s(\boldsymbol{c})_1}{\sqrt{s}} + \|\boldsymbol{c} - \boldsymbol{c}_s\|_2\right), \tag{2.8}$$

where the definitions of $\alpha, \rho, \sigma_s(\boldsymbol{c})_p$ are the same as in Theorem 2.3.

In practice, we only use one to three reweighted iterations, Lemma 3.4 in [35] is more useful (see Appendix A).

2.3. Chebyshev measure for Legendre polynomials

According to the gPC theory, Legendre polynomial is the appropriate basis for the uniformly distributed random variables [2]. Therefore, if the system involves uniform random variables the solution can be expanded in terms of *Legendre polynomials* and the selection of sampling points are based on the uniform distribution. With this setting, the mutual coherence $\mu(\Psi)$ converges to zero almost surely for asymptotically large random sample size m as pointed out in [21], and several theories were proposed to estimate the number of samples for different error level ϵ. Noticing the special structure of the measurement matrix Ψ consisting of Legendre polynomial, Rauhut and Ward [25] proposed a sampling strategy based on the Chebyshev measure $\mathrm{d}\nu(x) = \pi^{-1}(1-x^2)^{-1/2}\mathrm{d}x$ for recovering the Legendre sparse polynomials from a few samples for 1D (random space) problem. We know that if a random variable X is uniformly distributed on $[0,\pi]$, then $Y = \cos X$ is distributed according to the Chebyshev measure. Hence, in order to generate a sampling point based on Chebyshev measure, we first generate a sampling point x according to $U[0,\pi]$, then $y = \cos(x)$ is the point we need. Equation $(P_{1,\epsilon})$ is now modified as

$$(P_{1,\epsilon}^{Cheb}): \min_{\boldsymbol{c}} \|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \|A\boldsymbol{\Psi}\boldsymbol{c} - A\boldsymbol{u}\|_2 \leqslant \epsilon, \tag{2.9}$$

where A is an $m \times m$ diagonal matrix with $A_{j,j} = (\pi/2)^{1/2}(1-\xi_j^2)^{1/4}$ and $\xi_j,\ j = 1,2,\ldots,m$ are the sampling points based on Chebyshev measure, $\psi_{j,n} = L_{n-1}(\xi_j)$ and L_n is the nth order normalized Legendre polynomial. As was shown in [25], this strategy can improve the accuracy of the recovery for 1-D (random space) problem. Moreover, this method can be extended to a large class of orthogonal polynomials, including the Jacobi polynomials, of which the Legendre polynomials are a special case. The main theoretical conclusion is the following:

Definition 2.8 ([25]). Let $\psi_j, j \in [N] \overset{\text{def}}{=} \{1,2,\ldots,N\}$ be an orthonormal system of functions on $\mathcal{D}$, i.e.,

$$\int_{\mathcal{D}} \psi_j(x)\psi_k(x)\mathrm{d}\nu(x) = \delta_{j,k}, \quad j,k \in [N].$$

If this orthonormal system is uniformly bounded,

$$\sup_{j\in[N]}\|\psi_j\|_\infty = \sup_{j\in N}\sup_{x\in\mathcal{D}}|\psi_j(x)| \leqslant K \tag{2.10}$$

for some constant $K \geqslant 1$, we call the systems $\{\psi_j\}$ satisfying this condition *bounded orthonormal systems*.

Theorem 2.9 [25] RIP for bounded orthonormal systems. *Consider the matrix $\boldsymbol{\Psi} \in \mathbb{R}^{m\times N}$ with entries*

$$\Psi_{l,k} = \psi_k(x_l), \quad l \in [m], k \in [N],$$

formed by i.i.d. samples x_l drawn from the orthogonalization measure ν associated to the bounded orthonormal system $\{\psi_j, j \in [N]\}$ having uniform bound $K \geqslant 1$ in (2.10). If

$$m \geqslant C\delta^{-2}K^2 s\log^3(s)\log(N), \tag{2.11}$$

then with probability at least $1 - N^{-\gamma\log^3(s)}$, the restricted isometry constant δ_s of $\frac{1}{\sqrt{m}}\boldsymbol{\Psi}$ satisfies $\delta_s \leqslant \delta$. The constants $C, \gamma > 0$ are universal.

Remark 2.10 ([25]). Consider the functions

$$Q_n(x) = \sqrt{\frac{\pi}{2}}(1-x^2)^{1/4}L_n(x),$$

where L_n are nth order normalized Legendre polynomials. According to Lemma 5.1 in [25], $|Q_n(x)| < \sqrt{2+\frac{1}{n}}$ and $\{Q_n(x)\}$ forms a bounded orthonormal system with respect to Chebyshev measure. The matrix Φ with entries $\Phi_{j,n} = Q_{n-1}(x_j)$, which is used in the constraint $\|\Phi\boldsymbol{c} - A\boldsymbol{u}\| \leqslant \epsilon$, can be written as $\Phi = A\boldsymbol{\Psi}$ as in $(P_{1,\epsilon}^{Cheb})$. Then the system $\{Q_n\}$ is uniformly bounded on $[-1,1]$ and satisfies the bound $\|Q_n\|_\infty \leqslant \sqrt{3}$, i.e., $K = \sqrt{3}$ for this system. Therefore, according to Theorem 2.9, we can expect a high probability of good recovery based on a relatively small number of samples compared to regular ℓ_1 minimization.

We can extend the Chebyshev measure based sampling points strategy from 1-D to multi-D. Since $\xi_1, \xi_2, \ldots, \xi_d$ are i.i.d random variables and the basis functions used to present the solution are in the tensor product form (see Eq. (3.4)), we modify $(P_{1,\epsilon}^{Cheb})$ by setting $A_{j,j} = \prod_{k=1}^{d'}(\pi/2)^{1/2}(1-(\boldsymbol{\xi}_j)_k)^{1/4}$, i.e., we use the tensor product form in the weight as in [36]. Here $\boldsymbol{\xi}_j$ is the jth sampling point, which includes d entries and $(\boldsymbol{\xi}_j)_k$ is the k entry of it. Notice that we use d' instead of d here because according to our numerical tests, $d' = d$ does not necessarily provide good performance, hence it is better to take $d' < d$. This is consistent with the conclusion in [36], which shows that the Chebyshev measure may become less efficient in high-dimensional cases. An important reason is that the uniform bound K in Theorem 2.9 will increase as d increases, which implies that more samples are needed, see also [36]. More precisely, when generating a d dimensional sampling point $(\xi_1, \xi_2, \ldots, \xi_d)$, we generate the first d' entries $\xi_1, \xi_2, \ldots, \xi_{d'}$ according to Chebyshev measure and the remaining entries $\xi_{d'+1}, \ldots, \xi_d$ according to uniform distribution.

A brief summary of the theorems presented in this section is as follows: Theorem 2.9 and Remark 2.10 provide the estimate of the size of sampling points we need to obtain a measurement matrix with small RIP constant when the sampling points are based on Chebyshev measure. After obtaining observations and the measurement matrix, Theorem 2.7 shows that by employing reweighted iterations, we can expect a lower error bound than that in Theorem 2.3, which is the error estimate for regular ℓ_1 minimization. Alternatively, if the number of sampling points is fixed, Chebyshev points allow a larger s (but still relative small compared with N to guarantee sparsity) in the RIP condition, hence the upper bound of the error in Theorem 2.3 can be reduced. Then, with the reweighted iterations, this bound can be reduced further.

3. Problem set up

We use the same setting as in [21]: let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space where $\mathcal{P}$ is a probability measure on the σ-field $\mathcal{F}$. We consider the following SPDE defined on a bounded Lipschitz continuous domain $\mathcal{D} \subset \mathbb{R}^D$, $D = 1, 2, 3$, with boundary $\partial\mathcal{D}$,

$$\begin{aligned} &-\nabla \cdot (a(\boldsymbol{x};\omega)\nabla u(\boldsymbol{x};\omega)) = f(\boldsymbol{x}) \quad \boldsymbol{x} \in \mathcal{D}, \\ &u(\boldsymbol{x};\omega) = 0 \quad \boldsymbol{x} \in \partial\mathcal{D}, \end{aligned} \tag{3.1}$$

where ω is an "event" in probability space Ω. The diffusion coefficients is represented by a Karhunen–Loève (KL) expansion:

$$a(\boldsymbol{x};\omega) = \bar{a}(\boldsymbol{x}) + \sigma_a \sum_{i=1}^{d} \sqrt{\lambda_i}\phi_i(\boldsymbol{x})\xi_i(\omega), \tag{3.2}$$

where (λ_i, ϕ_i), $i = 1, 2, \ldots, d$ are eigenpairs of the covariance function $C_{aa}(\boldsymbol{x}_1, \boldsymbol{x}_2) \in L_2(\mathcal{D} \times \mathcal{D})$ of $a(\boldsymbol{x};\xi), \bar{a}(\boldsymbol{x})$ is the mean and σ_a is the standard deviation. ξ_i are i.i.d uniformly distributed random variables on $[-1, 1]$. Additional requirement on $a(\boldsymbol{x};\omega)$ are:

- For all $\boldsymbol{x} \in \mathcal{D}$, there exists constants $a_{\min}$ and $a_{\max}$ such that

  $$0 < a_{\min} \leqslant a(\boldsymbol{x};\omega) \leqslant a_{\max} \leqslant \infty \quad \text{a.s.}$$

- The covariance function $C_{aa}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is piecewise analytic on $\mathcal{D} \times \mathcal{D}$ [18], implying that there exist real constants c_1 and c_2 such that for $i = 1, \ldots, d$,

  $$0 \leqslant \lambda_i \leqslant c_1 e^{-c_2 i^{\kappa}},$$

where $\kappa := 1/D$ and $\boldsymbol{\alpha} \in \mathbb{N}^d$ is a fixed multi-index.

Here the second requirement guarantees the existence of a sparse solution for problem (3.1) [18]. With the setting of Eq. (3.2), a and u depend on $\boldsymbol{x}$ and ξ, so we omit ω and use the notation $a(\boldsymbol{x};\xi)$ and $u(\boldsymbol{x};\xi)$.

In the context of the gPC method, the solution of (3.1) is represented by an infinite series of the tensor form:

$$u(\boldsymbol{x};\xi) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_0^d} c_{\boldsymbol{\alpha}}(\boldsymbol{x})\psi_{\boldsymbol{\alpha}}(\xi), \tag{3.3}$$

where

$$\psi_{\boldsymbol{\alpha}}(\xi) = \psi_{\alpha_1}(\xi_1)\psi_{\alpha_2}(\xi_2)\cdots\psi_{\alpha_d}(\xi_d), \quad \alpha_i \in \mathbb{N} \cup \{0\}. \tag{3.4}$$

In practice we truncate this expression, e.g., up to order P:

$$u(\boldsymbol{x};\xi) \approx u_P^*(\boldsymbol{x};\xi) \overset{\text{def}}{=} \sum_{|\boldsymbol{\alpha}|=0}^{P} c_{\boldsymbol{\alpha}}(\boldsymbol{x})\psi_{\boldsymbol{\alpha}}(\xi). \tag{3.5}$$

By selecting m different sampling points $\xi_1, \xi_2, \ldots, \xi_m$ we obtain m different samples of $u : \boldsymbol{u} = (u_1, u_2, \ldots, u_m)$. Hence, we rewrite (3.5) as: $\boldsymbol{u} \approx \boldsymbol{\Psi c}$, where $\Psi_{ij} = \psi_{\boldsymbol{\alpha}_j}(\xi_i)$ and ψ is an $m \times N$ matrix with $N = \frac{(P+d)!}{P!d!}$. Now we can use the ℓ_1 minimization to seek a solution $\boldsymbol{c}$ satisfying:

$$(P_{1,\epsilon}) : \min_c \|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \|\boldsymbol{\Psi c} - \boldsymbol{u}\|_2 \leqslant \epsilon, \tag{3.6}$$

where ϵ is related to the truncation error. It is clear that the more terms we include in the expansion (3.5) the more accurate result we will get, which in turn means we can put smaller ϵ in (3.6). Notice that, if ϵ is too large we will only obtain a less accurate result while if ϵ is too small, the so called *over fitting* problem, will cause a less sparse solution, which is also not accurate. Several methods like CoSaMP [37] or iterative hard thresholding [38] avoid the *a priori* knowledge of the noise level ϵ in the context of signal processing or image processing.

