Numer. Math. (2016) 134:343–388
DOI 10.1007/s00211-015-0773-y

Numerische Mathematik

Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs

F. Nobile[1] · **L. Tamellini**[1,2] · **R. Tempone**[3]

Received: 6 March 2014 / Revised: 2 September 2015 / Published online: 30 October 2015

Abstract In this work we provide a convergence analysis for the quasi-optimal version of the sparse-grids stochastic collocation method we presented in a previous work: "On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods" (Beck et al., Math Models Methods Appl Sci 22(09), 2012). The construction of a sparse grid is recast into a knapsack problem: a profit is assigned to each hierarchical surplus and only the most profitable ones are added to the sparse grid. The convergence rate of the sparse grid approx-

The authors would like to recognize the support of King Abdullah University of Science and Technology (KAUST) AEA project "Predictability and Uncertainty Quantification for Models of Porous Media" and University of Texas at Austin AEA Rnd 3 "Uncertainty quantification for predictive mobdeling of the dissolution of porous and fractured media". F. Nobile and L. Tamellini have been partially supported by the Italian grant FIRB-IDEAS (Project n. RBID08223Z) "Advanced numerical techniques for uncertainty quantification in engineering and life science problems" and by the Swiss National Science Foundation under the Project No. 140574 "Efficient numerical methods for flow and transport phenomena in heterogeneous random porous media". They also received partial support from the Center for ADvanced MOdeling Science (CADMOS). R. Tempone is a member of the KAUST SRI Center for Uncertainty Quantification in Computational Science and Engineering. We acknowledge the usage of the Matlab® functions `patterson_rule.m` by J. Burkardt (http://people.sc.fsu.edu/~jburkardt/m_src/patterson_rule/patterson_rule.html) for the computation of Gauss–Patterson points and `lejapoints.m` by M. Caliari (http://profs.sci.univr.it/~caliari/software/lejapoints.m) for the computation of symmetrized Leja points.

✉ L. Tamellini
lorenzo.tamellini@epfl.ch; lorenzo.tamellini@unipv.it

1 CSQI-MATHICSE, École Polytechnique Fédérale Lausanne, Station 8, 1015 Lausanne, Switzerland

2 Dipartimento di Matematica, Università di Pavia, via Ferrata 1, 27100 Pavia, Italy

3 Applied Mathematics and Computational Science, 4700, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Kingdom of Saudi Arabia

imation error with respect to the number of points in the grid is then shown to depend on weighted summability properties of the sequence of profits. This is a very general argument that can be applied to sparse grids built with any univariate family of points, both nested and non-nested. As an example, we apply such quasi-optimal sparse grids to the solution of a particular elliptic PDE with stochastic diffusion coefficients, namely the "inclusions problem": we detail the convergence estimates obtained in this case using polynomial interpolation on either nested (Clenshaw–Curtis) or non-nested (Gauss–Legendre) abscissas, verify their sharpness numerically, and compare the performance of the resulting quasi-optimal grids with a few alternative sparse-grid construction schemes recently proposed in the literature.

Mathematics Subject Classification 41A10 · 65C20 · 65N12 · 65N30

1 Introduction

Sparse-grid polynomial approximation has recently emerged as one of the most appealing methods for approximating high-dimensional functions. Although it is as simple to use as a sampling strategy, it converges significantly faster if the function at hand presents some degree of differentiability. A number of "off-the-shelf" sparse-grid packages can be found on the web,[1] enhancing the spread of this technique among practitioners.

Yet, the sparse-grid technique is subject to a dramatic deterioration in performance as the number of random variables increases (i.e., it suffers from the so-called "curse of dimensionality effect", to which Monte Carlo sampling methods are instead essentially immune). To avoid or at least alleviate this undesirable feature, a number of approaches have recently been proposed. Among others, we mention the anisotropic sparse-grid technique [4,31] and the a posteriori adaptive strategy investigated in [10,13,19,21,26]; here we investigate further the quasi-optimal sparse-grid method proposed in [6]. Such method, here applied to elliptic PDEs with random coefficients, consists in reformulating the problem of the construction of a sparse grid as a knapsack problem as was first proposed in [10,19,21] (see [29] for a thorough analysis of the knapsack problem), and estimating the profit of each sparse-grid component (*hierarchical surplus*) using combined a priori/a posteriori information (i.e., providing a priori estimates that can be made quite sharp by a suitable numerical fitting procedure). The goal of this work is to present a convergence result for such "knapsack" sparse grids in terms of some weighted τ-summability of the profits: in particular, we extend and improve the preliminary estimates presented in [36]. Our result is general and can accommodate both the case of nested and non-nested abscissas. We mention that two other related works from 2013, [34] and [13], have addressed alternative convergence estimates for knapsack-type sparse grids. Also, see [31] for an older estimate of the convergence of Smolyak-type anisotropic sparse-grid approximations.

[1] See e.g. http://www.ians.uni-stuttgart.de/spinterp or http://dakota.sandia.gov.

As a specific application, we consider an elliptic PDE with random coefficients, namely the so-called "inclusions problem" already discussed in [4,8], whose solution falls in the class of analytic functions in polyellipses (see e.g. [3,8,14]). We will derive an estimate of the profits of the hierarchical surpluses for this family of functions and prove that such profits satisfy suitable weighted summability properties, thus allowing us to derive rigorous convergence results for the corresponding quasi-optimal sparse-grid approximation: in particular, we will show that it converges sub-exponentially with a rate comparable to that of the optimal ("best M-terms") L^2 approximation in the case of nested points and with half the rate for non-nested points, cf. [8, Theorem 16]. Finally we verify the sharpness of the estimates numerically, using Clenshaw–Curtis and Gauss–Legendre points as specific representatives of the two families of nested and non-nested points, and we compare the performances of the quasi-optimal sparse grids with those of a few other sparse-grid schemes recently proposed in the literature. Some numerical results obtained for the same test case by using sparse grids built with Leja points are presented in [30].

The rest of the work is organized as follows. Section 2 defines the general approximation problem and introduces the sparse-grid methodology. The construction of quasi-optimal sparse grids is explained in Sect. 3, where we also briefly explain the motivation for referring to the proposed methodology as "quasi-optimal"; the general convergence result in terms of the weighted τ-summability of the profits is then given in Sect. 4, see Theorem 1. Section 5 introduces the above-mentioned class of polyellipse-analytic Hilbert-space-valued functions and builds on the previous general theorem to derive rigorous convergence estimates for their quasi-optimal sparse-grid approximation with nested and non-nested collocation points, see Theorems 2 and 3. In particular, the theorems are stated at the beginning of the section, and the rest of the section is devoted to their proof. Section 6 discusses possible choices of univariate interpolation points that can be used to build sparse grids, namely Clenshaw–Curtis, Gauss–Legendre, Leja, and Gauss–Patterson points, while Sect. 7 introduces the "inclusion problem" and shows some numerical results that confirm the effectiveness of the proposed quasi-optimal strategy and the sharpness of the proposed convergence estimates. Finally, conclusions and final remarks are presented in Sect. 8.

2 Sparse-grid polynomial approximation of Hilbert-space-valued functions

We consider the problem of approximating a multivariate function $u(\mathbf{y}) : \Gamma \to V$, where Γ is an N-variate hypercube $\Gamma = \Gamma_1 \times \Gamma_2 \times \cdots \times \Gamma_N$ (with $\Gamma_n \subseteq \mathbb{R}$ and N possibly infinite) and V is a Hilbert space. Furthermore, we assume that each Γ_n is endowed with a probability measure $\varrho_n(y_n)dy_n$, so that $\varrho(\mathbf{y})d\mathbf{y} = \prod_{n=1}^{N} \varrho_n(y_n)$ is a probability measure on Γ, and we restrict our attention to functions in the Bochner space $L^2_\varrho(\Gamma; V)$, where

$$L^2_\varrho(\Gamma; V) = \left\{ u : \Gamma \to V \text{ s.t. } \int_\Gamma \|u(\mathbf{y})\|_V^2 \, \varrho(\mathbf{y})d\mathbf{y} < \infty \right\}.$$

Observe that, since V and $L^2_\varrho(\Gamma)$ are Hilbert spaces, $L^2_\varrho(\Gamma; V)$ can be equivalently understood as the tensor space $V \otimes L^2_\varrho(\Gamma)$, defined as the completion of formal sums $v = \sum_{k=1}^{k'} \phi_k \psi_k$, with $\phi_k \in V$ and $\psi_k \in L^2_\varrho(\Gamma)$, with respect to the inner product

$$(v, \widehat{v})_{V \otimes L^2_\varrho(\Gamma)} = \sum_{k,\ell} (\phi_k, \widehat{\phi}_\ell)_V \, (\psi_k, \widehat{\psi}_\ell)_{L^2_\varrho(\Gamma)}.$$

In particular, we aim at approximating $u(\mathbf{y})$ with global polynomials over Γ, which is a sound approach if u is a smooth V-valued function of $\mathbf{y}$. To introduce the polynomial subspace of $V \otimes L^2_\varrho(\Gamma)$ in which we will build our approximate solution, it is convenient to use a multi-index notation.[2] Let $w \in \mathbb{N}$ be an integer index indicating the level of approximation, and $\Lambda(w)$ a sequence of index sets in $\mathbb{N}^N$ such that $\Lambda(0) = \{\mathbf{0}\}$, $\Lambda(w) \subseteq \Lambda(w+1)$ for $w \geq 0$, and $\mathbb{N}^N = \bigcup_{w \in \mathbb{N}} \Lambda(w)$. Denoting by $\mathbb{P}_{\Lambda(w)}(\Gamma)$ the multivariate polynomial space

$$\mathbb{P}_{\Lambda(w)}(\Gamma) = span \left\{ \prod_{n=1}^{N} y_n^{p_n}, \quad \mathbf{p} \in \Lambda(w) \right\},$$

we will look for an approximation

$$u_w \in V \otimes \mathbb{P}_{\Lambda(w)}(\Gamma) = \left\{ \sum_j v_j q_j(\mathbf{y}), \quad v_j \in V, \quad q_j \in \mathbb{P}_{\Lambda(w)}(\Gamma) \right\}.$$

