# Compressed Sensing Petrov-Galerkin Approximations for Parametric PDEs

Jean-Luc Bouchot, Benjamin Bykowski, and Holger Rauhut
Chair for Mathematics C (Analysis)
RWTH Aachen University
Email: {bouchot,rauhut}@mathc.rwth-aachen.de
benjamin.bykowski@rwth-aachen.de

Christoph Schwab
Seminar of Applied Mathematics
ETH Zurich
Email: christoph.schwab@sam.math.ethz.ch

***Abstract*—We consider the computation of parametric solution families of high-dimensional stochastic and parametric PDEs. We review recent theoretical results on sparsity of polynomial chaos expansions of parametric solutions, and on compressed sensing based collocation methods for their efficient numerical computation.**

**With high probability, these randomized approximations realize best $N$-term approximation rates afforded by solution sparsity and are free from the curse of dimensionality, both in terms of accuracy and number of samples evaluations (i.e. PDE solves).**

**Through various examples we illustrate the performance of Compressed Sensing Petrov-Galerkin (CSPG) approximations of parametric PDEs, for the computation of (functionals of) solutions of intregral and differential operators on high-dimensional parameter spaces. The CSPG approximations allow to reduce the number of PDE solves, as compared to Monte-Carlo methods, while being likewise nonintrusive, and being "embarassingly parallel", unlike dimension-adaptive collocation or Galerkin methods.**

## I. Introduction

### A. Problem statement

We consider the problem of computing an (accurate) approximation of a parametric family $u$ defined on a bounded, "physical" domain $D \subset \mathbb{R}^d$ and depending on possibly countably many parameters $y = (y_j)_{j\geq 1} \in U = [-1,1]^{\mathbb{N}}$ implicitly through the parametric operator equation

$$A(y)u(y) = f. \tag{1}$$

Equation (1) is assumed to be well-posed in a reflexive, separable Banach space $V$, uniformly for all parameter sequences $y \in U$. From a Bochner space perspective, if the parameters $y$ are distributed according to a certain distribution $\mu$ on $U$, we are looking for strongly $\mu$-measurable maps $u \in \mathcal{L}(U, V; \mu)$. For simplicity of exposition, we assume here as in [1], [2] that $A(y)$ in (1) can be expressed as a linear combination of bounded linear operators $(A_j)_{j\geq 0} : V \to W'$, where $W$ is a suitable second (separable and reflexive) Banach space, such that $A(y) = A_0 + \sum_{j\geq 1} y_j A_j$ in $\mathcal{L}(V, W')$. We hasten to add that the principal conclusions of the present work apply to larger classes of countably-parametric operator equations (1) with sparse solution families. Given a parametric family $f(y) \in W'$, we wish to approximate numerically the corresponding parametric solution family $\{u(y) : y \in U\}$. To simplify the exposition, we restrict ourselves in this note to nonparametric data $f$; all results which follow apply verbatim in the general case. We consider the case which appears often in practice, of approximation of functionals of the parametric solution: given $\mathcal{G} \in \mathcal{L}(V, \mathbb{R})$, approximate numerically the map $F : U \to \mathbb{R},\ F(y) := \mathcal{G}(u(y))$.

### B. Tools and ideas

The present CSPG approach is motivated by recent sparsity results for the parametric solution maps in [3], [4] where sparsity results for coefficient sequences of *generalized polynomial chaos* (GPC) expansions of $u(y)$ were established. The term GPC expansion refers to $u(y) = \sum_{j\in\mathcal{F}} x_j \varphi_j(y)$ for some basis $(\varphi_j)_{j\in\mathcal{F}}$ of $L^2_\mu(U)$, indexed on the countable family $\mathcal{F}$. Due to the boundedness and linearity of $\mathcal{G}$, we can similarly write $F(y) = \sum_{j\in\mathcal{F}} g_j \varphi_j$. Knowing the compressibility of the coefficient sequence $(g_j)_{j\geq 0}$ (i.e., there exists $0 < p < 1$, such that $\|g\|_p < \infty$), the computational challenge is to recover only the most important ones using only as few as possible solves of (1), while ensuring a given accuracy. The sparsity of the GPC coefficient sequence together with Stechkin inequality ensures that $\sigma_s(g)_q \leq s^{1/q-1/p}\|g\|_p, 0 < p < q$, where $\sigma_s(g)_q$ denotes the best $s$-term approximation of $g$ in the $q$ norm. In other words, the coefficient sequence $g$ can be compressed with high accuracy and hence (see Theorem 1 in Section II-B) the function $F$ can be well approximated. To overcome the curse of dimensionality incurred by CS based methods (see [2, Eq.(1.16)]), we propose an extension based on $\ell_{p,\omega}$ spaces [5].

