MULTI-LEVEL COMPRESSED SENSING PETROV-GALERKIN DISCRETIZATION OF HIGH-DIMENSIONAL PARAMETRIC PDES

JEAN-LUC BOUCHOT, HOLGER RAUHUT, AND CHRISTOPH SCHWAB

Abstract. We analyze a novel multi-level version of a recently introduced compressed sensing (CS) Petrov-Galerkin (PG) method from [H. Rauhut and Ch. Schwab: Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations, Math. Comp. **304**(2017) 661–700] for the solution of many-parametric partial differential equations. We propose to use multi-level PG discretizations, based on a hierarchy of nested finite dimensional subspaces, and to reconstruct parametric solutions at each level from level-dependent random samples of the high-dimensional parameter space via CS methods such as weighted ℓ_1-minimization. For affine parametric, linear operator equations, we prove that our approach allows to approximate the parametric solution with (almost) optimal convergence order as specified by certain summability properties of the coefficient sequence in a general polynomial chaos expansion of the parametric solution and by the convergence order of the PG discretization in the physical variables. The computations of the parameter samples of the PDE solution is "embarrasingly parallel", as in Monte-Carlo Methods. Contrary to other recent approaches, and as already noted in [A. Doostan and H. Owhadi: A non-adapted sparse approximation of PDEs with stochastic inputs. JCP **230**(2011) 3015-3034] the optimality of the computed approximations does not require a-priori assumptions on ordering and structure of the index sets of the largest gpc coefficients (such as the "downward closed" property). We prove that under certain assumptions work versus accuracy of the new algorithms is asymptotically equal to that of one PG solve for the corresponding nominal problem on the finest discretization level up to a constant.

1. Introduction

Motivated in particular by uncertainty quantification, the numerical solution of parametric operator equations has gained significant attention in recent years. In many cases, the underlying parameter space is high dimensional or even infinite dimensional so that standard approximation methods are subject to the curse of dimensionality, see e.g. [17, 16]. Monte Carlo (MC) sampling, however, may be used in the context that the parametric model arises from a stochastic model and leads to a mean-square rate of $m^{-1/2}$ in terms of the number m of sample evaluations, with constants that are independent of the parameter dimension. The (dimension-independent) rate 1/2 is not improvable in MC methods, in general, and the challenge consists in developing methods that achieve a faster convergence rate and at the same time alleviate or even overcome the curse of dimensionality.

A number of computational approaches have emerged in recent years towards this end. Among these are adaptive stochastic Galerkin methods, as developed in [26, 25, 34], reduced basis approaches (see, eg., [6, 12]), adaptive Smolyak discretizations [49, 50], adaptive interpolation methods [14] as well as sampling methods [52]. Adaptive Galerkin methods [26, 25, 34] are intrusive in the sense that they cannot simply reuse a solver developed for the corresponding problem with fixed parameter. In contrast, the other above mentioned methods and algorithms are non-intrusive, but they rely on successive numerical solutions of the operator equations for various parameter instances that are chosen based on suitable precomputations. In contrast, (multilevel) Monte-Carlo ($MLMC$)[44], or Quasi-Monte Carlo approaches (QMC)[23] compute expectations or statistical moments of the (random parametric) solution via solutions for parameter instances chosen at random or "quasi-random", which allows to compute the "parameter snapshot" solutions in parallel.

Date: January 6, 2017.

CS supported in part by European Research Council (ERC) through the grant AdG 247277, JLB and HR supported in part by the ERC grant StG 258926. JLB and HR would like to thank the Hausdorff Research Institute for Mathematics, University of Bonn, where parts of this work have been performed during the trimester program Mathematics of Signal Processing. JLB thanks the support of the Clay Mathematics Institute for his visit to the CRM to attend the IRP Constructive Approximation and Harmonic Analysis where parts of this work has been done.

In this article, we build on a compressed sensing approach for numerically computing parametric solutions developed and analyzed in [46, 9] (see also [24, 45] for earlier work, and [15] for recent developments) and combine it with ideas originating from MLMC methods, see e.g. [33, 4]. For Petrov-Galerkin (PG) discretizations on a finite hierarchy of nested subspaces, ordered with respect to discretization levels, the presently proposed method "samples", in a judicious fashion, the parameter space and computes corresponding PG approximations for random choices of the parameter vector. As in MLMC-PG approaches, the number of such snapshot evaluations decreases with increasing discretization level (corresponding to increasing refinement of the discretization). In contrast to (ML)MC sampling, we employ a CS technique based on weighted ℓ_1-minimization [47, 1] or Iterative Hard Thresholding Pursuit [40, 7] in order to reconstruct the coefficients of a generalized polynomial chaos expansion of the difference of the parametric solution at two subsequent discretization levels. Finally, these differences are summed together to obtain a PG approximation of the parametric solution at the finest level. One contribution of this paper is to show that the generalized polynomial chaos (GPC) expansion of the differences of PG approximations of the parametric solution is approximately sparse by estimating the weighted ℓ_p-norm for $0 < p < 1$ of the sequence of Chebyshev coefficients by a term that depends in a controlled way on the discretization level. This fact makes the presently developed, multi-level version of the compressive sensing approach feasible. We provide dimension-independent convergence rates which exceed $1/2$ under certain sparsity assumptions on the parametric solution family of the operator equation and estimate the computational complexity for achieving such rates. Similar to MLMC methods, the workload for approximating the parametric solution is asymptotically the same as the one for computing one snapshot solution at the finest level up to a constant that depends only on smoothness parameters and $p \in (0, 1)$. However, in contrast to multilevel Monte Carlo, the convergence rates afforded by our scheme are practically independent of the dimension and only limited by the solutions' sparsity; in particular, they may significantly exceed $\mathcal{O}(m^{-1/2})$.

In mathematical terms, we consider linear, parametric operator equations of the generic form

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

Here the parameter vector $\boldsymbol{y} \in U$ lies in a high-dimensional space U making it challenging to computationally approximate the solution map $\boldsymbol{y} \mapsto u(\boldsymbol{y})$, due to the mentioned *curse of dimensionality*, a notion going back to R.E. Bellman [5], see [17, 16] for its relevance in the present context. Assuming that the parameter vector $\boldsymbol{y} = (y_j)_{j=1}^d$ takes values in finite intervals, we can consider, without loss of generality, $U = [-1, 1]^d$, where the parameter set dimension d may be finite or infinite.

In our setting, the parametric family of operators $A(\boldsymbol{y}) : \mathcal{X} \to \mathcal{Y}'$ maps from a reflexive Banach space $\mathcal{X}$ to the topological dual of, potentially, another reflexive Banach space $\mathcal{Y}$. A canonical example is the affine-parametric diffusion equation considered in [18, 19] and in the single-level version of the present work [46, 9]. For a bounded Lipschitz domain $D \subset \mathbb{R}^n$ (one should think of $n = 1, 2, 3$) and a parametric diffusion coefficient $a(\cdot, \boldsymbol{y}) \in L^\infty(D)$ that depends affinely on a parameter vector $\boldsymbol{y}$, i.e.,

$$a(x, \boldsymbol{y}) = \bar{a}(x) + \sum_{j \geq 1} y_j \psi_j(x), \quad x \in D, \tag{2}$$

we consider the model parametric, second order divergence form elliptic Dirichlet problem

$$A(\boldsymbol{y})u := -\nabla \cdot (a(\cdot, \boldsymbol{y})\nabla u) = f \quad \text{in } D, \quad u|_{\partial D} = 0. \tag{3}$$

The weak formulation of (3) in the Sobolev space $\mathcal{X} = \mathcal{Y} := H_0^1(D)$ reads: Given $f \in \mathcal{Y}'$, for every $\boldsymbol{y} \in U := [-1, 1]^{\mathbb{N}}$ find $u(\boldsymbol{y}) \in \mathcal{X}$ such that

$$\int_D a(x, \boldsymbol{y})\nabla u(x) \cdot \nabla v(x) \mathrm{d}x = \int_D f(x)v(x)\mathrm{d}x, \quad \text{for all } v \in \mathcal{Y}. \tag{4}$$

Eq. (3) is a particular example of an *affine-parametric operator equation* of the form

$$A(\boldsymbol{y}) := A_0 + \sum_{j \geq 1} y_j A_j, \quad \boldsymbol{y} = (y_j)_{j \geq 1} \in U := [-1; 1]^{\mathbb{N}}, \tag{5}$$

with $A_j := -\nabla \cdot (\psi_j \nabla), A_0 = -\nabla \cdot (\bar{a}\nabla)$. In (5), the operator $A_0 \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ is traditionally referred to as *nominal operator* or *mean field* while the operators $A_j \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$, for $j \geq 1$, are referred to as *fluctuations*. For the parametric problem to be well-posed uniformly with respect to the parameter $\boldsymbol{y} \in U$, we assume that $\sum_{j\geq 1}\|A_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')}<\infty$ in what follows. Further assumptions required for the convergence and applicability of our approach will be specified ahead. Parametric expansions such as (5) can be obtained e.g. by a Karhunen-Loève expansion of random input data for divergence-form partial differential equations, as explained in [51, 19].

In order to ensure well-posedness of the parametric diffusion problem (3) as in [19] we require the *uniform ellipticity assumption*: there exist constants $0<r\leq R<\infty$ such that

$$r \leq a(x,\boldsymbol{y}) \leq R, \qquad \text{for almost all } x\in D, \text{ for all } \boldsymbol{y}\in U\ . \tag{6}$$

The Lax-Milgram Lemma ensures that for every $\boldsymbol{y}\in U$, the weak formulation (4) admits a unique solution $u(\cdot,\boldsymbol{y})\in\mathcal{X}$ which satisfies the uniform a priori estimate

$$\sup_{\boldsymbol{y}\in U}\|u(\boldsymbol{y})\|_{\mathcal{X}} \leq r^{-1}\|f\|_{\mathcal{Y}'}.$$

Here and throughout the remainder, the term *"uniform"* refers to uniform with respect to the parameter sequence $\boldsymbol{y}\in U$.

For the sake of simplicity we detail here only the approximation of functionals $\mathcal{G}\in\mathcal{X}'$ of solutions to the parametric operator equation (1), i.e., we are interested in the numerical approximation of

$$F(\boldsymbol{y}) := \mathcal{G}(u(\boldsymbol{y})), \qquad \boldsymbol{y}\in U=[-1,1]^d,$$

pointwise with respect to $\boldsymbol{y}$. We expect that our approach can be generalized to the recovery of the vector-valued solution map $\boldsymbol{y}\mapsto u(\boldsymbol{y})$, but we postpone this generalization to later contributions. We are aiming at numerical schemes that are:

- Reliable: the convergence and accuracy should be verified and customizable;
- Parallelizable: parallel sampling as in Monte-Carlo methods should be allowed, with a convergence rate in terms of the number of samples which (up to possibly logarithmic terms) equal the best possible rate ensured by the compressibility of $F(\boldsymbol{y})$, i.e., by weighted ℓ_p-estimates of the Chebyshev coefficients of F;
- Non-intrusive: the approximation should use existing numerical solvers of the problem with fixed parameters, without any re-implementation of PDE solvers.

It is important to notice the difference to usual MC methods where the results obtained from random sampling usually hold in expectation. In contrast, our approach provides approximations that hold pointwise with respect to $\boldsymbol{y}$. We estimate the coefficients of a tensorized Chebyshev expansion; whence only matrix-vector multiplications are required in order to compute the solution $F(\boldsymbol{y})=\mathcal{G}(u(\boldsymbol{y}))$ for any given parameter vector $\boldsymbol{y}=(y_j)_{j=1}^d$ up to a prescribed accuracy. The computation scheme analyzed here differs from the single-level one introduced in [46] in the sense that computing the approximation is done in a more efficient and computationally tractable manner. To this end, an unknown function $u(\boldsymbol{y})$ is approximated by a telescopic sequence of so-called "details" at successively finer spatial resolutions: $u(\boldsymbol{y})\approx\sum_{l=1}^{L}(u^l(\boldsymbol{y})-u^{l-1}(\boldsymbol{y}))$ where u^l corresponds to a PG approximation on a discretization level l. This is analogous to MLMC methods, but is achieved here by *compressive sensing* of the parameters with a *level-dependent* number of parameter samples $\boldsymbol{y}^{(i)}$ on each discretization level in the physical domain.

We outline key ideas of the compressive sensing approach. We assume at our disposal a countable orthonormal basis $(\varphi_\nu)_{\nu\in\Lambda}$ of $L^2(U,\eta)$ with η denoting a probability measure on the parameter set U to be specified, and denote by $L^2(U,\eta;\mathcal{X})$ the Bochner space of strongly measurable maps from U to the (separable Hilbert) space $\mathcal{X}$ containing solution instances, which are square integrable w.r.t. η. We represent any function $u(\boldsymbol{y})$ with values in $\mathcal{X}$ as $u(\boldsymbol{y})=\sum_{\nu\in\Lambda}\alpha_\nu\varphi_\nu(\boldsymbol{y})$, where $\alpha=(\alpha_\nu)_{j\in\Lambda}$ denotes the unique sequence of coefficients $\alpha_\nu\in\mathcal{X}$. Hence, in order to compute an approximation of the parametric solution for any $\boldsymbol{y}$ it suffices to calculate an approximation of the coefficients α_ν. For a new input parameter $\boldsymbol{y}$, one evaluates the basis functions φ_ν at $\boldsymbol{y}$ and forms a linear combination to recover a direct estimation of the solution. Later on, we analyze the use of tensorized Chebyshev polynomials as orthonormal basis. The approximation is computed by evaluating the function at a few parameter points $\boldsymbol{y}^{(i)}$, $1\leq i\leq m$, and solving the linear system $\boldsymbol{g}=\Phi\alpha$, where $\boldsymbol{g}=(g_i)_{i=1}^m=\big(u(\boldsymbol{y}^{(i)})\big)_{i=1}^m$ and where Φ corresponds to the sensing matrix $\Phi\in\mathbb{R}^{m\times N}$ with entries $\Phi_{i,\nu}=\varphi_\nu(\boldsymbol{y}^{(i)})$, where N corresponds to the number of basis functions taken for the approximation. However at this stage the coefficients α_ν and the components g_i are elements in $\mathcal{X}$, and therefore, we first deal with the simpler case where a functional $\mathcal{G}$ (also known as the Quantity of Interest (QoI for short) in the uncertainty quantification literature) is applied to the solution, resulting in

$$b_i = \mathcal{G}(g_i) = \mathcal{G}\left(u(\boldsymbol{y}^{(i)})\right) = \sum_{\nu\in\Lambda} z_\nu \varphi_\nu(\boldsymbol{y}^{(i)}), \quad i = 1,\dots,m, \qquad z_\nu = \mathcal{G}(\alpha_\nu), \quad \nu \in \Lambda\ .$$

We are particularly interested in the situation that the number m of evaluations is smaller than the cardinality of Λ, so that the linear system $b = \Phi z$ is underdetermined. Approximate sparsity of the coefficient sequence (α_ν), and of (z_ν), allows to apply techniques from compressive sensing such as (weighted) ℓ_1-minimization or iterative hard thresholding (pursuit) in order to recover z accurately. In fact, approximate sparsity follows from the fact that $(\|\alpha_\nu\|_{\mathcal{X}})$ and (z_ν) are contained in weighted $\ell_p(\mathcal{F})$-spaces, as shown in [3, 16, 18, 19, 46] and, for a related coefficient sequence, in this paper.

We expect that an approximation of the full solution $u(\boldsymbol{y})$, $\boldsymbol{y} \in U$, taking values in the function space $\mathcal{X}$, can be computed by a variant of our compressive sensing scheme. One may use ideas from joint/block sparsity in order to recover the sequence (α_ν) with $\alpha_\nu \in \mathcal{X}$ via mixed ℓ_1/ℓ_2-minimization, see e.g. [27, 28, 30] (at least in the case that $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces). However, we postpone a detailed analysis to a later contribution and restrict ourselves here to the simpler case of recovering the real-valued function $\boldsymbol{y} \mapsto \mathcal{G}(u(\boldsymbol{y}))$.

The multi-level approximation scheme uses discretization levels $l = 1,\dots,L$, where the meshwidth at discretization level l is $2^{-l}h_0$, so that the finest discretization is $h_L = 2^{-L}h_0$. With n being the dimension of the domain D, we assume that the number of degrees of freedom at level l scales like $\mathcal{O}(2^{nl})$, and we further assume *available linear complexity, multilevel solvers for the approximate solution of the discretized linear system of equations* (uniformly with respect to the parameter $\boldsymbol{y}$) resulting in computational costs per PG solution $u^l(\boldsymbol{y}^{(i)})$ that scales linearly in the number of degrees of freedom: $\mathcal{O}(2^{nl})$.

The presently proposed multi-level extension of the CS PG approach from [46] proceeds analogous to MLMC (see, e.g., [39] or [33] and the references therein): for parameter choices $\{\boldsymbol{y}_l^{(i)}\}_{i=1,\dots,m_l}$ on discretization level l, compute PG solutions $u^l(\boldsymbol{y}_l^{(i)})$, $u^{l-1}(\boldsymbol{y}_l^{(i)})$ at two consecutive discretization levels l and $l-1$ (setting $u^0 \equiv 0$). From the differences $\mathfrak{d}u^l(\boldsymbol{y}_l^{(i)}) = u^l(\boldsymbol{y}_l^{(i)}) - u^{l-1}(\boldsymbol{y}_l^{(i)})$, we compute an approximation $\mathfrak{d}\widetilde{u}^l(\boldsymbol{y})$ via the single level compressive sensing approach of [46] for each $l = 1,\dots,L$. Finally, we combine the approximations at all levels similarly as in MLMC methods, i.e., $\widetilde{u}(\boldsymbol{y}) = \sum_{l=1}^{L} \mathfrak{d}\widetilde{u}^l(\boldsymbol{y})$, to obtain an approximation of the full parametric solution. The main result of this paper consists of an analysis of this method and provides, in its proof, a strategy on how to choose the number m_l of parameter points at each level l. Its precise statement, Theorem 9, is postponed to later in the exposition. To illustrate the type of results obtained here, we state now a version of Theorem 9 in the particular case of a linear, divergence form diffusion operator with affine dependence on the parameters (see Eqs. (2) and (3)). Ahead, we say that the weight sequence $\mathbf{v}$ is constant, if it is of the form $v_j = \beta$ for $j = 1,\dots,d$ for some $\beta > 1$ and $v_j = \infty$ for $j > d$, which corresponds to the case that the expansion (2) is finite (with d terms). We say that $\mathbf{v}$ has polynomial growth if $v_j = cj^\alpha$, $j \in \mathbb{N}$, for some $c > 1$, $\alpha > 0$. We refer to Section 5.3 for details on the weight sequences.

Theorem 1. *Let $L \in \mathbb{N}$ and $\gamma \in (0,1)$. Consider the diffusion equation (3) with affine parametric coefficient (2), forcing term $f \in H^{-1+t}(D)$ and functional $\mathcal{G} \in H^{-1+t'}(D)$, with the respective smoothness parameters $t, t' \geq 0$. Assume that in (2) holds $\bar{a} \in W^{t,\infty}(D)$ and that the fluctuations fulfill the weighted p-summability*[1]

$$\sum_{j\geq 1} \|\psi_j\|^p_{W^{t,\infty}(D)} v_j^{2-p} < \infty,$$

for a sequence $\mathbf{v} = (v_j)_{j\geq 1}$ of weights as well as the following stronger, weighted *version of the Uniform Ellipticity Assumption (6): there exists $0 < r \leq R < \infty$ such that*

$$\sum_{j\geq 1} v_j^{(2-p)/p} |\psi_j(x)| \leq \min\{\bar{a}(x) - r, R - \bar{a}(x)\}, \quad \text{for all } x \in D. \tag{7}$$

[1] To ease the presentation, here and throughout the paper, we have not highlighted the dependence of the summability parameter p on the regularity t of the right-hand-side f. It should be noted that the compressibility of the gpc expansion, the choice of the weight sequence, the number of samples per level all depend on the regularity of the data $\bar{a}$, ψ_j, D and f.