Clearly, the polynomial space $\mathbb{P}_{\Lambda(w)}(\Gamma)$ should be designed to have good approximation properties while keeping the number of degrees of freedom as low as possible. Although this is a problem-dependent choice, using the classical Tensor Product polynomial space $\mathbb{P}_{TP(w)}(\Gamma)$, with $TP(w) = \{\mathbf{q} \in \mathbb{N}^N : \max_n q_n \leq w\}$, is in general not a good choice, as its dimension grows exponentially fast with the number of random variables N (i.e., $\dim \mathbb{P}_{TP(w)}(\Gamma) = (1+w)^N$). Valid alternative choices that have been widely used in literature are: the Total Degree polynomial space $\mathbb{P}_{TD(w)}(\Gamma)$, $TD(w) = \{\mathbf{q} \in \mathbb{N}^N : \sum_n q_n \leq w\}$, see e.g. [20,27], that contains indeed only $\binom{N+w}{N}$ monomials but has approximation properties similar to the Tensor Product space; or the Hyperbolic Cross polynomial space $\mathbb{P}_{HC(w)}(\Gamma)$, $HC(w) = \{\mathbf{q} \in \mathbb{N}^N : \prod_n (q_n+1) \leq w+1\}$, see e.g. [1,28,35]. One could also introduce anisotropy in the approximation to enrich the polynomial space only in those variables y_n that contribute the most to the total variability of the solution, see e.g. [4]. Several methods can be used to compute the polynomial approximation u_w (projection, interpolation, regression): in this work we consider the sparse-grid approximation method, which we briefly review in the rest of this section.

[2] Throughout the rest of this work, $\mathbb{N}$ will denote the set of integer numbers including 0, and $\mathbb{N}_+$ that of integer numbers excluding 0. Moreover, $\mathbf{0}$ will denote the vector $(0, 0, \ldots, 0) \in \mathbb{N}^N$, $\mathbf{1}$ the vector $(1, 1, \ldots, 1) \in \mathbb{N}^N$, and $\mathbf{e}_j$ the j-th canonical vector in $\mathbb{R}^N$, i.e. a vector whose components are all zero but the j-th, whose value is one. Finally, given two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{N}^N$, $\mathbf{v} \leq \mathbf{w}$ if and only if $v_j \leq w_j$ for every $1 \leq j \leq N$.

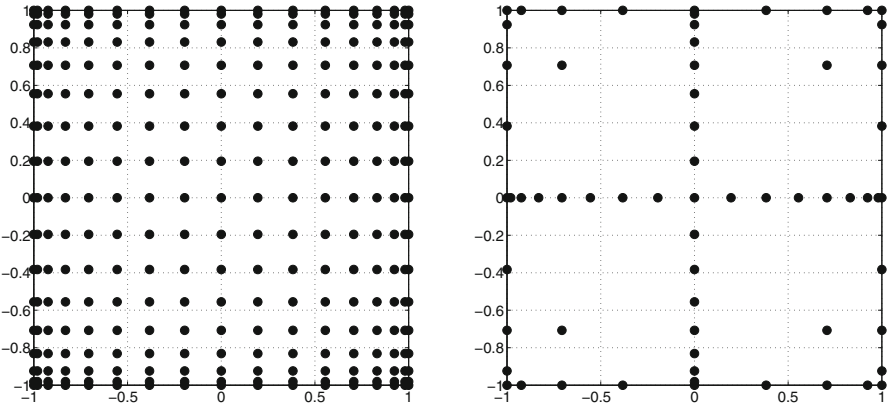

Fig. 1 *Left* tensor grid. *Right* sparse grid

2.1 The sparse-grid approximation method

For a given level of approximation $w \geq 0$, the sparse-grid approximation method (see e.g. [5,10] and references therein) consists in evaluating the function u in a set of W points $\mathbf{y}_1, \ldots, \mathbf{y}_W \in \Gamma$ and building a global polynomial approximation u_w (not necessarily interpolatory) in a suitable space $V \otimes \mathbb{P}_{\Lambda(w)}(\Gamma)$.

For each direction y_n, we introduce a sequence of one-dimensional polynomial Lagrangian interpolant operators of increasing degree, indexed by $i_n \geq 1$:

$$\forall\, i_n \geq 1, \quad \mathcal{U}_n^{m(i_n)} : C^0(\Gamma_n) \to \mathbb{P}_{m(i_n)-1}(\Gamma_n),$$

where $m(i_n)$ is the number of collocation points used to build the interpolant at level i_n and $\mathbb{P}_q(\Gamma_n)$ is the set of polynomials in y_n of degree at most q. We require the level-to-nodes function $m : \mathbb{N} \to \mathbb{N}$ to satisfy the following assumptions:

$$m(0) = 0, \quad m(1) = 1, \quad m(i_n) < m(i_n + 1), \quad i_n \geq 1.$$

In addition, let $\mathcal{U}_n^0[f] = 0,\ \forall f \in C^0(\Gamma_n)$. Next, we introduce the difference operators $\Delta_n^{m(i_n)} = \mathcal{U}_n^{m(i_n)} - \mathcal{U}_n^{m(i_n-1)}$ and consider a sequence of index sets $\mathcal{I}(w) \subset \mathbb{N}_+^N$, such that $\mathcal{I}(w) \subset \mathcal{I}(w+1), \mathcal{I}(0) = \{\mathbf{1}\}$, and $\bigcup_{w\in\mathbb{N}} \mathcal{I}(w) = \mathbb{N}_+^N$. We define the sparse-grid approximation of $u : \Gamma \to V$ at level w as

$$u_w(\mathbf{y}) = \mathcal{S}^m_{\mathcal{I}(w)}[u](\mathbf{y}) = \sum_{\mathbf{i}\in\mathcal{I}(w)} \bigotimes_{n=1}^{N} \Delta_n^{m(i_n)}[u](\mathbf{y})\,. \tag{1}$$

As pointed out in [19], it is desirable that the sum (1) has some telescopic properties. To ensure this, we have to impose some additional constraints on $\mathcal{I}$. Following [19] we say that a set $\mathcal{I}$ is *admissible*[3] if

[3] Also known as *lower sets* or *downward closed set*, see e.g. [15].

$$\forall \mathbf{i} \in \mathcal{I}, \quad \mathbf{i} - \mathbf{e}_j \in \mathcal{I} \quad \text{for} \quad 1 \le j \le N, \quad \text{such that} \quad i_j > 1. \tag{2}$$

We refer to this property as the *admissibility condition*, or *ADM* for short. Given a multi-index set $\mathcal{I}$, we will denote by $\mathcal{I}^{ADM}$ the smallest admissible set such that $\mathcal{I} \subset \mathcal{I}^{ADM}$. The set of all evaluation points needed by (1) is called a *sparse grid* (see Fig. 1) and we denote its cardinality by $W_{\mathcal{I}(w),m}$. Note that (1) is indeed equivalent to a linear combination of tensor grid interpolations, each of which uses only "few" interpolation points (see e.g. [42]):

$$\mathcal{S}^m_{\mathcal{I}(w)}[u](\mathbf{y}) = \sum_{\mathbf{i} \in \mathcal{I}(w)^{ADM}} c_{\mathbf{i}} \bigotimes_{n=1}^{N} \mathcal{U}_n^{m(i_n)}[u](\mathbf{y}), \qquad c_{\mathbf{i}} = \sum_{\substack{\mathbf{j} \in \{0,1\}^N \\ (\mathbf{i}+\mathbf{j}) \in \mathcal{I}(w)}} (-1)^{|\mathbf{j}|}. \tag{3}$$

Observe that many of the coefficients $c_{\mathbf{i}}$ in (3) may be zero: in particular, if $\mathcal{I}$ is admissible then $c_{\mathbf{i}}$ is zero whenever $\mathbf{i} + \mathbf{1} \in \mathcal{I}(w)$. For any sparse grid, one can associate a corresponding quadrature formula $\mathcal{Q}^m_{\mathcal{I}(w)}[\cdot]$,

$$\forall f \in C^0(\Gamma), \quad \int_\Gamma f(\mathbf{y})\varrho(\mathbf{y})d\mathbf{y} \approx \int_\Gamma \mathcal{S}^m_{\mathcal{I}(w)}[f](\mathbf{y})\varrho(\mathbf{y})d\mathbf{y} = \mathcal{Q}^m_{\mathcal{I}(w)}[f]$$
$$= \sum_{j=1}^{W^m_{\mathcal{I}(w)}} f(\mathbf{y}_j)\varpi_j,$$

for suitable weights $\varpi_j \in \mathbb{R}$. In particular, given $g \in V'$, where V' is the dual space of V, the expected value of $\langle g, u \rangle$ can be approximated as

$$\mathbb{E}[\langle g, u \rangle] \approx \mathcal{Q}^m_{\mathcal{I}(w)}[\langle g, u \rangle] = \sum_{j=1}^{W^m_{\mathcal{I}(w)}} \langle g, u(\mathbf{y}_j) \rangle \varpi_j.$$

The sequence of sets $\mathcal{I}(w)$, the level-to-nodes function m, and the family of collocation points to be used at each level characterize the sparse-grid operator $\mathcal{S}^m_{\mathcal{I}(w)}$ introduced in (1). The choice of $\mathcal{I}(w)$ will be the subject of the next section; anisotropic sets $\mathcal{I}(w)$ that enrich the approximation in specific directions of the parameter space Γ have been studied in [4,31]. As for the family of points, they should be chosen according to the probability measure $\varrho(\mathbf{y})d\mathbf{y}$ on Γ for optimal performance, e.g. the Gauss–Legendre points for the uniform measure, and the Gauss–Hermite points for the Gaussian measure, see e.g. [40]. For good uniform approximations on $\Gamma = [-1, 1]^N$, Clenshaw–Curtis points are also a good choice. In the following, when the set of points used to build the operator $\mathcal{U}_n^{m(i_n)}$ is a subset of the points of the operator $\mathcal{U}_n^{m(i_n+1)}$ it will be referred to as nested, and all other points will be referred to as *non-nested*. Nested quadrature formulae are well known for having a lower degree of exactness than Gaussian quadrature formulae when approximating integrals of functions of one variable; however, the accuracy of Clenshaw–Curtis points is similar to that of Gauss–Legendre points (cf. e.g. [39]), and nestedness allows for significant savings in the

construction of sparse grids. This distinction will also play a central role in the following sections.