The rest of the paper is organized as follows. Section II reviews the mandatory background on weighted compressed sensing, Petrov-Galerkin approximations, and their combined use for high-dimensional PDEs. Section III then introduces various model problems to numerically validate the theoretical results. We provide numerical experiments which show that the CSPG method can be easily implemented for various linear functionals $\mathcal{G}$, even with a large number of parameters. For the parametric model diffusion problems, we compare the convergence of the expectation with that of Monte-Carlo methods.

## II. Review on Compressed Sensing Petrov-Galerkin Approximations

### A. Generalized Polynomial Chaos (GPC) expansions and tensorized Chebyshev polynomials

GPC expansion of the parametric solution family $\{u(y) : y \in U\}$ of Eq. (1) are orthogonal expansions of $u(y)$ in terms of tensorized Chebyshev polynomials $(T_\nu)_{\nu \in \mathcal{F}}$, here $\mathcal{F}$ denotes the countable set of multiindices with finite support ($\mathcal{F} := \{\nu \in \mathbb{N}_0^{\mathbb{N}} : |\operatorname{supp}(\nu)| < \infty\}$):

$$u(y) = \sum_{\nu \in \mathcal{F}} d_\nu T_\nu(y). \tag{2}$$

Note that here the coefficients $d_\nu$ are functions in $V$ and that the expansion is in terms of the parameter sequences $y \in U$ and not in the spatial coordinates. The tensorized Chebyshev polynomials are defined as $T_\nu(y) = \prod_{j=1}^{\infty} T_{\nu_j}(y_j)$, $y \in U$, with $T_j(t) = \sqrt{2}\cos(j \arccos(t))$ and $T_0(t) \equiv 1$.

Defining $\sigma$ the probability measure on $[-1;1]$ as $\mathrm{d}\sigma(t) := \frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}$, Chebyshev polynomials form an orthonormal system in the sense that $\int_{-1}^{1} T_k(t) T_l(t) \mathrm{d}\sigma(t) = \delta_{k,l}$ for $k, l \in \mathbb{N}_0$. Similarly, with the product measure $\mathrm{d}\eta(y) := \bigotimes_{j \geq 1} \mathrm{d}\sigma(y_j)$, the tensorized Chebyshev polynomials are orthonormal in the sense that $\int_{y \in U} T_\mu(y) T_\nu(y) \mathrm{d}\eta(y) = \delta_{\mu,\nu}$, for $\mu, \nu \in \mathcal{F}$.

An important result proven in [6] ensures the $\ell_p$ summability, for some $0 < p \leq 1$, of the Chebyshev polynomial chaos expansion (2) of the solution of the model affine-parametric diffusion equation

$$-\nabla \cdot (a \nabla u) = f \ , \tag{3}$$

where $a$ admits the affine-parametric expansion

$$a(y; x) = a_0(x) + \sum_{j \geq 1} y_j a_j(x) \ . \tag{4}$$

Specifically, $\left\| (\|d_\nu\|_V)_{\nu \in \mathcal{F}} \right\|_p^p = \sum_{\nu \in \mathcal{F}} \|d_\nu\|_V^p < \infty$under the condition that the sequence of inifinity norms of the $a_j$ is itself $\ell_p$ summable: $\left\| (\|a_j\|_\infty)_{j \geq 1} \right\|_p < \infty$.

These results were extended to the weighted $\ell_p$ spaces (see Section II-B) for the more general parametric operator problem, Eq.(1) with linear parameters, in [2]: $(\|d_\nu\|_V)_{\nu \in \mathcal{F}} \in \ell_{\omega,p}(\mathcal{F})$ under certain assumptions, and with Eq. (6) defining the weighted $\ell_p$ spaces.