With probability at least $1-\gamma$, the function $F(\boldsymbol{y}) := \mathcal{G}(u(\boldsymbol{y}))$, $\boldsymbol{y} \in U$, can be approximated by L (weigthed) compressed sensing approximations based on a sequence of Galerkin projections into spaces of piecewise polynomials on regular, simplicial triangulations of meshwidth $h_\ell = 2^{-\ell} h_0$ from

$$m_l \gtrsim \max\{s_l \log^3(s_l)\log(N_l), \log(L/\gamma)\}$$

solution evaluations at discretization level l for $l = 1, \dots, L$, where $s_l \asymp 2^{(L-l)(t+t')p/(1-p)}$, and N_l is the size of the (level-dependent) active set of tensorized Chebyshev polynomials. The resulting approximation $F^\#$ satisfies

$$\|F - F^\#\|_\infty \leq C_p \|f\|_{H^{-1+t}(D)} \|\mathcal{G}\|_{H^{-1+t'}(D)} L 2^{-(t+t')L} h_0^{t+t'}$$
$$\|F - F^\#\|_2 \leq C_p' \|f\|_{H^{-1+t}(D)} \|\mathcal{G}\|_{H^{-1+t'}(D)} 2^{-(t+t')L} h_0^{t+t'}$$

Under the assumption that the computational cost of a single solve at level l scales linearly with respect to the number of degrees of freedom, i.e., is $\mathcal{O}(2^{nl})$ (for an n-dimensional domain D), this result is achieved with a total work for the computation of snapshot solutions that scales as $\mathcal{O}\left(\max\left\{2^{nL}, L^\beta 2^{L(t+t')p/(1-p)}\right\}\right)$, where $\beta = 4$ for constant weights $\mathbf{v}$ and $\beta = 5$ for polynomially growing weights. The constant hidden in the $\mathcal{O}$-notation includes a factor of $\log(d)$ in the case of constant weights.

We note in passing that in what follows, the estimates of the overall computational work do not account for the numerical solution of the convex programs required for the compressed sensing approximation of the mapping F. We justify this convention by the observation that the computational cost of ℓ_1-minimization is often of lower order compared to the total cost of evaluating the PDE samples.

Our theorem shows that in the case of sufficiently strong summability, i.e., $\frac{(t+t')p}{1-p} < n$, at a total cost that scales as a constant times a single PDE solve at the finest discretization level L, the multilevel CSPG (*MLCSPG*) strategy can approximate a fixed function F *for any* parameter vector $\boldsymbol{y} \in U$. This is analogous to what is afforded by MLMC methods, but the present MLCSPG strategy allows to achieve any convergence rate afforded by the gpc summability, and allows to approximate the full parametric dependence, while MLMC only yields expectations (or moments). Moreover, in our case the computational work scales favorably with decreasing p, which corresponds to better sparse approximation rates implied by the weighted p-summability of (norms of) polynomial chaos coefficients of the parametric solution. To be more precise, in the case of higher smoothness $t + t' > 0$, we obtain an approximation error that scales with $h_L^{t+t'}$. With a small enough value of p, we may exploit smoothness in the physical domain (allowing $t + t'$ such that $\frac{(t+t')p}{1-p} < n$) and balance approximation error for the PDE solves. In contrast, the computational work required by MLMC to achieve an expected approximation error scaling as h_L^t grows proportionally to 2^{2tL} when $2t \geq n$ (where t corresponds to the smoothness of the solution in the physical domain), see [4, Theorem 5.7], and there is no parameter p in MLMC whose tuning allows to avoid such growth.

Nevertheless, we note that t and p may not be tuned independently: in many instances increased smoothness t leads to a larger value of the summability parameter p.

We emphasize that the tools and results developed here do not require a particular structure on the support set of the best approximation. It is often the case (see e.g. [14, 43]), that proofs and/or methods require the sets of active indices in N-term gpc approximations be *downward closed*, their approximation properties then being, in particular, independent of the polynomial system adopted for implementation. In constrast, the presently proposed, compressed sensing based approach can recover (with high probability) *any support set of active multi-degrees of tensorized Chebyshev polynomial approximations* (only assuming very rough knowledge of its location as provided by weighted ℓ_p-estimates of polynomial chaos coefficients), yet still providing quasi-optimal rates of convergence. Moreover, apart from the ℓ_1-minimization part of the algorithm, all function evaluations can be done in parallel.

Theorem 1 is a particular case of our main Theorem 9 which we prove in Section 4 after recalling some basics about Petrov-Galerkin approximations in Section 2 combined with compressed sensing techniques in Section 3. Section 5 deals with pratical aspects such as truncating the dimension of the parameter space. The paper is finally concluded by numerical experiments to illustrate the theory in Section 6.

2. Petrov-Galerkin approximations

We deal with the pointwise numerical approximations of the countably-parametric operator equation Eq. (1). Numerically accessing the parametric solution map $U \ni \boldsymbol{y} \mapsto u(\boldsymbol{y})$ at a fixed parameter instance $\boldsymbol{y} \in U$ requires discretization of Eq. (1) also in "physical space". To this end, we introduce two dense, one-parameter families of discretization spaces $\{\mathcal{X}^h\}_{h>0} \subset \mathcal{X}$ and $\{\mathcal{Y}^h\}_{h>0} \subset \mathcal{Y}$ of equal finite dimensions $N^h := \dim(\mathcal{X}^h) = \dim(\mathcal{Y}^h)$ and assume that the parametric operator $A(\boldsymbol{y})$ fulfills the discrete and uniform $\inf - \sup$ conditions: there exists a $\mu > 0$ such that for any $h > 0$ and $\boldsymbol{y} \in U$

$$\begin{cases} \inf_{0\neq v^h\in\mathcal{X}^h} \sup_{0\neq w^h\in\mathcal{Y}^h} & \frac{\langle A(\boldsymbol{y})v^h, w^h\rangle}{\|v^h\|_\mathcal{X}\|w^h\|_\mathcal{Y}} \geq \mu > 0 \\ \inf_{0\neq w^h\in\mathcal{Y}^h} \sup_{0\neq v^h\in\mathcal{X}^h} & \frac{\langle A(\boldsymbol{y})v^h, w^h\rangle}{\|v^h\|_\mathcal{X}\|w^h\|_\mathcal{Y}} \geq \mu > 0 \,. \end{cases} \tag{8}$$

The PG projections are defined as the solution to the following weak variational problems:

$$\text{Find } u^h(\boldsymbol{y}) := G^h(\boldsymbol{y})(u(\boldsymbol{y})), \text{ such that } \langle A(\boldsymbol{y})u^h(\boldsymbol{y}), v^h\rangle = \langle f, v^h\rangle \quad \text{for all } v^h \in \mathcal{Y}^h \,. \tag{9}$$

We recall the following classical result (see for example [8, Chapter 6]).

Proposition 1. *Let $\mathcal{X}^h$ and $\mathcal{Y}^h$ be discretization spaces for the PG method, such that the uniform discrete* $\inf - \sup$ *conditions (8) are fulfilled and assume that the bilinear operator $\mathcal{X} \times \mathcal{Y} \ni (u, w) \mapsto \langle A(\boldsymbol{y})u, w\rangle$ is continuous, uniformly with respect to $\boldsymbol{y} \in U$.*

Then the PG projections $G^h(\boldsymbol{y}) : \mathcal{X} \to \mathcal{X}^h$ are well-defined linear operators, whose norms are uniformly bounded with respect to the parameters $\boldsymbol{y}$ and h, i.e.,

$$\sup_{\boldsymbol{y}\in U} \sup_{h>0} \|u^h(\boldsymbol{y})\|_\mathcal{X} \leq \frac{1}{\mu}\|f\|_{\mathcal{Y}'}, \tag{10}$$

$$\sup_{\boldsymbol{y}\in U} \sup_{h>0} \|G^h(\boldsymbol{y})\|_{\mathcal{L}(\mathcal{X})} \leq \frac{C}{\mu} \tag{11}$$

The Galerkin projections are uniformly quasi-optimal: for every $\boldsymbol{y} \in U$ we have the a-priori error bound

$$\|u(\boldsymbol{y}) - u^h(\boldsymbol{y})\|_\mathcal{X} \leq \left(1 + \frac{C}{\mu}\right) \inf_{v^h\in\mathcal{X}^h} \|u(\boldsymbol{y}) - v^h\|_\mathcal{X} \,. \tag{12}$$

As is classical in the theory of polynomial approximation (see, e.g. [20, 48]), we use a holomorphic extension of the parametric operator family $A(\boldsymbol{y})$ to complex parameter sequences $\boldsymbol{z} \in \mathcal{O} \supset U$, where $\mathcal{O}$ is some suitable subset of the complex plane. Here, when dealing with extensions of operators and solutions to parameters taking values in the complex domain, we identify the function spaces $\mathcal{X}$ and $\mathcal{Y}$ with their complexifications $\mathcal{X} \otimes \{1, \boldsymbol{i}\}$ and $\mathcal{Y} \otimes \{1, \boldsymbol{i}\}$ for the sake of simplicity. We require the parametric operator $\mathcal{O} \ni \boldsymbol{z} \mapsto A(\boldsymbol{z})$ to be holormorphic with respect to any finite set of variables and to be boundedly invertible. Hereby, a Banach-space valued mapping $z \mapsto R(z) \in E$ of a single complex variable is said to be holomorphic (in some open domain $\mathcal{O}$) if

$$\lim_{h\to 0} \frac{R(z_0 + h) - R(z_0)}{h}$$

exists in E for any $z_0 \in \mathcal{O}$, with $h \to 0$ understood in $\mathbb{C}$. Note that our assumption on A requires it to be holomorphic with respect to any component z_j of $\boldsymbol{z}$ independently. Joint holomorphy with respect to an arbitrary, finite subset of variables $\boldsymbol{z}' = (z_j)_{j\in\Lambda}$ with $|\Lambda| < \infty$ of $\boldsymbol{z}$ then follows from Hartogs' theorem. In the sequel, we will often assume that the open set $\mathcal{O}$, on which $\boldsymbol{z} \mapsto A(\boldsymbol{z})$ is holomorphic, contains the product of Bernstein ellipses $\mathcal{E}_\rho = \bigotimes_{j\geq 1} \mathcal{E}_{\rho_j}$ with $\mathcal{E}_\sigma := \{(z + z^{-1})/2, z \in \mathbb{C} : |z| = \sigma\}$. In the case of complex-parametric operators, the bounded invertibility of $A(\boldsymbol{z})$ is equivalent to the *complex discrete* $\inf - \sup$ *conditions*: there exists a constant $\mu_\mathbb{C} > 0$ such that for any $h > 0$ and $\boldsymbol{z} \in \mathcal{O}$

$$\begin{cases} \inf_{0\neq v^h\in\mathcal{X}^h} \sup_{0\neq w^h\in\mathcal{Y}^h} & \mathrm{Re}\frac{\langle A(\boldsymbol{z})v^h, w^h\rangle}{\|v^h\|_\mathcal{X}\|w^h\|_\mathcal{Y}} \geq \mu_\mathbb{C} > 0, \\ \inf_{0\neq w^h\in\mathcal{Y}^h} \sup_{0\neq v^h\in\mathcal{X}^h} & \mathrm{Re}\frac{\langle A(\boldsymbol{z})v^h, w^h\rangle}{\|v^h\|_\mathcal{X}\|w^h\|_\mathcal{Y}} \geq \mu_\mathbb{C} > 0. \end{cases} \tag{13}$$

Approximation results on discretization spaces are usually combined with prior knowledge of the regularity of the data. For this, we assume that there exists a $0 < t \leq \bar{t}$ such that the parametric family $A(\boldsymbol{y}) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$

is regular in given smoothness scales $\{\mathcal{X}_t\}_{t\geq 0}$, resp. $\{\mathcal{Y}_t\}_{t\geq 0}$, satisfying:

$$\mathcal{X} = \mathcal{X}_0 \supset \mathcal{X}_1 \supset ... \supset \mathcal{X}_t \ , \quad \mathcal{Y} = \mathcal{Y}_0 \supset \mathcal{Y}_1 \supset ... \supset \mathcal{Y}_t \ . \tag{14}$$

Here, the smoothness index $t \geq 0$ denotes, for example, a differentiation order in a scale of Sobolev or Besov spaces. These spaces are defined by interpolation for non integer indices. We shall also require a corresponding scale on the dual side, with $\mathcal{X}'_t := (\mathcal{X}')_t$, and $\mathcal{Y}'_t := (\mathcal{Y}')_t$:

$$\mathcal{X}' = \mathcal{X}'_0 \supset \mathcal{X}'_1 \supset ... \supset \mathcal{X}'_t \ , \quad \mathcal{Y}' = \mathcal{Y}'_0 \supset \mathcal{Y}'_1 \supset ... \supset \mathcal{Y}'_t \ . \tag{15}$$

Note carefully that with this notation, $(\mathcal{X}_t)'$ **generally differs** from $\mathcal{X}'_t$. For example, in the case of the diffusion equation (3), one may choose $\mathcal{X} = H_0^1(D)$ and $\mathcal{X}_t = H_0^{1+t}(D)$. In this case, $\mathcal{X}' = H^{-1}(D) = (H_0^1(D))'$ and $(\mathcal{X}_t)' = H^{-1-t}(D) \neq H^{-1+t}(D) = (\mathcal{X}')_t =: \mathcal{X}'_t$.

A first statement of solution regularity in the scales (14), (15) takes the form of uniform bounded invertibility of the family of parametric operators $A(\boldsymbol{y})$:

$$A(\boldsymbol{y}) \in \mathcal{L}(\mathcal{X}_t, \mathcal{Y}'_t), \quad \text{for all } \boldsymbol{y} \in U, \quad \text{and } \sup_{\boldsymbol{y}\in U} \|A(\boldsymbol{y})^{-1}\|_{\mathcal{L}(\mathcal{Y}'_t,\mathcal{X}_t)} < \infty. \tag{16}$$

For the PG discretization, we assume at hand two one-parameter families $\{\mathcal{X}^h\}_{h>0}$ and $\{\mathcal{Y}^h\}_{h>0}$ of $\mathcal{X}$ and of $\mathcal{Y}$, respectively, with finite, equal dimension: $N_h = \dim(\mathcal{X}^h) = \dim(\mathcal{Y}^h) < \infty$. We assume furthermore that $\{\mathcal{X}^h\}_{h>0}$ and $\{\mathcal{Y}^h\}_{h>0}$ are dense in $\mathcal{X}$ and in $\mathcal{Y}$, respectively. Here the discretization parameter $h > 0$ usually stands for the meshwidth in finite element discretizations of fixed polynomial degree, on a quasiuniform triangulation of the physical bounded, polyhedral domain D. We assume that these spaces admit *approximation properties* in the smoothness scales (14), (15),

$$\begin{array}{ll} \inf_{w^h\in\mathcal{X}^h} \|w - w^h\|_{\mathcal{X}} \leq C_t h^t \|w\|_{\mathcal{X}_t}, & \text{for all } w \in \mathcal{X}_t, \\ \inf_{v^h\in\mathcal{Y}^h} \|v - v^h\|_{\mathcal{Y}} \leq C_{t'} h^{t'} \|v\|_{\mathcal{Y}_{t'}}, & \text{for all } v \in \mathcal{Y}_{t'}. \end{array} \tag{17}$$

Together with the bounded invertibility of the family of operators $A(\boldsymbol{y})$, it holds:

$$\|u(\boldsymbol{y}) - u^h(\boldsymbol{y})\|_{\mathcal{X}} \overset{\text{Eq. (12)}}{\leq} C \inf_{v^h\in\mathcal{X}^h} \|u(\boldsymbol{y}) - v^h\|_{\mathcal{X}} \overset{\text{Eq. (17)}}{\leq} c_t h^t \|u(\boldsymbol{y})\|_{\mathcal{X}_t} \overset{\text{Eq. (16)}}{\leq} C_t h^t \|f\|_{\mathcal{Y}'_t}. \tag{18}$$

Here, the constant C_t depends on a uniform bound on the inverse of the parametric operator in the appropriate smoothness space: $\sup_{\boldsymbol{y}\in U} \|A(\boldsymbol{y})^{-1}\|_{\mathcal{L}(\mathcal{Y}'_t,\mathcal{X}_t)}$, and on the smoothness parameter t, but not on the discretization parameter h.

Moreover, as we confine the exposition to functionals of solutions $F(\boldsymbol{y}) = \mathcal{G}(u(\boldsymbol{y}))$ for some $\mathcal{G}(\cdot) \in \mathcal{X}'$, we assume *adjoint regularity*, i.e., there exists $t' \geq 0$, such that $\mathcal{G} \in \mathcal{X}'_{t'}$, and such that the parametric adjoint solution $w_{\mathcal{G}}(\boldsymbol{y}) \in \mathcal{Y}'$ of the problem

$$A(\boldsymbol{y})^* w_{\mathcal{G}}(\boldsymbol{y}) = \mathcal{G} \tag{19}$$

satisfies $w_{\mathcal{G}}(\boldsymbol{y}) \in \mathcal{Y}'_{t'}$ uniformly with respect to $\boldsymbol{y}$:

$$\sup_{\boldsymbol{y}\in U} \|w_{\mathcal{G}}(\boldsymbol{y})\|_{\mathcal{Y}'} \leq C\|\mathcal{G}\|_{\mathcal{X}'_{t'}} \ . \tag{20}$$

Under the adjoint regularity (20), the uniform parametric discrete inf-sup condition (8) and the approximation property (17), an Aubin-Nitsche duality argument as, e.g., in [41], implies superconvergence: for any $\boldsymbol{y} \in U$, with F^h the functional applied to the parametric PG solution $u^h(\boldsymbol{y})$ defined in (9) on discretization spaces of parameter h,

$$|F(\boldsymbol{y}) - F^h(\boldsymbol{y})| \leq C_{t+t'} h^{t+t'} \|f\|_{\mathcal{Y}'_t} \|\mathcal{G}\|_{\mathcal{X}'_{t'}} \ . \tag{21}$$

3. Single-Level Compressed Sensing Petrov-Galerkin approximations

The multi-level compressed sensing PG (MLCSPG) discretization is a generalization of the single-level algorithms and results developed in [46]. Analogous to MLMC path simulations (see e.g. [33] and the references there) or MLMC Finite Element discretizations (see e.g. [4]) the MLCSPG method described here considers a sampling scheme from [46] with a number of sampling points depending on the discretization level.

Such compressed sensing reconstruction techniques have already shown promise in the context of numerical solutions of PDEs on high-dimensional parameter spaces: we refer, for example, to [55, 24, 45, 46, 9].