Finally, we point out (see also [4]) that given any polynomial space $\mathbb{P}_\Lambda(\Gamma)$, one can always find a sparse grid that delivers approximations in that space simply by taking $m(i) = i$ and $\mathcal{I} = \{\mathbf{i} \in \mathbb{N}_+^N : \mathbf{i} - \mathbf{1} \in \Lambda\}$. Conversely, given a sparse-grid approximation $\mathcal{S}^m_{\mathcal{I}(w)}$, the underlying polynomial space is $\mathbb{P}_\Lambda(\Gamma)$ with $\Lambda = \{\mathbf{q} \in \mathbb{N}^N : \mathbf{q} \leq m(\mathbf{i}) - \mathbf{1} \text{ for some } \mathbf{i} \in \mathcal{I}\}$. Moreover, the sparse-grid approximation is exact on such $\mathbb{P}_\Lambda(\Gamma)$.

3 Construction of quasi-optimal sparse grids

We now summarize the procedure for the construction of quasi-optimal sparse grids, using the approach introduced in our previous work [6] (see also [10,19,21]); a discussion on the meaning of the expression "quasi-optimal" can be found at the end of this section. We begin by introducing the concept of work (or computational cost) associated with the construction of a given sparse-grid approximation of u. To this end, we make the following Assumption that will hold throughout this paper:

Assumption A0 *The work needed to construct a sparse-grid approximation is proportional to the total number of points used, $W_{\mathcal{I}(w),m}$.*

In other words, we assume that the computation of the points of the sparse grid itself is negligible and that evaluating the target function u at each point of the sparse grid has the same cost (in the example we will consider in Sect. 7, this means that the cost of solving the PDE associated to any value of the random parameters is always the same). For notational convenience, we will use the same symbol, $W_{\mathcal{I}(w),m}$, to denote both the work of the sparse-grid approximation of u and the cardinality of the sparse grid. Next, for each multi-index $\mathbf{i} \in \mathcal{I}(w)$, we introduce the *hierarchical surplus* operator

$$\Delta^{m(\mathbf{i})} = \bigotimes_{n=1}^{N} \Delta^{m(i_n)},$$

such that the sparse-grid approximation (1) can actually be seen as a sum of hierarchical surplus operators applied to u. The quasi-optimal sparse grid relies on the concept of *profit* of a hierarchical surplus: to this end, we associate an *error contribution* and a *work contribution* with each hierarchical surplus, i.e. the contribution to the total error (respectively cost) that can be ascribed to a specific hierarchical surplus composing a sparse grid.

We thus begin by introducing the quantity

$$\delta E(\mathbf{i}) = \left\| \Delta^{m(\mathbf{i})}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \tag{4}$$

associated to the hierarchical surplus $\Delta^{m(\mathbf{i})}$; observe that for any multi-index set $\mathcal{J}$ such that $\mathbf{i} \notin \mathcal{J}$ and $\mathcal{J}$, $\{\mathcal{J} \cup \mathbf{i}\}$ are both admissible according to definition (2), we have

$$\left(u - \mathcal{S}^m_{\{\mathcal{J}\cup\mathbf{i}\}}[u]\right) - \left(u - \mathcal{S}^m_{\mathcal{J}}[u]\right) = \mathcal{S}^m_{\{\mathcal{J}\cup\mathbf{i}\}}[u] - \mathcal{S}^m_{\mathcal{J}}[u] = \Delta^{m(\mathbf{i})}[u],$$

such that $\delta E(\mathbf{i})$ can be considered as a good indicator of the error reduction due to the addition of $\Delta^{m(\mathbf{i})}$ to any sparse-grid approximation of u (in other words, $\delta E(\mathbf{i})$ is independent of $\mathcal{J}$). We can then naturally define the *error contribution* as any upper bound $\Delta E(\mathbf{i})$ for $\delta E(\mathbf{i})$,

$$\delta E(\mathbf{i}) \le \Delta E(\mathbf{i}),$$

and observe that the total error satisfies

$$\begin{aligned}\left\|u - \mathcal{S}^m_{\mathcal{I}}[u]\right\|_{V\otimes L^2_\varrho(\Gamma)} &= \left\|\sum_{\mathbf{i}\notin\mathcal{I}} \Delta^{m(\mathbf{i})}[u]\right\|_{V\otimes L^2_\varrho(\Gamma)} \\ &\le \sum_{\mathbf{i}\notin\mathcal{I}} \left\|\Delta^{m(\mathbf{i})}[u]\right\|_{V\otimes L^2_\varrho(\Gamma)} \le \sum_{\mathbf{i}\notin\mathcal{I}} \Delta E(\mathbf{i}). \end{aligned} \tag{5}$$

Defining a work contribution $\Delta W(\mathbf{i})$ is a more delicate issue. One could define the quantity $\delta W(\mathbf{i}) = W_{\{\mathcal{J}\cup\mathbf{i}\},m} - W_{\mathcal{J},m}$ as the work contribution of the hierarchical surplus $\Delta^{m(\mathbf{i})}$; however, in general $\delta W(\mathbf{i})$ depends on the starting set $\mathcal{J}$ unless nested points are used (see Example 1 next).

In the case of nested points, upon defining

$$d(i_n) = m(i_n) - m(i_n - 1), \tag{6}$$

we can safely define an "exact" *work contribution* $\Delta W(\mathbf{i})$ as

$$\Delta W(\mathbf{i}) = \prod_{n=1}^{N} d(i_n) \tag{7}$$

decomposing the total work as $W_{\mathcal{I}(w),m} = \sum_{\mathbf{i}\in\mathcal{I}(w)} \Delta W(\mathbf{i})$.

On the other hand, for non-nested points, we consider the following definition of the work contribution:

$$\Delta W(\mathbf{i}) = \prod_{n=1}^{N} m(i_n), \tag{8}$$

i.e., the cost of the tensor grid associated to $\mathbf{i}$, so that $W_{\mathcal{I}(w),m} \le \sum_{\mathbf{i}\in\mathcal{I}(w)} \Delta W(\mathbf{i})$. This work contribution estimate is reasonable if one builds the (non-nested) sparse grid "incrementally", i.e. starting from $\mathcal{I} = \{\mathbf{1}\}$, adding one multi-index $\mathbf{i} \in \mathbb{N}_+^N$ to $\mathcal{I}$ at a time, and immediately evaluating the function u on the corresponding tensor grid $\bigotimes_{n=1}^N \mathcal{U}_n^{m(i_n)}$, cf. Eq. (3). By doing this, one does not exploit the fact that many tensor grids in the final formula (3) are multiplied by zero coefficients and therefore, $W_{\mathcal{I}(w),m} \le \sum_{\mathbf{i}\in\mathcal{I}(w)} \Delta W(\mathbf{i})$.

Example 1 To show that $\delta W(\mathbf{i})$ is not uniquely defined when non-nested points are used, we take as an example the case of sparse grids built over $\Gamma = [-1, 1]^2$ using Gauss–Legendre points. We set $\mathbf{i} = (1, 5)$ and consider the multi-index sets $\mathcal{J}_1$ and $\mathcal{J}_2$, shown in the top-left and top-right plots of Fig. 2. We then consider four different sparse grids: $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$, built using the multi-index sets $\mathcal{J}_1$, $\{\mathcal{J}_1 \cup \mathbf{i}\}$, $\mathcal{J}_2$, and $\{\mathcal{J}_2 \cup \mathbf{i}\}$, respectively.

Comparing $\mathcal{S}_1$ and $\mathcal{S}_2$ (bottom-left plot of Fig. 2), we can see that adding $\mathbf{i}$ to $\mathcal{J}_1$ results in a sparse grid with 9 additional points (i.e., $\delta W(\mathbf{i}) = 9$). Conversely, the comparison of $\mathcal{S}_3$ and $\mathcal{S}_4$ (bottom-right plot of Fig. 2) shows that adding $\mathbf{i}$ to $\mathcal{J}_2$ does not change the number of points of the sparse grid (i.e., $\delta W(\mathbf{i}) = 0$), because 4 new points are added but 4 points are no longer present.