### B. Weighted compressed sensing

The last decade has seen the emergence of compressed sensing [7] as a method to solve underdetermined linear systems $Ax = y$ under sparsity constraints $\|x\|_0 := |\operatorname{supp}(x)| \leq s$. Some work has focused on finding conditions on $A$ such that the problem

$$x^\# = \operatorname{argmin} \|x\|_1, \quad \text{s.t. } \|Ax - y\|_2 \leq \eta \tag{5}$$

yields an exact and unique solution. One known condition is that the matrix $A$ should fulfill a *Restricted Isometry Property* (RIP) of order $s$ and constant $\delta$ for reasonable values of $s$ and $\delta$. $A$ is said to have RIP$(s, \delta)$ if, for any $s$- sparse vector $x$, $|\|Ax\|_2^2 - \|x\|_2^2| \leq \delta \|x\|_2^2$. In other words, for small $\delta$, the matrix $A$ behaves almost like an isometry on sparse vectors.

In parallel, the community also focused on finding matrices obeying the RIPs. While the problem of finding deterministic matrices with interesting RIPs is still open, it has been proven that random matrices are good candidates [8].

More recently, results from compressed sensing have been generalized to handle weighted sparsity [5]. In a similar manner, given a weight sequence $(\omega_j)_{j \in \Lambda}$, $\omega_j \geq 1$, one can derive weighted sparsity of a vector $x$ as $\|x\|_{\omega,0} := \sum_{j \in \operatorname{supp}(x)} \omega_j^2$. Similarly, weighted $\ell_p$ norms are defined, for $0 < p \leq 2$ as $\|x\|_{\omega,p} = \left( \sum_{j \in \Lambda} \omega_j^{2-p} |x_j|^p \right)^{1/p}$ and the associated spaces are defined as

$$\ell_{p,\omega}(\Lambda) := \left\{ x \in \mathbb{R}^\Lambda : \|x\|_{p,\omega} < \infty \right\} . \tag{6}$$

Weigthed $\ell_p$ can be further extended to functions $f = \sum_{j \geq 1} x_j \varphi_j$ expanded in certain basis elements $(\varphi_j)_{j \geq 1}$ as $|||f|||_{\omega,p} := \|x\|_{\omega,p}$. This in particular is interesting for function interpolation from very few samples:

**Theorem 1** (Complete statement in [5])**.** *For $(\varphi_j)_{j \in \Lambda}$ a countable orthonormal system and a sequence of weights $\omega_j \geq \|\varphi_j\|_\infty$, there exists a finite subset $\Lambda_0$ with $N = |\Lambda_0|$ such that drawing $m \geq c_0 s \log^3(s) \log(N)$ samples at random ensures that, with high probability,*

$$\|f - f^\#\|_\infty \leq |||f - f^\#|||_{\omega,1} \leq c_1 \sigma_s(f)_{\omega,1}, \tag{7}$$

$$\|f - f^\#\|_2 \leq d_1 \sigma_s(f)_{\omega,1} / \sqrt{s} \tag{8}$$

*where $f := \sum_{j \in \Lambda} x_j \varphi_j$ is the function to recover, and $f^\#(y) := \sum_{j \in \Lambda_0} x_j^\# \varphi_j(y)$ is the solution obtained via weighted $\ell_1$ minimization.*

In particular, the previous theorem offers error bounds in both $L_2$ and $L_\infty$ norms.

### C. Petrov-Galerkin Approximations

*Petrov-Galerkin* (PG) approximations of a function $u \in V$ defined on $D$ are based on one-parameter families of nested and finite-dimensional subspaces $\{V^h\}_h$ and $\{W^h\}_h$ which are dense in $V$ and $W$, respectively (with the discretization parameter $h$ denoting e.g. meshwidth of triangulations in Finite Element methods, or the reciprocal of the spectral order in spectral methods). Given $y \in U$, the PG projection of $u(y)$ onto $V^h$, $u^h(y) := G_h(u(y))$, is defined as the unique solution of the finite-dimensional, parametric variational problem: *Find* $u^h \in V^h$ such that