The key idea in these works is to decompose the solution of Eq. (1) via its (tensorized Chebyshev or Legendre) polynomial chaos expansion with respect to the parameter vector $\boldsymbol{y}$. A strongly measurable mapping $u : U \to \mathcal{X} : \boldsymbol{y} \mapsto u(\boldsymbol{y})$ which is square (Bochner-) integrable with respect to the Chebyshev measure $\mathrm{d}\eta$ over U can be represented as a gpc expansion, i.e.,

$$u(\boldsymbol{y}) = \sum_{\nu \in \mathcal{F}} u_\nu T_\nu(\boldsymbol{y}), \tag{22}$$

where in this case the coefficients in this expansion are functions $u_\nu \in \mathcal{X}$. Here $\mathcal{F} := \{\nu \in \mathbb{N}_0^{\mathbb{N}} : |\operatorname{supp}(\nu)| < \infty\}$ is the set of multi-indices with finite support. The tensorized Chebyshev polynomials are defined as

$$T_\nu(\boldsymbol{y}) = \prod_{j=1}^{\infty} T_{\nu_j}(y_j) = \prod_{j \in \operatorname{supp}(\nu)} T_{\nu_j}(y_j), \quad \boldsymbol{y} \in U, \quad \nu \in \mathcal{F}, \tag{23}$$

with the univariate Chebyshev polynomials defined by

$$T_j(t) = \sqrt{2}\cos\left(j \arccos(t)\right), \qquad \text{and } T_0(t) \equiv 1 \,. \tag{24}$$

Defining the probability measure σ on $[-1;1]$ as $\mathrm{d}\sigma(t) := \frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}$, the univariate Chebyshev polynomials T_j defined in (24) form an orthonormal system in $L^2([-1,1];\sigma)$ in the sense that

$$\int_{-1}^{1} T_k(t) T_l(t) \mathrm{d}\sigma(t) = \delta_{k,l}, \quad k, l \in \mathbb{N}_0 \,.$$

Similarly, with the product measure

$$\mathrm{d}\eta(\boldsymbol{y}) := \bigotimes_{j \geq 1} \mathrm{d}\sigma(y_j) = \bigotimes_{j \geq 1} \frac{\mathrm{d}y_j}{\pi\sqrt{1-y_j^2}},$$

the tensorized Chebyshev polynomials (23) are orthonormal with respect to η in the sense that

$$\int_{\boldsymbol{y} \in U} T_\mu(\boldsymbol{y}) T_\nu(\boldsymbol{y}) \mathrm{d}\eta(\boldsymbol{y}) = \delta_{\mu,\nu}, \quad \mu, \nu \in \mathcal{F} \,.$$

A result proven in [37] ensures the ℓ_p summability, for some $0 < p \leq 1$, of the polynomial chaos expansion (22) for the diffusion case, Eq. (3):

$$\left\| (\|u_\nu\|_{\mathcal{X}})_{\nu \in \mathcal{F}} \right\|_p^p = \sum_{\nu \in \mathcal{F}} \|u_\nu\|_{\mathcal{X}}^p < \infty$$

under the uniform ellipticity assumption (6) and the condition that the sequence of infinity norms of the ψ_j is itself ℓ_p summable:

$$\left\| (\|\psi_j\|_\infty)_{j \geq 1} \right\|_p^p = \sum_{j \geq 1} \|\psi_j\|_\infty^p < \infty.$$

Recent results by [3] show that these conditions can be improved by considering pointwise convergence of the series $\sum_{j \geq 1} |\psi_j|$ instead of infinity norms in the whole domain D. This takes advantage of the local structure of the basis elements ψ_j, e.g. when only few of them are overlapping, as is the case for wavelets. In particular, ℓ_p summability of Legendre coefficients can be obtained when $(\|\psi_j\|_\infty)_j \in \ell_q$ for $q := 2p/(2-p)$ provided that the interiors of the supports of the ψ_j do not overlap. The summability results from [37] concerning Chebyshev expansions were extended to weighted ℓ_p estimates for the general parametric operator problem (1) with affine dependence as in (5), in [46] under slightly stronger assumptions. This result is particularly important for us as it ensures the recovery of the coefficients u_ν (or any functional thereof) via CS methods.

The results on the approximation via an MLCSPG framework rely on the single-level results developed in [47, 46], where functions are approximated via a weighted-sparse expansion in an appropriate basis. We review here the main ideas. Given a (finite) orthonormal system $(\phi_\nu)_{\nu \in \Lambda}$, with $|\Lambda| = N < \infty$ for $L^2(U, \eta)$ where η is a probability measure, for any fixed function $f : U \to \mathbb{R}$, there exists a unique sequence of coefficients $\mathbf{f} = (f_\nu)_{\nu \in \Lambda}$ such that

$$f(\boldsymbol{y}) = \sum_{\nu \in \Lambda} f_\nu \phi_\nu(\boldsymbol{y}), \ \forall y \in U \,. \tag{25}$$

We define an ℓ_p norm associated with this expansion as $|||f|||_p := \|\mathbf{f}\|_p$.

In particular, a function f is said to be sparse (or compressible) if its sequence of coefficients in expansion (25) is sparse (or compressible) itself. Our goal is to recover this said sequence from seemingly few evaluations of the function f at certain (here random) sample points $\boldsymbol{y}^{(i)}$, for $1 \leq i \leq m$. This can be done by CS methods: after introducing the sensing matrix Φ as $\Phi_{i,j} := \phi_j(\boldsymbol{y}^{(i)})$ and letting $g_i := f(\boldsymbol{y}^{(i)})$, it holds

$$\boldsymbol{g} = \Phi \mathbf{f} \ .$$

Hence, assuming that the expansion is sparse, and the number of samples rather small, we are dealing with the by-now classical problem of recovering a sparse vector from few linear measurements, by solving, for instance, the convex program

$$\min_{\boldsymbol{z} \in \mathbb{R}^N} \|\boldsymbol{z}\|_1, \quad \text{subject to } \Phi \boldsymbol{z} = \boldsymbol{g}. \tag{26}$$

In our context, it is beneficial to use a *weighted framework* which has been developed recently in [47]. Introducing a sequence of positive weights $(\omega_\nu)_{\nu\in\Lambda}$ with $|\omega_\nu| \geq 1$ for all ν, a weighted ℓ_p (quasi-)norm (henceforth indexed $\ell_{\omega,p}$ when appropriate) can be defined as

$$\|\mathbf{f}\|_{\omega,p}^p := \sum_{\nu\in\Lambda} \omega_\nu^{2-p} |f_\nu|^p, \quad 0 < p \leq 2.$$

In particular, it holds $\|\mathbf{f}\|_{\omega,2} = \|\mathbf{f}\|_2$ and $\|\mathbf{f}\|_{\omega,1} = \|\mathbf{f} \odot \omega\|_1$, where $\odot$ defines the pointwise multiplication. Moreover, choosing the constant weight $\omega_\nu = 1$ yields the original definitions of ℓ_p norms. Formally letting $p \downarrow 0$ motivates the introduction of the *weighted sparsity* measure

$$\|\mathbf{f}\|_{\omega,0} := \sum_{\nu\in\Lambda, f_\nu \neq 0} \omega_\nu^2 \ .$$

A vector $\mathbf{x}$ is therefore called *weighted s-sparse* (with respect to a weight sequence ω) if $\|\mathbf{x}\|_{\omega,0} \leq s$. We may therefore define the error of best weighted s-term approximation as

$$\sigma_{\omega,s}(f) = \sigma_{\omega,s}(\mathbf{f}) := \inf_{\boldsymbol{z}:\|\boldsymbol{z}\|_{\omega,0}\leq s} \|\mathbf{f} - \boldsymbol{z}\|_{\omega,p}.$$

With these weighted error measures at hand, the Basis Pursuit problem (26) can be generalized to include a-priori information encoded in the sequences ω of weights, as

$$\min_{\boldsymbol{z} \in \mathbb{R}^N} \|\boldsymbol{z}\|_{\omega,1}, \quad \text{subject to } \Phi \boldsymbol{z} = \boldsymbol{g}. \tag{27}$$

More details on such weighted spaces and weighted sparse approximations can be found in [47] where the following fundamental result is also proved.

Theorem 2. *Suppose $(\phi_\nu)_{\nu\in\Lambda}$ is a finite orthonormal system with $|\Lambda| = N < \infty$ and that weights $\omega_\nu \geq \|\phi_\nu\|_\infty$ are given. For a (weighted) sparsity $s \geq 2\|\omega\|_\infty^2$, draw*

$$m \geq C s \log^3(s) \log(N) \tag{28}$$

sample points $\boldsymbol{y}^{(i)}$ at random, $1 \leq i \leq m$, according to the orthonomalization measure η. The constant $C > 0$ in (28) is universal, i.e., independent of all other quantities including s, m and N.

Then, with probability at least $1 - N^{-\log(s)^3}$, any function $f = \sum f_\nu \phi_\nu$ can be approximated by the function $\widehat{f} := \sum \widehat{f}_\nu \phi_\nu$, where $\widehat{\mathbf{f}}$ is the solution to the weighted basis pursuit problem (27). The approximation holds in the following sense:

$$\|f - \widehat{f}\|_\infty \leq \left|\left|\left| f - \widehat{f} \right|\right|\right|_{\omega,1} \leq c_1 \sigma_s(f)_{\omega,1}, \quad \text{and} \quad \|f - \widehat{f}\|_2 \leq d_1 \sigma_s(f)_{\omega,1}/\sqrt{s} \ .$$

In particular, using the weighted Stechkin inequality from [47]

$$\sigma_s(\mathbf{f})_{\omega,q} \leq \left(s - \|\omega\|_\infty^2\right)^{1/q-1/p} \|\mathbf{f}\|_{\omega,p}, \quad p < q \leq 2, \quad \|\omega\|_\infty^2 < s \ , \tag{29}$$

we obtain the estimates, for $p < 1$, in terms of the sparsity

$$\|f - \widehat{f}\|_\infty \leq c s^{1-1/p} \|\mathbf{f}\|_{\omega,p}, \quad \|f - \widehat{f}\|_2 \leq d s^{1/2-1/p} \|\mathbf{f}\|_{\omega,p}. \tag{30}$$

Choosing $m \asymp s\ln^3(s)\log(N)$ relates the reconstruction error and the number m of samples as

$$\|f-\widehat{f}\|_\infty \le c\left(\frac{\log^3(m)\log(N)}{m}\right)^{1/p-1}\|\mathbf{f}\|_{\omega,p}, \quad \|f-\widehat{f}\|_2 \le d\left(\frac{\log^3(m)\log(N)}{m}\right)^{1/p-1/2}\|\mathbf{f}\|_{\omega,p}.$$

Remark 1. *Recent works [11, 38] on restricted isometry constants for subsampled Fourier matrices suggest that the factor* $\log^3(s)$ *in (28) can be reduced to* $\log^2(s)$*, but details remain to be worked out.*

4. Multi-level Compressed Sensing Petrov-Galerkin approximations

4.1. **A multi-level framework.** We extend the foregoing CS methods to sweeping the parameter domain to multi-level ("ML" for short) discretizations of the parametric problems, in the spirit of the ML MC methods for numerical treatment of operator equations with random inputs as developed in [39, 33, 4]. There, the solution of the parametric operator equation (1) is approximated on a sequence of partitions of the physical domain D of widths $\{h_l\}_{l=1}^L$ for a prescribed, maximal refinement level $L \in \mathbb{N}$. To simplify the exposition, we assume dyadic refinement, i.e. $h_{l+1} = h_l/2 = 2^{-l}h_0$ for a given, small enough, initial resolution $h_0 > 0$.

For a given parameter sequence $\boldsymbol{y}$, we may write the Galerkin projection $u^L(\boldsymbol{y}) \in \mathcal{X}^{h_L}$ of $u(\boldsymbol{y})$ as

$$u^L(\boldsymbol{y}) = \sum_{l=1}^{L} u^l(\boldsymbol{y}) - u^{l-1}(\boldsymbol{y}), \tag{31}$$

where we define $u^0(\boldsymbol{y}) \equiv 0$ (note that we will equivalently parametrize the approximations and spaces by l or h_l). The idea behind our MLCSPG approach is to estimate every difference between consecutive levels of approximation (the *details*) via a single level CSPG as presented above. For the remaining, we let

$$\mathfrak{d}u^l(\boldsymbol{y}) := u^l(\boldsymbol{y}) - u^{l-1}(\boldsymbol{y}), \quad 1 \le l \le L, \tag{32}$$

denote the difference between two scales of approximation.

As already outlined in the introduction, our method produces pointwise numerical approximations $\widehat{\mathfrak{d}u^l}(\boldsymbol{y})$ of $\mathfrak{d}u^l(\boldsymbol{y})$ via a (single level) CSPG method. For each level l, we choose a number m_l of parameter vectors $\boldsymbol{y}_l^{(1)}, \dots, \boldsymbol{y}_l^{(m_l)}$, compute the PG approximations $u^l(\boldsymbol{y}_l^{(i)})$ and $u^{l-1}(\boldsymbol{y}_l^{(i)})$ by solving the corresponding finite dimensional linear systems, and form the samples $\mathfrak{d}u^l(\boldsymbol{y}_l^{(i)}) = u^l(\boldsymbol{y}_l^{(i)}) - u^{l-1}(\boldsymbol{y}_l^{(i)})$, $i = 1, \dots, m_l$. From these samples, one approximates the coefficients in the tensorized Chebyshev expansion of $\mathfrak{d}u^l$ via weighted ℓ_1-minimization (or any sparse recovery method). This yields approximations $\widehat{\mathfrak{d}u^l}(\boldsymbol{y})$, $\ell = 1, \dots, L$, and

$$u^L_{\mathrm{MLCS}}(\boldsymbol{y}) := \sum_{l=1}^{L} \widehat{\mathfrak{d}u^l}(\boldsymbol{y})$$

then provides an approximation of the targeted parametric solution $u = u(\boldsymbol{y})$. The convergence of our MLCSPG framework can be estimated via the triangle inequality,

$$\|u(\boldsymbol{y}) - {u^L}_{\mathrm{MLCS}}(\boldsymbol{y})\|_{\mathcal{X}} \le \|u(\boldsymbol{y}) - u^L(\boldsymbol{y})\|_{\mathcal{X}} + \sum_{l=1}^{L} \left\|\mathfrak{d}u^l(\boldsymbol{y}) - \widehat{\mathfrak{d}u^l}(\boldsymbol{y})\right\|_{\mathcal{X}}. \tag{33}$$

For simplicity, we constrain our considerations to a functional $\mathcal{G} \in \mathcal{X}'$ applied to the solution, leading to the real-valued QoI $F(\boldsymbol{y}) = \mathcal{G}(u(\boldsymbol{y}))$ to be approximated. The above considerations apply verbatim when replacing $u(\boldsymbol{y})$ by $F(\boldsymbol{y})$, and $\mathfrak{d}u^l$ by $\mathfrak{d}F^l$, the levelwise PG approximation, and $\widehat{\mathfrak{d}u^l}(\boldsymbol{y})$ by $\widehat{\mathfrak{d}F^l}(\boldsymbol{y})$. The triangle inequality leads to the error estimate

$$\left|F(\boldsymbol{y}) - {F^L}_{\mathrm{MLCS}}(\boldsymbol{y})\right| \le |F(\boldsymbol{y}) - F^L(\boldsymbol{y})| + \sum_{l=1}^{L} \left|\mathfrak{d}F^l(\boldsymbol{y}) - \widehat{\mathfrak{d}F^l}(\boldsymbol{y})\right|. \tag{34}$$

The first term on the right hand side of Eq. (33) can be estimated using the uniform parametric regularity (16), the uniform parametric inf-sup condition (8) and the approximation property (17): for a regularity parameter $0 < t \le \bar{t}$ of the data f,

$$\|u(\boldsymbol{y}) - u^L(\boldsymbol{y})\|_{\mathcal{X}} \le C_t h_L^t \|f\|_{\mathcal{Y}'_t}. \tag{35}$$

Passing to the functional $\mathcal{G} \in \mathcal{X}'_{t'}$, we obtain a superconvergence error bound for the Petrov-Galerkin-Finite Element Method (PG-FEM) via a classical Aubin-Nitsche duality argument [41]:

$$|F(\boldsymbol{y}) - F^L(\boldsymbol{y})| \leq C_{t+t'} h_L^{t+t'} \|f\|_{\mathcal{Y}'_t} \|\mathcal{G}\|_{\mathcal{X}'_{t'}}.$$

Our goal is to verify that the single-level result, Theorem 2, applies to all levels $l = 1, \dots L$, and to obtain error bounds similar to the one in Eq. (21). We consider the Chebyshev expansions of the differences,

$$\mathfrak{d}u^l(\boldsymbol{y}) = \sum_{\nu \in \mathcal{F}} \mathfrak{d}u^l_\nu T_\nu(\boldsymbol{y}), \tag{36}$$

$$\mathfrak{d}F^l(\boldsymbol{y}) = \sum_{\nu \in \mathcal{F}} \mathfrak{d}F^l_\nu T_\nu(\boldsymbol{y}). \tag{37}$$

Assuming summability of the expansion in $\ell_{\omega,p}(\mathcal{F})$, we can apply Theorem 2 with a number of samples $m_l \gtrsim s_l \log(s_l)^3 \log(N_l)$, for suitable choices of $s_1, \dots, s_L$, and in particular we can use the error estimate (30) in terms of the (weighted) sparsity s_l for each level of approximation. This results in the bound

$$|\mathfrak{d}F^l(\boldsymbol{y}) - \widehat{\mathfrak{d}F^l}(\boldsymbol{y})| \leq C \left\| \left(\mathfrak{d}F^l_\nu\right)_\nu \right\|_{\omega,p} s_l^{1-1/p}, \quad \text{for all } 1 \leq l \leq L, \tag{38}$$

where $C > 0$ is a universal constant (independent of s_l, $\boldsymbol{y}$, l). Theorem 2 applies only to finite orthonormal systems. Thus, for each $l = 1, 2, ..., L$, the countably infinite index set $\mathcal{F}$ has to be truncated to a finite, but possibly large, subset Γ_l of $N_l := |\Gamma_l| < \infty$ many indices of the relevant (few) essential Chebyshev coefficients in the parametric solution's gpc expansion. We describe a strategy for selecting the index sets Γ_l depending on s_l in Section 5.1. A good choice for the s_l turns out to be $s_l \asymp 2^{(L-l)(t+t')p/(1-p)}$, as will be derived ahead.

Finally, summing up the contributions from all discretization levels and drawing

$$m_l \gtrsim 2^{(L-l)(t+t')p/(1-p)} (L-l)^3 \log(N_l)$$

sample points per level will imply the error bounds in Theorem 9. The choice of this number of sampling points is justified in Section 4.3 and by the following result, whose proof is the purpose of the next section.

Theorem 3. *Let $\{A(\boldsymbol{y}) : \boldsymbol{y} \in U\}$ be a parametric family of operators as defined in (5). Assume that the operator A_0 is* $\inf - \sup$ *stable. For $B_j := A_0^{-1} A_j$ and for $0 \leq t \leq \bar{t}$, introduce the sequence*

$$\boldsymbol{b}_t := (b_{t,j})_{j \geq 1} \quad \textit{with} \quad b_{t,j} := \|B_j\|_{\mathcal{L}(\mathcal{X}_t)} = \|B_j^*\|_{\mathcal{L}(\mathcal{Y}_t)}. \tag{39}$$

Let $\mathbf{v} := (v_j)_{j \geq 1}$ be a sequence of weights with $v_j \geq 1$ such that, for some $p < 1$,

$$\sum_{j \geq 1} b_{t,j} v_j^{(2-p)/p} \leq \kappa_{\mathbf{v},p} < 1, \quad \textit{and} \tag{40}$$

$$\sum_{j \geq 1} b_{t,j}^p v_j^{2-p} < \infty. \tag{41}$$

Let ρ be a δ-admissible sequence of polyradii, with $\delta = (1 - \kappa_{\mathbf{v},p})/2$, i.e., such that

$$\sum_{j \geq 1} (\rho_j - 1) b_{t,j} \leq \delta. \tag{42}$$

Then the family of operators $A(\boldsymbol{y})$ is uniformly $\inf - \sup$ *stable. Assume in addition that $A_j \in \mathcal{L}(\mathcal{X}_t, \mathcal{Y}'_t)$, $j \geq 0$, are defined on the scale of smoothness spaces $\mathcal{X}_t$ and that the approximation property (17) holds. Assume moreover that $A_0 : \mathcal{X}_t \to \mathcal{Y}'_t$ is boundedly invertible and that the sequence $\boldsymbol{b}_t$ is small, and that the polyradius ρ is $\boldsymbol{b}_t$ admissible, i.e.*

$$\sum_{j \geq 1} b_{t,j} \leq \kappa_t < 1, \tag{43}$$

$$\sum_{j \geq 1} (\rho_j - 1) b_{t,j} \leq \delta_t, \tag{44}$$

for $\delta_t < 1 - \kappa_t$.