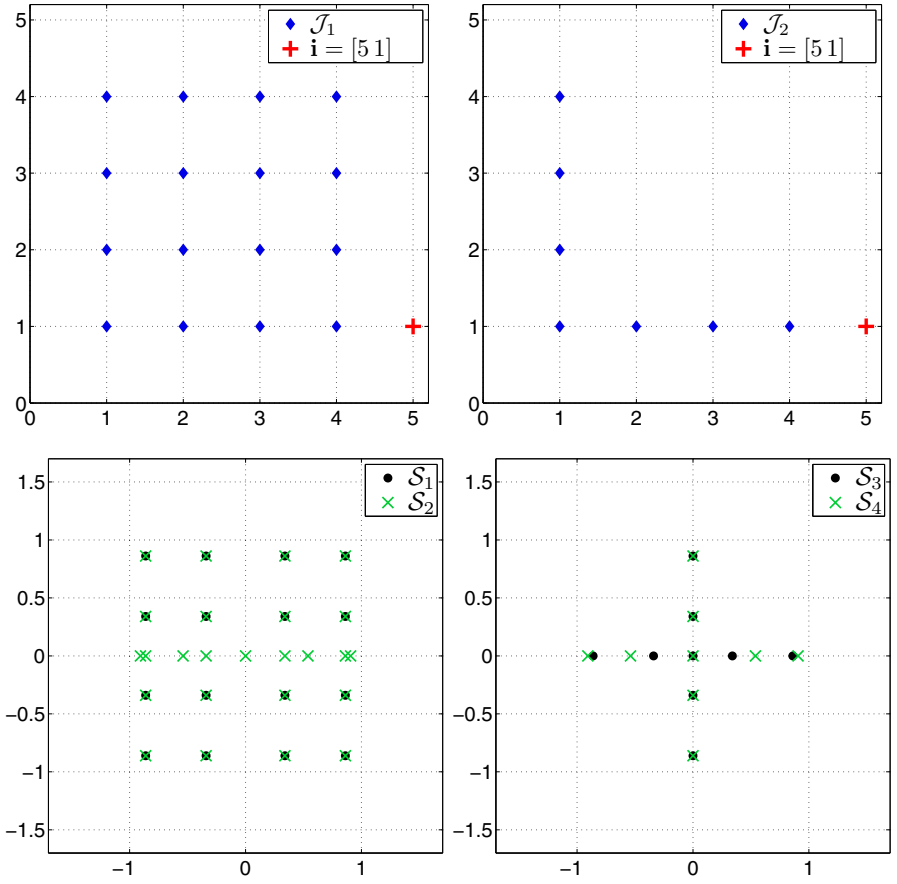

Fig. 2 The *top row* shows the two different multi-index sets considered in Example 1, i.e., $\mathcal{J}_1$ (*left*) and $\mathcal{J}_2$ (*right*), as well as the multi-index $\mathbf{i}$ (*red cross*). The *bottom row* shows the resulting sparse grids $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$, corresponding to the sets $\mathcal{J}_1$, $\{\mathcal{J}_1 \cup \mathbf{i}\}$, $\mathcal{J}_2$, and $\{\mathcal{J}_2 \cup \mathbf{i}\}$. The sparse grid $\mathcal{S}_2$ has 9 points more than $\mathcal{S}_1$ (*left*), while $\mathcal{S}_3$ and $\mathcal{S}_4$ have the same number of points (*right*) (color figure online)

Next, we introduce the *estimated profit* of a hierarchical surplus,

$$P(\mathbf{i}) = \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})}, \tag{9}$$

and the sequence of decreasingly ordered profits, $\{P_j^{ord}\}_{j\in\mathbb{N}_+}$

$$P_j^{ord} \geq P_{j+1}^{ord}.$$

It is also convenient to introduce a function that assigns the corresponding multi-index to the j-th ordered profit: we will denote such function as $\mathbf{i}(j)$, i.e. $P_j^{ord} = P(\mathbf{i}(j))$. Incidentally, note that as soon as two hierarchical surpluses have equal estimated profits, the map $\mathbf{i}(j)$ is not unique: in this case, any criterion to select a specific sequence can be used.

We can now define a quasi-optimal sparse-grid approximation at level w as the sparse grid including in sum (1) only the set of w hierarchical surpluses with the highest profit, in the spirit of a knapsack problem (cf. [29]; we will return on this aspect at the end of this section), possibly made admissible according to condition (2):

$$\mathcal{I}(w) = \mathcal{J}(w)^{ADM}, \quad \mathcal{J}(w) = \{\mathbf{i}(1), \mathbf{i}(2), \ldots, \mathbf{i}(w)\}. \tag{10}$$

Equivalently, one can automatically enforce the admissibility of the set $\mathcal{I}(\omega)$ by introducing the auxiliary profits (see also [12])

$$P^*(\mathbf{i}) = \max_{\mathbf{j}\geq\mathbf{i}} P(\mathbf{j}), \tag{11}$$

considering the decreasingly ordered sequence $\{P_j^{*,ord}\}_{j\in\mathbb{N}_+}$

$$P_j^{*,ord} \geq P_{j+1}^{*,ord}, \tag{12}$$

and the new ordering $\mathbf{i}^*(j)$, such that $P_j^{*,ord} = P^*(\mathbf{i}^*(j))$ so that (10) can be rewritten as

$$\mathcal{I}(w) = \{\mathbf{i}^*(1), \mathbf{i}^*(2), \ldots, \mathbf{i}^*(w)\}. \tag{13}$$

Definition 1 A sequence of profits $\{P(\mathbf{i})\}_{\mathbf{i}\in\mathbb{N}_+^N}$ is *monotone* if

$$\forall \quad \mathbf{i}, \mathbf{j} \in \mathbb{N}_+^N, \quad \mathbf{i} \leq \mathbf{j} \Rightarrow P(\mathbf{i}) \geq P(\mathbf{j}).$$

Notice that since P^* is a monotone sequence by construction, set (13) will always be admissible.

The idea of constructing a sparse grid based on the profit of each hierarchical surplus (or other suitable "optimality indicators") has been first proposed in a series of works [10,19,21], see also [26]. In particular, a possible approach could be an adaptive "greedy-type" algorithm in which the set $\mathcal{I}$ is constructed iteratively: given a set $\mathcal{I}^{(k)}$

at the k-th iteration, one looks at the "neighborhood" (or "margin") $\mathcal{M}^{(k)}$ of $\mathcal{I}^{(k)}$ and adds to the set $\mathcal{I}^{(k)}$ the most profitable hierarchical surplus in $\mathcal{M}^{(k)}$,

$$\mathcal{I}^{(k+1)} = \mathcal{I}^{(k)} \cup \{\mathbf{i}\}, \quad \mathbf{i} = \underset{\mathbf{j}\in\mathcal{M}^{(k)}}{\operatorname{argmax}} P^*(\mathbf{j}),$$

see e.g [19,26] for a similar algorithm based however on optimality indicators other than profits. Clearly, this methodology implicitly assumes some kind of decay of the profits, or, equivalently, that the next most profitable multi-index always belongs to the margin of the current set $\mathcal{I}^{(k)}$. Moreover, the exploration of the margin $\mathcal{M}^{(k)}$ can be expensive in high dimensions. Therefore, in the context of elliptic PDEs with random coefficients, in [6] we proposed adding hierarchical surpluses based on a priori error and work contribution estimates with numerically tuned constants (hybrid "a priori"/"a posteriori" estimates) that we observed to be quite sharp and thus effective in reducing the cost of the construction of the sparse grid. Note that an analogous fully "a priori" approach has been considered in [10,21] in the context of a wavelet-type approximation of high-dimensional deterministic PDEs.

We close this section by discussing the meaning of the expression "quasi-optimal"; analogous considerations can also be found in [10, Section 3.2].

We begin by noting that the proposed sparse grids construction algorithm (10)-(13) is the same as the Dantzig algorithm for the resolution of the linear programming relaxation (cf. [29]) of the binary knapsack problem. In more details, the binary knapsack problem consists in determining the set of objects that maximizes the total revenue (in our case the sum of the error contributions of the optimal set, cf. Eq. (5)) under a constraint on the available capacity (in our case the maximum work allowed for the construction of the sparse grid), i.e.

$$\begin{aligned} &\max \sum_{\mathbf{i}\in\mathbb{N}_+^N} \Delta E(\mathbf{i}) x_{\mathbf{i}} \\ &\text{s.t.} \sum_{\mathbf{i}\in\mathbb{N}_+^N} \Delta W(\mathbf{i}) x_{\mathbf{i}} \le W_{max}, \\ &\quad x_{\mathbf{i}} \in \{0,1\}, \end{aligned}$$

and the optimal index set $\mathcal{I}$ is given by $\mathcal{I} = \{\mathbf{i} \in \mathbb{N}_+^N : x_{\mathbf{i}} = 1\}$. The *linear programming relaxation* of such a problem allows us to include instead fractions of objects in the selected set, i.e. $x_{\mathbf{i}} \in [0,1]$, and the Dantzig solution of such a relaxed problem consists in ordering the objects in a decreasing manner, according to the revenue/volume ratio (i.e., the profits $P(\mathbf{i})$), and picking them in the resulting order. Note that only the last object included in the selection can possibly be taken not entirely (i.e., with $x_{\mathbf{i}} < 1$), whereas all previous objects are taken entirely (i.e. with $x_{\mathbf{i}} = 1$).

It can be shown that the Dantzig solution is optimal for the relaxed problem, see e.g. [29] or [24, Lemma 2.1], and that the optimal value obtained is at most twice as large as that of the initial binary problem, see again [29] (hence, loosely speaking, "quasi-optimal"). Moreover, if the Dantzig solution is integer (i.e., no fractions of objects enter the selection), then it is also the optimal solution of the initial binary

knapsack problem. In the context of the sparse-grid approximation, this amounts to say that if one does not have to meet an actual constraint on the maximum number of function evaluations (i.e. W_{max} is actually arbitrary), the sequence of sparse grids generated by the proposed method can be claimed as optimal in the "knapsack sense".

However, to state the construction of a sparse grid as a knapsack problem, we had to express the sparse-grid error in terms of a sum of error contributions and, because the error contributions are not orthogonal in general, we had to do so by using the triangular inequality, cf. again Eq. (5), which is "non-optimal".[4] Also the fact that the set of the w hierarchical surpluses may be non-admissible could introduce some additional work, i.e., some degree of "non-optimality" (although by redefining the profits as in Eq. (11) the algorithm still applies). Finally, as will be clearer in the following sections, we will use in practice *estimated profits* with numerical tuning rather than exact profits (which may not even be well defined in the case of non-nested quadrature points, cf. Example 1). For all these reasons we say that the construction of the sparse grids based on estimated profits is "quasi-optimal".