$$(A(y) u^h(y))(w^h) = f(w^h), \quad \forall w^h \in W^h \ . \tag{9}$$

Note that for $A(y)$ boundedly invertible and under some (uniform w.r. to $y \in U$ and w.r. to $h$) inf-sup conditions (see, e.g. [1, Prop. 4]) the PG projections $G_h$ in (9) are well-defined and quasioptimal *uniformly w.r. to* $y \in U$: there exists a constant $C > 0$ such that for all $y \in U$ csand $h > 0$ sufficiently small it holds

$$\|u(y) - u^h(y)\|_V \leq C \inf_{v^h \in V^h} \|u(y) - v^h\|_V \ . \tag{10}$$

### D. Review of the CSPG algorithm

We review the CSPG algorithm of [2]. It relies on the evaluation of $m$ solutions $b^{(l)} := \mathcal{G}(u(y^{(l)}))$ of Eq. (1) given $m$ randomly chosen parameter instances $y^{(l)} = (y_j^{(l)})_{j\geq 1}$, $1 \leq l \leq m$. These real-valued solutions $b^{(l)}$ are then used in a weighted $\ell_1$ minimization program in order to determine the most significant coefficients in expansion (2), see Algorithm 1[1].

**Data**: Weights $(v_j)_{j\geq 1}$, an accuracy $\varepsilon$ and sparsity $s$ parameters, a multiindex set $\mathcal{J}_0^s$ with $N = |\mathcal{J}_0^s| < \infty$, a compressibility parameter $0 < p < 1$, and a number of samples $m$.
**Result**: $\left(g_\nu^\#\right)_{\nu\in\mathcal{J}_0^s}$, a CS-based approximation of the coefficients $(g_\nu)_{\nu\in\mathcal{J}_0^s}$.
**for** $l = 1, \cdots, m$ **do**
  Draw $y^{(l)}$ at random;
  Numerically estimate $b^{(l)}$;
**end**

$$(\Phi)_{l,\nu} \leftarrow T_\nu(y^{(l)}), \quad \nu \in \mathcal{J}_0^s, 1 \leq l \leq m$$
$$\omega_\nu \leftarrow 2^{\|\nu\|_0/2} \prod_{j:\nu_j\neq 0} v_j^{\nu_j}, \quad \nu \in \mathcal{J}_0^s \tag{11}$$

Compute $g^\#$ as the solution of

$$\min \|g\|_{\omega,p}, \qquad \text{s.t. } \|\Phi g - b\|_2 \leq 2\sqrt{m}\varepsilon \tag{12}$$

**Algorithm 1:** Pseudo-code for the CSPG Algorithm

Before we justify the use of this algorithm, it is worth mentioning the roles of the different parameters: $\varepsilon$ is an accuracy parameter that will have an impact on the discretization parameter $h$ and on the number of samples required. The importance and impact of the choice of the weight sequence $(v_j)_{j\geq 1}$ is still not clear and is left for future research. The index set $\mathcal{J}_0^s \subset \mathcal{F}$ (see definition below) acts as an estimation of the optimal set containing the best $s$-(weighted) sparse approximation of $F$.

**Theorem 2** ([2])**.** *Let $u(y)$ be the solution to Eq.* (1) *such that $A_0$ is boundedly invertible and $\sum_{j\geq 1}\beta_{0,j} \leq \kappa \in (0,1)$ where $\beta_{0,j} := \|A_0^{-1}A_j\|_{\mathcal{L}(V,V)}$, and the scales of smoothness spaces have a parametric regularity property (see [2, Assumption 2.3]). If the sequence $(v_j)_{j\geq 1}$ is such that $\sum_j \beta_{0,j} v_j^{(2-p)/p} \leq \kappa_{v,p} \in (0,1)$ and $(\beta_{0,j})_j \in \ell_{v,p}$ then $g \in \ell_{\omega,p}$ with $\omega$ as in Eq.* (11)*. Define an accuracy $\varepsilon > 0$ and sparsity $s$ such that*

$$\sqrt{5}\ 4^{1-1/p} s^{1/2-1/p}\|g\|_{\omega,p} \leq \varepsilon \leq C_2 s^{1/2-1/p}\|g\|_{\omega,p}, \tag{13}$$