Then the affine-parametric family of operators $\{A(\boldsymbol{y}) : \boldsymbol{y} \in U\}$ is uniformly boundedly invertible in $\mathcal{L}(\mathcal{X}_t, \mathcal{Y}'_t)$, and there hold bounds on the Chebyshev gpc coefficients

$$\|\mathfrak{d}u^l_\nu\|_{\mathcal{X}} \le C h_l^t \|f\|_{\mathcal{Y}'_t} \rho^{-\nu} , \quad \text{and} \quad |\mathfrak{d}F^l_\nu| \le C h_l^{t+t'} \|\mathcal{G}\|_{\mathcal{X}'_{t'}} \|f\|_{\mathcal{Y}'_t} \rho^{-\nu} \quad \text{for all} \quad \nu \in \mathcal{F}. \tag{45}$$

Moreover, for each $\nu \in \mathcal{F}$, there exists a δ-admissible sequence $\rho = \rho(\nu)$ satisfying (44) such that the sequence with components $\rho(\nu)^{-\nu} = \prod_{j\ge1} \rho(\nu)_j^{-\nu_j}$, $\nu \in \mathcal{F}$, satisfies $(\rho(\nu)^{-\nu})_{\nu\in\mathcal{F}} \in \ell_{\omega,p}$, where

$$\omega_\nu := \theta^{\|\nu\|_0} \mathbf{v}^\nu = \theta^{\|\nu\|_0} \prod_{j\ge1} v_j^{\nu_j}, \quad \nu \in \mathcal{F} . \tag{46}$$

We want to stress once again that the result presented above is written without the explicit dependence of the weight sequence $\mathbf{v}$ on the regularity parameter t. Moreover, we note that the conditions (39) - (44) are, for $t > 0$, strictly stronger than the summability conditions which were required in the single-level PG analysis in [46].

4.2. **Summability of the Chebyshev expansions.** This section provides the proof of the core result of the present paper, Theorem 3. We show that under general assumptions, the parametric solution's sequence of Chebyshev coefficients $\left(\mathfrak{d}F^l_\nu\right)_{\nu\in\mathcal{F}} \in \ell_{p,\omega}$, and in particular that the following a priori estimate holds:

$$\left\| \left(\mathfrak{d}F^l_\nu\right)_{\nu\in\mathcal{F}} \right\|_{\omega,p} \le C h_l^{t+t'} \|f\|_{\mathcal{Y}'_t} \|\mathcal{G}\|_{\mathcal{X}'_{t'}} \|(\rho(\nu)^{-\nu})_{\nu\in\mathcal{F}}\|_{\omega,p}. \tag{47}$$

The novel point of this estimate is the scaling of the right hand side with $h_l^{t+t'}$. The proof of this assertion is structured in three main steps, analogously to [19, 46]. First we show that the difference between levels is holomorphic in polydiscs. Then, this holomorphy is used to bound the norm of any Chebyshev coefficient. This norm depends on a sequence of radii of holomorphy ρ. Finally, we construct a sequence of radii and weights such that the sequence of coefficients is $\ell_{\omega,p}$ summable.

4.2.1. *Holomorphy.* This first part shows that, under some uniform invertibility assumption of the family of (complexified) operators $A(\boldsymbol{z})$ (which are satisfied in particular for the affine-parametric family considered here), the solutions are holomorphic with respect to any finite set of variables. This then allows to use Cauchy's integral formula to estimate the norm of the Chebyshev coefficients.

Theorem 4. *For some $\mathcal{O} \subset \mathbb{C}^{\mathbb{N}}$ with $\mathcal{O} \supset U$, assume that the complex* $\inf - \sup$ *conditions (13) hold with constant $\mu_{\mathbb{C}}$ uniformly for $\boldsymbol{z} \in \mathcal{O}$. If the solution map $\mathcal{O} \ni \boldsymbol{z} \to u(\boldsymbol{z}) \in \mathcal{X}$ is holomorphic with respect to any finite set of parameters, then*

(1) for any level l of PG discretization (corresponding to the discretization parameter $h_l = 2^{-l}h_0$ for a given $h_0 > 0$ sufficiently small), the parametric Galerkin projections $\mathcal{O} \ni \boldsymbol{z} \to u^l(\boldsymbol{z}) \in \mathcal{X}^l$ are holomorphic with respect to any finite subset of the sequence $\boldsymbol{z} \in \mathcal{O}$, with domains of holomorphy whose size is independent of l, i.e. of the discretization parameter h_l,

(2) the Petrov-Galerkin projections are quasi-optimal, uniformly with respect to the level of approximation l and the vector of (complex) parameters $\boldsymbol{z} \in \mathcal{O}$:

$$\|u(\boldsymbol{z}) - u^l(\boldsymbol{z})\|_{\mathcal{X}} \le \left(1 + \frac{C}{\mu_{\mathbb{C}}}\right) \inf_{v^l\in\mathcal{X}^l} \|u(\boldsymbol{z}) - v^l\|_{\mathcal{X}} .$$

Proof. The holomorphy follows from the linearity of the PG approximation as stated in Proposition 1. The quasi optimality is obtained in the same way as in the real case. □

The next corollary which uses the notation (32) follows directly.

Corollary 1. *Under the conditions above, if in addition the approximation property of the discretization spaces holds for complex parameters $\boldsymbol{z} \in \mathcal{O}$, then for any two consecutive discretization levels l and $l+1$, $l \ge 0$, the mappings $\mathcal{O} \ni \boldsymbol{z} \mapsto \mathfrak{d}u^l(\boldsymbol{z}) \in \mathcal{X}$ are holomorphic with respect to any finite set of variables and satisfy the uniform bound*

$$\sup_{\boldsymbol{z}\in\mathcal{O}} \|\mathfrak{d}u^l(\boldsymbol{z})\|_{\mathcal{X}} \le C'_{t,\mu_{\mathbb{C}}} h_l^t \sup_{\boldsymbol{z}\in\mathcal{O}} \|u(\boldsymbol{z})\|_{\mathcal{X}_t}.$$

Proof. The statement is a consequence of the previous results and the triangle inequality:

$$\begin{aligned}\sup_{\boldsymbol{z}\in\mathcal{O}}\|\mathfrak{d}u^l(\boldsymbol{z})\|_{\mathcal{X}} &\le \left(1+\frac{C}{\mu_{\mathbb{C}}}\right)\sup_{\boldsymbol{z}\in\mathcal{O}}\left(\inf_{v^l\in\mathcal{X}^l}\|u(\boldsymbol{z})-v^l\|_{\mathcal{X}}+\inf_{v^{l-1}\in\mathcal{X}^{l-1}}\|u(\boldsymbol{z})-v^{l-1}\|_{\mathcal{X}}\right)\\ &\le \left(1+\frac{C}{\mu_{\mathbb{C}}}\right)\sup_{\boldsymbol{z}\in\mathcal{O}}C_t\left(h_l^t\|u(\boldsymbol{z})\|_{\mathcal{X}_t}+h_{l-1}^t\|u(\boldsymbol{z})\|_{\mathcal{X}_t}\right)=C'_{t,\mu_{\mathbb{C}}}h_l^t\sup_{\boldsymbol{z}\in\mathcal{O}}\|u(\boldsymbol{z})\|_{\mathcal{X}_t}.\end{aligned}$$

□

4.2.2. *Nominal inf-sup conditions imply uniform inf-sup conditions.* The preceding result, Theorem 4, requires the validity of a *uniform discrete inf-sup condition* for the PG discretization; here, uniformity is understood with respect to the discretization parameter $h > 0$ and with respect to the parameter sequence $\boldsymbol{z} \in \mathcal{O}$ in Theorem 4 or with respect to $\boldsymbol{y} \in U$ in (8), respectively. In what follows, we assume that the two one-parameter families of dense subspaces $\{\mathcal{X}^h\}_{h>0} \subset \mathcal{X}$ and $\{\mathcal{Y}^h\}_{h>0} \subset \mathcal{Y}$ are of equal, finite dimension $N^h = \dim(\mathcal{X}^h) = \dim(\mathcal{Y}^h)$ and are stable for the nominal operator $A_0 \in \mathcal{L}(\mathcal{X},\mathcal{Y}')$ in (5), i.e., the discrete inf-sup conditions hold

$$\inf_{0\neq w^h\in\mathcal{X}^h}\sup_{0\neq v^h\in\mathcal{Y}^h}\frac{\langle A_0w^h,v^h\rangle}{\|w^h\|_{\mathcal{X}}\|v^h\|_{\mathcal{Y}}}\ge\mu_0>0\,,\quad \inf_{0\neq v^h\in\mathcal{Y}^h}\sup_{0\neq w^h\in\mathcal{X}^h}\frac{\langle A_0w^h,v^h\rangle}{\|w^h\|_{\mathcal{X}}\|v^h\|_{\mathcal{Y}}}\ge\mu_0>0\,. \tag{48}$$

Theorem 5. *Suppose that the parametric operators $A(\boldsymbol{y})$, $A(\boldsymbol{z})$ are affine-parametric, as in (5). Assume further that for $t \ge 0$ the sequences $\boldsymbol{b}_t = (b_{t,j})_{j\ge1}$ in (39) are small, in the sense that (43) holds. Then, (43) with $t = 0$ implies that the discrete inf-sup conditions (8) hold uniformly with respect to $\boldsymbol{y} \in U$.*

Moreover, if the sequence of polyradii $\rho = (\rho_j)_{j\ge1}$ is admissible, in the sense that (42) holds for $t = 0$ and for some $\delta < 1-\kappa_0$, then the complex-parametric $A(\boldsymbol{z})$ satisfies the uniform inf-sup conditions (13) for $\boldsymbol{z} \in \mathcal{D}_\rho = \bigotimes_{j\ge1}\mathcal{D}_{\rho_j}$, where $\mathcal{D}_{\rho_j} := \{z \in \mathbb{C} : |z| \le \rho_j\}$.

Similarly, $A(\boldsymbol{z})$ is invertible in $\mathcal{L}(\mathcal{X}_t,\mathcal{Y}'_t)$ uniformly for $\boldsymbol{z} \in \mathcal{D}_\rho$ if ρ is δ-admissible w.r.t. the sequence $\boldsymbol{b}_t = (b_{t,j})_{j\ge1}$, where $b_{t,j} := \|A_0^{-1}A_j\|_{\mathcal{L}(\mathcal{X}_t)}$.

Proof. Let $\boldsymbol{b}_0$ be such that condition (43) holds with $t = 0$. Since A_0 is assumed to be boundedly invertible, we can write $A(\boldsymbol{y}) = A_0\left(I + \sum_{j\ge1} y_jA_0^{-1}A_j\right)$ and estimate

$$\left\|\sum_{j\ge1}y_jA_0^{-1}A_j\right\|_{\mathcal{L}(\mathcal{X})}\le\sum_{j\ge1}|y_j|b_{0,j}\le\sum_{j\ge1}b_{0,j}:=\kappa_0<1.$$

It follows from a perturbation (Neumann series) argument that the operator $A(\boldsymbol{y})$ is uniformly boundedly invertible. The discrete inf $-$ sup conditions hold with $\mu \le \mu_0(1-\kappa_0)$.

One may extend this argument to the complexified operator $A(\boldsymbol{z})$ defined for $\boldsymbol{z} \in \mathcal{D}_\rho$. This yields

$$\left\|\sum_{j\ge1}z_jA_0^{-1}A_j\right\|_{\mathcal{L}(\mathcal{X})}\le\sum_{j\ge1}|z_j|b_j\le\sum_{j\ge1}\rho_jb_j:=\delta+\kappa_0<1\,.$$

Therefore, the complex inf $-$ sup conditions (13) hold with constant $\mu_{\mathbb{C}} \le \mu_0(\delta+\kappa)$.

The proof of the uniform invertibility in $\mathcal{L}(\mathcal{X}_t,\mathcal{Y}'_t)$ follows by the same Neumann series argument. □

4.2.3. *Norm bounds on the Chebyshev gpc coefficients.* We now estimate the magnitudes of the Chebyshev coefficients. These estimates are used in the next section to show the $\ell_{\omega,p}$ summability of the sequence of Chebyshev coefficients. We recall that $\mathcal{E}_\rho = \bigotimes_{j\ge1}\mathcal{E}_{\rho_j}$ is a product of Bernstein ellipses $\mathcal{E}_{\rho_j} := \{(z + z^{-1})/2, z \in \mathbb{C} : 1 \le |z| \le \rho_j\}$.

Theorem 6. *Let $\nu \in \mathcal{F}$. Assume that the discretization spaces have the approximation property (17). Additionaly, assume that there exists a sequence $\rho = (\rho_j)_{j\ge1}$, with $\rho_j > 1$ such that the complex extension $\boldsymbol{z} \mapsto \mathfrak{d}u^l(\boldsymbol{z})$ is holomorphic with respect to any finite set of variables on $\mathcal{E}_\rho$ and with $A(\boldsymbol{z}) \in \mathcal{L}(\mathcal{X}_t,\mathcal{Y}'_t)$*

being uniformly boundedly invertible for every $\boldsymbol{z} \in \mathcal{E}_\rho$. *Then the Chebyshev coefficients of the difference* $\mathfrak{d}u^l = u^l - u^{l-1}$ *can be estimated as*

$$\|\mathfrak{d}u^l_\nu\|_{\mathcal{X}} \le Ch_l^{t'}\|f\|_{\mathcal{Y}'_t}\rho^{-\nu} .$$

If in addition we assume smoothness for the functional, i.e. $\mathcal{G} \in \mathcal{X}'_{t'}$ *for some* $0 < t' \le \bar{t}$, *then it holds*

$$|\mathfrak{d}F^l_\nu| \le Ch_l^{t+t'}\|f\|_{\mathcal{Y}'_t}\|\mathcal{G}\|_{\mathcal{X}'_{t'}}\rho^{-\nu}, \tag{49}$$

where the constants depend on the smoothness parameters t *and* t' *but not on* h_l.

Proof. The proof is similar to the one in [46] with appropriate modifications due to the introduction of the levels. The tensorized Chebyshev polynomials being orthogonal, it holds

$$\mathfrak{d}u^l_\nu = \int_U \mathfrak{d}u^l(\boldsymbol{y})T_\nu(\boldsymbol{y})\mathrm{d}\eta(\boldsymbol{y}).$$

Consider the multi-index $\nu = n\mathbf{e}_1 = (n, 0, 0\cdots) \in \mathcal{F}$ and split the parameter space as $U = [-1, 1] \times U'$, then any parameter sequence $\boldsymbol{y}$ can be written as $\boldsymbol{y} = (y_1, \boldsymbol{y}')$ with $y_1 \in [-1, 1]$. Thus

$$\mathfrak{d}u^l_{n\mathbf{e}_1} = \int_{U'}\int_{-1}^{+1} T_n(t)\mathfrak{d}u^l(t, \boldsymbol{y}')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}\mathrm{d}\eta(\boldsymbol{y}'). \tag{50}$$

With the change of variables $t = \cos(\phi)$ we obtain

$$\int_{-1}^{+1} T_n(t)\mathfrak{d}u^l(t, \boldsymbol{y}')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}} = \frac{\sqrt{2}}{\pi}\int_0^{\pi} \cos(n\phi)\mathfrak{d}u^l(\cos(\phi), \boldsymbol{y}')\mathrm{d}\phi = \frac{1}{\sqrt{2}\pi}\int_{-\pi}^{+\pi} \cos(n\phi)\mathfrak{d}u^l(\cos(\phi), \boldsymbol{y}')\mathrm{d}\phi.$$

This gives

$$\begin{aligned}\int_{-1}^{+1} T_n(t)\mathfrak{d}u^l(t, \boldsymbol{y}')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}} &= \frac{1}{\sqrt{2}\pi i}\int_{|z|=1} \frac{z^n + z^{-n}}{2}\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2}, \boldsymbol{y}'\right)\frac{\mathrm{d}z}{z} \\ &= \frac{1}{2\sqrt{2}i\pi}\int_{|z|=1} z^{n-1}\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2}, \boldsymbol{y}'\right)\mathrm{d}z + \frac{1}{2\sqrt{2}i\pi}\int_{|z|=1} z^{-n-1}\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2}, \boldsymbol{y}'\right)\mathrm{d}z.\end{aligned}$$

Due to the assumption that the extension $\boldsymbol{z} \to \mathfrak{d}u^l(\boldsymbol{z})$ to $\mathcal{E}_\rho$ is holomorphic, the mappings

$$z \mapsto z^{n-1}\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2}, \boldsymbol{y}'\right), \quad \text{and} \quad z \mapsto z^{-n-1}\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2}, \boldsymbol{y}'\right)$$

are analytic on $\mathcal{E}_{\rho_1}$. By Cauchy's theorem it follows, for $1 < \sigma < \rho_1$, that

$$\begin{aligned}\int_{-1}^{+1} T_n(t)\mathfrak{d}u^l(t, \boldsymbol{y}')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}} &= \frac{1}{2\sqrt{2}i\pi}\int_{|z|=\sigma^{-1}} z^{n-1}\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2}, \boldsymbol{y}'\right)\mathrm{d}z \\ &\quad + \frac{1}{2\sqrt{2}i\pi}\int_{|z|=\sigma} z^{-n-1}\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2}, \boldsymbol{y}'\right)\mathrm{d}z.\end{aligned}$$

Now notice that $z \mapsto \mathfrak{d}u^l(z, \boldsymbol{y}')$ is bounded by $C'h_l^t\|f\|_{\mathcal{Y}'_t}$ (in $\mathcal{X}$) in a polydisc contained in $\mathcal{E}_\rho$. Indeed, the approximation property of the discretization spaces, see Corollary 1, together with the bounded invertibility in the smoothness spaces, ensures

$$\sup_{\boldsymbol{z}\in\mathcal{E}_\rho}\|\mathfrak{d}u^l(\boldsymbol{z})\|_{\mathcal{X}} = \sup_{\boldsymbol{z}\in\mathcal{E}_\rho}\|u^l(\boldsymbol{z}) - u^{l-1}(\boldsymbol{z})\|_{\mathcal{X}} \le Ch_l^t\sup_{\boldsymbol{z}\in\mathcal{E}_\rho}\|u(\boldsymbol{z})\|_{\mathcal{X}_t} \le C'h_l^t\|f\|_{\mathcal{Y}'_t}. \tag{51}$$

It follows that

$$\begin{aligned}\left\|\int_{-1}^{+1} T_n(t)\mathfrak{d}u^l(t,\boldsymbol{y}')\frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}\right\|_{\mathcal{X}} &\leq \frac{1}{2\sqrt{2}\pi}\int_{|z|=\sigma^{-1}} |z^{n-1}| \left\|\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2},\boldsymbol{y}'\right)\right\|_{\mathcal{X}} \mathrm{d}z \\ &\quad + \frac{1}{2\sqrt{2}\pi}\int_{|z|=\sigma} |z^{-n-1}| \left\|\mathfrak{d}u^l\left(\frac{z+z^{-1}}{2},\boldsymbol{y}'\right)\right\|_{\mathcal{X}} \mathrm{d}z \\ &\leq \frac{1}{2\sqrt{2}\pi\sigma^{n-1}}2\pi C' h_l^t \|f\|_{\mathcal{Y}_t'}\sigma^{-1} + \frac{1}{2\sqrt{2}\pi\sigma^{n+1}}2\pi\sigma C' h_l^t\|f\|_{\mathcal{Y}_t'} \\ &= \sqrt{2}C'h_l^t\|f\|_{\mathcal{Y}_t'}\sigma^{-n}. \end{aligned} \tag{52}$$

This bound is valid for any $\sigma < \rho_1$ and hence holds up to $\sigma = \rho_1$.