4 Convergence estimate for the quasi-optimal sparse-grid approximation

We now state and prove a convergence result of the quasi-optimal sparse-grid approximation built according to (13); see [36] for earlier versions and [13,34] for an alternative estimate derived in the context of elliptic PDEs with random diffusion coefficients. We first recall a technical result, the so-called Stechkin Lemma, see e.g. [14, eq. (3.13)] for a proof.

Lemma 1 (Stechkin) *Let $0 \leq p \leq q$ and let $\{a_j\}_{j\in\mathbb{N}_+}$ be a positive decreasing sequence. Then, for any $M > 0$*

$$\left(\sum_{j>M} (a_j)^q\right)^{1/q} \leq M^{-\frac{1}{p}+\frac{1}{q}} \left(\sum_{j\in\mathbb{N}_+} (a_j)^p\right)^{1/p}.$$

Theorem 1 (Convergence estimate of the quasi-optimal sparse grid) *If the auxiliary profits* (11) *satisfy the weighted summability condition*

$$\left(\sum_{\mathbf{i}\in\mathbb{N}_+^N} P^*(\mathbf{i})^\tau \Delta W(\mathbf{i})\right)^{1/\tau} = C_P(\tau) < \infty \tag{14}$$

for some $0 < \tau \leq 1$, then

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V\otimes L^2_\varrho(\Gamma)} \leq W^{1-1/\tau}_{\mathcal{I}(w),m} C_P(\tau). \tag{15}$$

[4] As opposed to the Parseval identity, which is an equality and therefore an "optimal decomposition" in the case of orthogonal hierarchical surpluses.

where $\mathcal{I}(w)$ is given by (13), and ΔW is given by (7) for grids with nested points, and by (8) for grids with non-nested points.

Proof We start by introducing the following auxiliary sequences:

- $\{\Delta W_j\}_{j\in\mathbb{N}_+}$ is the sequence of work contributions arranged using the same order as the sequence of the auxiliary profits (12). Note that this sequence will not be ordered in general.
- $\{Q_j\}_{j\in\mathbb{N}_+}$ is the sequence of the sum of the first j work contributions, i.e.

$$Q_0 = 0,\ Q_j = \sum_{k=1}^{j} \Delta W_k\,.$$

- $\{\Delta E_j\}_{j\in\mathbb{N}_+}$ is the sequence of error contributions arranged using the same order as the sequence of the auxiliary profits (12). Again, this sequence will not be ordered in general.
- $\{\Delta\widetilde{E}_k\}_{k\in\mathbb{N}_+}$ is a modification of the error contributions sequence $\{\Delta E_j\}_{j\in\mathbb{N}_+}$, where each ΔE_j is repeated the same number of times as the corresponding work contribution. More precisely

$$\{\Delta\widetilde{E}_k\}_{k\in\mathbb{N}_+} = \left\{ \underbrace{\Delta E_1,\ \Delta E_1, \ldots}_{\Delta W_1\ times}, \underbrace{\Delta E_2,\ \Delta E_2, \ldots}_{\Delta W_2\ times}, \ \ldots \right\},$$

  i.e. $\Delta\widetilde{E}_{Q_{j-1}+s} = \Delta E_j$, for $s = 1, \ldots, \Delta W_j$.
- $\{\widetilde{P}_k\}_{k\in\mathbb{N}_+}$ is the analogously modified sequence of auxiliary profits,

$$\{\widetilde{P}_k\}_{k\in\mathbb{N}_+} = \left\{ \underbrace{\frac{\Delta E_1}{\Delta W_1}, \frac{\Delta E_1}{\Delta W_1}, \ldots}_{\Delta W_1\ times}, \underbrace{\frac{\Delta E_2}{\Delta W_2}, \frac{\Delta E_2}{\Delta W_2}, \ldots}_{\Delta W_2\ times}, \ \ldots \right\} \tag{16}$$

  i.e. $\widetilde{P}_{Q_{j-1}+s} = P_j^{*,ord}$, for $s = 1, \ldots, \Delta W_j$.

Using error decomposition (5) and the fact that $|\mathcal{I}(w)| = w$, cf. Eq. (13), we get

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V\otimes L^2_\varrho(\Gamma)} \leq \sum_{j>w} \Delta E_j.$$

Next, we recast the previous sum of error contributions in terms of the auxiliary sequence $\widetilde{P}_k$ in (16), i.e.

$$\sum_{j>w} \Delta E_j = \sum_{j>w} \sum_{s=1}^{\Delta W_j} \frac{\Delta\widetilde{E}_{Q_{j-1}+s}}{\Delta W_j} = \sum_{k>Q_w} \widetilde{P}_k\,.$$

We now apply Lemma 1 with $q = 1$ and $p = \tau$ to obtain

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq \sum_{k > Q_w} \widetilde{P}_k \leq Q_w^{-1/\tau+1} \left(\sum_{k>0} \widetilde{P}_k^\tau \right)^{1/\tau};$$

observe that

$$\left(\sum_{k>0} \widetilde{P}_k^\tau \right)^{1/\tau} = \left(\sum_{j>0} \left(P_j^{*,ord} \right)^\tau \Delta W_j \right)^{1/\tau} = \left(\sum_{\mathbf{i} \in \mathbb{N}_+^N} P^*(\mathbf{i})^\tau \Delta W(\mathbf{i}) \right)^{1/\tau}$$

and that

$$Q_w = \sum_{k=1}^{w} \Delta W_k \geq W_{\mathcal{I}(w),m},$$

due to the definitions of $\Delta W(\mathbf{i})$ in (7) and (8). Then,

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq Q_w^{-1/\tau+1} \left(\sum_{\mathbf{i} \in \mathbb{N}_+^N} P^*(\mathbf{i})^\tau \Delta W(\mathbf{i}) \right)^{1/\tau} \leq C_P(\tau) W_{\mathcal{I}(w),m}^{1-1/\tau},$$

which concludes the proof. □

Remark 1 An alternative approach would consist in sorting the hierarchical surpluses according to error contribution estimates $\Delta E(\mathbf{i})$ rather than according to profits. Following the lines of the theorem above, one could derive the following convergence estimate for the resulting sparse grid:

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq w^{1-1/\tau} C_E(\tau), \quad C_E(\tau) = \left(\sum_{\mathbf{i} \in \mathbb{N}_+^N} \Delta E(\mathbf{i})^\tau \right)^{1/\tau}.$$

However, recasting this estimate in terms of computational cost to make it comparable with (15) would require further assumptions on the shape of the optimal set $\mathcal{I}(w)$.

5 Quasi-optimal sparse-grid approximation of analytic functions on polyellipses

In this section, we apply convergence Theorem 1 to a particular class of Hilbert-space-valued functions, which contains in particular the solution of some linear elliptic

equations with random diffusion coefficients, as will be shown in Sect. 7. More precisely, given any $\zeta_1, \zeta_1, \ldots, \zeta_N > 1$, we introduce the Bernstein polyellipse

$$\mathcal{E}_{\zeta_1,\ldots,\zeta_N} = \prod_{n=1}^{N} \mathcal{E}_{n,\zeta_n},$$

where $\mathcal{E}_{n,\zeta_n}$ denotes the univariate Bernstein ellipse

$$\mathcal{E}_{n,\zeta_n} = \left\{ z_n \in \mathbb{C} : \quad \mathfrak{Re}\,(z) \leq \frac{\zeta_n + \zeta_n^{-1}}{2} \cos\phi, \right.$$
$$\left. \mathfrak{Im}\,(z) \leq \frac{\zeta_n - \zeta_n^{-1}}{2} \sin\phi, \quad \phi \in [0, 2\pi) \right\},$$

and consider the class of functions $u : \Gamma = [-1, 1]^N \to V$ satisfying the following assumption:

Assumption A1 ("Polyellipse analyticity") *There exist $\zeta_1^*, \zeta_2^*, \ldots, \zeta_N^* > 1$ such that the function $u : \Gamma \to V$, $\Gamma = [-1, 1]^N$, admits a complex continuation $u^* : \mathbb{C}^N \to V$ which is a V-valued holomorphic function in the Bernstein polyellipse $\mathcal{E}_{\zeta_1,\ldots,\zeta_N}$ for every $1 < \zeta_n < \zeta_n^*$, $n = 1, 2, \ldots, N$, with $\sup_{\mathbf{z}\in\mathcal{E}_{\zeta_1,\ldots,\zeta_N}} \|u^*(\mathbf{z})\|_V \leq B_u$ and $B_u = B_u(\zeta_1, \zeta_2, \ldots, \zeta_N) \to \infty$ as $\zeta_n \to \zeta_n^*$, $n = 1, \ldots, N$.*

As we mentioned in the Introduction, sparse polynomial approximations are particularly suitable for this kind on functions, see e.g. [8]. We now state the convergence results for the quasi-optimal sparse-grid approximation of functions satisfying Assumption A1, and devote the rest of this section to their proof. For this, we need to introduce a few more definitions and assumptions concerning the sequences of univariate collocation points.