*with $C_2 > \sqrt{5}4^{1-1/p}$ independent of $s$ and define $\mathcal{J}_0^s := \{\nu \in \mathcal{F} : 2^{\|\nu\|_0}\prod_{j:\nu_j\neq 0} v_j^{2\nu_j} \leq s/2\}$ with $N := |\mathcal{J}_0^s|$. Pick $m \asymp C s\log^3(s)\log(N)$ ($C$ a universal constant) samples i.i.d at random w.r.t. to the Chebyshev distribution and define $\widetilde{F}(y) := \sum_{\nu\in\mathcal{J}_0^s} g_\nu^\#(y)$ with $g^\#$ solution of Eq.* (12)*. Then, there exists a (universal) constant $C'$ such that, with probability at least $1 - 2N^{-\log^3(s)}$, it holds:*

$$\|F - \widetilde{F}\|_2 \leq C'\|g\|_{\omega,p}s^{1/2-1/p} \leq C_g''\left(\frac{\log^3(m)\log(N)}{m}\right)^{1/p-1/2}, \tag{14}$$

$$\|F - \widetilde{F}\|_\infty \leq C'\|g\|_{\omega,p}s^{1-1/p} \leq C_g''\left(\frac{\log^3(m)\log(N)}{m}\right)^{1/p-1}, \tag{15}$$

*where $C_g''$ depends on $C'$ and $\|g\|_{\omega,p}$.*

## III. Numerical examples

For the empirical validation of the theories introduced in the previous sections, we consider various use cases in 1-dimensional physical domains. We are mostly interested in the high-dimensionality of the parameter space $U$. The physical domain of the operator equation (1) renders the PG discretization more complicated but does not affect the CS access to the parameter space $U$.

In the numerical examples we consider model affine-parametric diffusion equations (3), (4). As stated in the introduction, we consider $y = (y_j)_{j\geq 1} \in U$ and $x \in D = [0,1]$. For the practical numerical results, we use trigonometric polynomials as basis functions $a_j$ in (4), for a constant $\alpha$:

$$a_{2j-1}(x) = \frac{\cos(j\pi x)}{j^\alpha}, \quad a_{2j}(x) = \frac{\sin(j\pi x)}{j^\alpha}. \tag{16}$$

We used the nominal field $a_0 = 0.1 + 2\zeta(\alpha,1)$ where $\zeta(\alpha,1) := \sum_{n\geq 1}\frac{1}{n^\alpha}$.The remaining parameters are set as:

- $f \equiv 10$,
- For $j \geq 1$, $\widetilde{v_j} = \gamma j^\tau$, where $\gamma = 1.015$ and $\tau = \log(\lfloor\sqrt{s}/2\gamma\rfloor)/\log(d)$, then $v_{2j-1} = v_{2j} = \widetilde{v_j}$, $j \geq 1$,[2]
- $\alpha = 2$, $\varepsilon = 0.5\cdot 10^{-9}$,
- $m = 2s\log(s)\log(N)^3$ with $\mathcal{J}_0^s$ denoting the initial set of important coefficients of the GPC, and $N := |\mathcal{J}_0^s|$. $s$ varies from 30 to 500, see Table I for examples.[4]

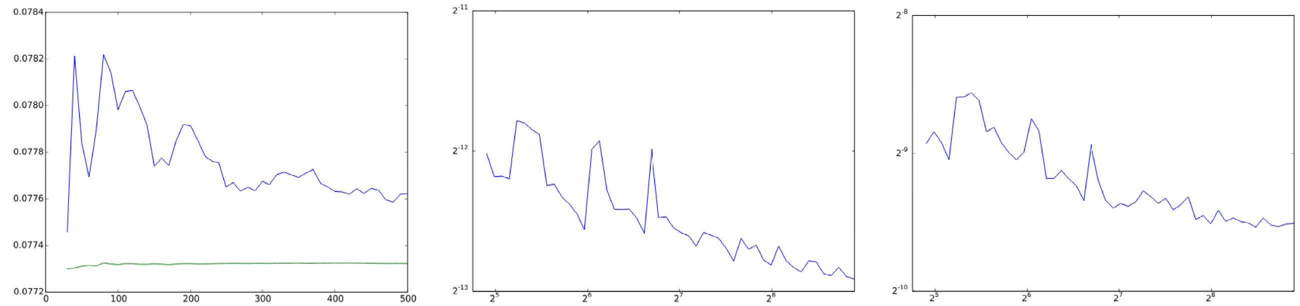
Note that no values are assigned for the compressibility parameter $p$. While it is of prime importance in the theoretical analysis and for the understanding of the algorithm, it has no effect on the implementation. For information purposes, we have also given the bounds on the error in Table I for a compressibility parameter $p = 1/2$ (i.e. $\text{err}^{1/p-1/2}$ and $\text{err}^{1/p-1}$, resp. for the $L^2$ and $L^\infty$ norms, where $\text{err} = \frac{\log^3(m)\log(N)}{m}$). Be aware that these estimations do not include the constant $C_g''$ in front in Eqs. (14) and (15). Another remark regarding the implementation is that the number of variables has been trunctaed to $d = 20$ parameters (i.e. $k$ stops at 10 in the expansion of the diffusion coefficient).