In order to prevent the optimization from biasing toward the non-zero entries in $\boldsymbol{c}$ whose corresponding columns in ψ have large norms, a weighted matrix W can be included in the ℓ_1 norm [26], thus we have

$$(P_{1,\epsilon}^W) : \min_c \|W\boldsymbol{c}\|_1 \quad \text{subject to} \quad \|\boldsymbol{\Psi c} - \boldsymbol{u}\|_2 \leqslant \epsilon, \tag{3.7}$$

where W is a diagonal matrix whose $[j,j]$ entry is the ℓ_2 norm of the jth column of ψ. For the original ℓ_1 minimization, $W = \mathrm{I}$, i.e., the identity matrix. This can also be considered as a special case of reweighted ℓ_1 method in that we put a non-identity weight matrix in the first step and employ no additional iterations. Once the coefficients are computed, we obtain the gPC expansion of the solution (3.5), and then we can estimate the statistics of the solution, e.g., $\mathbb{E}(u) = c_0, Var(u) = \sum_{|\boldsymbol{\alpha}|=1}^{P} c_{\boldsymbol{\alpha}}^2$ since $\psi_{\boldsymbol{\alpha}}$ are orthonormal with respect to the distribution of ξ. We summarize the above descriptions in Algorithm 1. In this paper we use "a trial" to denote completing Algorithm 1 once.

Algorithm 1. Reweighted ℓ_1 minimization method for elliptic equation (3.1)

1: Generate m sampling points $\xi_1, \xi_2, \ldots, \xi_m$ based on the distribution of ξ (or based on the Chebyshev measure if Chebyshev sampling points are employed). Run the deterministic solver to solve (3.1) for each ξ_i to obtain m samples of the solution $u_1, u_2, \ldots, u_m$. Denote $\boldsymbol{u} = (u_1, u_2, \ldots, u_m)$ and it is the "observation" in $(P_{1,\epsilon})$. The "measurement matrix" $\boldsymbol{\Psi}$ in $(P_{1,\epsilon})$ is $\Psi_{ij} = \psi_{\boldsymbol{\alpha}_j}(\xi_i)$, where $\psi_{\boldsymbol{\alpha}}$ are the basis functions in (3.3). The size of $\boldsymbol{\Psi}$ is $m \times N$, where N is the total number of basis functions depending on P in (3.5).

2: Set the tolerance ϵ in $(P_{1,\epsilon})$, set the iteration count $l = 0$ and weight $w_i^{(0)} = 1,\ i = 1, 2, \ldots, N$. If $(P_{1,\epsilon}^W)$ instead of $(P_{1,\epsilon})$ is implemented in the first step, select appropriate w_i based on the corresponding $(P_{1,\epsilon}^W)$ algorithm. Select the maximum number of the iterations $l_{\max}$, usually 2 or 3 is enough.

3: Solve the weighted ℓ_1 minimization problem

$$\boldsymbol{c}^{(l)} = \arg\min \|W^{(l)}\boldsymbol{c}\|_1 \quad \text{subject to} \quad \|\boldsymbol{\Psi}\boldsymbol{c} - \boldsymbol{u}\|_2 \leqslant \epsilon,$$

where W is a diagonal matrix with $W_{jj}^{(l)} = w_j^{(l)}$. If the Chebyshev sampling points are employed, solve

$$\boldsymbol{c}^{(l)} = \arg\min \|W^{(l)}\boldsymbol{c}\|_1 \quad \text{subject to} \quad \|A\boldsymbol{\Psi}\boldsymbol{c} - A\boldsymbol{u}\|_2 \leqslant \epsilon$$

instead.

4: Update the weights: for each $i = 1, 2, \ldots, N$,

$$w_i^{(l+1)} = \frac{1}{|c_i^{(l)}| + \tau}.$$

5: Terminate upon convergence or when l attains a specified maximum number of iterations $l_{\max}$. Otherwise, increment l and go to step 3.

6: Compute statistics of the solution after obtaining $\boldsymbol{c}$. For instance, $\mathbb{E}(u) = c_0, Var(u) = \sum_{i=1}^{N-1} c_i^2$, etc.

4. Numerical tests

In this section, we start with a 1D problem and then discuss multi-dimensional problems, where the *dimension* here refers to the random space. More precisely, in the problems we consider below, we refer dimension to "d" in Eq. (3.2). We will investigate the relative error in the mean $\varepsilon_m \overset{\text{def}}{=} |c_{|\boldsymbol{\alpha}|=0} - \mathbb{E}(u)|/|\mathbb{E}(u)|$ and the standard deviation $\varepsilon_s \overset{\text{def}}{=} |\left(\sum_{i=1}^{P} c_{|\boldsymbol{\alpha}|=i}^2\right)^{1/2} - \sigma(u)|/|\sigma(u)|$ of the solution at specific spatial point. Also, we will check the L_2 error of the numerical solution $\varepsilon_u = \|\tilde{u}_P^* - u\|_2/\|u\|_2$ at specific spatial point, where $\tilde{u}_P^*$ is the truncated gPC expansion with coefficients recovered by our method. The integral is calculated by the sparse grids method. All the ℓ_1 minimizations in this paper are achieved by the `SPGL1` package [39].

4.1. One-dimensional problem

We consider the problem:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}x}\left(a(x;\xi)\frac{\mathrm{d}}{\mathrm{d}x}u(x,\xi)\right) = -1, \quad x \in (0,1), \\ &u(0) = u(1) = 0. \end{aligned} \tag{4.1}$$

where ξ is uniformly distributed on $[-1, 1]$. We set $a(x;\xi) = a(\xi) = 1 + 0.5\xi$. The exact solution for this problem is

$$u(x;\xi) = \frac{x(1-x)}{2a(\xi)} = \frac{x(1-x)}{2+\xi}.$$

We investigate the solution at $x = 0.5$ and omit the symbol x unless confusion arises. The expectation of the solution is $\mathbb{E}(u) = \frac{1}{8}\ln 3$ and the standard deviation is $\sigma(u) = \sqrt{\frac{1}{48} - \frac{1}{64}(\ln 3)^2}$. The coefficients in the gPC expansion

$$u(\xi) = \sum_{j=0}^{\infty} c_j L_j(\xi) \tag{4.2}$$

can be computed analytically:

$$c_j = \int_{-1}^{1} u(\xi) L_j(\xi) \mathrm{d}v(\xi),$$

where L_j is jth order *normalized* Legendre polynomial. Fig. 1 presents the absolute values of c_j while those below 10^{-14} are neglected. We can see that the solution is *nearly sparse* as only eight coefficients are larger than 10^{-5}, which means that very few basis make considerable contribution to the solution. We also point out that it does not matter if we change the order of the basis since the compressive sensing method will detect the pattern of sparsity automatically. Moreover, the threshold 10^{-5} used here is only for demonstration purposes to present the sparsity of the coefficients. For different problems, the threshold can be different.

4.1.1. Uniformly distributed sampling points

We first employ the sampling points ξ_i based on uniform distribution on $[-1, 1]$ (will be called "uniform points"). The purpose of this test is to demonstrate the benefit of using reweighted iterations. In our test, ϵ and τ in the reweighted procedure are set empirically. A better choice of ϵ can be obtained by, e.g., cross-validation, which we adopt in the multi-dimensional tests below. We truncate the solution of (4.1) with $P = 80$, hence, $N = P + 1 = 81$, and try to recover the coefficients of the following gPC expansion:

$$u(\xi) \approx u_{80}^* = \sum_{j=0}^{80} c_j L_j(\xi). \tag{4.3}$$

In this test the truncation error with $P = 80$ is negligible, hence the L_2 error of the solution can be reflected by the l_2 error of the gPC coefficients. More precisely, since $\|u - u_P^*\|_2$ is negligible, $\|u - \tilde{u}_P^*\|_2$ is the same as $\|u_P^* - \tilde{u}_P^*\|_2$ which equals $\left(\sum_{j=0}^{P} |\tilde{c}_j - c_j|^2\right)^{1/2}$ (ℓ_2 error of gPC coefficients) since the gPC basis is orthonormal. Here $\tilde{u}_P^*$ is the approximation of u_P^* with coefficients $\tilde{\boldsymbol{c}}$ obtained by our method. We present the l_2 error by the gPC coefficients by using 10, 15 and 20 uniform points with and without reweighted iterations in Fig. 2. In all the tests we set $l_{\max}$ to be 2 in Algorithm 1. In order to obtain the histograms of the recovery error, we run 10,000 trials for each test and set $\epsilon = 10^{-4}, \tau = 8 \times 10^{-2}$. From Fig. 2(a)–(c) we observe that without using the reweighted iterations, 20 uniform points yield the best recovery in that bins denoting smaller error are higher as the number of samples increases while bins denoting larger error are lower. When the reweighted iterations are employed, i.e., (g)–(i), the recovery is much better, which can be observed by comparing (a) and (g), (b) and (h), (c) and (i), respectively. Also, comparing (b) with (g), we observe that with the reweighted iterations, using 10 uniform points can render a comparable result than using 15 uniform points without reweighting. Similarly, comparing (c) with (h), we see that by using 15 uniform points with reweighted iterations we obtain much better results than using 20 uniform points without reweighting.

Fig. 3 presents the improvement from the reweighted iterations by considering the ℓ_2 reconstruction error: $\|\boldsymbol{c} - \boldsymbol{c}^{(2)}\|_2 / \|\boldsymbol{c} - \boldsymbol{c}^{(0)}\|_2$, where $\boldsymbol{c}$ is the vector of exact coefficients. Notice that the upper bound of $\|\boldsymbol{c} - \boldsymbol{c}^{(0)}\|_2$ is given in Theorem 2.3 and the upper bound of $\|\boldsymbol{c} - \boldsymbol{c}^{(2)}\|_2$ is close to the one in Theorem 2.7. Since we set $l_{\max} = 2$, i.e., we compute $\boldsymbol{c}^{(0)}, \boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}$, where $\boldsymbol{c}^{(0)}$ is the results by ℓ_1 minimization; we only use two additional ℓ_1 optimizations. We observe that 56.2% of the 10-sample tests, 72.0% of the 15-sample tests and 68.9% of the 20-sample tests show a reduction of ℓ_2 reconstruction error up to 50% or more, i.e., $\|\boldsymbol{c} - \boldsymbol{c}^{(2)}\|_2 / \|\boldsymbol{c} - \boldsymbol{c}^{(0)}\|_2 \leqslant 0.5$. If we are more aggressive to check the percentage of reduction of ℓ_2 reconstruction error up to 90% or more, i.e., $\|\boldsymbol{c} - \boldsymbol{c}^{(2)}\|_2 / \|\boldsymbol{c} - \boldsymbol{c}^{(0)}\|_2 \leqslant 0.1$, the answer is 12.4% for the 10-sample

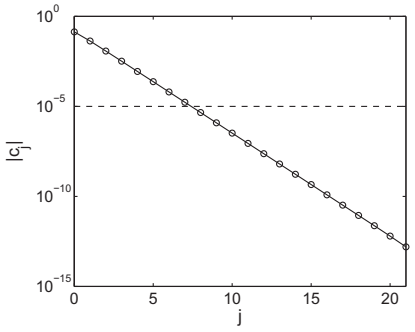


Fig. 1. 1D example: absolute values of 22 coefficients with the largest magnitude in the gPC expansion for Eq. (4.1) at $x = 0.5$ with $a = 1 + 0.5\xi$ and $\xi \sim U[-1, 1]$. Only 8 of these absolute values are larger than 10^{-5}.