Finally, inserting Eq. (52) back into Eq. (50) after integrating over $\boldsymbol{y}' \in U'$ with respect to the probability measure $\mathrm{d}\eta(\boldsymbol{y}')$ yields

$$\|\mathfrak{d}u^l_{n\mathbf{e}_1}\|_{\mathcal{X}} \leq C'h_l^t\|f\|_{\mathcal{Y}_t'}\rho^{-n}.$$

Similarly, given any $\nu \in \mathcal{F}$, it follows that

$$\|\mathfrak{d}u^l_\nu\|_{\mathcal{X}} \leq C'h_l^t\|f\|_{\mathcal{Y}_t'}\rho^{-\nu},$$

by applying Cauchy's integral formula in $\mathbb{C}$ with respect to each variable z_j for $j \in \{j : \nu_j \neq 0\}$.

The Chebyshev coefficients of the functional are estimated in a similar manner, using (21),

$$|\mathfrak{d}F^l_\nu| \leq C'h_l^{t+t'}\|f\|_{\mathcal{Y}_t'}\|\mathcal{G}\|_{\mathcal{X}_{t'}'}\rho^{-\nu}. \tag{53}$$

□

4.2.4. *Summability of the sequence of Chebyshev gpc coefficients.* It remains to prove the existence of a δ-admissible polyradii ρ (depending on ν) and to verify the $\ell_{\omega,p}$-summability of the right hand side of (49) with respect to $\nu \in \mathcal{F}$, i.e., of the sequence $(\rho(\nu)^{-\nu})_{\nu\in\mathcal{F}}$. Hereby, we identify suitable weights $\omega = (\omega_\nu)_{\nu\in\mathcal{F}}$ as well. In contrast to unweighted ℓ_p-summability [18, 19], weighted $\ell_{\omega,p}$-summability – considered first in [46] – requires stronger assumptions on the sequence $(b_{0,j})_{j\in\mathbb{N}}$ used as base for the δ-admissibility (42). Namely, with $\mathbf{v} = (v_j)_{j\in\mathbb{N}}$ and $v_j \geq 1$, we ask for properties (40) and (41) to be valid.

Theorem 7. *Let $\mathbf{v}$ be a sequence of weights fulfilling the summability conditions (40) and (41) and let $\omega_\nu := \theta^{\|\nu\|_0}\mathbf{v}^\nu$ for any $\nu \in \mathcal{F}$ and some $\theta \geq 1$. There exists a sequence of polyradii $(\rho(\nu))_{\nu\in\mathcal{F}}$ such that*

- i) *for each $\nu \in \mathcal{F}$, $\rho = \rho(\nu)$ is δ-admissible, with $\delta = (1-\kappa_{v,p})/2$, and*
- ii) $\| \left(\rho(\nu)^{-\nu}\right)_{\nu\in\mathcal{F}} \|_{\omega,p} \leq K_{\theta,p} < \infty$.

Proof. Full details of the argument can be found in [46]; here, we only indicate the main steps, in particular the construction of a sequence of weights ω and an associated, admissible sequence of polyradii.

For the weights $\mathbf{v}$ and a constant $\theta \geq 1$, we define the sequence of weights

$$\omega_\nu(\theta) := \theta^{\|\nu\|_0}\mathbf{v}^\nu = \theta^{\|\nu\|_0}\prod_{j:\nu_j\neq 0} v_j^{\nu_j}, \quad \text{for all } \nu \in \mathcal{F}.$$

Because of (40), there exists a finite set $E \subset \mathbb{N}$ such that, with $F := \mathbb{N}\backslash E$,

$$\sum_{j\in F} v_j^{(2-p)/p} b_{0,j} \leq \frac{\delta}{8\theta^{(2-p)/p}}.$$

For a given constant $\alpha > 1$ with $(\alpha-1)\sum_{j\in E} v_j^{(2-p)/p} b_{0,j} < \delta/2$, we define the sequences of polyradii (generally depending on ν) as

$$\rho_j(\nu) = \begin{cases} \alpha v_j^{(2-p)/p}, & j \in E, \\ \max\left\{v_j^{(2-p)/p}, \frac{\nu_j}{2|\nu_F|b_j}\right\}, & j \in F, \end{cases} \tag{54}$$

where we used the notation $|\nu_F| := \sum_{j\in F} \nu_j$. The δ-admissibility of this sequence, as well as its $\ell_{\omega,p}$ summability, ensuring the summability of the Chebyshev expansion of the differences, have been proved in [46, Theorem 4.2]. □

Combining the estimate (53) with the $\ell_{\omega,p}$ summability of the sequence ρ yields

$$\|\mathfrak{d}F^l\|_{\omega,p} \leq C h_l^{t+t'} \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} \| \left(\rho^{-\nu}\right)_{\nu\in\mathcal{F}} \|_{\omega,p} \,. \tag{55}$$

Consequently, with Eq. (38) it follows

$$|\mathfrak{d}F^l(\boldsymbol{y}) - \widehat{\mathfrak{d}F^l}(\boldsymbol{y})| \leq C s_l^{1-1/p} h_l^{t+t'} \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} \| \left(\rho(\nu)^{-\nu}\right)_{\nu\in\mathcal{F}} \|_{\omega,p} \,.$$

Theorem 3 is a direct consequence of the results in this section. Indeed, the bounded invertibility in the smoothness spaces of $A_0 \in \mathcal{L}(\mathcal{X}_t, \mathcal{Y}_t')$ together with the summability (44) implies the uniform bounded invertibility of the operator $A(\boldsymbol{z}) \in \mathcal{L}(\mathcal{X}_t, \mathcal{Y}_t')$, via a perturbation argument as stated in Theorem 5. This ensures the applicability of Theorem 6 (which itself depends on the two previous theorems). Theorem 7 finally proves the existence of both a positive weight sequence ω and a sequence of polyradii ρ as well as the $\ell_{\omega,p}$ summability.

4.3. **Rate of convergence of the MLCSPG method.** To simplify the exposition, we only derive the bounds for the approximation of a functional of the parametric solution. The results can be applied mutatis mutandis to derive the convergence rates for the full solution $u(\boldsymbol{y})$, – once the details of the (single level) compressive sensing scheme for the approximation of the full solution are worked out. We continue the estimate in (34) as follows:

$$\begin{aligned}
|F(\boldsymbol{y}) - F^L_{\text{MLCS}}(\boldsymbol{y})| &\leq |F(\boldsymbol{y}) - F^L(\boldsymbol{y})| + \sum_{l=1}^{L} \left|\mathfrak{d}F^l(\boldsymbol{y}) - \widehat{\mathfrak{d}F^l}(\boldsymbol{y})\right| \\
&\leq C h_L^{t+t'} \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} + \sum_{l=1}^{L} C s_l^{1-1/p} |||\mathfrak{d}F^l|||_{\omega,p} \\
&\leq C \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} \left(h_L^{t+t'} + \sum_{l=1}^{L} s_l^{1-1/p} h_l^{t+t'} \| \left(\rho(\nu)^{-\nu}\right)_{\nu\in\mathcal{F}} \|_{\omega,p} \right).
\end{aligned}$$

We absorb the norm $\| \left(\rho(\nu)^{-\nu}\right)_{\nu\in\mathcal{F}} \|_{\omega,p}$ into the constant $C > 0$, yielding

$$|F(\boldsymbol{y}) - F^L_{\text{MLCS}}(\boldsymbol{y})| \leq C \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} \left(h_L^{t+t'} + \sum_{l=1}^{L} s_l^{1-1/p} h_l^{t+t'} \right) \,.$$

Using that the levels are related via $h_l = h_{l-1}/2$ we obtain

$$|F(\boldsymbol{y}) - F^L_{\text{MLCS}}(\boldsymbol{y})| \leq C \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} h_L^{t+t'} \left(1 + \sum_{l=1}^{L} s_l^{1-1/p} 2^{(L-l)(t+t')} \right) \,.$$

We balance sampling and discretization errors on each mesh level l in this bound. Thus the choice

$$s_l \asymp 2^{(L-l)(t+t')p/(1-p)} = 2^{(L-l)\sigma_p(t+t')}, \qquad \text{with } \sigma_p(t) = \frac{tp}{1-p}, \tag{56}$$

implies an overall error bound of

$$|F(\boldsymbol{y}) - F^L_{\text{MLCS}}(\boldsymbol{y})| \leq C \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} (L+1) h_L^{t+t'} = C \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} (|\log(h_L)| + 1) h_L^{t+t'} \,.$$

The choice of the sparsities (56) together with (28) yields a number of samples per level scaling as

$$m_l \asymp s_l \log^3(s_l) \log(N_l) \asymp 2^{(L-l)\sigma_p(t+t')} \left(L-l\right)^3 \log(N_l) \,. \tag{57}$$

Note that the size N_l of the initial index set Γ_l may depend on s_l and on the choice of weights ω. More details are given in the next section. The global error in L^2 is bounded as in Eq. (34),

$$\|F - F^L_{\text{MLCS}}\|_2 \leq \|F - F^L\|_2 + \sum_{l=1}^{L} \|\mathfrak{d}F^l - \widehat{\mathfrak{d}F^l}\|_2 \,. \tag{58}$$

The first term is computed using the uniform bound (21) and the fact that η is a probability measure. To compute the sum, it suffices to apply the ℓ_2-error bound in (30) to the L details, $\|\mathfrak{d}F^l - \widehat{\mathfrak{d}F^l}\|_2 \leq D s_l^{1/2-1/p}\|\mathfrak{d}F^l\|_{\omega,p}$. Hence, applying (21) to the first term and combining the ℓ_2 bound in (30) with the prior estimate (47) and the number of samples (56) in the terms in the sum yields

$$\|F - F^L_{\text{MLCS}}\|_2 \leq C\|f\|_{\mathcal{Y}'_t}\|\mathcal{G}\|_{\mathcal{X}'_{t'}} h_L^{t+t'} \left(1 + \sum_{l=1}^{L} 2^{(t+t')(L-l)\frac{p-2}{2(1-p)}} 2^{(L-l)(t+t')}\right) = C'\|f\|_{\mathcal{Y}'_t}\|\mathcal{G}\|_{\mathcal{X}'_{t'}} h_L^{t+t'} \,.$$

Alternatively, one can also balance the number of samples with the discretization error to reach a prescribed L_2 error of $\mathcal{O}(h_L^{t+t'})$ by combining Eq. (58) with the compressed sensing approximation (30):

$$\begin{aligned}\|F - F^L_{\text{MLCS}}\|_2 &\leq C h_L^{t+t'}\|f\|_{\mathcal{Y}'_t}\|\mathcal{G}\|_{\mathcal{X}'_{t'}} + C\sum_{l=1}^{L} s_l^{1/2-1/p} h_l^{t+t'}\|f\|_{\mathcal{Y}'_t}\|\mathcal{G}\|_{\mathcal{X}'_{t'}}\left\|\left(\rho(\nu)^{-\nu}\right)_{\nu\in\mathcal{F}}\right\|_{\omega,p} \\ &\leq C\|f\|_{\mathcal{Y}'_t}\|\mathcal{G}\|_{\mathcal{X}'_{t'}} h_L^{t+t'}\left(1 + \sum_{l=1}^{L} s_l^{1/2-1/p} 2^{(L-l)(t+t')}\right) \,.\end{aligned}$$

In this case, choosing

$$s_l \asymp 2^{\frac{(L-l)(t+t')2p}{2-p}} \tag{59}$$

ensures the L_2 error bound

$$\|F - F^L_{\text{MLCS}}\|_2 \leq C h_L^{t+t'}(1 + |\log h_L|)\|f\|_{\mathcal{Y}'_t}\|\mathcal{G}\|_{\mathcal{X}'_{t'}} \,. \tag{60}$$

5. Implementation Aspects

This section describes several aspects that are relevant for the numerical applicability of the theoretical approach introduced above. In particular, we investigate the truncation of the (potentially infinite) sequence of parameters to a finite subset, and specify initial choices of finite index sets $\Lambda \subset \mathcal{F}$ that are guaranteed to contain the support of the best (weighted) s-term approximation of the solution and can be used within weighted ℓ_1-minimization or other CS algorithms.

5.1. **Dimension truncation.** So far, we have worked on a purely theoretical level, where the parameter vector is potentially infinite (but countable). To ensure the applicability of the results, we have to verify that truncating the parameter vector to a finite dimensional space (yet allowing this truncation to be rather large) still allows for reliable approximations.

We consider the weak solutions of the truncated version of Eq. (1):

$$\text{Find } u^{(B)} \in \mathcal{X}, \text{ such that } \langle A^{(B)}(\boldsymbol{y})u^{(B)}, v\rangle = \langle f, v\rangle \quad \text{for all } v \in \mathcal{Y}, \tag{61}$$

where the operator $A^{(B)}(\boldsymbol{y})$ is defined, for a finite $B \in \mathbb{N}$, as $A(y_1, y_2, \cdots, y_B, 0, 0, \cdots)$.

In particular, we assume some decay of the *energy* of the operator $A(\boldsymbol{y})$ (i.e. assuming a certain order on the parameters) such that for any $\varepsilon > 0$, there exists $B := B(\varepsilon, A)$ with

$$\|A(\boldsymbol{y}) - A^{(B)}(\boldsymbol{y})\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} \leq \varepsilon\mu, \quad \forall \boldsymbol{y} \in U, \tag{62}$$

where μ is the constant appearing in the $\inf - \sup$ conditions (8).

In this case, the following generalization of results in [22] holds.

Proposition 2. *Assume the operator A satisfies the (continuous)* $\inf - \sup$ *conditions* (8) *and the decay property* (62)*. Then for any accuracy parameter ε, there exists a truncation parameter $B \in \mathbb{N}$ such that the solutions to the truncated problem* (61) *and to the original problem* (1) *are close to each other in the following sense*

$$\|u^{(B)}(\boldsymbol{y}) - u(\boldsymbol{y})\|_{\mathcal{X}} \leq \frac{C\varepsilon}{\mu}\|f\|_{\mathcal{Y}'}, \tag{63}$$

where $u^{(B)}(\boldsymbol{y})$ is the solution of the truncated problem (61).

Proof. The weak solutions are characterized by

$$\begin{aligned} \text{Find } u(\boldsymbol{y}) \text{ such that} \quad & \langle A(\boldsymbol{y})u(\boldsymbol{y}), v\rangle = \langle f, v\rangle \quad \text{for all } v \in \mathcal{Y}, \\ \text{Find } u^{(B)}(\boldsymbol{y}) \text{ such that} \quad & \langle A^{(B)}(\boldsymbol{y})u^{(B)}(\boldsymbol{y}), v\rangle = \langle f, v\rangle \quad \text{for all } v \in \mathcal{Y}. \end{aligned}$$

Since these equalities hold for all v, they imply the orthogonality conditions

$$\langle A(\boldsymbol{y})u(\boldsymbol{y}) - A^{(B)}(\boldsymbol{y})u^{(B)}(\boldsymbol{y}), v\rangle = 0 \quad \text{for all } v \in \mathcal{Y}.$$

Rearranging the terms yields $\langle A(\boldsymbol{y})\left(u(\boldsymbol{y}) - u^{(B)}(\boldsymbol{y})\right), v\rangle = -\langle \left(A(\boldsymbol{y}) - A^{(B)}(\boldsymbol{y})\right) u^{(B)}(\boldsymbol{y}), v\rangle$ for all $v \in \mathcal{Y}$. This means that $u(\boldsymbol{y}) - u^{(B)}(\boldsymbol{y})$ is the weak solution to the operator equation (1) with forcing term $\left(A(\boldsymbol{y}) - A^{(B)}(\boldsymbol{y})\right) u^{(B)}(\boldsymbol{y})$. Consequently, using the $\inf - \sup$ conditions twice and the decay property (62), we obtain

$$\|u(\boldsymbol{y}) - u^{(B)}(\boldsymbol{y})\|_{\mathcal{X}} \leq \frac{C}{\mu}\left\|A(\boldsymbol{y}) - A^{(B)}(\boldsymbol{y})\right\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')}\|u^{(B)}(\boldsymbol{y})\|_{\mathcal{X}} \leq \frac{\varepsilon C}{\mu}\|f\|_{\mathcal{Y}'},$$

which concludes the proof. □

Consequently, it is sufficient to draw the m_l samples per level at random according to the truncated distribution. As a concrete example let us consider the case of linear dependence on the parameters as described in [46] and in Eq. (5). Assuming that $A_0 : \mathcal{X} \to \mathcal{Y}'$ is invertible (which was required in Theorem 3) and that $(b_{0,j})_j \in \ell_1$ (which is weaker than the conditions in the previous section) the fluctuations A_j, $j \geq 1$ are arranged in nonincreasing order, i.e., such that $b_{0,j} \geq b_{0,k}$ for $1 \leq j \leq k$, then the operator (5) satisfies the following dimension truncation error bound

$$\|A(\boldsymbol{y}) - A^{(B)}(\boldsymbol{y})\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} = \|\sum_{j>B} y_j A_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} = \|A_0 \sum_{j>B} y_j B_j\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} \leq \|A_0\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} \sum_{j>B} b_{0,j},$$

for any $\boldsymbol{y} \in U$. Moreover (see [46, Thm 2.9],[41, Thm 5.1]), the tail can be estimated by

$$\sum_{j>B} b_{0,j} \leq \min\left\{\frac{1}{1/p-1}, 1\right\} \|(b_{0,j})_j\|_p B^{-(1/p-1)}$$

for some $p < 1$. Consequently, choosing $B \geq h_L^{-(t+t')p/(1-p)}$ yields a global approximation (accounting for the truncation error, the PG approximation error, and the CS error) in $\mathcal{O}(h_L^{t+t'})$.

5.2. **Initial set of candidate vectors.** As detailed in the discussion before Theorem 3, the results are, so far, developed for an infinite Chebyshev expansion. To render the problem computationally feasible, we truncate to a finite-dimensional, parametric expansion, where the truncation dimension is at our disposal and therefore can be considered a discretization parameter. Let the sums (36) and (37) be truncated to a finite set $\Gamma_l \subset \mathcal{F}$. Some strategies for selecting such a set Γ_l were already described in [46], which was based on the work in [47]. We have the following analog to Theorem 2 (proven in [47]) in the case of expansions in terms of a countable sequence of parameters.

Theorem 8. *Let $\gamma \in (0, 1)$. Let $F(\boldsymbol{y}) = \sum_{\nu\in\mathcal{F}} F_\nu T_\nu(\boldsymbol{y})$ be a function with $\|\mathbf{F}\|_{\omega,p} = \|(F_\nu)_\nu\|_{\omega,p} < \infty$ for some $p < 1$ and some weights $\omega_\nu \geq \|T_\nu\|_\infty$ for all $\nu \in \mathcal{F}$. For a given sparsity $s_l \geq 1$, define the initial set of indices as*

$$\Gamma_l := \{\nu \in \mathcal{F} : \omega_\nu^2 \leq s_l/2\}. \tag{64}$$

Furthermore, assume that $N_l := |\Gamma_l|$ is finite and draw

$$m_l \geq c_0 s_l \max\{\log^3(s_l)\log(N_l), \log(1/\gamma)\} \tag{65}$$

sampling points $\boldsymbol{y}^{(i)}$ independently and identically distributed according to the orthogonalization measure η. Let $\widehat{\mathbf{F}}$ be the solution of

$$\min \|\mathbf{H}\|_{\omega,1} \quad \textit{subject to } \|A\mathbf{H}-b\|_2 \leq 2^{1-p}\tau\sqrt{m_l}s_l^{1/2-1/p}\|\mathbf{F}\|_{\omega,p},$$

for some $\tau \geq 1$ and set $\widehat{F} = \sum_{\nu\in\Gamma_l}\widehat{\mathbf{F}}_{\boldsymbol{\nu}}T_{\boldsymbol{\nu}}$. Then, with probability at least $1-\gamma$

$$\|F-\widehat{F}\|_\infty \leq \|\mathbf{F}-\widehat{\mathbf{F}}\|_{\omega,1} \leq c_\tau s_l^{1-1/p}\|\mathbf{F}\|_{\omega,p},$$
$$\|F-\widehat{F}\|_2 = \|\mathbf{F}-\widehat{\mathbf{F}}\|_2 \leq d_\tau s_l^{1/2-1/p}\|\mathbf{F}\|_{\omega,p}\ .$$

A drawback of the recovery based on an optimization problem is that it requires the knowledge (or an approximation) of the norm of the unknown vector $\mathbf{F}$. This can be overcome in practical applications by applying the recovery to various estimations (similar to a cross validation in the machine learning literature [53]) or by using greedy methods, e.g. [13, 29].