[1]Due to space constraints some details are omitted. We refer to [2] for details.

[2]doubling the $v_j$ ensures that the sin and cos functions with the same frequencies are given the same weights.

[3]This number of samples is different than the one suggested in the theorem. This is motivated by results on nonuniform recovery from random matrices in compressed sensing - see [7, Ch.9.2].

[4]The size of the initial set depends on the sparsity and on the weights. The numbers given here are only valid for the setting described above.



(a) Expectation in the $L^2$ norm of the MC estimation (blue) and CSPG (green) methods.

(b) Estimation of the convergence of the $L^2$ error norm for CSPG.

(c) Estimation of the convergence of the $L^\infty$ error norm for CSPG.

Fig. 1. Numerical results of the average of the solution of the operator equation (1) (see text for details).

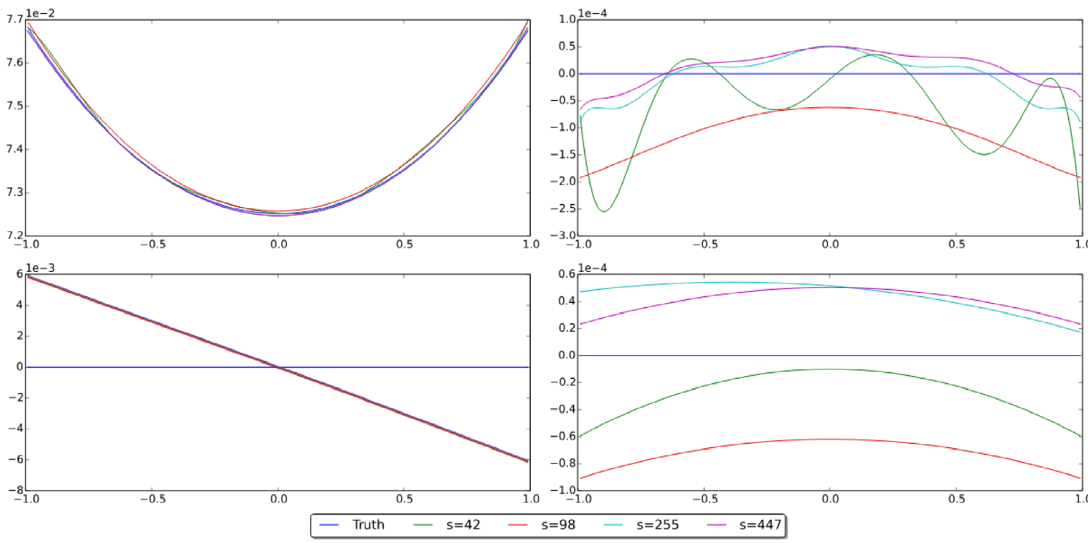



Fig. 2. Convergence pointwise for various sparsity levels.