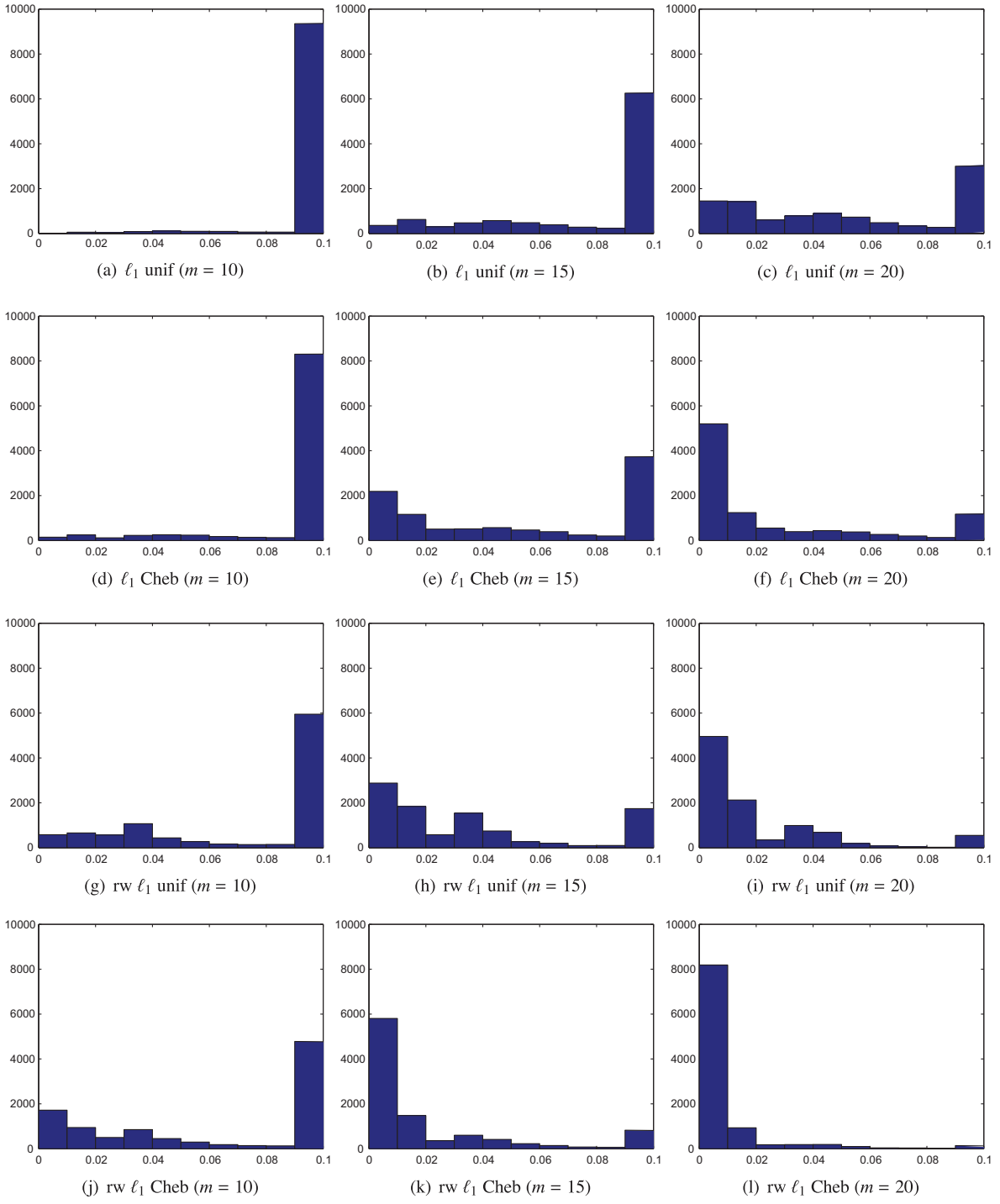


Fig. 2. 1D example: comparison of the relative error of the ℓ_2 norm of the gPC coefficients ($\|\boldsymbol{c} - \tilde{\boldsymbol{c}}\|_2/\|\boldsymbol{c}\|_2$, where $\boldsymbol{c}$ is the exact solution and $\tilde{\boldsymbol{c}}$ is the numerical solution) by using uniform points ("unif") and Chebyshev points ("Cheb") with ℓ_1 minimization and reweighted ℓ_1 minimization ("rw"). Number of basis $N = 81$, number of samples $m = 10$ (left column), $m = 15$ (middle column), $m = 20$ (right column). Total number of trials is 10,000. $l_{\max} = 2, \epsilon = 10^{-4}, \tau = 8 \times 10^{-2}$. x-axis presents the range of the relative error and the histograms demonstrate the number of trials with the relative error in specific ranges.

tests, 22.3% for the 15-sample tests and 17.9% for the 20-sample tests. Notice that with the change of the number of the sampling points m, the property of the measurement matrix $\boldsymbol{\Psi}$ changes as well. More precisely, the RIP condition ($\delta_{2s} < \sqrt{2} - 1$) in Theorems 2.3 and 2.7 will be satisfied for different s. (If m is too small, then this condition may not be satisfied for any

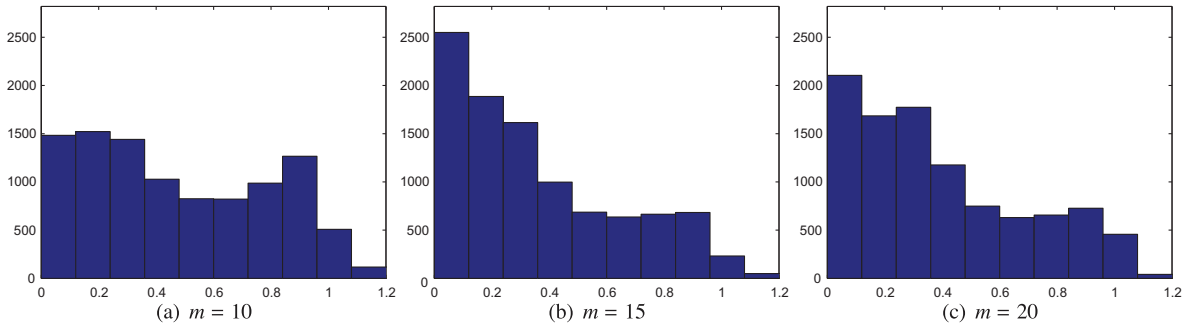


Fig. 3. 1D example: improvement of the reweighted iterations with uniform points by checking the ℓ_2 error: $\|\boldsymbol{c} - \boldsymbol{c}^{(2)}\|_2 / \|\boldsymbol{c} - \boldsymbol{c}^{(0)}\|_2$, where $\boldsymbol{c}$ is the vector of the exact coefficients. Total number of trials is 10,000. Number of basis $N = 81$, number of samples, $m = 10$ in (a), $m = 15$ in (b), $m = 20$ in (c). $l_{\max} = 2, \epsilon = 10^{-4}, \tau = 8 \times 10^{-2}$. x-axis presents the range of the improvement and the histograms demonstrate the number of trials with the improvement in specific ranges.

$s \leqslant N/2$). Hence, the error bound in Theorems 2.3 and 2.7 will be different, and the ratios of improvement are different as well. Finally, we also point out that in a few tests (less than 0.4%), the reweighted iterations show worse results.

The observations for the 1-D test are consistent with the conclusions of the ℓ_1 minimization and reweighted ℓ_1 minimization in [22] for the deterministic cases in that: (1) more sampling points will provide a higher probability of more accurate recovery; (2) reweighted iterations can enhance the sparsity of the solution, which in turn can improve the accuracy but there are also very small chances that the result is worse than ℓ_1 minimization.

4.1.2. Chebyshev measure based sampling points

Now we use the sampling points based on Chebyshev measure (will be called "Chebyshev points") to repeat the experiments in the last subsection. Fig. 2 also presents the relative error in the coefficients by using 10, 15 and 20 Chebyshev points with and without reweighted iterations. Comparing the first two rows in this figure, where we only use the ℓ_1 minimization and keep increasing the number of sampling points, we observe that without reweighted iterations Chebyshev points render better recovery than uniform points. This is consistent with Theorem 2.9, which claims that by using Chebyshev points we can reduce the number of sampling points to obtain a measurement matrix satisfying the RIP condition. Comparing the third and the fourth row, where reweighted iterations are employed, we observe that when the Chebyshev points are used, the reweighted iterations still improve the accuracy. We can also fix the number of the sampling points and compare the recovery accuracy with different sampling strategies. Let us consider the middle column for example, where the number of sampling points is fixed to be 15. We can observe from the pattern of the histogram that for this test case, reweighted iterations provide greater enhancement than using the Chebyshev points as (h) (where uniform points and reweighted iterations are employed) shows better results than (e) (where Chebyshev points and ℓ_1 minimization are employed) and the combination of these two ideas yields remarkable improvement as shown in (k). This phenomenon implies that with the same number of sampling points, Chebyshev points allows a larger s to satisfy the RIP condition (Theorem 2.9) and reweighted iterations further decreases the upper bound of the recovering error of $\boldsymbol{c}$ (Theorem 2.7). Moreover, the result in (k) (15 Chebyshev points with reweighted ℓ_1) is much better than that in (c) (20 uniform points with ℓ_1) and it is also better than the results in (f) (20 Chebyshev points with ℓ_1) and (i) (20 unform points with reweighted ℓ_1). Therefore, the combination of Chebyshev points and the reweighted iterations has the potential to perform good recovery with fewer sampling points compared with the ℓ_1 minimization method.

Since the inherent combinatorial nature of the RIP makes it impossible to directly compute the RIP constant of a matrix [40], we present the mutual coherences of the measurement matrices for different cases in Table 1 for comparison. We can observe that the mutual coherence is smaller when the Chebyshev points are used.

Moreover, the improvement of the accuracy of recovering the coefficient by using reweighted iterations with Chebyshev points is presented in Fig. 4. We observe that similar to the conclusion in section 4.1.1, reweighted iterations enhance the accuracy of recovery. Also, we notice that more than 1/4 of the tests with 20 Chebyshev points show almost no improvement, which is different from the result in Fig. 3(c), where the uniformly distributed sampling points are employed. This

Table 1
Mutual coherence μ of the measurement matrices of 1D tests. Means and standard deviations ("s.d.") of μ in tests with different m are presented.

	$m = 10$		$m = 15$		$m = 20$	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Uniform points	0.9282	0.0293	0.8467	0.0461	0.7785	0.0546
Chebyshev points	0.9166	0.0315	0.8193	0.0479	0.7355	0.0536

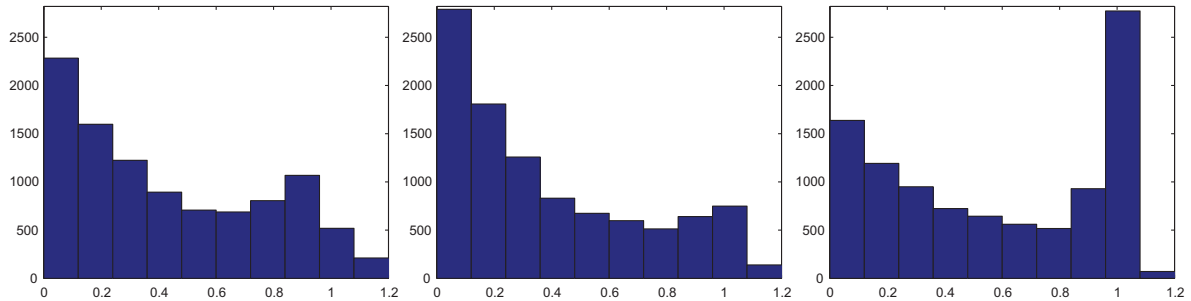


Fig. 4. 1D example: improvement of the reweighted iterations with Chebyshev points by checking the ℓ_2 error $\|\boldsymbol{c} - \boldsymbol{c}^{(2)}\|_2/\|\boldsymbol{c} - \boldsymbol{c}^{(0)}\|_2$, where $\boldsymbol{c}$ is the vector of the exact coefficients. Total number of trials is 10,000. Number of basis $N = 81$, number of samples, $m = 10$ in (a), $m = 15$ in (b), $m = 20$ in (c). $l_{\max} = 2, \epsilon = 10^{-4}, \tau = 8 \times 10^{-2}$. x-axis presents the range of the improvement and the histograms demonstrate the number of trials with the improvement in specific ranges.

difference means that 20 is a relative large number for the Chebyshev points. It is clear that if m is very large, e.g., $m = N$ we will obtain a very accurate recovery by ℓ_1 minimization, hence, reweighted iterations will provide very little improvement.