Definition 2 For a given family of collocation points, let $\mathbb{M}_n^{m(i_n)}$ be the norm of the interpolation operator $\mathcal{U}_n^{m(i_n)} : C^0(\Gamma_n) \to L^2_\varrho(\Gamma_n)$,

$$\mathbb{M}_n^{m(i_n)} = \sup_{\|f\|_{L^\infty(\Gamma_n)}=1} \left\| \mathcal{U}_n^{m(i_n)}[f] \right\|_{L^2_\varrho(\Gamma_n)},$$

and let

$$\overline{\mathbb{M}}_n^{m(i_n)} = \max_{j=1,\ldots,i_n} \mathbb{M}^{m(j)}.$$

Assumption A2 *For nested collocation points, there exists a constant $C_{\mathbb{M}} > 0$ such that*

$$\frac{\overline{\mathbb{M}}_n^{m(i_n)}}{d(i_n)} \leq C_{\mathbb{M}}, \quad \forall i_n \in \mathbb{N}_+,$$

with $d(\cdot)$ as in Eq. (6), while for non-nested collocation points, there exists a constant $C_{\mathbb{M}} > 0$ such that

$$\frac{\overline{\mathbb{M}}_n^{m(i_n)}}{m(i_n)} \leq C_{\mathbb{M}}, \quad \forall i_n \in \mathbb{N}_+.$$

Assumption A3 *There exists a constant $C_m > 0$ such that there holds*

$$\frac{d(i_n+1)}{d(i_n)} \leq C_m, \quad \forall i_n \in \mathbb{N}_+.$$

Theorem 2 (Convergence of nested quasi-optimal sparse grids) *Let u be a function satisfying Assumption A1. If the collocation points used are nested and satisfy Assumptions A2 and A3, then the quasi-optimal sparse grid approximation built according to the profits*

$$P(\mathbf{i}) = C_E \prod_{n=1}^{N} \frac{e^{-g_n m(i_n-1)} \overline{\mathbb{M}}_n^{m(i_n)}}{d(i_n)}, \quad g_n < \log \zeta_n^*, \tag{17}$$

where C_E is a constant depending exponentially on N that will be specified in Lemma 5, converges as

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq \inf_{0<\tau<1} \mathcal{C}(N,\tau) W^{1-1/\tau}_{\mathcal{I}(w),m},$$

where

$$\mathcal{C}(N,\tau) = \left(C_E^\tau (C^\tau_{\mathbb{M}} \widehat{C}_m)^N \prod_{n=1}^{N} \frac{e^{\tau g_n}}{e^{\tau g_n} - 1} \right)^{1/\tau},$$
$$\widehat{C}_m = \max\{1, C_m\}.$$

Moreover, letting $g_m = \sqrt[N]{\prod_{n=1}^N g_n}$ denote the geometric mean of $g_1, \ldots, g_N$, and assuming without loss of generality that $g_1 \leq g_2 \leq \cdots \leq g_N$, there exist some constants α_L, β_L, and C_{log} that will be specified in Lemmas 3 and 4 such that

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq C_E C^N_{\mathbb{M}} \exp\left(\left(C_{log} - \frac{g_m \delta}{\widehat{C}_m e} \right) N \sqrt[N]{W_{\mathcal{I}(w),m}} \right), \tag{18}$$

which holds for $0 < \delta < 1 - \frac{1}{e}$ and for

$$W_{\mathcal{I}(w),m} > \left(\frac{g_N e \widehat{C}_m}{g_m(\alpha_L - \delta \beta_L)} \right)^N. \tag{19}$$

Theorem 3 (Convergence of non-nested quasi-optimal sparse grids) *Let u be a function satisfying Assumption A1. If the collocation points used are non-nested and satisfy*

Assumptions A2 and A3, then the quasi-optimal sparse grids approximation built according to

$$P(\mathbf{i}) = C_E \prod_{n=1}^{N} \frac{e^{-g_n m(i_n-1)} \overline{\mathbb{M}}_n^{m(i_n)}}{m(i_n)}, \quad g_n < \log \zeta_n^*, \tag{20}$$

where C_E a constant depending exponentially on N that will be specified in Lemma 5, converges as

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq \inf_{0<\tau<1} \mathcal{C}(N,\tau) W^{1-1/\tau}_{\mathcal{I}(w),m},$$

where

$$\mathcal{C}(N,\tau) = \left((C_E C_{\mathbb{M}}^N)^\tau \prod_{n=1}^{N} \left(\widehat{C}_m \frac{e^{\tau g_n}}{e^{\tau g_n} - 1} + \frac{2}{\tau g_n e} \frac{e^{\tau g_n/2}}{e^{\tau g_n/2} - 1} \right) \right)^{1/\tau},$$
$$\widehat{C}_m = \max\{1, C_m\}.$$

Moreover, letting $g_m = \sqrt[N]{\prod_{n=1}^N g_n}$ denote the geometric mean of $g_1, \ldots, g_N$, and assuming without loss of generality that $g_1 \leq g_2 \leq \cdots \leq g_N$, there exist some constants α_L, β_L, and C_{log} that will be specified in Lemmas 3 and 4 such that

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq C_E C_{\mathbb{M}}^N \exp\left(\left(2C_{log} - \mathcal{K} g_m \right) N \sqrt[2N]{W_{\mathcal{I}(w),m}} \right), \tag{21}$$

with $\mathcal{K} = (\delta + 1 - \log 2)/(2\sqrt{e})$ which holds for $0 < \delta < 1 - \frac{1}{e}$ and for

$$W_{\mathcal{I}(w),m} > \max\left\{ \left(\frac{2\sqrt{e} g_n}{g_m(\alpha_L - \delta\beta_L)} \right)^{2N}, \left(\frac{\widehat{C}_m e^{3/2} g_n}{2 g_m} \right)^{2N} \right\}. \tag{22}$$

Remark 2 Assumption A2 can actually be weakened. Indeed, the previous theorems would still hold even in the case of a polynomial growth of the ratio $\overline{\mathbb{M}}_n^{m(i_n)}/m(i_n)$, i.e. when $C_{\mathbb{M}}$ is actually a function, $C_{\mathbb{M}} = C_{\mathbb{M},n}(i_n)$, at the price of changing the rates g_n in the statements of the theorems with $g_n^b = g_n(1-\epsilon)$ for every $\epsilon > 0$, such that $C_{\mathbb{M}}(i_n) e^{-g_n m(i_n-1)} \leq C_{\mathbb{M},\epsilon} e^{-g_n^b m(i_n-1)}$ for every $i_n \in \mathbb{N}_+$, where $C_{\mathbb{M},\epsilon}$ approaches infinity as $\epsilon \to 0$.

Convergence estimates (18) and (21) that we just provided share the same structure of the result obtained for the optimal ("best M-terms") L^2 approximation of u in [8, Theorem 16]. In particular,

1. they show that the convergence of the quasi-optimal sparse-grid approximation is essentially sub-exponential with a rate comparable to that of the optimal L^2 approximation in the case of nested points and with half the rate for non-nested points. This difference can be ascribed to the fact that the construction of sparse

grids on non-nested collocation points is not as efficient as its nested counterpart and that the non-nested work contribution estimate is actually pessimistic.
2. the predicted convergence rate is only obtained after a sufficiently large amount of work (collocation points in this case, polynomial terms added in the expansion in the case of the optimal L^2 approximation) has been performed.
3. both convergence rate and minimal work depend non-trivially on the choice of the parameters δ and C_{log}.

A detailed discussion on the interplay between δ and C_{log} can be found in [8, Remarks 17, 19]; here we only mention that the two expressions in (22) are almost equivalent; for example, if the quadrature points are such that $\widehat{C}_m = 2$ (see Sect. 6), then the first term is larger than the second if $\frac{2}{e} \geq \alpha_L - \delta\beta_L$, i.e. if δ is larger than approximately 0.45, which is well inside the range of feasible values of δ, cf. Lemma 3. In Sect. 7 we will show numerically that a convergence estimate of the form

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq \mathcal{A}_n \exp\left(-\mathcal{B}_n N \sqrt[N]{W_{\mathcal{I}(w),m}} \right)$$

for nested points, and

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq \mathcal{A}_{nn} \exp\left(-\mathcal{B}_{nn} N \sqrt[2N]{W_{\mathcal{I}(w),m}} \right)$$

for non-nested points, with $\mathcal{A}_n, \mathcal{A}_{nn} \in \mathbb{R}$ and $\mathcal{B}_n, \mathcal{B}_{nn} \in \mathbb{R}_+$, appropriately capture the behavior of the computational error.

Finally, let us point out that other estimates for best M-term approximations of polyellipse analytic functions can be found in [14]. In those estimates, the dependence on N has been removed by assuming that the parameters $\{\zeta_n\}_{n=1}^N$ go to infinity with n at a certain rate. In this work, we have not made any assumption on $\{\zeta_n\}_{n=1}^N$, therefore convergence rates that degenerate with respect to N have to be expected.

5.1 Preliminary results

We start the proofs of Theorems 2 and 3 by introducing the Chebyshev expansion of u (see e.g. [40]) and estimate the decay of its coefficients. To this end, we introduce the Chebyshev polynomials of the first kind $\Psi_q(t)$ on $[-1, 1]$, which are defined as the unique polynomials satisfying

$$\Psi_q(\cos\vartheta) = \cos(q\vartheta), \quad 0 \leq \vartheta \leq \pi, \quad q \in \mathbb{N}.$$

As a consequence, $|\Psi_q(t)| \leq 1$ on $[-1, 1]$, with $\Psi_q(1) = 1$ and $\Psi_q(-1) = (-1)^q$; moreover, they are orthogonal with respect to the weight $\rho_C(t) = 1/\sqrt{1-t^2}$:

$$\int_{-1}^{1} \Psi_q(t)\Psi_\kappa(t)\rho_C(t)dt = \begin{cases} 0 & \kappa \neq q \\ \pi & \kappa = q = 0 \\ \pi/2 & \kappa = q \neq 0. \end{cases}$$

Lemma 2 *Let $\Psi_{q_n}(y_n)$ be the family of Chebyshev polynomials of the first kind on $\Gamma_n = [-1, 1]$, and let*

$$\mathbf{\Psi}_{\mathbf{q}}(\mathbf{y}) = \prod_{n=1}^{N} \Psi_{q_n}(y_n), \quad \mathbf{q} = (q_1, q_2, \ldots, q_N) \in \mathbb{N}^N.$$

be the generic N-variate Chebyshev polynomial. If the function u satisfies Assumption A1, then it admits a Chebyshev expansion

$$u(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{q} \in \mathbb{N}^N} u_{\mathbf{q}}(\mathbf{x}) \mathbf{\Psi}_{\mathbf{q}}(\mathbf{y}),$$

with coefficients

$$u_{\mathbf{q}}(\mathbf{x}) = \frac{1}{\int_\Gamma \mathbf{\Psi}_{\mathbf{q}}^2(\mathbf{y}) \varrho_C(\mathbf{y}) d\mathbf{y}} \int_\Gamma u(\mathbf{x}, \mathbf{y}) \mathbf{\Psi}_{\mathbf{q}}(\mathbf{y}) \varrho_C(\mathbf{y}) d\mathbf{y},$$

which converges in $C^0(\Gamma, V)$, and whose coefficients $u_{\mathbf{q}}(\mathbf{x})$ are such that

$$\|u_{\mathbf{q}}\|_V \leq C_{Cheb}(\mathbf{q}) \prod_{n=1}^{N} e^{-g_n q_n}, \quad g_n = \log \zeta_n \tag{23}$$

with $1 < \zeta_n < \zeta_n^*$, $C_{Cheb}(\mathbf{q}) = 2^{\|\mathbf{q}\|_0} B_u(\zeta_1, \ldots, \zeta_N)$, where $\|\mathbf{q}\|_0$ denotes the number of non-zero elements of $\mathbf{q}$, and $B_u(\zeta_1, \ldots, \zeta_N)$ as in Assumption A1.