TABLE I

DEPENDENCY OF THE NUMBER OF SAMPLES ON THE SPARSITY AND BOUNDS OF THE $L^2$ AND $L^\infty$ ERROR BOUNDS FOR $p = 1/2$.

| $s$ | 30 | 100 | 200 | 300 | 400 | 500 |
|---|---|---|---|---|---|---|
| $N$ | 1554 | 5712 | 10006 | 12690 | 15601 | 17043 |
| $m$ | 441 | 1731 | 3685 | 5670 | 7725 | 9744 |
| $L^2$ | 1.9396 | 1.4393 | 1.1765 | 1.0373 | 0.9469 | 0.8802 |
| $L^\infty$ | 3.7619 | 2.0717 | 1.3843 | 1.0759 | 0.8967 | 0.7747 |

While we would ideally like to compare the approximations $\widetilde{F}$ with the true functions $F$ in the $L^2$ and $L^\infty$ norms, this is unrealistic. We provide a lower bound on the $L^\infty$ norm based on independent draws of test samples $z^{(l)}$, $1 \leq l \leq m_{\text{test}} := 10000$. The error in the $L^2$ norm can be accurately estimated (through Bernstein inequality) by sums of differences.

On the other hand, a comparison with usual Monte Carlo methods can be done only when comparing the first coefficient $g_0$ of the expansion (2), the expectation of the function, and the Monte-Carlo estimate. Indeed, for $F(y) := \sum_{\nu \in \mathcal{F}} g_\nu T_\nu(y)$, for $T_\nu$ the tensorized Chebyshev polynomials, it holds $\mathbb{E}[F] = g_0$. Monte-Carlo estimation is done as the empirical mean $E_m[F] := \frac{1}{m}\sum_{l=1}^m F(y^{(l)})$. It is known (see [9, Lemma 4.1]) that for a number of samples $m$ drawn at random it holds

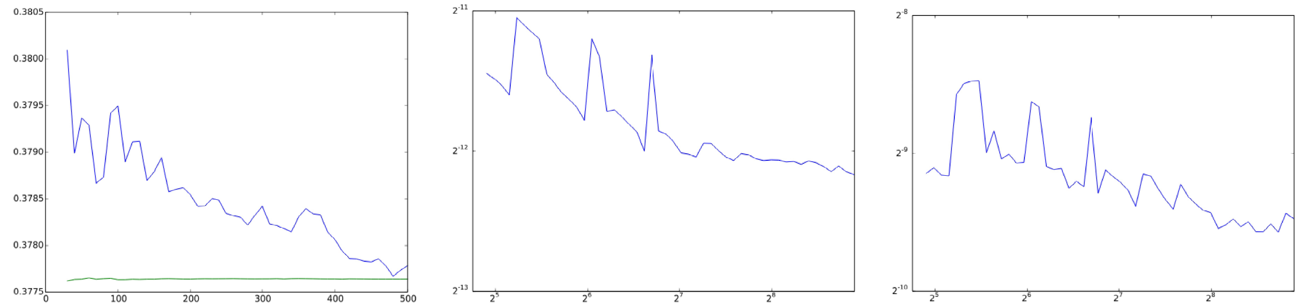
$$\|E_m[F] - \mathbb{E}[F]\|_{L^2(U,\eta)} \leq m^{-1/2}\|F\|_{L^2(U,\eta)} \tag{17}$$

### A. Average

For this first use case, we consider the functional $I : V \to \mathbb{R}$ defined as $I(u(y)) := \int_{x\in D} u(y;x)\mathrm{d}x$, and $D = [0,1]$. Fig. 1(a) illustrates the convergence of both the Monte-Carlo and the CSPG method. The $x$-axis corresponds to the sparsity parameter $s$, or, seen differently, to the number of samples used (see Table I for the correspondance).

The CSPG method shows a much better convergence of the expectation. The other interesting point to note and that has not been studied so far, is the stability of the first coefficient. While the number of samples required to ensure small $L^2$ and $L^\infty$ errors is important (as illustrated in Figs. 1(b) and 1(c) respectively) it seems that the reliable estimation of the first coefficient is not drastically affected by strong undersampling.

Figs. 1(b) and 1(c) respectively illustrate the estimation of the empirical $L^2$ and the $L^\infty$ errors. These values have to be compared with the last two rows of Table I.




(a) Expectation of the exponential functional

(b) $L^2$ error estimation

(c) $L^\infty$ error estimation

Fig. 3. Numerical results of the exponential weighted integral.

We can also look at the pointwise convergence of the approximation in $U$ with respect to the sparsity (or number of samples). Fig. 2 shows the values of $t \mapsto \widetilde{F}(te_k)$, where $e_k = (\delta_{kj})_j$, for $-1 \leq t \leq 1$. The upper left corner shows the computed solution of $t \mapsto \widetilde{F}(te_1)$ for various sparsity levels. The remaining three quadrants correspond to the (pointwise) error for $k = 1, 2, 3$. The CSPG approximated solutions converge, as $s$ increases, to the true solution.