To conclude, in the numerical tests of the 1-D problem, we firstly verify the benefit of reweighted iterations and Chebyshev points for the test problem as the results are consistent with those in the corresponding references for deterministic cases, e.g., [22,25]. We then obtain that the reweighted iterations can make more contribution in the improvement. Finally, the combination of these two techniques can provide remarkable enhancement of the accuracy in the recovery. Notice that this 1-D problem is for demonstration purposes, and it can be solved more accurately by other methods, e.g., gPC with Galerkin projection or a sparse grid method with less computational cost. In the next subsection we will discuss high-dimensional cases.

4.2. Multi-dimensional problems

We consider the following elliptic equation which is 1-D in physical space but multi-D in random space:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}x}\left(a(x;\xi)\frac{\mathrm{d}}{\mathrm{d}x}u(x;\xi)\right) = -1, \quad x \in (0,1), \\ &u(0) = u(1) = 0. \end{aligned} \tag{4.4}$$

where the stochastic diffusion coefficients $a(x;\xi)$ is given by the KL expansion in Eq. (3.2). Here, $\{\lambda_i\}_{i=1}^d$ and $\{\phi_i(x)\}_{i=1}^d$ are, respectively, the d largest eigenvalues and corresponding eigenfunctions of the Gaussian covariance kernel

$$C_{aa}(x_1, x_2) = \exp\left[-\frac{(x_1 - x_2)^2}{l_c^2}\right], \tag{4.5}$$

in which l_c is the correlation length of $a(x;\xi)$ that dictates the decay of the spectrum of C_{aa}. The random variables $\{\xi_i\}_{i=1}^d$ are assumed to be independent and uniformly distributed on $[-1,1]$. The coefficient σ_a controls the variability of $a(x;\xi)$ and we consider two cases here: $(l_c, d) = (1/5, 14), (l_c, d) = (1/14, 40)$ and set $\bar{a} = 0.1, \sigma_a = 0.03; \bar{a} = 0.1, \sigma_a = 0.021$, respectively. Therefore, the requirements for coefficient $a(\boldsymbol{x};\xi)$ in Section 3 are satisfied. Both the sparse grid method with Clenshaw–Curtis abscissas and Monte Carlo method are tested. For each sampling point ξ_i in random space, the deterministic second-order ODE is solved by integrating Eq. (4.4) to obtain

$$u'(x) = \frac{a(0)u'(0) - x}{a(x)}. \tag{4.6}$$

Again, integrating Eq. (4.6) and letting $M = a(0)u'(0)$ we have

$$u(x) = u(0) + \int_0^x \frac{M - s}{a(s)}\mathrm{d}s = \int_0^x \frac{M - s}{a(s)}\mathrm{d}s. \tag{4.7}$$

By imposing the boundary condition $u(1) = 0$, we can compute M. In order to compute the integrals in Eq. (4.7), we split the domain $[0,1]$ into 2000 equi-distance subintervals and use 3 Gaussian quadratures in each subinterval. The reference solution is obtained by level 7 sparse grids method for $d = 14$ and level 4 sparse grids method for $d = 40$ since this accuracy is sufficient for the demonstrations in this paper. For $d = 14$ we set $P = 3$, i.e., gPC basis up to 3rd-order are employed and the total number of basis is $N = 680$. For $d = 40$ we set $P = 2$, i.e., gPC basis up to 2rd-order are employed and the total number of basis is $N = 861$.

4.2.1. Effect of ϵ

We first study the effect of ϵ in $(P_{1,\epsilon})$. It dictates the distance between the exact solution and the approximated one through the projection to the pre-selected basis. In practice, we truncate the right-hand-side of Eq. (3.3) to obtain an approximation, therefore there is always an error since the basis is not complete. As pointed out in [21], ideally, we would like to choose $\epsilon \approx \|\boldsymbol{\Psi c} - \boldsymbol{u}\|_2$. Fig. 5 presents the error in the mean and the standard deviation of the approximated solution with the coefficients recovered from the ℓ_1 minimization, i.e., $(P_{1,\epsilon})$ for $d = 14$. Here ϵ varies from 10^{-4} to 10^{-1} with a fixed ratio and three trials (denoted by different colors) with 120 uniform points each. These three trials are randomly selected from 1000 such trials to demonstrate the effect of ϵ. It is clear that, as ϵ decreases, the error in both mean and standard deviation decreases considerably in the range of $\epsilon \in [10^{-2}, 10^{-1}]$. However, when ϵ continues to decrease, the improvement is very little or there may not be an improvement. This is because ϵ has reached the level of truncation error, and therefore the only way to improve the accuracy in the current setting is to include more terms in the gPC expansion of the solution or increase m. This conclusion, which is an intrinsic property of ℓ_1 minimization method, holds for general cases as mentioned in Section 3.

These tests imply that for the current problem setting, the selection of ϵ plays an essential role in a specific range before it descends to the vicinity of the truncation error while very little or even no improvement can be achieved by continuing decreasing ϵ when it is already very close to the truncation error. Moreover, weighted ℓ_1 minimization $(P^W_{1,\epsilon})$ improves the results only a little or has no benefit (not presented here). Therefore, in order to considerably reduce the error further we need other techniques. In the rest part of this paper, we will only use $(P_{1,\epsilon})$ in the first step of Algorithm 1.

4.2.2. Reweighted ℓ_1 minimization

As shown for the 1-D test case, reweighted ℓ_1 minimization has the potential of further increasing the sparsity of the solution and therefore reducing the recovery error. Fig. 6 presents the results of repeating the same trials as in the last subsection but with reweighted ℓ_1 minimizations for the uniform points with $d = 14$. Different colors denote different trials. Dash lines are the result by the ℓ_1 minimization $\epsilon = 10^{-4}$. For the mean, the reweighted ℓ_1 method reduces the relative error by 95% (blue), 70% (black) and 71% (red). For the standard deviation, three different trials show reduction of the error to be 74% (blue), 49% (black), 43% (red). Similar results for $d = 40$ are presented in Fig. 7. Notice that, ϵ, τ are chosen randomly here. We will employ cross-validation method as in [21] to select parameters in the next subsection. To our knowledge, so far the best theoretical result to estimate the error bound is Theorem 2.7 and its related lemmas and theorems in [35].

Quantitative comparisons of results in Fig. 6 and Fig. 7 are presented in Tables 2 and 3 for $d = 14$ and $d = 40$, respectively. We can observe that for some trials, the improvement is dramatic, e.g., for trial 1 of $d = 14$ case, the error for the mean is reduced by more than 90% while the error for the standard deviation is reduced by 75%. For some trials, the improvement is not that impressive with the reduction of error ranging from 30% to 50%.

Next, we also employ the cross-validation method to obtain a relatively better choice of ϵ (details can be found in Appendix B). We run 1000 trials of each experiment and the histograms of the L_2 error of the solution at $x = 0.5$ are shown in Figs. 8 and 9. It is clear that the reweighted ℓ_1 minimization reduces the L_2 error of the solution when uniform points or Chebyshev points are employed. A quantitative comparison of the error of statistics at $x = 0.5$ is shown in Tables 4 (for $d = 14$) and 5 (for $d = 40$). As a comparison, e.g., for $d = 14$, the error by level 1 sparse grids method (29 samples) is $9.4387e-4$ for the mean and $5.0068e-2$ for the standard deviation while the corresponding data for level 2 sparse grids method (421 samples) is $3.2826e-5$ and $1.4532e-3$. We can observe that, for the estimate of the mean (i.e., c_0), neither methods guarantee that the accuracy is better than level 1 sparse grids method and few trials reach the accuracy of level 2 sparse grids method. However, for higher requirement of the accuracy, reweighted ℓ_1 minimization performs well. For example, in Table 4, 75% tests with reweighted ℓ_1 method show an error in the mean less than $5e{-}4$, which is about half of the error by the level 1 sparse grid method. Without the reweighted iterations, this percentage is only 7%. Similar conclusions can be drawn for the comparisons of the error of the standard deviation.

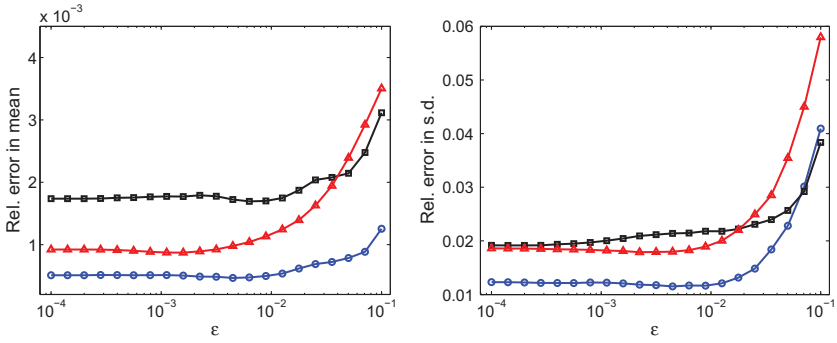


Fig. 5. 14D example: relative error in the mean (left) and the standard deviation (right) of the approximated solution with the coefficients recovered from ℓ_1 minimization $(P_{1,\epsilon})$. Three trials with 120 uniform points each are presented by different symbols; ϵ varies from 10^{-4} to 10^{-1}. $\bar{a} = 0.1, \sigma_a = 0.03, d = 14$.

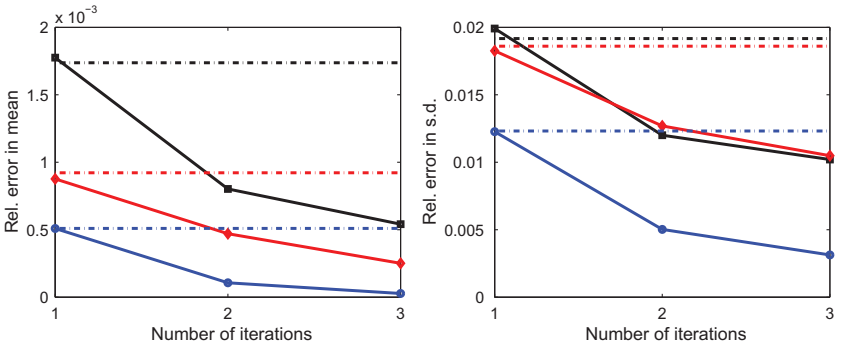


Fig. 6. 14D example: relative error in the mean and the standard deviation of the approximated solution with the coefficients recovered from reweighted ℓ_1 minimization. Three different trials with 120 uniform points each are tested. Different colors denote different trials, solid lines are the results by the reweighted ℓ_1 minimization and the dash lines are the result with $\epsilon = 10^{-4}$ in ℓ_1 minimization as in Fig. 5. Here $\epsilon = 10^{-3}, \tau = 10^{-3}$ for all the tests and $\bar{a} = 0.1, \sigma_a = 0.03, d = 14$.

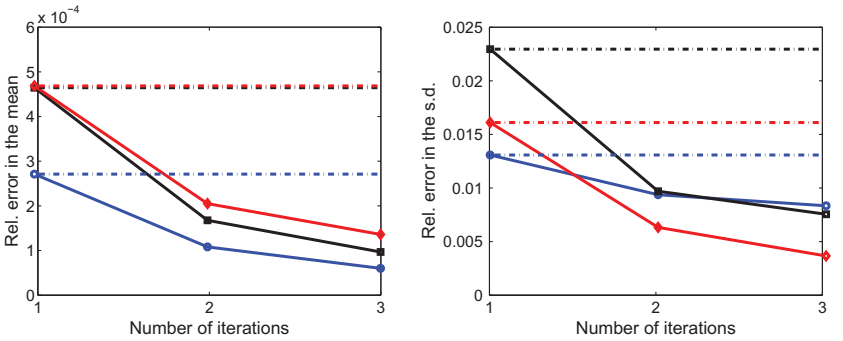


Fig. 7. 40D example: relative error in the mean and the standard deviation of the approximated solution with the coefficients recovered from reweighted ℓ_1 minimization. Three different trials with 200 uniform points each are tested. Different colors denote different data sets, solid lines are the results by the reweighted ℓ_1 minimization while the dash lines are the result with $\epsilon = 10^{-4}$ in ℓ_1 minimization. Here $\epsilon = 5 \times 10^{-3}, \tau = 10^{-3}$ for all the tests and $\bar{a} = 0.1, \sigma_a = 0.021, d = 40$.