The cardinality N_l of the set

$$\Gamma_l = \{\nu\in\mathcal{F} : \omega_\nu^2 \leq s_l/2\} = \{\nu\in\mathcal{F} : \|\nu\|_0\log(\theta) + \sum_{j\in\operatorname{supp}\nu} 2\log(v_j)\nu_j \leq \log(s_l/2)\},$$

where the weights ω_ν are chosen as in (46), influences the number m_l of samples in (65) (and the computational complexity of the weighted ℓ_1-minimization problem). Obviously, N_l depends on s_l as well as on the weight sequence (v_j) used in the definition (46) of (ω_ν). We recall the following estimates from [46].

Proposition 3. *Let $\omega_\nu = \theta^{\|\nu\|}\mathbf{v}^\nu$, $\nu\in\mathcal{F}$, for a sequence $\mathbf{v}=(v_j)_{j\geq1}$ specified below and assume $s_l \geq 1$.*

(1) For $v_j = \beta$ for $1\leq j\leq d$ and $v_j=\infty$ for $j>d$ (i.e., we consider constant weights for the first d dimensions and ignore the remaining ones), we have

$$N_l = |\Gamma_l| \leq \begin{cases} \left(\left(1+\frac{1}{\log_2(\beta^2)}\right)ed\right)^{\log_{2\beta^2}(s_l/2)}, & s_l < 2^{d+1}\beta^{2d}, \\ (\log_{\beta^2}(\beta^2 s_l/2))^d, & s_l \geq 2^{d+1}\beta^{2d}\ . \end{cases}$$

(2) For polynomially growing weights $v_j = cj^\alpha$ with $c>1$ and $\alpha>0$, there holds subexponential growth

$$N_l \leq C_{\alpha,c}s_l^{\gamma_{\alpha,c}\log(s_l)}$$

for some constants $C_{\alpha,c}>0$ and $\gamma_{\alpha,c}>0$ depending only on c and α.

Inserting these bounds into Condition (65) on the number of required samples (assuming that the $\log(1/\gamma)$-term does not exceed the other logarithmic terms) shows that the following choices of m_l are valid:

- For constant weights $v_j=\beta$ for $1\leq j\leq d$ and $v_j=\infty$ for $j>d$, we can chose

$$m_l \asymp \begin{cases} \log(d)s_l\log^4(s_l), & s_l < 2^{d+1}\beta^{2d}, \\ ds_l\log^3(s_l)\log(\log(s_l)), & s_l\geq 2^{d+1}\beta^{2d}. \end{cases} \tag{66}$$

- For polynomially growing weights $v_j = cj^\alpha$ with $c>1$ and $\alpha>0$, we can chose

$$m_l \asymp s_l\log^5(s_l)\ . \tag{67}$$

The case of exponentially growing weights has been analyzed in [46] and yields situations where $N_l\leq m_l$. In this situation, compressed sensing techniques should not be used, as least-squares methods are expected to perform better [43].

We note that in the case of constant weights, the first case in (66) is the most relevant. In fact, with the choice of s_l as in (56), i.e., $s_l = \mathcal{C}2^{(L-l)(t+t')p/(1-p)}$ for some proportionality constant $\mathcal{C}>0$, if $c := \frac{d+1+2d\log_2(\beta)}{(t+t')p}(1-p) - \frac{\log_2(\mathcal{C})(1-p)}{(t+t')p}$ is large enough (for instance $c\geq L$, which is true whenever $\mathcal{C}\leq s^{d+1}\beta^{2d}/2^{L(t+t')p/(1-p)}$) then only the first case of (66) will occur for all $l=1,\dots,L$. In particular, with all the parameters (β, t, t', and p) fixed, a larger number d of active variables will lead to a larger c. It is therefore reasonable to assume that this corresponds to the main regime.

5.3. **Computational Cost.** In the ensuing work bounds, we assume at our disposal multilevel solvers as described, e.g. in [35, 54]. These solvers compute approximate solutions of the Galerkin equations at cost scaling linearly in the number of unknowns of the mesh. This gives rise to the following complexity estimates, where we treat the case of constant and polynomially growing weights.

Proposition 4. *Under the assumptions (64), (65) as well as (40), (41) for some $0 < p < 1$ and smoothness parameters t, t', the function $\boldsymbol{y} \mapsto \mathcal{G}(u(\boldsymbol{y}))$ can be approximated in $L^2(U, \eta)$ to accuracy $\mathcal{O}(h_L^{t+t'})$ via a MLCSPG discretization with L levels and with total work W_L^T scaling as*

$$W_L^T \lesssim \begin{cases} \dfrac{\log(d)\sigma_p(\tau)^4 L^4 2^{L\sigma_p(\tau)}}{\sigma_p(\tau) - n}, & \sigma_p(\tau) > n, \quad (v_j) \text{ constant} \\ \log(d)\sigma_p(\tau)^4 L^5 2^{nL}, & \sigma_p(\tau) = n, \quad (v_j) \text{ constant} \\ \dfrac{\log(d)\sigma_p(\tau)^4 (2^{nL} - 2^{\sigma_p(\tau)L})}{(n - \sigma_p(\tau))^4}, & \sigma_p(\tau) < n, \quad (v_j) \text{ constant} \\ \dfrac{\sigma_p(\tau)^5 L^5 2^{L\sigma_p(\tau)}}{\sigma_p(\tau) - n}, & \sigma_p(\tau) > n, \quad (v_j) \text{ polynomial} \\ \sigma_p(\tau)^5 L^6 2^{nL}, & \sigma_p(\tau) = n, \quad (v_j) \text{ polynomial} \\ \dfrac{\sigma_p(\tau)^5 (2^{nL} - 2^{\sigma_p(\tau)L})}{(n - \sigma_p(\tau))^5}, & \sigma_p(\tau) < n, \quad (v_j) \text{ polynomial} \end{cases} \tag{68}$$

where $\sigma_p(\tau) = \tau p/(1-p)$ with $\tau = t + t'$ and where n denotes the spatial dimension.

Proof. Multigrid solvers have a computational complexity scaling linearly with the number $w_l \asymp 2^{nl}$ of unknowns at level l which implies that the work at level l is on the order of $W_l = m_l \cdot w_l$, $1 \le l \le L$.

Assuming we are given constant weights $v_j = \beta$, for $1 \le j \le d$, and $s_l < 2^{d+1}\beta^{2d}$, and that d is sufficiently large, we can chose m_l as in the first row of Eq. (66). Thus, omitting constants,

$$\begin{aligned} W_L^T &= \sum_{l=1}^L W_l \lesssim \sum_{l=1}^L \log(d) s_l \log^4(s_l) 2^{nl} \lesssim \sum_{l=1}^L \log(d) 2^{(L-l)\sigma_p(\tau)} \log^4\left(2^{(L-l)\sigma_p(\tau)}\right) 2^{nl} \\ &\lesssim \log(d)\sigma_p(\tau)^4 2^{nL} \sum_{l=1}^L (L-l)^4 2^{(L-l)(\sigma_p(\tau)-n)} = \log(d)\sigma_p(\tau)^4 2^{nL} \sum_{j=1}^{L-1} j^4 2^{j(\sigma_p(\tau)-n)}. \end{aligned} \tag{69}$$

We can bound $S := \sum_{j=1}^{L-1} j^4 2^{j(\sigma_p(\tau)-n)} \le \int_0^L 2^{x(\sigma_p(\tau)-n)} x^4 \mathrm{d}x$. If $\sigma_p(\tau) = n$, it follows that $S \le L^5/5$. Otherwise, with $K = (\sigma_p(\tau) - n)\ln(2)$, an integration by part yields

$$S \le \frac{L^4 e^{LK}}{K} - \frac{4}{K} \int_0^L x^3 e^{xK} \mathrm{d}x. \tag{70}$$

If $K > 0$, i.e. $\sigma_p(\tau) > n$, the remaining integral is positive and thus $S \le \frac{L^4 e^{LK}}{K} = \frac{L^4 2^{L(\sigma_p(\tau)-n)}}{\ln(2)(\sigma_p(\tau)-n)}$. If $K < 0$, repeated integration by parts leads to

$$S \le \frac{L^4 e^{LK}}{K} - \frac{4L^3 e^{LK}}{K^2} + \frac{12L^2 e^{LK}}{K^3} - \frac{24L e^{LK}}{K^4} + \frac{24}{K^4} \int_0^L e^{xK} \mathrm{d}x. \tag{71}$$

Noticing that $\frac{L^4 e^{LK}}{K} - \frac{4L^3 e^{LK}}{K^2} + \frac{12L^2 e^{LK}}{K^3} - \frac{24L e^{LK}}{K^4} < 0$, it follows that

$$S \le \frac{24}{K^4} \int_0^L e^{xK} \mathrm{d}x = 24 \frac{e^{LK} - 1}{K^5} = \frac{24(1 - 2^{(\sigma_p(\tau)-n)L})}{(n - \sigma_p(\tau))^5 \ln(2)^5}. \tag{72}$$

The result for polynomially growing weight sequences (v_j) is shown in a similar fashion (with appropriate changes in exponents). □

Remark 2. *Recalling that the workload for the computation of one solution at the finest discretization level L is $w_L \asymp 2^{nL}$, the previous result means that for $\sigma_p(t + t') < n$, the total work is bounded only by a multiple of the cost of one PDE solve at the finest level, where the multiplicative constant involves a factor of $\log(d)$ in the case of constant weights and in addition only depends on n, p, t, t'.*

Combining Theorem 3 together with Proposition 4 about the computation costs and Proposition 2 regarding the truncation of the operator, we are finally able to state our main theorem. To this end we first summarize the assumptions on the parametric operator $A(\boldsymbol{y}) = A_0 + \sum_{j\geq 1} y_j A_j$.

- The nominal operator A_0 is inf-sup stable, i.e.,

$$\inf_{0\neq w^h\in\mathcal{X}^h} \sup_{0\neq v^h\in\mathcal{Y}^h} \frac{\langle A_0 w^h, v^h\rangle}{\|w^h\|_{\mathcal{X}}\|v^h\|_{\mathcal{Y}}} \geq \mu_0 > 0, \quad \inf_{0\neq v^h\in\mathcal{Y}^h} \sup_{0\neq w^h\in\mathcal{X}^h} \frac{\langle A_0 w^h, v^h\rangle}{\|w^h\|_{\mathcal{X}}\|v^h\|_{\mathcal{Y}}} \geq \mu_0 > 0.$$

- For some $0 < p < 1$ and some weight sequence $\mathbf{v} = (v_j)_{j\in\mathbb{N}}$ with $v_j \geq 1$, the sequence $\boldsymbol{b}_0$ with components $b_{0,j} = \|A_0^{-1}A_j\|_{\mathcal{L}(\mathcal{X})}$, $j\geq 1$, satisfies

$$\kappa_{\mathbf{v},p} := \sum_{j\geq 1} b_{0,j} v_j^{(2-p)/p} < 1 \quad \text{and} \quad \sum_{j\geq 1} b_{0,j}^p v_j^{2-p} < \infty.$$

- For some $t\in(0,\bar{t})$, the operators A_j, $j\geq 0$, are defined as operators from $\mathcal{X}_t$ into $\mathcal{Y}_t'$ the sequence $\mathbf{b}_t$ with components $b_{t,j} = \|A_0^{-1}A_j\|_{\mathcal{L}(\mathcal{X}_t)}$ satisfies

$$\kappa_t := \sum_{j\geq 1} b_{t,j} \leq 1\,, \quad \boldsymbol{b}_t \in \ell^{p_t}.$$

Theorem 9. *Let $\gamma\in(0,1)$ and $L\in\mathbb{N}$ be a number of discretization levels. Let $A(\boldsymbol{y})$ be an affine-parametric operator and let $\mathbf{v} = (v_j)_{j\geq\mathbb{N}}$ be a sequence of weights with $v_j \geq 1$. Assume that $A_0 \in \mathcal{L}(\mathcal{X}_t, \mathcal{Y}_t')$ is boundedly invertible and that the sequence $\boldsymbol{b}_t = (b_{t,j})_{j\geq 1}$ are such that the summability conditions (43) and (41) hold true for some $0<p<1$. Then, for any discretization level $1\leq l\leq L$, the sequence of Chebyshev coefficients of $\mathfrak{d}u^l$ with respect to the parameter vector (36) is (weighted) compressible, i.e., for a sequence of weights $\omega = (\omega_\nu)_{\nu\in\mathcal{F}}$ with $\omega_\nu = \theta^{\|\nu\|_0}\mathbf{v}^\nu$ there holds $\sum_{\nu\in\mathcal{F}} \omega^{2-p}\|\mathfrak{d}u_\nu^l\|_{\mathcal{X}}^p < \infty$.*

Moreover, if we are interested in a functional of the solution $F(\boldsymbol{y}) = \mathcal{G}(u(\boldsymbol{y}))$ and if the operators $A(\boldsymbol{y}) \in \mathcal{L}(\mathcal{X}_t, \mathcal{Y}_t')$ are boundedly invertible in the smoothness scales $(\mathcal{X}_t, \mathcal{Y}_t)$ in (14), (15) and if $\mathcal{G}\in\mathcal{X}_{t'}'$, then the function $F(\boldsymbol{y}) = \sum_{\nu\in\mathcal{F}} F_\nu T_\nu(\boldsymbol{y})$ can be approximated by $F_{MLCS}^L(\boldsymbol{y}) := \sum_{l=1}^L \widehat{\mathfrak{d}F^l}(\boldsymbol{y})$ where $\widehat{\mathfrak{d}F^l}(\boldsymbol{y})$ is a single-level CSPG approximation from $m_l \asymp s_l \max\{\log^3(s_l)\log(N_l), \log(L/\gamma)\}$ sampling points with $s_l \asymp 2^{(L-l)(t+t')p/(1-p)}$, where $N_l = |\Gamma_l|$ for $\Gamma_l = \{\nu\in\mathcal{F} : \omega_\nu^2 \leq s_l/2\}$.

Then, with probability at least $1-\gamma$, this approximation fulfills the bounds

$$\|F - F_{MLCS}^L\|_\infty \leq C h_L^{t+t'} \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'}(L+1), \tag{73}$$

$$\|F - F_{MLCS}^L\|_2 \leq C' h_L^{t+t'} \|f\|_{\mathcal{Y}_t'} \|\mathcal{G}\|_{\mathcal{X}_{t'}'} \tag{74}$$

and can be computed in a total work that scales as

$$W_L^T \lesssim \begin{cases} 2^{nL}, & \sigma_p(\tau) < n, \\ L^{\gamma+1}2^{nL}, & \sigma_p(\tau) = n, \\ L^\gamma 2^{L\sigma_p(\tau)}, & \sigma_p(\tau) > n, \end{cases} \tag{75}$$

where $\gamma = 4$ or 5 for constant or polynomially growing weights $\mathbf{v}$, respectively.

Proof. This theorem comes by applying L times Theorem 8 with probability γ/L at each level. The bound (73) follows from Theorem 3 and using the calculations from Section 4.3. ☐

6. Numerical results

We illustrate our theoretical findings with some numerical examples. We consider the diffusion problem from Eq. (3) and apply the QoI $F(\boldsymbol{y}) = \int_{x\in D} u(x,\boldsymbol{y})\mathrm{d}x$. We first introduce the problem of piecewise constant diffusion in $n=1$ and $n=2$ spatial dimensions, and then look into a trigonometric expansion of the diffusion coefficient. As a numerical solver, we have used the tools developed via the FEniCS project [2, 42].[2] Before introducing our results, we want to stress out that the dimension truncation introduced in Section 5.1 is irrelevant here as we consider only finite dimensional expansions of the diffusion coefficient.

[2]Note that all the code for reproducible research and further use is available from one of the authors' github page: https://github.com/jlbouchot/CSPDEs

TABLE 1. Number of sampling points and cardinality of active set for continuous, piecewise linear FE approximation of the diffusion problem in one spatial dimension with uniform weights $v_j = 1.05$, $1 \leq j \leq d = 6$.

$L-l=$	0	1	2
N_l	637	3619	17066
m_l	259	659	1560
s_l	20	40	80

6.1. **Error incurred by compressed sensing approximation.** The first example illustrates that the error incurred by the multiple CS approximations is controlled. We consider both cases $n = 1$ and $n = 2$ (corresponding to one and two spatial dimensions), and present results for the piecewise constant diffusion coefficient case. We further consider a constant forcing term f and use a basis of degree two polynomials for the finite element method, which computes an exact solution (for every fixed $\boldsymbol{y}$). In particular, the error in Eq. (34) is only generated by the compressed sensing components of the algorithm. Note that within this *component*, are included 1) the truncation of the operator (presented in Section 5.1), 2) the truncation of the (level-dependent) polynomial space to Γ_l, and 3) the compressed sensing approximation. active parameters, the two other ones are unavoidable. We prevent error 1) from occurring by considering a finite dimensional expansion (avoiding truncation). However, errors 2) and 3) and unavoidable when using the compressed sensing framework

6.1.1. *One spatial dimension.* We consider the bounded interval $D = (0,1)$ with equispaced partition $\overline{D} = \bigcup_{i=1}^d \overline{D_i}$ into subintervals $D_i = (x_{i-1}, x_i)$ where $x_i = i/d$, for some $d \in \mathbb{N}$. We let $a(x, \boldsymbol{y}) = \bar{a} + \sum_{j=1}^d y_j c_j \chi_{D_j}(x)$ with $\bar{a}$ being a constant independent of x, $\{c_j\}_{j=1}^d$ a predefined (fixed) sequence such that the (weighted) uniform ellipticity assumption (6) holds, and χ_{D_j} the indicator function of the set D_j. Fig. 1 is spltted into four main quadrants, each divided in two graphs. The upper graph of each quadrant shows the pointwise estimation of the functions $t \mapsto F^L_{\text{MLCS}}(t\mathbf{e}_k)$, for $-1 \leq t \leq 1$, and $k = 1$ (upper left quadrant), $k = 2$ (upper right), $k = 3$ (lower left), and $k = 4$ (lower right), and for $L = 1, 2, 3$ (green, red, and light blue curves respectively). The bottom graph shows the pointwise error $F^L_{\text{MLCS}}(t\mathbf{e}_k) - F(t\mathbf{e}_k)$ for the various k. To generate these plots,

we selected the parameters to be $d = 6$ and picked uniform (small) local variations as $c_j = 1/6$, for $1 \leq j \leq d$. The uniform weights v_j are selected as $v_j = 1.05$ for all j. The sparse approximation was done via a weighted version of the Hard Thresholding Pursuit algorithm [31], picked for its proven fast convergence [10]. We also set the forcing term $f \equiv 1$ to be constant. Then, *for any* $\boldsymbol{y} \in U$ *the solution to the diffusion equation is continuous and piecewise quadratic.* The (level dependent) number of samples and sparsities have been chosen as

$$m_l = 2 \cdot s_l \log(N_l), \tag{76}$$

$$s_l = 20 \cdot 2^{L-l}\ . \tag{77}$$

The choice of m_l differs slightly from the theoretically justified choice in Eq. (57). The selection (76) refers to the usual rule of thumb in compressed sensing which is justified by non-uniform recovery results with random matrices, see [32, Ch.9.2] for details. While the choice (76) of CS sample numbers m_l is below what is sufficient by our theoretical results, we shall see in the numerical examples ahead that even this optimistic selection of sample number is more than sufficient for our problems. The choice (77) of s_l corresponds to Eq. (56) where the proportionality constant is chosen as 20 and the regularity assumption of the solution is taken as $\sigma_p(t + t') = (t + t')p/(1 - p) = 1$ to simplify the exposition.