### B. Exponential convolution kernel

In this subsection, we study the accuracy of the CSPG method when we consider an exponential weighted integral centered at a point $x_0 = 0.5$: $F(y) := u(y) * K_{x_0}$ where the convolution kernel is given, for a constant $\mathcal{N}$, by

$$K_{x_0}(x) := \frac{1}{\mathcal{N}} e^{-\frac{|x-x_0|}{\sigma}}. \tag{18}$$

Again we can notice the faster convergence rate of the CSPG method. Moreover, the bounds derived in Theorem 2 are very crude and easily achieved numerically. Fig. 3 illustrates the convergence of the CSPG algorithm. In the left graph, the expectation is shown with respect to the (weighted) sparsity. The blue curve illustrates the convergence of the MC method, while the rather flat constant line implies that the CSPG method computes the first coefficient of the GPC expansion $g_0$ well very soon. The two following log-log plots show the decrease of the $L^2$ and $L^\infty$ errors. As the sparsity increases, the results are more accurate and, seem to be more stable.

## IV. Conclusion

This paper shows the practical and numerical use of a compressed-sensing based Petrov-Galerkin approximation method for computing the functionals of solutions of high-dimensional parametric elliptic equations. We have empirically verified the theoretical results and compared them, when possible, to MC methods, where the convergence appears to be slower than the proposed approach. Although only model, affine-parametric problems were considered here, we emphasize that the presently proposed CSPG approach is nonintrusive, and exploits sparsity of the GPC expansion; therefore, its range of application extends to non-affine parametric operator equations, such as PDEs in uncertain domains. The presently proposed CS based GPC recovery does not, a-priori, impose specific structure on the sets of active coefficients as required in some adaptive collocation methods, see for instance [1], [10]. It is also important to put our results in perspective with those of [11] where probabilistic error bounds were derived.

The promising results presented here require further analysis as to the quantitative effect of the different parameters, in particular regarding the weights in Eq. (11) (and associated $v_j$'s), are not well understood. Moreover, it will be of interest to later derive methods based on these techniques to recover the full, parametric solution of (1), rather than only a functional.

## Acknowledgment

J.-L. B. and H. R. are funded by the European Research Council through the Starting Grant StG 258926 (SPALORA).

C. S. acknowledges support from the Swiss National Science Foundation.

## References

[1] C. Schwab, "Qmc galerkin discretization of parametric operator equations," in *Monte Carlo and Quasi-Monte Carlo Methods 2012*. Springer, 2013, pp. 613–629.

[2] H. Rauhut and C. Schwab, "Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations," arXiv preprint arXiv:1410.4929, 2014.

[3] A. Cohen, R. Devore, and C. Schwab, "Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's," *Analysis and Applications*, vol. 9, no. 01, pp. 11–47, 2011.

[4] A. Cohen, R. DeVore, and C. Schwab, "Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs," *Foundations of Computational Mathematics*, vol. 10, no. 6, pp. 615–646, 2010.

[5] H. Rauhut and R. Ward, "Interpolation via weighted $\ell_1$ minimization," arXiv preprint arXiv:1308.0759, 2013.

[6] M. Hansen and C. Schwab, "Analytic regularity and nonlinear approximation of a class of parametric semilinear elliptic PDEs," *Mathematische Nachrichten*, vol. 286, no. 8–9, pp. 832–860, 2013.

[7] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.

[8] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[9] A. Barth, C. Schwab, and N. Zollinger, "Multi-level monte carlo finite element method for elliptic pdes with stochastic coefficients," *Numerische Mathematik*, vol. 119, no. 1, pp. 123–161, 2011.

[10] C. Schillings and C. Schwab, "Sparse, adaptive smolyak quadratures for bayesian inverse problems," *Inverse Problems*, vol. 29, no. 6, p. 065011, 2013.

[11] A. Doostan and H. Owhadi, "A non-adapted sparse approximation of pdes with stochastic inputs," *Journal of Computational Physics*, vol. 230, no. 8, pp. 3015–3034, 2011.