Table 2
14D example: comparison of ℓ_1 and reweighted ℓ_1 minimization with 3 iterations ($l_{\max} = 2$) for the three trials in Fig. 6. 120 uniform points are employed in each trial. $\bar{a} = 0.1, \sigma_a = 0.03, d = 14, \epsilon = 10^{-3}, \tau = 10^{-3}$.

	Relative error in the mean ε_m		Relative error in the s.d. ε_s	
	ℓ_1	rw ℓ_1	ℓ_1	rw ℓ_1
Trial 1	5.0925e−4	2.7083e−5	1.2268e−2	3.1292e−3
Trial 2	1.7753e−3	5.4107e−4	1.9909e−2	1.0200e−2
Trial 3	8.7641e−4	2.5095e−4	1.8257e−2	1.0487e−2

Table 3
40D example: comparison of ℓ_1 and reweighted ℓ_1 minimization with 3 iterations ($l_{\max} = 2$) for the three trials in Fig. 7. 200 uniform points are employed in each trial. $\bar{a} = 0.1, \sigma_a = 0.021, d = 40, \epsilon = 10^{-3}, \tau = 10^{-3}$.

	Relative error in the mean ε_m		Relative error in the s.d. ε_s	
	ℓ_1	rw ℓ_1	ℓ_1	rw ℓ_1
Trial 1	2.7113e−4	6.1976e−5	1.3064e−2	8.3692e−3
Trial 2	4.6284e−4	9.8414e−4	2.2842e−2	7.5808e−3
Trial 3	4.6676e−4	1.3727e−4	1.6061e−2	3.7268e−3

Moreover, we consider the global error of statistics (mean and the standard deviation) over the physical domain $x \in [0, 1]$ by selecting equi-distance points $x_i = 0.05 \times i,\ i = 1, \ldots, 19$ and computing the error of statistics at each point. Specifically, we compute the l_2 error of the statistics at these points. For example, to investigate the global error of the mean, we compute $\left(\sum_{i=1}^{19} |\mathbb{E}(u)|_{x=x_i} - \mathbb{E}(\tilde{u}_p^*)|_{x=x_i}|^2\right)^{1/2} / \left(\sum_{i=1}^{19} |\mathbb{E}(u)|_{x=x_i}|^2\right)^{1/2}$, where u is the reference solution and $\tilde{u}_p^*$ is the approximated solution.

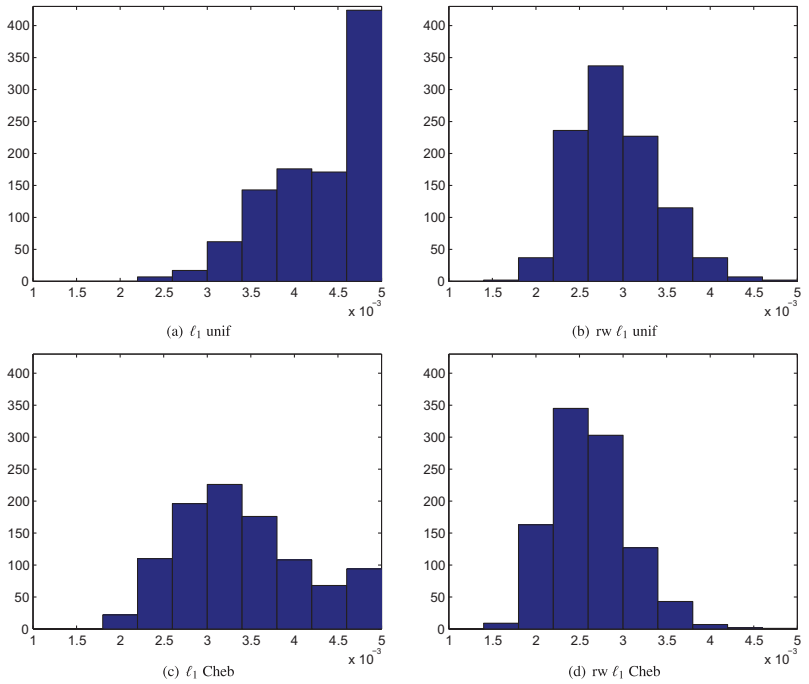


Fig. 8. 14D example: L_2 error of the solution at $x = 0.5$. x-axis is the range of the error and the histograms demonstrate the number of trials with the error in specific ranges. In each trial, 120 sampling points are employed and 1000 trials are tested to present the histograms. "ℓ_1" denotes the standard ℓ_1 minimization; "rw ℓ_1" denotes reweighted ℓ_1 minimization; "unif" denotes that only uniform points are used; "Cheb" denotes that Chebshev points are employed in the sampling points. In (a) and (b) the sampling points are uniform points; in (c) and (d) the first d' dimension of the sampling points are Chebyshev and the remainings are uniform. ϵ is obtained by cross-validation and $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.03, d = 14, d' = 6, \tau = 10^{-3}$.

The results are presented in Fig. 10 (for $d = 14$) and Fig. 11 (for $d = 40$). We observe that the reweighted iterations improve the results when uniform points are employed. We list in Table 6 the 95% confidence interval of the mean of the global error of statistics by Monte Carlo method (without post processing) and by reweighted ℓ_1 minimization. It is clear that the reweighted ℓ_1 minimization reduces the error of the mean by about 85% for both $d = 14$ and $d = 40$. It reduces the error of the standard deviation by about 65% for $d = 14$ and 70% for $d = 40$. The performance is slightly different for the two cases. There are two main reasons that affect the efficiency of the method in our test cases: (1) for these two different cases, the "sparsity" of the coefficients is different; (2) we expand the solution of $d = 14$ case with $P = 3$ while the solution of $d = 40$ case with $P = 2$ due to computational limitations, which is not adequate for higher accuracy. Since the standard deviation depends on the coefficients of all the Legendre polynomials except for c_0 (which is the mean), the accuracy not only depends on the performance of ℓ_1 minimization but also relies on the gPC expansion of the solution.

Remark 4.1. In these tests, there are analytical results guaranteeing that the solutions are sparse in the random space no matter which physical point is considered, hence, we obtain a good global recovery. For problems in which the sparsity varies at different locations, our method is effective on the area with sparse solution while less effective on the area with less sparse solution. This is a multiscale problem and requires further investigation.

4.2.3. Reweighted ℓ_1 minimization combined with Chebyshev points

Similar to the tests of the 1-D case, we also compare the results by different sampling strategies with or without reweighted iterations. Fig. 8 presents the results for $d = 14$ case. It is clear that Chebyshev points help to improve the result as (c) shows better results than (a), and (d) shows better results than (b). A very important point is that instead of using Chebyshev point in all the dimensions, we only use for the first d' ($< d$) dimensions and in the remaining dimensions we still use uniform points. We notice that this is the most effective way to apply this sampling strategy in this case. Our numerical tests show that $3 \leqslant d' \leqslant 7$ are good choices, and we present the results for $d' = 6$ in this paper. For $d = 40$, as shown in Fig. 9, we observe a different phenomenon, i.e., Chebyshev points may not improve the estimate, e.g., at $x = 0.5$, especially when reweighted ℓ_1 is employed. Our tests show that for the estimate of the solution on the entire physical domain, $1 \leqslant d' \leqslant 3$ are good choices, and we present the result of $d' = 3$ for this test case.

Tables 7 and 8 present comparison between ℓ_1 minimization and reweighted ℓ_1 minimization when Chebyshev points are employed. It is clear that the reweighted iterations improve the accuracy of the recovery. Comparing these two tables with Tables 4 and 5, we observe that the benefit of employing Chebyshev points is clear for $d = 14$ and the estimate of the stan-

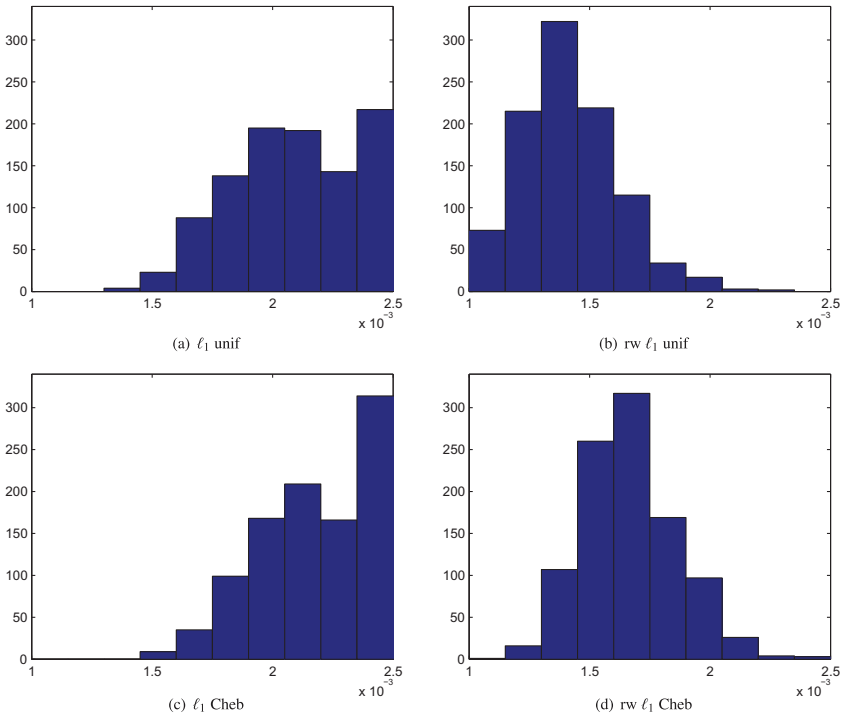


Fig. 9. 40D example: L_2 error of the solution at $x = 0.5$. x-axis is the range of the error and the histograms demonstrate the number of trials with the error in specific ranges. In each trial, 120 sampling points are employed and 1000 trials are tested to present the histograms. "ℓ_1 denotes the standard ℓ_1 minimization; "rw" denotes reweighted ℓ_1 minimization; "unif" denotes that only uniform points are used; "Cheb" denotes that Chebshev points are employed in the sampling points. In (a) and (b) the sampling points are uniform points; in (c) and (d) the first d' dimension of the sampling points are Chebyshev and the remainings are uniform. ϵ is obtained by cross-validation and $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.021, d = 40, d' = 3, \tau = 10^{-3}$.

Table 4
14D example: number of trials out of 1000 with corresponding error at $x = 0.5$ for different methods for $d = 14$. 120 uniform points are employed in each trial. $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.03, d = 14, \tau = 10^{-3}$. As a comparison, the error of level 1 sparse grids method (29 samples) 9.4387e−4 for the mean and 5.0068e−2 for the standard deviation while the corresponding data of level 2 sparse grids method (421 samples) is 3.2826e−5 and 1.4532e−3.