The initial mesh size was set to $h_0 = 5 \cdot 10^{-4}$ but further numerical tests – not included in this paper, but available online – have shown that this parameter has, in this case, little to no influcence over the results (since the numerical solver is exact).

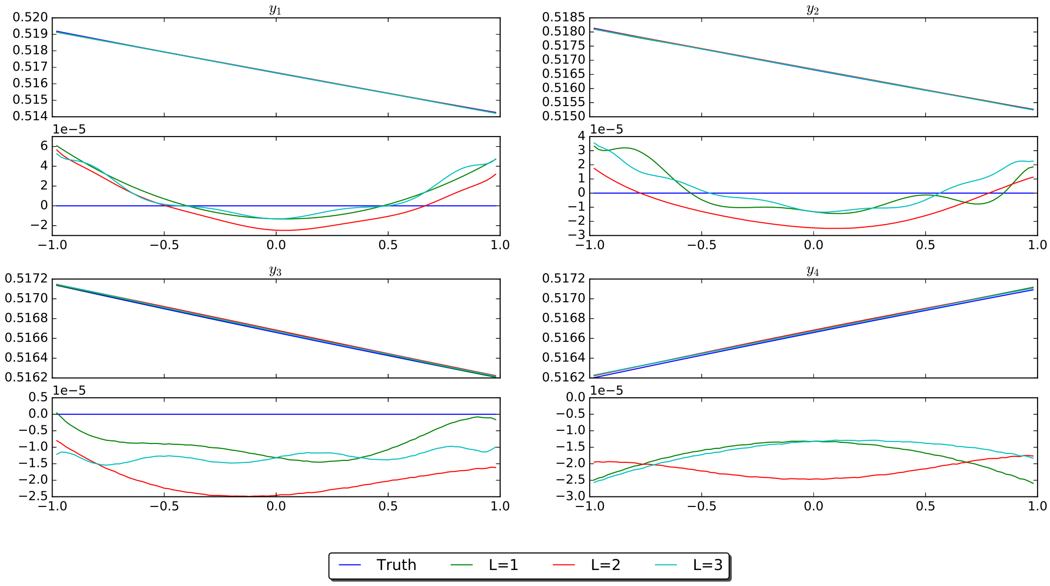

FIGURE 1. Convergence of the MLCSPG method for the piecewise constant diffusion problem, $c_j = 1/d$, $1 \leq j \leq d = 6$. The graphs show $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j)$ (upper graphs) and $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j) - F(t\mathbf{e}_j)$ (lower graphs), for $j = 1$ (upper left quadrant), $j = 2$ (upper right quadrant), $j = 3$ (bottom left quadrant) and $j = 4$ (bottom right quadrant), and $-1 \leq t \leq 1$. See text for more details.

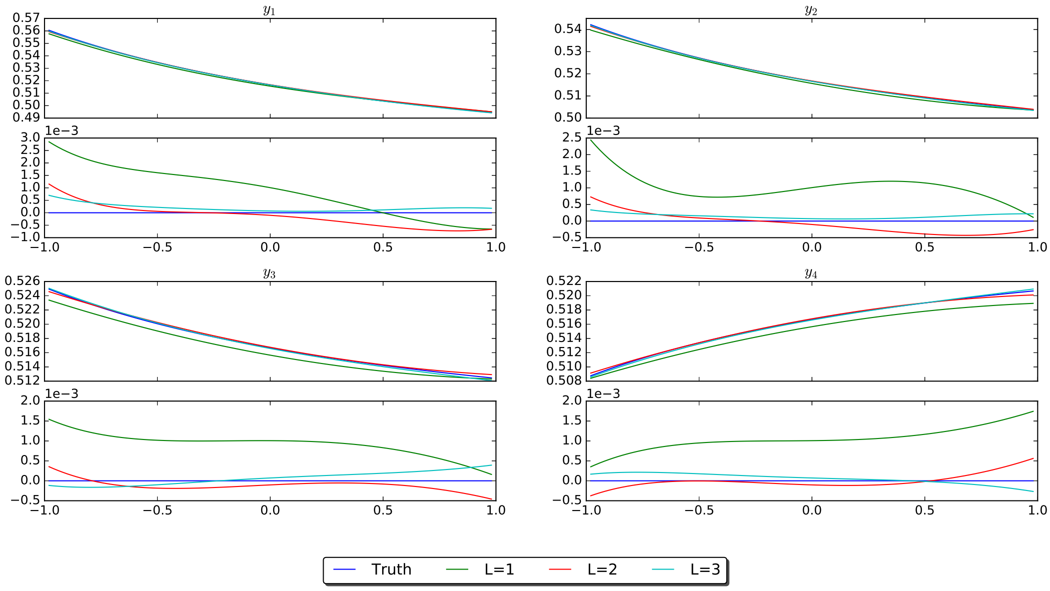

FIGURE 2. Convergence of the MLCSPG method for the piecewise constant diffusion problem, $c_j = 2$, $1 \leq j \leq d = 6$. Each quadrant is divided into two graphs. The top one corresponds to the pointwise estimation $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j)$, with $-1 \leq t \leq 1$, for $j = 1$ (upper left quadrant), $j = 2$ (upper right), $j = 3$ (bottom left) and $j = 4$ (bottom right). The lower graph in each quadrant corresponds to the pointwise error $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j) - F(t\mathbf{e}_j)$.

The value of the local fluctuations can be increased without violating the Uniform Ellipticity Assumption (thanks to the non-overlapping support of the ψ_j). This is shown for instance in Fig. 2, where the c_j are chosen uniformly as $c_j = 2$. A similar behavior as in the previous scenario is noticed, showing that the methods is robust, even when the local variations are large compared to the mean field.

Table 1 shows the number of sampling points and the size of the active set $N_l = |\Gamma_l|$. This set has been computed according to (64). In particular, one can see that considering relatively small sparsities yields a fairly large starting set (which likely contains the support of the best weighted s-sparse approximation) and keeps the number of sampling points (i.e., PDE solves) low.

We compare the convergence of our algorithms with other methods: Monte-Carlo sampling and least squares (ℓ_2 recovery) [43]. These results were obtained using the piecewise constant linear diffusion problem above, with small variations ($c_j = 1/6$ uniformly). We have however slightly increased the values of the constant weights $v_j = 1.07$ for these experiments. The estimation of the Chebyshev coefficients are displayed in Fig. 3, where the magnitudes of the Chebyshev coefficients of the (functional of the) parametric solution are displayed on a logarithmic (base 10) scale. The x-axes corresponds to an enumeration of the multi-index of the Chebyshev coefficient, whereby the larger ones (in magnitude, according to the ℓ_2 recovery) are first. The least squares solution is obtained as follows. We first build the active set of candidates for the truncated polynomial space as predicted by Theorem 8, i.e. $\Gamma = \cup_{l=1}^{3} \Gamma_l$. This set has total dimension $N = 12171$. Note that this value is smaller than the one in the previous experiment due to the increase of the weigths v_j. Then $m = 24342$ sampling points $\boldsymbol{y}^{(i)}$ are chosen at random, and the values $b_i = F(\boldsymbol{y}^{(i)})$ are computed and stacked into a vector $\boldsymbol{y} = (b_i)_i$. Finally, the coefficients $(F_\nu^{\ell_2})_{\nu \in \Gamma}$ are computed as the minimizer of the least squares problem

$$\min_F \|\boldsymbol{b} - \mathbf{A}F\|_2,$$

where $\mathbf{A}_{i,\nu} = T_\nu(\boldsymbol{y}^{(i)})$, with $1 \le i \le m$ and $\nu \in \Gamma$. To display our results on Fig. 3, we have an (implicit) enumeration $\pi : \{1, \dots, 12171\} \to \Gamma$ such that $|F_{\pi(1)}^{\ell_2}| \ge |F_{\pi(2)}^{\ell_2}| \ge \cdots \ge |F_{\pi(12171)}^{\ell_2}|$. For this experiment, we computed the solutions to the weighted ℓ_1-minimization problems using the SCP optimization procedure from the CVXPY package [21] with accuracy for the numerical optimization set to 10^{-6}. The downward triangles are the results using our suggested MLCSPG method with the multiplicative constant in Eq. (77) equal to 5 (red curve) and 15 (blue curve). The selection of the constant equal to 15 corresponds to $m_1 = 2258$ solves at the coarsest level, $m_2 = 968$ at the second, and to $m_3 = 394$ solves at the finest discretization level $L = 3$. $m_1 = 576$, $m_2 = 228$, and $m_3 = 73$ samples, for the red curve. The crosses correspond to the MC simulations, where we have used $m = 2.5 \cdot 10^7$ (red curve) and $m = 2.5 \cdot 10^9$ (blue curve) samples for the estimation of the Chebyshev coefficients. Noting that the y values of the graphs corresponding to the $\log_{10}$ of the magnitude of the coefficients, we see that the accuracy of the MC estimations is limited by the mean square convergence rate $m^{-1/2}$. The purple circles correspond to the ℓ_2 estimation described above (this corresponds to an oversampling ratio of 2, which is far below theoretical results). In this example, our approach (as illustrated by the downward triangle curves in Fig. 3) produces reliable approximations of gpc coefficients which are large in magnitude with a number of samples orders of magnitudes smaller than both the ℓ_2 and the MC approaches. The limitation of the MC method to a square root convergence rate requires a prohibitive number of samples for more complicated PDEs. It is important to note also that the accuracy of the recovered coefficients via our MLCSPG method are constrained by the accuracy of the numerical solver for the weighted ℓ_1 minimization. Finally, the yellow curve corresponds to the (negative, for illustrative purposes) total degree of the multi-index of the associated Chebyshev coefficient while the black curve corresponds to the (negative of the) maximum degree in the tensor product (23). It is interesting to notice that the magnitude seems to be smaller as the degree of the multi-index increases.

6.1.2. *Behavior in two spatial dimensions.* We consider now the two dimensional domain $D = (0,1)^2$ which we split into a uniform 4×4 grid with cells $D_{ij} = (\frac{i-1}{\sqrt{d}}, \frac{i}{\sqrt{d}}) \times (\frac{j-1}{\sqrt{d}}, \frac{j}{\sqrt{d}})$, for $1 \le i, j \le \sqrt{d}$, and $d = 4^2$. We assume the coefficient diffusion to be locally constant on each cell, with local variations $c_i = 2$, where the numbering used is shown in Table 2. The nominal (constant) field is considered to be $\bar{a} = 5$ as before.

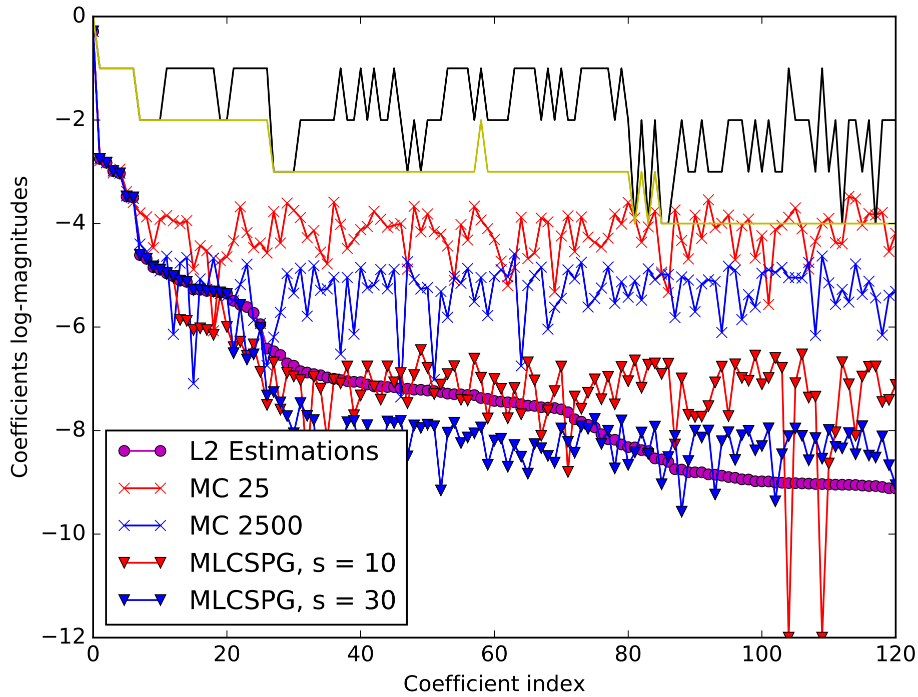

FIGURE 3. Estimations of the 120 largest coefficients (in magnitude) of the gpc of the piecewise constant diffusion problem (see text for details on the parameters) reordered by decreasing magnitude, according to their estimations via a least squares method. The least squares estimation computed 12171 coefficients from 24342 random samples; the downward blue triangles corresponds to a constant 15 in Eq. (77), while the downward red triangles correspond to the constant 5. The *MC* 25 curve corresponds to Monte Carlo estimations with $2.5 \cdot 10^7$ samples while the *MC* 2500 curve is based on $2.5 \cdot 10^9$ samples.

TABLE 2. Numbering of the parameters for the 2 dimensional problem

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

TABLE 3. Number of sampling points and cardinality of active set for continuous, piecewise linear FE approximation of the diffusion problem in two spatial dimensions. Uniform constant weights $v_j = 1.07$, $d = 16$.

$L - l =$	0	1	2
N_l	1977	18073	136373
m_l	304	785	1892
s_l	20	40	80

To compensate for the increase in parametric dimensionality, we now set the constant weights associated with the local variations to $v_i = 1.1$, for all $1 \leq i \leq d$. The multiplicative constant in (77) is set to 20. Fig. 4

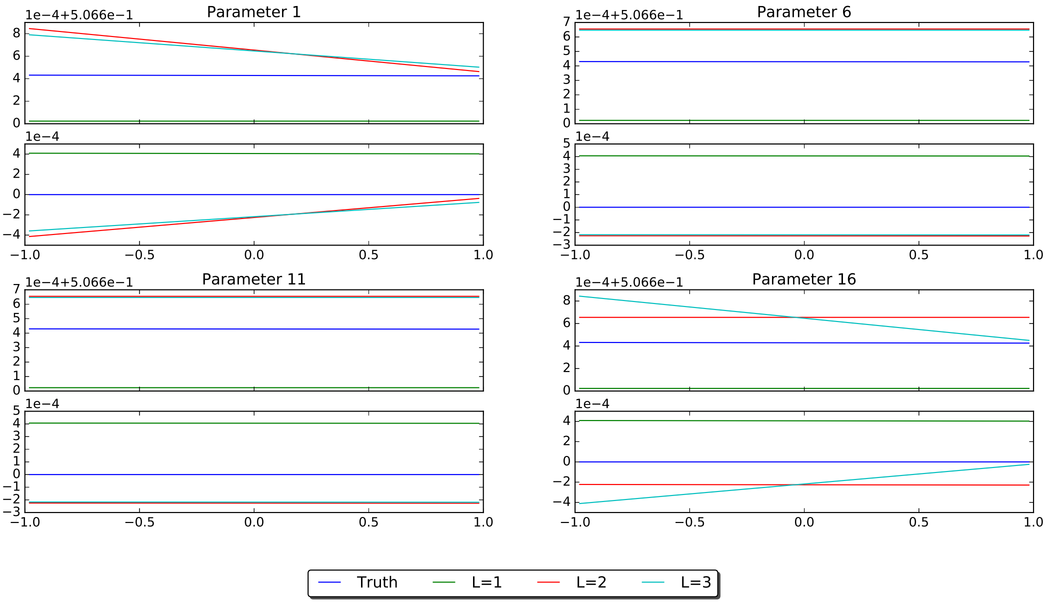

FIGURE 4. Convergence of the MLCSPG method in two spatial dimensions for the piecewise constant diffusion problem, $c_j = 2$, $1 \leq j \leq d = 16$. Each quadrant contains both the estimation $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j)$ (top graphs) and the pointwise errors $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j) - F(t\mathbf{e}_j)$ (lower graphs), for $-1 \leq t \leq 1$. We considered $j = 1$ (upper left quadrant), $j = 6$ (upper right), $j = 11$ (lower left), and $j = 16$ (lower right).

shows the pointwise convergence of the function $t \mapsto F(t\mathbf{e}_k)$ for k corresponding to the local variations along the diagonal cells of the 4×4 grid, where the parameters have been numbered according to Table 2.

6.2. **Behavior in higher prametric dimensions.** Next, we investigate the applicability of the proposed framework in the presence of high-dimensional expansions. We want to illustrate that 1) it is indeed possible to deal with fairly large dimensional problems and that 2) the global error scales as expected by Theorem 9. To this end, for an even number of parameters d, let us consider again a diffusion process similar to Eq. (3) where the ψ_j are chosen to be trigonometric polynomials:

$$a(x, \boldsymbol{y}) := \bar{a} + \sum_{j=1}^{d/2} y_{2j-1} \frac{\cos(j\pi x)}{j^\alpha} + y_{2j} \frac{\sin(j\pi x)}{j^\alpha}. \tag{78}$$

Fig. 5 shows the pointwise convergence of the first 4 cosine fluctuations when dealing with $d = 20$ parameters (i.e. 10 cosine and 10 sine components). To generate this figure, the mean field was taken constant in the spatial domain $\bar{a} = 4.3$ and the decay of the fluctuations is set to $\alpha = 2$. Once again we have chosen uniform weights $v_j = 1.1$ for all j (an example of polynomially growing weights is given in the next section). As in the previous scenario, the pointwise approximations as well as the pointwise errors are computed. The original discretization mesh width was set to $h_0 = 5 \cdot 10^{-4}$, and as we can see on the graphs, the final error is in the order of 10^{-3} (this error is larger than the estimated order of h_L due to the multiplicative constants entering in the calculations). In contrast to the previous case, as the number of levels increases, the pointwise error is steadily decreasing. This behavior is also noticeable when increasing the number of parameters to $d = 30$ (15 sines and 15 cosines in the expansion), as depicted in Fig.6.

6.3. **High-dimensions, polynomial weights, and target accuracy.** In the last subsection concerning numerical results, we look once again at the diffusion problem (3) with trigonometric expansions (78). The parameters related to the expansion are kept the same ($\bar{a} = 4.3$, $\alpha = 2$), and the weighted ℓ_1-minimization is

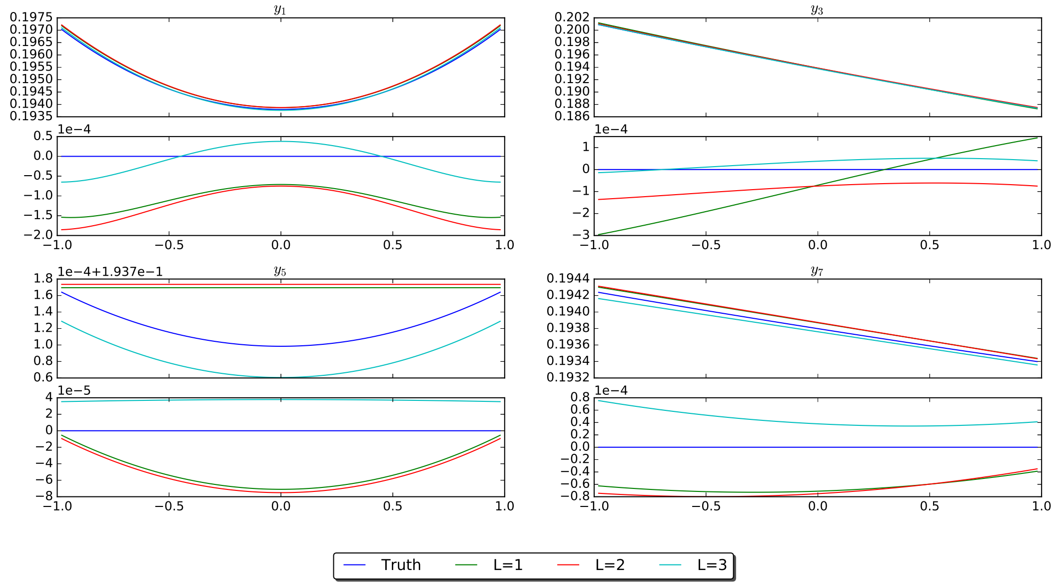

FIGURE 5. Estimations $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j), -1 \le t \le 1$, for $j \in \{1,3,5,7\}$ (i.e. corresponding to the first four cosine components, displayed repectively in the upper left, upper right, bottom left, and bottom right quadrants) via the MLCSPG method in *moderately high* dimensions, $d = 20$ parameters.