Proof The proof is a straightforward extension to the N-dimensional case of the argument in [16, Chapter 7, Theorem 8.1]; see also [3]. □

Remark 3 An analogous bound could be proved for the decay of the coefficients of the Legendre expansion of u, namely

$$\|u_{\mathbf{q}}\|_V \leq B_u(\zeta_1, \ldots, \zeta_N) \prod_{n=1}^{N} \frac{r_n^{q_n}}{\tau_n \zeta_n} \left(\sqrt{1 - r_n^2} + \mathcal{O}\left(\frac{1}{q_n^{1/3}} \right) \right),$$

with τ_n to be chosen in $(0, 1)$ and

$$r_n = r_n(\tau_n, \zeta_n) = \frac{1}{1 + \zeta_n(1 - \tau_n) + \sqrt{\zeta_n^2(1 - \tau_n)^2 + 2\zeta_n(1 - \tau_n)}},$$

see [2,22]. Hence, the same analysis presented in this work could still be performed using the Legendre expansion of u instead of the Chebyshev one.

Remark 4 Lemma 2 and Remark 3 state that the convergence of the coefficients of both Chebyshev and Legendre expansions is essentially exponential with respect to

the degree of approximation of each parameter. Our numerical experience shows that such a bound is actually sharp, at least for the inclusion problem that will be discussed in Sect. 7; see also [8] for more examples on the Legendre expansion.

We close this section with some technical lemmas that will be needed in the following analysis.

Lemma 3 *For* $0 < \epsilon < \frac{e-1}{e} = \epsilon_{max} \approx 0.63$, *there holds*

$$\frac{1}{1-e^{-x}} \leq \frac{(1-\epsilon)e}{x}, \quad 0 < x \leq x_{cr}(\epsilon).$$

Moreover, the function $x_{cr}(\epsilon)$ *is concave and can be bounded from below as*

$$\alpha_L - \beta_L \epsilon \leq x_{cr}(\epsilon), \quad 0 < \epsilon < \epsilon_{max}$$

with $\alpha_L \approx 2.49$ *and* $\beta_L = \alpha_L/\epsilon_{max}$.

Proof See [8, Lemma 13]. □

Lemma 4 *Given any* $C_{log} \in (0, 1/e]$, *and denoting by* $\bar{t}$ *the largest root of the equation* $\log t = C_{log} t$, *then* $\forall\, N > 0$ *there holds*

$$M \leq e^{C_{log} N \sqrt[N]{M}} \quad \forall\, M > \bar{t}^N, \quad \forall\, N > 0.$$

If $C_{log} = 1/e$, *the bound holds for any* $M > 0$.

Proof See [8, Lemma 14]. □

5.2 Estimates of hierarchical surplus error contributions

Lemma 5 *If* u *satisfies Assumption* A1, *then for each hierarchical surplus operator* $\Delta^{m(\mathbf{i})}$ *there holds*

$$\delta E(\mathbf{i}) \leq \Delta E(\mathbf{i}) = C_E e^{-\sum_{n=1}^N g_n m(i_n - 1)} \prod_{n=1}^N \overline{\mathbb{M}}_n^{m(i_n)},$$

with $C_E = 4^N B_u(\zeta_1, \ldots, \zeta_N) \prod_{n=1}^N \frac{1}{1-e^{-g_n}}$, g_n *as in Lemma* 2 *and* $B_u(\zeta_1, \ldots, \zeta_N)$ *as in Assumption* A1.

Proof Let us consider the Chebyshev expansion of u. From the definition (4) of $\delta E(\mathbf{i})$ we have

$$\delta E(\mathbf{i}) = \left\| \Delta^{m(\mathbf{i})}[u] \right\|_{V\otimes L^2_\varrho(\Gamma)} = \left\| \Delta^{m(\mathbf{i})} \left[\sum_{\mathbf{q}\in\mathbb{N}^N} u_{\mathbf{q}} \boldsymbol{\Psi}_{\mathbf{q}} \right] \right\|_{V\otimes L^2_\varrho(\Gamma)}$$
$$= \left\| \sum_{\mathbf{q}\in\mathbb{N}^N} u_{\mathbf{q}} \Delta^{m(\mathbf{i})}[\boldsymbol{\Psi}_{\mathbf{q}}] \right\|_{V\otimes L^2_\varrho(\Gamma)}$$

Observe now that by construction of hierarchical surplus there holds $\Delta^{m(\mathbf{i})}[\boldsymbol{\Psi}_{\mathbf{q}}] = 0$ for all Chebyshev polynomials $\boldsymbol{\Psi}_{\mathbf{q}}$ such that $\exists n : q_n < m(i_n - 1)$. Therefore, the previous sum reduces to the multi-index set $\mathbf{q} \geq m(\mathbf{i} - \mathbf{1})$, and we have

$$\delta E(\mathbf{i}) = \left\| \sum_{\mathbf{q}\geq m(\mathbf{i}-\mathbf{1})} u_{\mathbf{q}} \Delta^{m(\mathbf{i})}[\boldsymbol{\Psi}_{\mathbf{q}}] \right\|_{V\otimes L^2_\varrho(\Gamma)}$$
$$\leq \sum_{\mathbf{q}\geq m(\mathbf{i}-\mathbf{1})} \left\| u_{\mathbf{q}} \right\|_V \left\| \Delta^{m(\mathbf{i})}[\boldsymbol{\Psi}_{\mathbf{q}}] \right\|_{L^2_\varrho(\Gamma)}.$$

Next, using the definition of $\Delta^{m(\mathbf{i})}$ we bound

$$\left\| \Delta^{m(\mathbf{i})}[\boldsymbol{\Psi}_{\mathbf{q}}] \right\|_{L^2_\varrho(\Gamma)} = \prod_{n=1}^{N} \left\| \Delta^{m(i_n)}[\Psi_{q_n}] \right\|_{L^2_{\varrho_n}(\Gamma_n)}$$
$$\leq \prod_{n=1}^{N} 2\,\overline{\mathbb{M}}_n^{m(i_n)} \left\| \Psi_{q_n} \right\|_{L^\infty(\Gamma_n)} = \prod_{n=1}^{N} 2\,\overline{\mathbb{M}}_n^{m(i_n)}.$$

Recalling estimate (23) for the decay of the Chebyshev coefficients of u, one obtains

$$\delta E(\mathbf{i}) \leq \sum_{\mathbf{q}\geq m(\mathbf{i}-\mathbf{1})} \left\| u_{\mathbf{q}} \right\|_V \prod_{n=1}^{N} 2\overline{\mathbb{M}}_n^{m(i_n)} \leq 2^N \sum_{\mathbf{q}\geq m(\mathbf{i}-\mathbf{1})} C_{Cheb}(\mathbf{q}) \prod_{n=1}^{N} e^{-g_n q_n} \overline{\mathbb{M}}_n^{m(i_n)}$$
$$\leq 4^N B_u(\zeta_1, \ldots, \zeta_N) \prod_{n=1}^{N} \overline{\mathbb{M}}_n^{m(i_n)} \sum_{q_n \geq m(i_n-1)} e^{-g_n q_n}$$
$$\leq 4^N B_u(\zeta_1, \ldots, \zeta_N) \prod_{n=1}^{N} \overline{\mathbb{M}}_n^{m(i_n)} \frac{e^{-g_n m(i_n-1)}}{1 - e^{-g_n}}.$$

□

Remark 5 This bound had been already proposed without proof in [6], using the norm of the interpolation operator $\mathcal{U}_n^{m(i_n)} : C^0(\Gamma_n) \to L^\infty(\Gamma_n)$, i.e. the standard Lebesgue constant associated to $\mathcal{U}_n^{m(i_n)}$, instead of $\overline{\mathbb{M}}_n^{m(i_n)}$.

5.3 Convergence result: nested case

We now focus on the case of nested sequences of collocation points. Observe that the profits (17) are derived by the profit definition (9) combining the work contribution (7) and Lemma 5.

Lemma 6 *Under Assumption A2, the auxiliary profits*

$$P^b(\mathbf{i}) = C_E C_{\mathbb{M}}^N \prod_{n=1}^{N} e^{-g_n m(i_n-1)}$$

are such that

$$P(\mathbf{i}) \le P^b(\mathbf{i}), \quad \forall \ \mathbf{i} \in \mathbb{N}_+^N, \tag{24}$$

where $P(\mathbf{i})$ are the profits in (17). Moreover, the sequence $\{P^b(\mathbf{i})\}_{\mathbf{i}\in\mathbb{N}_+^N}$ is monotone according to Definition 1, i.e.

$$P^{b,*}(\mathbf{i}) = \max_{\mathbf{j}\ge\mathbf{i}} P^b(\mathbf{j}) = P^b(\mathbf{i}), \quad \forall \mathbf{i} \in \mathbb{N}_+^N$$

and, under Assumption A3, it satisfies the weighted τ-summability condition (14) for every $0 < \tau < 1$. In particular, there holds

$$\sum_{\mathbf{i}\in\mathbb{N}_+^N} P^b(\mathbf{i})^\tau \Delta W(\mathbf{i}) \le C_E^\tau (C_{\mathbb{M}}^\tau \widehat{C}_m)^N \prod_{n=1}^{N} \frac{e^{\tau g_n}}{e^{\tau g_n} - 1},$$

with C_E as in Lemma 5 and $\widehat{C}_m = \max\{1, C_m\}$.