	Relative error in the mean ε_m			Relative error in the s.d. ε_s	
	ℓ_1	rw ℓ_1		ℓ_1	rw ℓ_1
<9.4e−4	380	975	<5e−2	1000	1000
<5e−4	71	747	<1e−2	167	850
<3.3e−5	0	42	<1.5e−3	2	60

dard deviation for $d = 40$. However, for the estimate of the mean for $d = 40$ case when reweighted ℓ_1 is used, there is no improvement by employing Chebyshev points. The global error of statistics over the physical domain is presented in Fig. 12 and Fig. 13 for $d = 14$ and $d = 40$, respectively. We can observe the improvement by applying the reweighted iterations. Also, comparing these results with those in Figs. 10 and 11, where only uniform points are employed, we notice that when standard ℓ_1 minimization is used, Chebyshev points improve the results for both $d = 14$ and $d = 40$ case. When reweighted iterations are employed, Chebyshev points improve the results for $d = 14$. However, for $d = 40$ there is improvement for the estimate of the standard deviation but no improvement for the mean. This findings implies that for mildly high dimension, e.g., $d = 14$, the combination of reweighted ℓ_1 and Chebyshev points works well while for higher dimensions, e.g., $d = 40$ only applying reweighted ℓ_1 minimization is sufficient. We list in Table 9 the 95% confidence interval of the mean of the global error of statistics by Monte Carlo method (without post processing) with uniform points and by reweighted ℓ_1 minimization with Chebyshev points. We observe that the reweighted ℓ_1 minimization reduces the error of the mean by nearly 90% for both $d = 14$ and $d = 40$. It reduces the error of the standard deviation by nearly 80% for $d = 14$ and 74% for $d = 40$. Comparing the results in Table 9 with those in Table 6, we notice that the combination of Chebyshev points and reweighted ℓ_1 minimization is the best choice for $d = 14$, but this is not true for $d = 40$. The reweighted ℓ_1 with uniform points provides better estimate of the mean while the reweighted ℓ_1 with Chebyshev points shows better estimate of the standard deviation. These results are consistent with the comparison of Fig. 11 (b) and Fig. 13 (b) as well as comparison of Fig. 11 (d) and Fig. 13 (d). In this specific case, the exact mean is more than 50 times larger than the exact standard

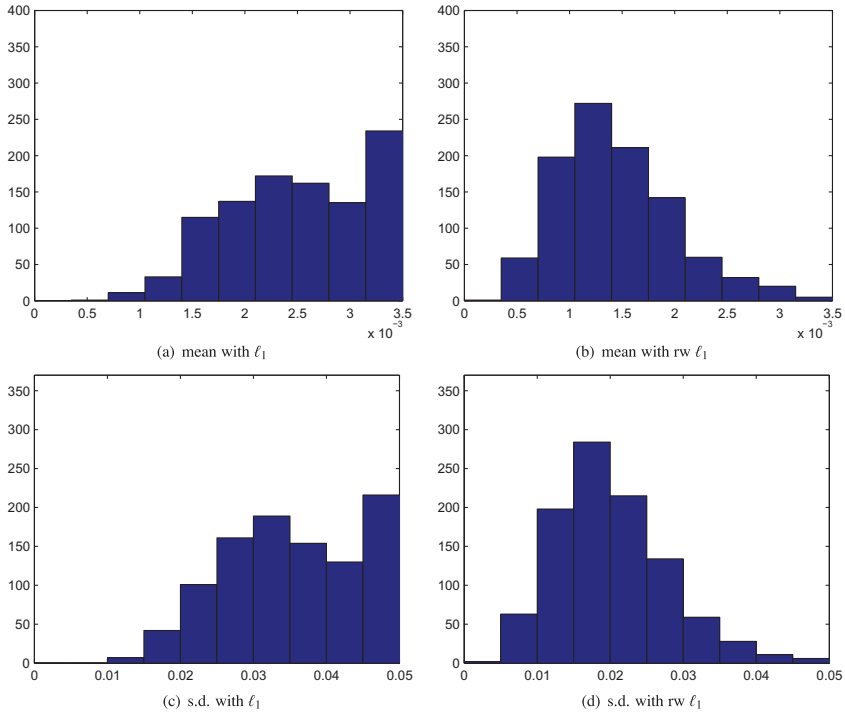


Fig. 10. 14D example: relative (global) error in the mean and the standard deviation ("s.d.") of the approximated solution over the physical domain with the coefficients $\boldsymbol{c}$ recovered from ℓ_1 or reweighted ("rw") ℓ_1 minimization with *uniform* points. In each trial, 120 sampling points are employed and 1000 trials are tested. ϵ is obtained by cross-validation and $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.03, d = 14, \tau = 10^{-3}$.

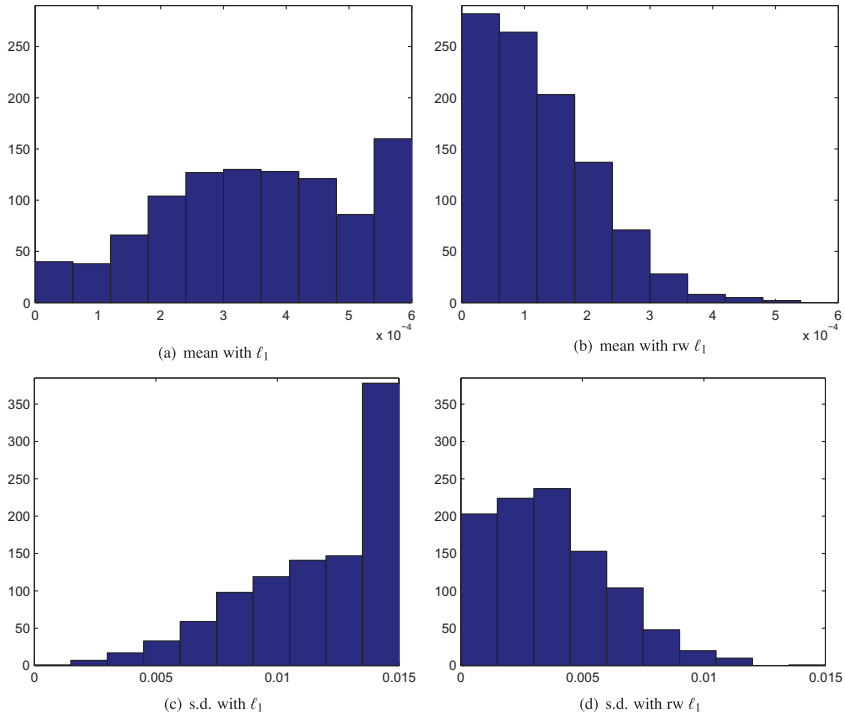


Fig. 11. 40D example: relative error (global) in the mean and the standard deviation ("s.d.") of the approximated solution over the physical domain with the coefficients $\boldsymbol{c}$ recovered from ℓ_1 or reweighted ("rw") ℓ_1 minimization with *uniform* points. In each trial, 200 sampling points are employed and 1000 trials are tested. ϵ is obtained by cross-validation and $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.021, d = 40, \tau = 10^{-3}$.

Table 5
40D example: number of trials out of 1000 with corresponding error at $x = 0.5$ for different methods for $d = 40$. 200 uniform points are employed in each trial. $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.021, d = 40, \tau = 10^{-3}$. As a comparison, the error of level 1 sparse grids method (81 samples) $4.3655e-4$ for the mean and $6.4767e-2$ for the standard deviation while the corresponding data of level 2 sparse grids method (3281 samples) is $1.9131e-5$ and $3.6733e-3$.

	Relative error in the mean ε_m			Relative error in the s.d. ε_s	
	ℓ_1	rw ℓ_1		ℓ_1	rw ℓ_1
<4.4e−4	682	993	<6.5e−2	1000	1000
<1.0e−4	63	463	<1.0e−2	287	983
<1.9e−5	16	87	<3.7e−3	13	534

Table 6
95% confidence interval of the mean of relative error of statistics over the entire physical domain when uniform points are employed. "MC" denotes the result without any post processing, "rw ℓ_1" denotes reweighted ℓ_1 minimization.

	Relative error in the mean ε_m		Relative error in the s.d. ε_s	
	MC	rw ℓ_1	MC	rw ℓ_1
$d = 14$	[8.8e−3, 9.5e−3]	[1.4e−3, 1.5e−3]	[5.8e−2, 6.1e−2]	[2.0e−2, 2.1e−2]
$d = 40$	[3.6e−3, 3.8e−3]	[4.7e−4, 4.9e−4]	[4.3e−2, 4.6e−2]	[1.3e−2, 1.4e−2]

Table 7
14D example: number of trials out of 1000 with corresponding error at $x = 0.5$ for different methods for $d = 14$. 120 Chebyshev points are employed in each trial. $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.03, d = 14, d' = 6, \tau = 10^{-3}$. As a comparison, the error of level 1 sparse grids method (29 samples) $9.4387e-4$ for the mean and $5.0068e-2$ for the standard deviation while the corresponding data of level 2 sparse grids method (421 samples) is $3.2826e-5$ and $1.4532e-3$.

	Relative error in the mean ε_m			Relative error in the s.d. ε_s	
	ℓ_1	rw ℓ_1		ℓ_1	rw ℓ_1
<9.4e−4	920	975	<5e−2	1000	1000
<5e−4	690	792	<1e−2	841	958
<3.3e−5	59	68	<1.5e−3	181	238

Table 8
40D example: number of trials out of 1000 with corresponding error at $x = 0.5$ for different methods for $d = 40$. 200 Chebyshev points are employed in each trial. $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.021, d = 40, d' = 3, \tau = 10^{-3}$. As a comparison, the error of level 1 sparse grids method (81 samples) $4.3655e-4$ for the mean and $6.4767e-2$ for the standard deviation while the corresponding data of level 2 sparse grids method (3281 samples) is $1.9131e-5$ and $3.6733e-3$.

	Relative error in the mean ε_m			Relative error in the s.d. ε_s	
	ℓ_1	rw ℓ_1		ℓ_1	rw ℓ_1
<4.4e−4	795	988	<6.5e−2	1000	1000
<1.0e−4	207	414	<1.0e−2	655	998
<1.9e−5	43	70	<3.7e−3	92	754

deviation. Hence when we compare the L_2 error at fixed spatial points, the reweighted ℓ_1 method with uniform points performs better. This is consistent with the comparison of Fig. 9 (b) and (d). The above results imply that for moderately high dimensional (~ 10) problems, combination of reweighted ℓ_1 and Chebyshev points is a good choice while for higher dimensional problems, reweighted ℓ_1 only might be a better choice than combining it with Chebyshev points.

Remark 4.2. In this section we use reweighted ℓ_1 minimization, Chebyshev points and the combination of these two approaches to improve the accuracy of the recovery of the coefficients in the gPC expansion of the solution of SPDEs. We point out that in the optimization problem $(P_{1,\epsilon})$ we set the bound of the "noise" by the cross-validation method, for which we do not tune the parameters precisely since our purpose is to demonstrate the improvement by the new method. Therefore, for some trials, the result by cross-validation may not be as accurate as using a randomly or empirically chosen parameter.

Remark 4.3. When applying Chebyshev points to high dimensional problems, we use an empirical parameter $d' < d$ because $d' = d$ is not a good choice. We observe that when using Chebyshev points in ℓ_1 minimization, the results are better than the standard ℓ_1 minimization where only uniform points are employed. As pointed out in [36], when Chebyshev points are applied to high dimensional problems, the number of samples to accurately recover the coefficients scales as 2^d. We reduce this by selecting $d' < d$, that is, employing Chebyshev points in only a few dimensions. Hence we can still take advantage of this sampling strategy especially in moderately high dimensional (~ 10) problems. In our test, we know *a priori* that the first

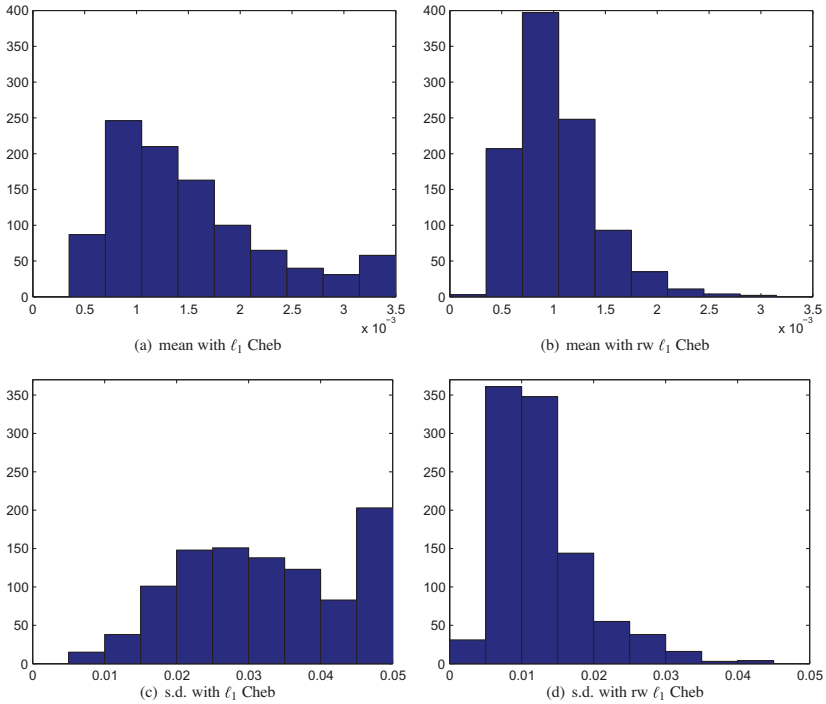


Fig. 12. 14D example: relative (global) error in the mean and the standard deviation ("s.d.") of the approximated solution over the physical domain with the coefficients $\boldsymbol{c}$ recovered from ℓ_1 or reweighted ("rw") ℓ_1 minimization with *Chebyshev* points. In each trial, 120 sampling points are employed and 1000 trials are tested. ϵ is obtained by cross-validation and $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.03, d = 14, d' = 6, \tau = 10^{-3}$.