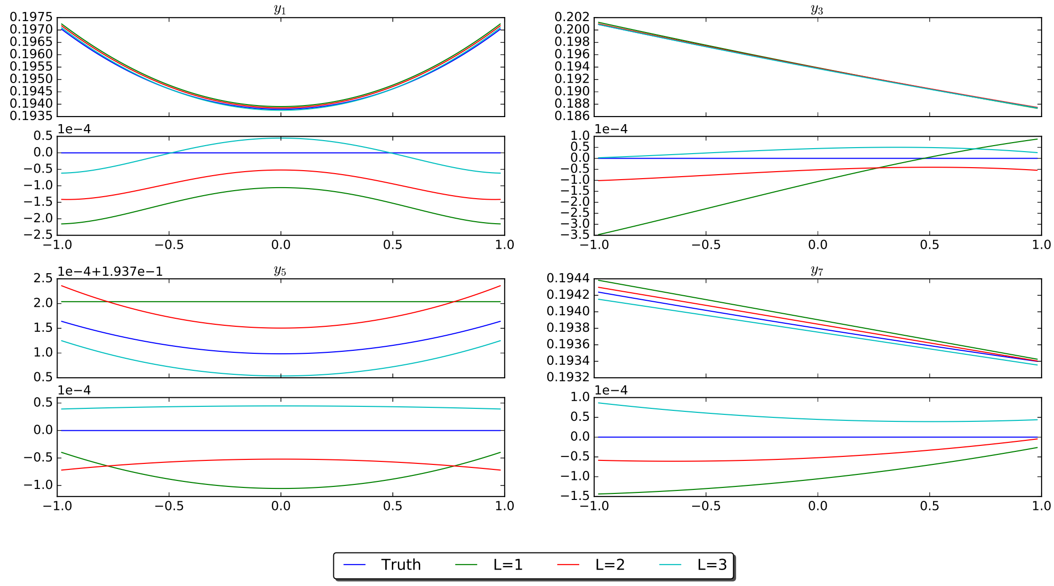

FIGURE 6. Estimations $t \mapsto F^L_{\mathrm{MLCS}}(t\mathbf{e}_j), -1 \le t \le 1$, and pointwise errors $t \mapsto$ for $j \in \{1,3,5,7\}$ (i.e. corresponding to the first four cosine components, displayed repectively in the upper left, upper right, bottom left, and bottom right quadrants) via the MLCSPG method in *moderately high* dimensions, $d = 30$ parameters.

TABLE 4. Approximation error of the MLCSPG method for the approximation of a diffusion problem with trigonometric expansion of the diffusion coefficients. See text for details on the parameters.

$L =$	1	2	3	4
L_E^1	$1.836 \cdot 10^{-3}$	$2.923 \cdot 10^{-4}$	$4.192 \cdot 10^{-4}$	$3.252 \cdot 10^{-5}$
L_E^2	$2.361 \cdot 10^{-3}$	$3.700 \cdot 10^{-4}$	$1.305 \cdot 10^{-4}$	$4.192 \cdot 10^{-5}$
L_E^∞	$9.707 \cdot 10^{-3}$	$1.523 \cdot 10^{-3}$	$6.647 \cdot 10^{-4}$	$2.995 \cdot 10^{-4}$
$L - l =$	0	1	2	3
N_l	211	802	2200	6200
m_l	172	428	986	2236
s_l	16	32	64	128

computed via CVXPY. We consider $d = 40$ parameters (20 sines and 20 cosines) and consider a multiplication constant $c = 1.02$. Finally, in Table 4 we report on the empirical errors obtained by testing the MLCSPG model against $N_{\text{test}} = 10000$ i.i.d. samples of the parameter vector $\boldsymbol{y}$, as well as the numbers of samples per level, and the sizes of the active sets Γ_l. Note that the mutiplicative constant for the value of s_l (see Eq. (77)) is set to 16.[3] It is important to notice that the total number of solves remains much smaller than the size of the active set, and yet the approach yields very accurate results. In this series of tests, the final target accuracy is $h_L = 5 \cdot 10^{-5}$ and we let the number of coarser levels vary from 1 to 4. The empirical errors are calculated as

$$
\begin{aligned}
L_E^1 &:= \frac{1}{N_{\text{test}}} \sum_{1 \le j \le N_{\text{test}}} |F(\boldsymbol{y}^{(j)}) - F^{\text{CSPG}}(\boldsymbol{y}^{(j)})| \, , \\
L_E^2 &:= \sqrt{\frac{1}{N_{\text{test}}} \sum_{1 \le j \le N_{\text{test}}} |F(\boldsymbol{y}^{(j)}) - F^{\text{CSPG}}(\boldsymbol{y}^{(j)})|^2} \, , \\
L_E^\infty &:= \max_{1 \le j \le N_{\text{test}}} |F(\boldsymbol{y}^{(j)}) - F^{\text{CSPG}}(\boldsymbol{y}^{(j)})| \, .
\end{aligned}
$$

It is interesting to point out that, even though the same goal accuracy has been provided in all the cases, adding the contribution from coarser levels helps in reducing the overall error.

The pointwise convergence for the levels $L = 2, 3, 4$ are plotted in Fig. 7. In opposition to the graphs in the previous sections, this one shows the result for a prescribed target accuracy $h_L = 5 \cdot 10^{-4}$ and not for a fixed starting discretization level. Here again, we can see that increasing the number of levels actually improves the accuracy. This is partially due to the fact that the multi-level approach allows to adjust CS estimations of coefficients that may have been inaccurately estimated at a coarser level.

7. CONCLUSIONS

For a class of abstract, affine-parametric, linear operator equations depending on sequences $\boldsymbol{y}$ of parameters, we have introduced a multi-level generalization of the CS approach from [46] to efficiently scan the high-dimensional parameter space. For the approximate solution of (instances of) the parametric operator equations, we stipulated available inf-sup stable, Petrov-Galerkin ("PG" for short) discretizations of the "nominal" operator $A_0 = A(\mathbf{0})$; in particular, (48) holds. The small perturbation hypothesis (43) at $t = 0$ implies uniform (w.r.t. $\boldsymbol{y} \in U$) inf-sup stability (8) of the PG discretization (Thm. 5). Admissible PG discretizations comprise, in particular, all classical primal or mixed Finite Element Methods (FEM for short), as well as spectral and collocation methods for elliptic and certain linear, parabolic evolution equations. Throughout, we used multi-level Finite Element Galerkin discretizations in $D \subset \mathbb{R}^n$ with isotropic mesh refinements, responsible for the $\mathcal{O}(2^{nl})$ scaling in the proof of Proposition 4. Anisotropic, "sparse-grid"

[3]For the sake of compactness, we have chosen not to insist on this multiplicative constant. The numerical results seem to be reliable and stable within a range between 10 and 30.

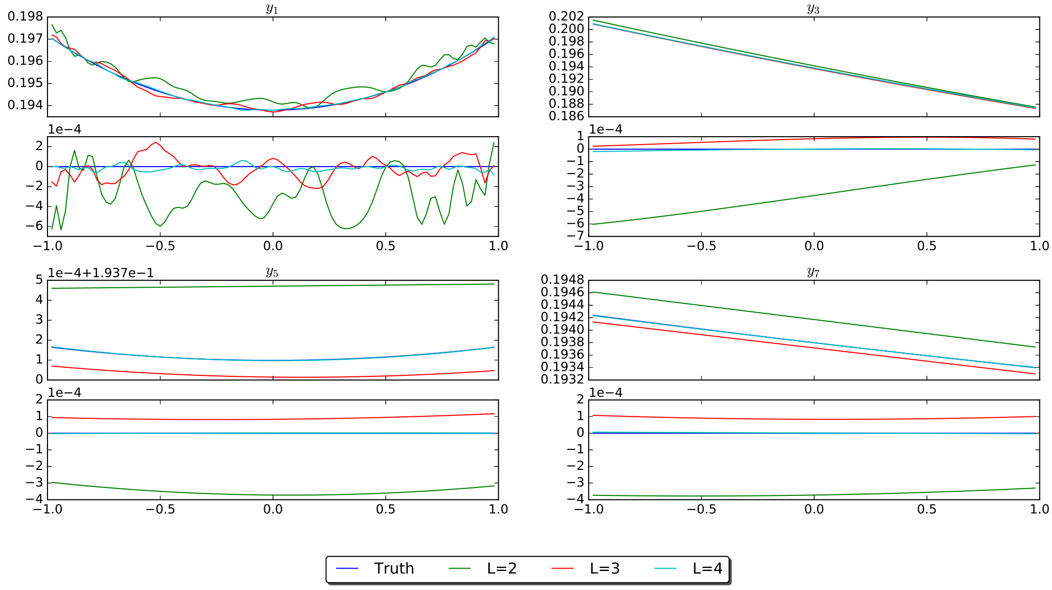

FIGURE 7. Estimations of $t \mapsto F^L_{\text{MLCS}}(t\mathbf{e}_j), -1 \le t \le 1$ (upper graph) and the errors $t \mapsto F^L_{\text{MLCS}}(t\mathbf{e}_j) - F(t\mathbf{e}_j)$, for $j \in \{1, 3, 5, 7\}$ (i.e. corresponding to the first four cosine components) via the MLCSPG method in *moderately high* dimensions, $d = 40$ parameters, using polynomial weights.

discretizations of the parametric problems in D would result, with analogous analysis, in so-called "multi-index" compressed sensing PG methods, analogous to multi-index MC in [36], with $\mathcal{O}(l^{n-1}2^l)$ in place of $\mathcal{O}(2^{nl})$. We analyzed error vs. work of the multi-level extension of the combined, CS-PG algorithm and showed that it affords improved, as compared to the single-level variant from [46, 9], error vs. work bounds with convergence rates that are independent of the dimension of the space parameters which are active in the approximation, while being *"nonintrusive"*, i.e. accessing an available solver at each discretization level. This is analogous to what is known from multi-level Monte-Carlo ("MLMC" for short) sampling methods, as surveyed e.g. in [33]. Contrary to MLMC methods whose convergence rate is limited by the (mean-square) rate 1/2 afforded by MC methods, and the recently proposed sparse-grid methods in [14] which rely on a particular ("downward closed") structure of the sets of active polynomials, however, the presently proposed approach yields dimension-independent convergence rates (potentially far beyond 1/2) in the sup-norm with respect to the parameters, exploiting any sparsity in the gpc coefficient sequence of the parametric solutions, *without* strong, a-priori structural assumptions on the active polynomial degrees. At the same time, the MLCSPG approach is nonintrusive and intrinsically parallel as MLMC methods. If a-priori information on the structure of sets of active indices (such as "downward closedness") is available, corresponding accelerations of the SLCS approach have recently been investigated in [15]. This is afforded by adopting Chebyshev gpc expansions which are orthonormal with respect to a probability measure which underlies the CS method, whereas sparse-grid methods as in [14] afford greater flexibiliy as regards the choice of gpc system.

We remark that although here only affine-parametric operator equations were considered, the key results of the present paper require merely *sparsity of Chebyshev gpc expansions* (as expressed, e.g., in summability of sequences of $\mathcal{X}_t$-norms of gpc expansion coefficients in the conditions (40) - (42), rather than the weaker summability of $\mathcal{X}$-norms in the SLCSPG considered in [46]) of the parametric solutions, and some (possibly crude) bounds of these coefficients which enter the weight sequence ω, and a family of uniformly inf-sup stable PG discretization methods. Such results are available for rather general, holomorphic-parametric, nonlinear operator equations in [16]. In case that the ψ_j in (2), (3) have supports which are localized to subdomains of D with controlled overlap, higher summability for the Chebyshev gpc expansion coefficients holds; we refer to [3] for details. The presently proposed MLCSPG algorithms are able to exploit better summability of Chebyshev gpc expansion coefficients without any modification in the algorithm.

References

[1] B. Adcock. Infinite-dimensional ℓ^1 minimization and function approximation from pointwise data. arXiv preprint arXiv:1503.02352, March 2015.

[2] M. S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The FEniCS project version 1.5. *Archive of Numerical Software*, 3(100), 2015.

[3] M. Bachmayr, A. Cohen, and G. Migliorati. Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients. *ESAIM Math. Model. Numer. Anal.*, 2016.

[4] A. Barth, C. Schwab, and N. Zollinger. Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, 2011.

[5] R. Bellman. *Dynamic programming*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2010. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus.

[6] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Anal.*, 43(3):1457–1472, 2011.

[7] T. Blumensath and M. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Trans. Information Theory*, 55(4):1872–1882, 2009.

[8] D. Boffi, F. Brezzi, and M. Fortin. *Mixed finite element methods and applications*. Springer, 2013.

[9] J.-L. Bouchot, B. Bykowski, H. Rauhut, and C. Schwab. Compressed sensing Petrov-Galerkin approximations for parametric PDEs. In *International Conference on Sampling Theory and Applications (SampTA)*, 2015.

[10] J.-L. Bouchot, S. Foucart, and P. Hitczenko. Hard thresholding pursuit algorithms: number of iterations. *Appl. Comput. Harmon. Anal.*, 41(2):412–435, 2016.

[11] J. Bourgain. An improved estimate in the restricted isometry problem. In *Geometric Aspects of Functional Analysis*, pages 65–70. Springer, 2014.

[12] A. Buffa, Y. Maday, A. Patera, C. Prud'homme, and G. Turinici. *A priori* convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM Math. Model. Numer. Anal.*, 46(3):595–603, 2012.

[13] B. Bykowski. Weighted ℓ_1 methods for high-dimensional parametric operator equations in the context of uncertainty. Master's thesis (Chair C for Mathematics, RWTH Aachen), 2015.

[14] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Foundations of Computational Mathematics*, 14(4):601–633, 2014.

[15] A. Chkifa, N. Dexter, H. Tran, and C. G. Webster. Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. arXiv preprint arXiv:1602.05823, February 2016.

[16] A. Cohen, A. Chkifa, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric pdes. *Journ. Math. Pures et Appliquees*, 103(2):400–428, 2015.

[17] A. Cohen and R. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numerica*, 2015.

[18] 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*, 10(6):615–646, 2010.

[19] A. Cohen, R. Devore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Analysis and Applications*, 9(01):11–47, 2011.

[20] P. J. Davis. *Interpolation and Approximation*. Blaisdell Pub, 1963.

[21] S. Diamond and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[22] J. Dick, F. Y. Kuo, Q. T. Le Gia, D. Nuyens, and C. Schwab. Higher order QMC Petrov-Galerkin discretization for affine parametric operator equations with random field inputs. *SIAM Journal on Numerical Analysis*, 54(4):2541–2568, 2016.

[23] J. Dick, F. Y. Kuo, Q. T. LeGia, and C. Schwab. Multi-level higher order QMC Galerkin discretization for affine parametric operator equations. *SIAM J. Numer. Anal.*, 54(4):2541–2568, 2016.

[24] A. Doostan and H. Owhadi. A non-adapted sparse approximation of PDEs with stochastic inputs. *Journal of Computational Physics*, 230(8):3015–3034, 2011.

[25] M. Eigel, C. Gittelson, C. Schwab, and E. Zander. Adaptive stochastic Galerkin FEM. *Comp. Meth. Appl. Mech. Engg.*, 270:247–269, 2014.

[26] M. Eigel, C. J. Gittelson, C. Schwab, and E. Zander. A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes. *ESAIM: M2AN*, 49(5):1367–1398, 2015.

[27] Y. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inform. Theory*, 55(11):5302–5316, 2009.

[28] Y. Eldar and H. Rauhut. Average case analysis of multichannel sparse recovery using convex relaxation. *IEEE Trans. Inform. Theory*, 56(1):505–519, 2010.

[29] J. Fell and H. Rauhut. Sparse approximation from random Fourier samples of wavelet expansions via weighted iterative hard thresholding. in preparation, 2015.

[30] M. Fornasier and H. Rauhut. Recovery algorithms for vector-valued data with joint sparsity constraints. *SIAM Journal on Numerical Analysis*, 46(2):577–613, 2008.

[31] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563, 2011.
[32] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Springer, 2013.
[33] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.
[34] C. Gittelson. Adaptive wavelet methods for elliptic partial differential equations with random operators. *Numerische Mathematik*, 126:471–513, 2013.
[35] W. Hackbusch. *Multigrid methods and applications*, volume 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1985.
[36] A.-L. Haji-Ali, F. Nobile, and R. Tempone. Multi-index Monte Carlo: when sparsity meets sampling. *Numer. Math.*, 132(4):767–806, 2016.
[37] M. Hansen and C. Schwab. Analytic regularity and nonlinear approximation of a class of parametric semilinear elliptic PDEs. *Mathematische Nachrichten*, 286(8–9):832–860, 2013.
[38] I. Haviv and O. Regev. The restricted isometry property of subsampled fourier matrices. arXiv preprint arXiv:1507.01768, July 2015.
[39] S. Heinrich. Monte carlo complexity of global solution of integral equations. *Journal of Complexity*, 14(2):151–175, 1998.
[40] J. Jo. Iterative hard thresholding for weighted sparse approximation. arXiv preprint arXiv:1312.3582, 2013.
[41] F. Y. Kuo, C. Schwab, and I. H. Sloan. Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients. *SIAM Journal on Numerical Analysis*, 50(6):3351–3374, 2012.
[42] A. Logg, K.-A. Mardal, G. N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method.* Springer, 2012.
[43] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Analysis of discrete l^2 projection on polynomial spaces with random evaluations. *Foundations of Computational Mathematics*, 14(3):419–456, 2014.
[44] S. Mishra, C. Schwab, and J. Sukys. Multi-level monte carlo finite volume methods for uncertainty quantification of acoustic wave propagation in random heterogeneous layered medium. *Journal of Computational Physics*, 312:192–217, 2016.
[45] J. Peng, J. Hampton, and A. Doostan. A weighted ℓ_1-minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics*, 267:92–111, 2014.
[46] H. Rauhut and C. Schwab. Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations. *Math. Comp.*, 86(304):661–700, 2017. Report 2014-14, Seminar for Applied Mathematics, ETH Zürich.
[47] H. Rauhut and R. Ward. Interpolation via weighted ℓ_1 minimization. *Appl. Comput. Harmonic Anal.*, 40(2):321–351, 2016.
[48] T. J. Rivlin. *An introduction to the approximation of functions*. Blaisdell Publishing, 1969.
[49] C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6), 2013.
[50] C. Schillings and C. Schwab. Sparsity in Bayesian inversion of parametric operator equations. *Inverse Problems*, 30(6), 2014.
[51] C. Schwab and R. A. Todor. Karhunen–Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, 2006.
[52] G. Tang and G. Iaccarino. Subsampled Gauss quadrature nodes for estimating polynomial chaos expansions. *SIAM/ASA J. Uncertain. Quantif.*, 2(1):423–443, 2014.
[53] R. Ward. Compressed sensing with cross validation. *Information Theory, IEEE Transactions on*, 55(12):5773–5782, 2009.
[54] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Rev.*, 34(4):581–613, 1992.
[55] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using ℓ_1-minimization. *International Journal for Uncertainty Quantification*, 2(3), 2012.

Chair for Mathematics C (Analysis), RWTH Aachen University, Germany
E-mail address: bouchot@mathc.rwth-aachen.de

Chair for Mathematics C (Analysis), RWTH Aachen University, Germany
E-mail address: rauhut@mathc.rwth-aachen.de

Seminar for Applied Mathematics, ETH Zürich, Switzerland
E-mail address: christoph.schwab@sam.math.ethz.ch