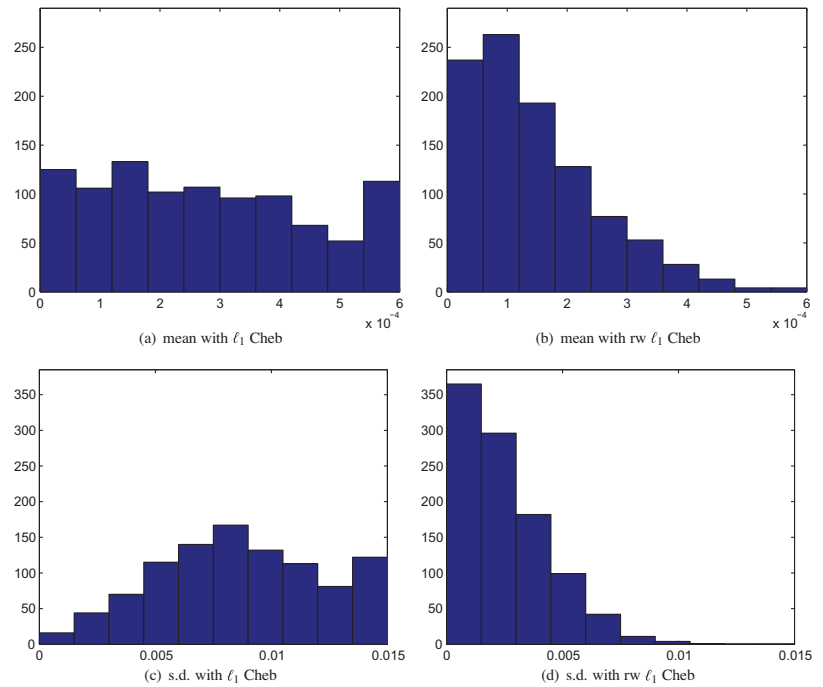


Fig. 13. 40D example: relative (global) error in the mean and the standard deviation ("s.d.") of the approximated solution over the physical domain with the coefficients $\boldsymbol{c}$ recovered from ℓ_1 or reweighted ("rw") ℓ_1 minimization with *Chebyshev* points. In each trial, 200 sampling points are employed and 1000 trials are tested. ϵ is obtained by cross-validation and $l_{\max} = 2, \bar{a} = 0.1, \sigma_a = 0.021, d = 40, d' = 3, \tau = 10^{-3}$.

Table 9
95% confidence interval of the mean of relative error of statistics over the entire physical domain. "MC unif" denotes the results by Monte Carlo method with uniform points and without any post processing; "rw ℓ_1 Cheb" denotes the results by employing reweighted ℓ_1 minimization method and Chebyshev points.

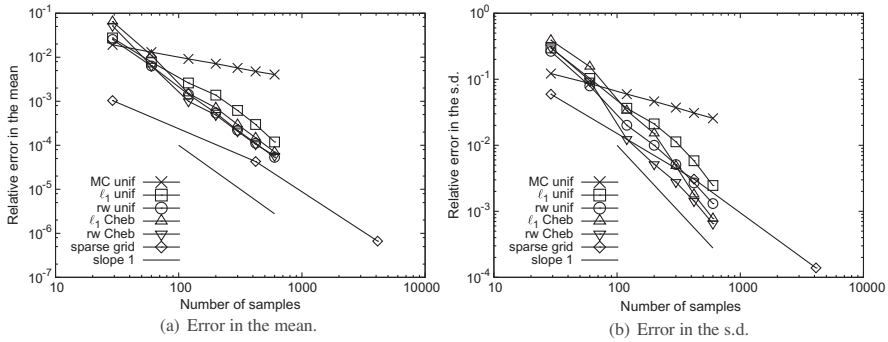
	Relative error in the mean ε_m		Relative error in the s.d. ε_s	
	MC unif	rw ℓ_1 Cheb	MC unif	rw ℓ_1 Cheb
$d = 14$	[8.8e−3, 9.5e−3]	[9.9e−4, 1.0e−3]	[5.8e−2, 6.1e−2]	[1.2e−2, 1.3e−2]
$d = 40$	[3.6e−3, 3.8e−3]	[4.2e−4, 4.4e−4]	[4.3e−2, 4.6e−2]	[1.1e−2, 1.2e−2]

several dimensions are more important due to the expression of diffusion coefficients (represented by the hierarchical Karhunen–Loève expansion), hence we employ Chebyshev points in these dimensions. For the reweighted ℓ_1 method, Chebyshev points only help in the moderately high dimensional (~ 10) case. An important reason is that we employed a lower order gPC expansion, e.g., $P = 2$ for $d = 40$. The upper bound K of the L^∞ norm of the basis is $\sqrt{5}$, which is already very low. Since the Chebyshev points improve the recovery by controlling K (see Remark 2.10), it contributes only a little to the case with very low gPC order. This contribution is negligible compared with the improvement from reweighted iterations.

4.3. Comparison of the effectiveness

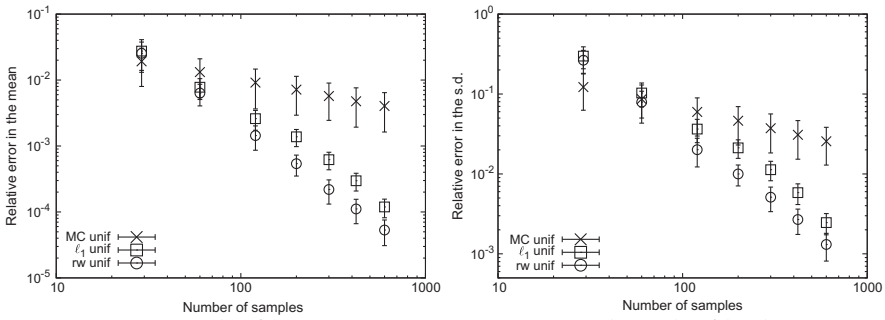
We now consider an increasing number of random samples to evaluate the solution and compare the global relative error with different methods, including Monte Carlo method and (reweighted) ℓ minimization (with uniform points or Chebyshev points). For each case we compute the results at $x = 0.1, 0.2, \ldots, 0.9$ on the physical domain to obtain the global error. For $d = 14$, we consider $m = \{29, 60, 120, 200, 300, 421, 600\}$. We note that sample size $m = 29$ and $m = 421$ correspond to the number of sparse grid method of level 1 and 2, respectively. In this test, for sample size $m = \{29, 60, 120\}$ we attempt to recover the gPC coefficients of the 3rd-order gPC expansion. For larger sample size m, we also include the first 320 basis function from the 4th-order gPC expansion, resulting in $m = 1000$ (see also Case I in [21]). We test 1000 trials for each sample size m and present the mean of the relative error in Fig. 14. We observe that when a sufficient number of solution samples is available, which in Fig. 14 is 120, ℓ_1 minimization shows an advantage over the Monte Carlo method. If reweighted iterations are employed, the ℓ_1 minimization begins to show an advantage with a smaller number of samples as pointed out in [22]. As the number of samples increases, we observe a more considerable improvement over the Monte Carlo method. For example, when $m = 600$, the mean of the relative error in the mean by reweighted ℓ_1 (both uniform points and Chebyshev points) is only about 1.5% of the error by the Monte Carlo method, i.e., our method increases the accuracy by about two orders in the estimate of the mean. Also, for this test case, we notice that with the same number of samples ($m = 421$), the accuracy of the estimate of the mean by our method is close to that by sparse grid method of level 2 while the estimate of the standard deviation by our method is better. We also present a line of slope 1 in Fig. 14 to compare the behavior of our method. Theorem 3.6 in [21] provides an upper bound of the L_2 error of the solution, which is approximately $\mathcal{O}(m^{-1/2})$. Fig. 14 implies that in practice we can expect faster convergence in the estimate of statistics, which is also reflected in Figs. 3 and 5 in [21]. Moreover, we notice that the reweighted iterations do not change the rate of convergence substantially, but reduce the error by around $50\% \sim 60\%$ for the mean and $60\% \sim 70\%$ for the standard deviation. We present the error bars of the 1000 trials for different sample size by the Monte Carlo method and (reweighted) ℓ_1 minimization method in Fig. 15. This figure also implies that reweighted iterations improve the estimate of the solutions, when a sufficient number of samples is available.

Similar results for $d = 40$ are presented in Figs. 16 and 17. In this test we chose $m = \{81, 120, 200, 400, 600, 800, 1000\}$. For sample size $m = \{81, 120, 200\}$, we attempt to recover the gPC coefficients of the 2nd-order gPC expansion. For larger



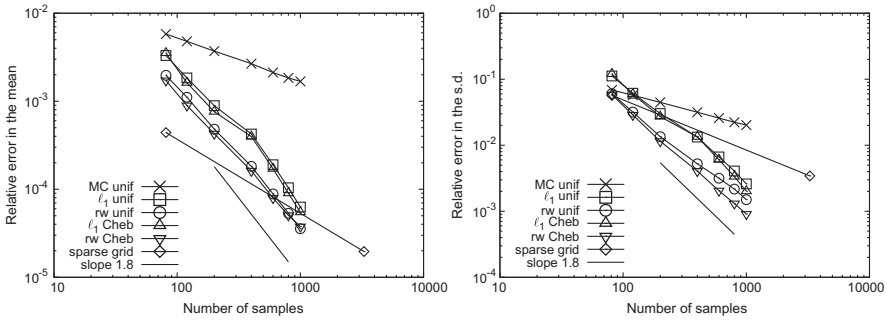
(a) Error in the mean. (b) Error in the s.d.

Fig. 14. 14D example: mean of the relative error (global) in the mean and the standard deviation ("s.d.") of the approximated solution over the physical domain by different methods. "unif" means uniform points, "Cheb" means Chebyshev points, "rw" means reweighted method.



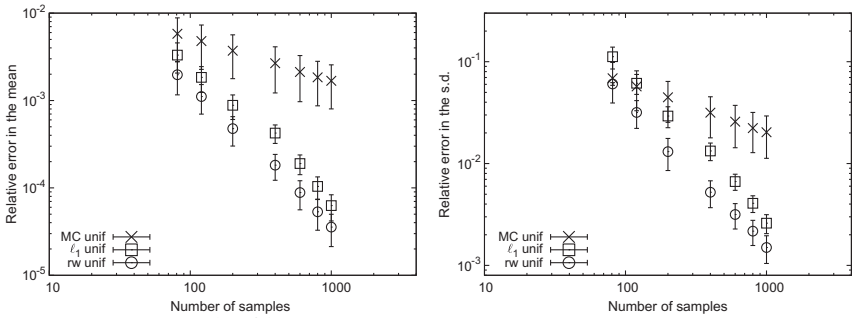
(a) Error bar of the mean. (b) Error bar of the s.d.

Fig. 15. 14D example: error bars (mean and standard deviation) of the relative error (global) in the mean and the standard deviation (“s.d.”) of the approximated solution over the physical domain by different methods. “unif” means uniform points, “rw” means reweighted method.



(a) Error in the mean. (b) Error in the s.d.

Fig. 16. 40D example: mean of the relative error (global) in the mean and the standard deviation (“s.d.”) of the approximated solution over the physical domain by different methods. “unif” means uniform points, “Cheb” means Chebyshev points, “rw” means reweighted method.



(a) Error bar of the mean. (b) Error bar of the s.d.

Fig. 17. 40D example: error bars (mean and standard deviation) of the relative error (global) in the mean and the standard deviation (“s.d.”) of the approximated solution over the physical domain by different methods. “unif” means uniform points, “rw” means reweighted method.

sample size m, we also include the first 639 basis function from the 3rd-order gPC expansion, hence $m = 1500$ (see also Case II in [21]). Fig. 16 implies that $m = 200$, is the smallest size of the samples in our test which guarantees that ℓ_1 minimization performs better than the Monte Carlo method. However, we note that when the reweighted ℓ_1 minimization is employed, $m = 120$ is sufficient to enable a better estimate. Also, similar to the $d = 14$ case, the reweighted ℓ_1 minimization method, even the standard ℓ_1 minimization, shows an advantage over the sparse grid method when sufficient samples are available. Moreover, we observe an even faster convergence rate than the $d = 14$ case as shown by a line of slope 1.8 for comparison. This again implies that in practice, we may expect a much faster convergence rate than 0.5. We also notice that the reweighted iterations do not change this rate substantially as in $d = 14$ case.

5. Summary

In this paper we have applied the reweighted ℓ_1 minimization method to solve elliptic PDEs with random coefficients. Assuming that the solution is “sparse” when we expand it in terms of a gPC basis, we use techniques for compressive sensing

to achieve an accurate recovery with the number of solution samples significantly smaller than the cardinality of the gPC basis. Specifically, we have applied in this paper, for the first time in this context, reweighted ℓ_1 minimization and combined it with Chebyshev points to SPDEs up to 40 dimensional (in random space) problems. The numerical tests show significant improvement over standard ℓ_1 minimization. That is, with the same number of solution samples, our method achieves better recovery of the gPC coefficients $\boldsymbol{c}$ (see e.g., Eq. (2.6)); hence, we can obtain better estimate of statistics, e.g., the mean and the standard deviation of the solution. Although here we have only tested random coefficients depending on uniform random variables, Chebyshev points are also suitable for other Jacobi polynomials in the gPC expansion [25]. Hence, Chebyshev points can also be employed for random variables associated with different types of Jacobi polynomials. The reweighted ℓ_1 minimization we employ in the paper is the most basic one and it is very probable that more sophisticated versions, e.g., [22,35,23,24] will result in better recovery. We can also combine ideas like multi-element method [4], quasi-Monte Carlo sampling points, sparse grid points [36], ANOVA points [12,14], etc., with our method. This will be reported in future work.

One limitation of this method is that we should know *a priori* that the solution is "sparse", otherwise it is impossible to obtain accurate results by means of compressive sensing. Also, with very small probability, the reweighted ℓ_1 minimization may render worse results than ℓ_1 minimization.

Finally, this investigation suggests that it may be possible to obtain a reasonably accurate estimate of statistics for those problems, where the deterministic solver is very expensive, and hence only a small number of simulations can be afforded. Our method reuses the data available and has the potential to obtain good accuracy with very limited data. As is well known, when the Monte Carlo method is employed, the magnitude of error is $\mathcal{O}(\frac{1}{\sqrt{m}})$. Due to the sparsity of the solution, in the numerical tests presented in this paper, the convergence rate of the ℓ_1 minimization method in estimating the statistics is no less than 1, which is higher than the result of the current theoretical upper bound for the L_2 error. The reweighted iterations do not change this rate but help to reduce the error by about 50%. This implies that we have enhanced the efficiency by two times compared with the standard ℓ_1 minimization method. If we compare the new method with the Monte Carlo method without post processing, the improvement of the efficiency can be more than 100 times. A more rigorous analysis of the convergence rate of the new method is a rather difficult task and it will be presented in future work.

Acknowledgements

We would like to acknowledge support by MURI/AFOSR, NSF, and by the Applied Mathematics program of the US DOE Office of Advanced Scientific Computing Research. We would like to thank Professor Jianfeng Lu and Dr. Xiaodong Li for fruitful discussions.

Appendix A. Single reweighted ℓ_1-minimization

Lemma A.1. (Single reweighted $_1$ minimization [35]). *Assume that $\boldsymbol{\Psi}$ satisfies the RIP condition with $\delta_{2s} < \sqrt{2}-1$. Let $\boldsymbol{c}$ be an arbitrary vector with noisy measurements $\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{c} + \mathbf{e}$, where $\|\mathbf{e}\|_2 < \epsilon$. Let $\boldsymbol{z}$ be a vector such that $\|\mathbf{z} - \boldsymbol{c}\|_\infty \leqslant B$ for some constant B. Denote by $\boldsymbol{c}_s$ the vector consisting of the s (where $s \leqslant |supp(\boldsymbol{c})|$) largest coefficients of $\boldsymbol{c}$ in absolute value. Let η be the smallest coordinate of $\boldsymbol{c}_s$ in absolute value, and set $b = \|\boldsymbol{c} - \boldsymbol{c}_s\|_\infty$. Then when $\eta \geqslant B$ and $\rho C_1 < 1$, the approximation from reweighted $_1$ minimization using weights $w_i = 1/(z_i + \tau)$ satisfies*

$$\|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_2 \leqslant D_1\epsilon + D_2\frac{\sigma_s(\boldsymbol{c})_1}{\tau}, \tag{A.1}$$

where

$$D_1 = \frac{(1+C_1)\alpha}{1-\rho C_1}, \quad D_2 = C_2 + \frac{(1+C_1)\rho C_2}{1-\rho C_1}, \quad C_1 = \frac{B+\tau+b}{\eta - B + \tau}, \quad C_2 = \frac{2(B+\tau+b)}{\sqrt{s}},$$

and $\rho, \alpha, \sigma_s(\boldsymbol{c})_p$ are as in Theorem 2.3.

Appendix B. Cross-validation

The cross-validation method we use in this paper follows the description in Section 3.5 of [21]. We first divide the m available solution samples to m_r reconstruction and m_v validation samples such that $m = m_r + m_v$. Then repeat the ℓ_1 minimization $(P_{1,\epsilon})$ (not the reweighted iteration) on the reconstruction samples and with multiple values of truncation error tolerance ϵ_r. In this paper we test 11 different ϵ_r from $[10^{-4}, 10^{-2}]$ with constant ratio. Next, set $\epsilon = \sqrt{m/m_r}\hat{\epsilon}_r$, where $\hat{\epsilon}_r$ is such that the corresponding truncation error on the m_v validation samples is minimum. Finally, we repeat the above cross-validation algorithm for multiple replications of the reconstruction and validation samples. The estimate of $\epsilon = \sqrt{m/m_r}\hat{\epsilon}_r$ is then based on the values of $\hat{\epsilon}_r$ for which the average of the corresponding truncation errors ϵ_v, over all replications of the validation samples, is minimum. In this paper we set $m_r \approx 3m/4$ and performed the cross-validation for four replications. More details can be found in [21].

References

[1] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, New York, 1991.
[2] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[3] X. Wan, G.E. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures, SIAM J. Sci. Comput. 28 (3) (2006) 901–928.
[4] X. Wan, G.E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, J. Comput. Phys. 209 (2) (2005) 617–642.
[5] L. Mathelin, M. Hussaini, A stochastic collocation algorithm for uncertainty analysis, Tech. Rep. Technical Report NAS 1.26:212153; NASA/CR-2003-212153, NASA Langley Research Center, 2003.
[6] D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (3) (2005) 1118–1139.
[7] F. Nobile, R. Tempone, C.G. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (5) (2008) 2309–2345.
[8] J. Foo, X. Wan, G.E. Karniadakis, The multi-element probabilistic collocation method (ME-PCM): error analysis and applications, J. Comput. Phys. 227 (22) (2008) 9572–9595.
[9] X. Ma, N. Zabaras, An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations, J. Comput. Phys. 228 (8) (2009) 3084–3113.
[10] F. Nobile, R. Tempone, C.G. Webster, An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (5) (2008) 2411–2442.
[11] Z. Zhang, M. Choi, G.E. Karniadakis, Error estimates for the ANOVA method with polynomial chaos interpolation: tensor product functions, SIAM J. Sci. Comput. 34 (2) (2012) A1165–A1186.
[12] J. Foo, G.E. Karniadakis, Multi-element probabilistic collocation method in high dimensions, J. Comput. Phys. 229 (5) (2010) 1536–1557.
[13] X. Ma, N. Zabaras, An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations, J. Comput. Phys. 229 (10) (2010) 3884–3915.
[14] X. Yang, M. Choi, G. Lin, G.E. Karniadakis, Adaptive ANOVA decomposition of stochastic incompressible and compressible flows, J. Comput. Phys. 231 (4) (2012) 1587–1614.
[15] D. Lucor, G.E. Karniadakis, Adaptive generalized polynomial chaos for nonlinear random oscillators, SIAM J. Sci. Comput. 26 (2) (2004) 720–735.
[16] D. Venturi, X. Wan, G.E. Karniadakis, Stochastic low-dimensional modelling of a random laminar wake past a circular cylinder, J. Fluid Mech. 606 (2008) 339–367.
[17] R.A. Todor, C. Schwab, Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients, IMA J. Numer. Anal. 27 (2) (2007) 232–261.
[18] M. Bieri, C. Schwab, Sparse high order FEM for elliptic SPDEs, Comput. Methods Appl. Mech. Eng. 198 (13–14) (2009) 1149–1170.
[19] M. Hansen, C. Schwab, Analytic regularity and best *N*-term approximation of high dimensional parametric initial value problems, Tech. Rep. 2011-64, Seminar for Applied Mathematics, ETHZ, 2011.
[20] V.H. Hoang, C. Schwab, *N*-term Galerkin Wiener chaos approximations of elliptic PDEs with lognormal Gaussian random inputs, Tech. Rep. 2011-59, Seminar for Applied Mathematics, ETHZ, 2011.
[21] A. Doostan, H. Owhadi, A non-adapted sparse approximation of PDEs with stochastic inputs, J. Comput. Phys. 230 (8) (2011) 3015–3034.
[22] E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, J. Fourier Anal. Appl. 14 (5-6) (2008) 877–905.
[23] M. Khajehnejad, W. Xu, A. Avestimehr, B. Hassibi, Improved sparse recovery thresholds with two-step reweighted ℓ_1 minimization, in: 2010 IEEE International Symposium on Information Theory Proceedings (ISIT), 2010, pp. 1603 –1607.
[24] W. Xu, M. Khajehnejad, A. Avestimehr, B. Hassibi, Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the grassmann angle framework, in: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2010, pp. 5498–5501.
[25] H. Rauhut, R. Ward, Sparse legendre expansions via ℓ_1-minimization, J. Approx. Theory 164 (5) (2012) 517–533.
[26] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (1) (2009) 34–81.
[27] R.A. DeVore, V.N. Temlyakov, Some remarks on greedy algorithms, Adv. Comput. Math. 5 (2–3) (1996) 173–187.
[28] V.N. Temlyakov, Greedy algorithms and *m*-term approximation with regard to redundant dictionaries, J. Approx. Theory 98 (1) (1999) 117–145.
[29] V.N. Temlyakov, Weak greedy algorithms, Adv. Comput. Math. 12 (2–3) (2000) 213–227.
[30] D.L. Donoho, M. Elad, V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory 52 (1) (2006) 6–18.
[31] E.J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (12) (2005) 4203–4215.
[32] E.J. Candès, The restricted isometry property and its implications for compressed sensing, C.R. Math. Acad. Sci. Paris 346 (9–10) (2008) 589–592.
[33] S. Foucart, A note on guaranteed sparse recovery via ℓ_1-minimization, Appl. Comput. Harmon. Anal. 29 (1) (2010) 97–103.
[34] S. Foucart, M.-J. Lai, Sparsest solutions of underdetermined linear systems via lq-minimization for $0 < q \leqslant 1$, Appl. Comput. Harmon. Anal. 26 (3) (2009) 395–407.
[35] D. Needell, Noisy signal recovery via iterative reweighted ℓ_1 minimization, in: Proceedings Asilomar Conference on Signal Systems and Computers, Pacific Grove, CA, 2009.
[36] L. Yan, L. Guo, D. Xiu, Stochastic collocation algorithms using ℓ_1-minimization, Int. J. Uncertainty Quantif. 2 (3) (2012) 279–293.
[37] D. Needell, J.A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.
[38] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (3) (2009) 265–274.
[39] E. van den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912.
[40] J.D. Blanchard, C. Cartis, J. Tanner, Compressed sensing: how sharp is the restricted isometry property?, SIAM Rev 53 (1) (2011) 105–125.