Proof Inequality (24) follows from Assumption A2, while the fact that the sequence $\{P^b(\mathbf{i})\}_{\mathbf{i}\in\mathbb{N}_+^N}$ is monotone is a straightforward consequence of its definition. As for the summability property, we start by observing that we can actually write the weighted sum $\sum_{\mathbf{i}\in\mathbb{N}_+^N} P^b(\mathbf{i})^\tau \Delta W(\mathbf{i})$ as a product of series depending on i_n only,

$$\sum_{\mathbf{i}\in\mathbb{N}_+^N} C_E^\tau \prod_{n=1}^{N} \left[\left(C_{\mathbb{M}} e^{-g_n m(i_n-1)} \right)^\tau d(i_n) \right] = C_E^\tau C_{\mathbb{M}}^{\tau N} \prod_{n=1}^{N} \sum_{i_n=1}^{\infty} \left(e^{-g_n m(i_n-1)} \right)^\tau d(i_n), \tag{25}$$

so that we only need to study the summability of

$$\mathbb{S}_n = \sum_{i_n=1}^{\infty} \left(e^{-g_n m(i_n-1)} \right)^\tau d(i_n), \quad n = 1, \ldots, N.$$

We begin by taking out of the sum the term for $i_n = 1$ and using Assumption A3

$$\mathbb{S}_n = 1 + \sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} d(i_n) \leq 1 + C_m \sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} d(i_n - 1). \tag{26}$$

Next, observe that

$$\begin{aligned} e^{-\tau g_n m(i_n-1)} d(i_n - 1) &= e^{-\tau g_n m(i_n-1)} \left(m(i_n - 1) - m(i_n - 2)\right) \\ &\leq \sum_{j_n=m(i_n-2)+1}^{m(i_n-1)} e^{-j_n \tau g_n}, \end{aligned}$$

such that

$$\sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} d(i_n - 1) \leq \sum_{j_n=1}^{\infty} e^{-\tau g_n j_n}. \tag{27}$$

Therefore, going back to (26) we obtain

$$\mathbb{S}_n \leq 1 + C_m \sum_{i_n=1}^{\infty} e^{-\tau g_n i_n} \leq \max\{1, C_m\} \sum_{i_n=0}^{\infty} e^{-\tau g_n i_n} = \widehat{C}_m \frac{e^{\tau g_n}}{e^{\tau g_n} - 1}.$$

and the proof is concluded by substituting this bound in (25). □

We are now ready to give the full proof of Theorem 2.

Proof of Theorem 2 In the case of nested collocation points, the quasi-optimal sparse grid is built using the profits from (17). First, we observe that, due to Lemma 6, it holds that

$$P^*(\mathbf{i}) = \max_{\mathbf{j} \geq \mathbf{i}} P(\mathbf{i}) \leq \max_{\mathbf{j} \geq \mathbf{i}} P^b(\mathbf{i}) = P^b(\mathbf{i}).$$

Therefore, the profits $P^*(\mathbf{i})$ have (at least) the same τ-summability properties as $P^b(\mathbf{i})$, and thus from Theorem 1 we have

$$\begin{aligned} \left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} &\leq W^{1-1/\tau}_{\mathcal{I}(w),m} \left(\sum_{\mathbf{i} \in \mathbb{N}_+^N} P^*(\mathbf{i})^\tau \Delta W(\mathbf{i}) \right)^{1/\tau} \\ &\leq W^{1-1/\tau}_{\mathcal{I}(w),m} \left(\sum_{\mathbf{i} \in \mathbb{N}_+^N} P^b(\mathbf{i})^\tau \Delta W(\mathbf{i}) \right)^{1/\tau}. \end{aligned}$$

Now, due to Lemma 6, the profits P^b satisfy the weighted τ-summability condition for every τ in $(0, 1)$, hence we can use any τ in the range $0 < \tau < 1$ to obtain a

valid bound for the error of the sparse grid. Thus, we can choose the smallest bound by minimizing the error estimate over τ: to this end, we closely follow the argument in [8, Theorem 16]. We have

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq W^{1-1/\tau}_{\mathcal{I}(w),m} \left(C^\tau_E (C^\tau_{\mathbb{M}} \widehat{C}_m)^N \prod_{n=1}^N \frac{e^{\tau g_n}}{e^{\tau g_n} - 1} \right)^{1/\tau}, \tag{28}$$

and we want to minimize the right-hand side with respect to τ. We do not solve this minimization problem exactly, but rather take $\tau = e\mathcal{K}$, with $\mathcal{K}^N = \dfrac{\widehat{C}^N_m}{W_{\mathcal{I}(w),m} \prod_{n=1}^N g_n}$. The motivation for this choice is the following: if τ is small, we can approximate $\frac{e^{\tau g_n}}{e^{\tau g_n}-1} \approx \frac{1}{\tau g_n}$ and rewrite the right-hand side of (28) as

$$\begin{aligned}
& W^{1-1/\tau}_{\mathcal{I}(w),m} \left(C^\tau_E (C^\tau_{\mathbb{M}} \widehat{C}_m)^N \prod_{n=1}^N \frac{e^{\tau g_n}}{e^{\tau g_n} - 1} \right)^{1/\tau} \\
& \quad \approx W^{1-1/\tau}_{\mathcal{I}(w),m} \left(C^\tau_E (C^\tau_{\mathbb{M}} \widehat{C}_m)^N \frac{1}{\tau^N \prod_{n=1}^N g_n} \right)^{1/\tau} \\
& \quad = W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{\widehat{C}^N_m}{\tau^N W_{\mathcal{I}(w),m} \prod_{n=1}^N g_n} \right)^{1/\tau} \\
& \quad = W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{\mathcal{K}^N}{\tau^N} \right)^{1/\tau}.
\end{aligned}$$

Therefore, we enforce

$$\begin{aligned}
0 = \frac{d}{d\tau} \left(\frac{\mathcal{K}^N}{\tau^N} \right)^{1/\tau} &= \frac{d}{d\tau} \exp\left(-\frac{1}{\tau} \log \frac{\tau^N}{\mathcal{K}^N} \right) \\
&= \exp\left(-\frac{1}{\tau} \log \frac{\tau^N}{\mathcal{K}^N} \right) \left[\frac{1}{\tau^2} \log \frac{\tau^N}{\mathcal{K}^N} - \frac{1}{\tau} \frac{N}{\tau} \right]
\end{aligned}$$

that is

$$0 = \frac{N}{\tau^2} \left[\log \frac{\tau}{\mathcal{K}} - 1 \right], \tag{29}$$

resulting in $\tau = e\mathcal{K}$. We now insert this choice of τ in the original bound (28) obtaining

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{\widehat{C}^N_m}{W_{\mathcal{I}(w),m}} \prod_{n=1}^N \frac{e^{g_n e\mathcal{K}}}{e^{g_n e\mathcal{K}} - 1} \right)^{1/\tau}.$$

Next, we bound each of the factors $e^{eg_n\mathcal{K}}/(e^{eg_n\mathcal{K}} - 1)$ by Lemma 3 (with $x = eg_n\mathcal{K}$) to obtain

$$\begin{aligned}
&\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V\otimes L^2_\varrho(\Gamma)} \\
&\quad \leq W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{\widehat{C}^N_m}{W_{\mathcal{I}(w),m}} \prod_{n=1}^{N} (1-\epsilon_n) \frac{e g_m \sqrt[N]{W_{\mathcal{I}(w),m}}}{e g_n \widehat{C}_m}, \right)^{1/(e\mathcal{K})}, \\
&\quad = W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\prod_{n=1}^{N} (1-\epsilon_n) \right)^{1/(e\mathcal{K})}. \qquad (30)
\end{aligned}$$

Note that the latter equation holds true for ϵ_n and $W_{\mathcal{I}(w),m}$ satisfying $e g_n \mathcal{K} \leq x_{cr}(\epsilon_n)$ (cf. again Lemma 3), which in turn is satisfied if we choose ϵ_n using the lower bound in Lemma 3, i.e.

$$e g_n \mathcal{K} = \frac{e g_n \widehat{C}_m}{\sqrt[N]{W_{\mathcal{I}(w),m}} g_m} = \alpha_L - \beta_L \epsilon_n \Rightarrow \epsilon_n = \left(\alpha_L - \frac{g_n e \widehat{C}_m}{g_m \sqrt[N]{W_{\mathcal{I}(w),m}}} \right) \frac{1}{\beta_L}.$$

Moreover, we also have to enforce $\epsilon_n > 0$ to ensure convergence of estimate (30), thus obtaining a constraint on $W_{\mathcal{I}(w),m}$. Namely, taken any $0 < \delta < \epsilon_{max}$ we require $\epsilon_n > \delta$, which implies

$$\delta < \left(\alpha_L - \frac{g_n e \widehat{C}_m}{g_m \sqrt[N]{W_{\mathcal{I}(w),m}}} \right) \frac{1}{\beta_L} \qquad \Rightarrow \qquad W_{\mathcal{I}(w),m} > \left(\frac{g_n e \widehat{C}_m}{g_m (\alpha_L - \delta \beta_L)} \right)^N.$$

Since we have assumed that the coefficients g_n are ordered increasingly, this condition has to be checked for $n = N$ only, hence (19). With this choice of $\epsilon_{W_{\mathcal{I}(w),m},n}$, Eq. (30) further simplifies to

$$\begin{aligned}
\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V\otimes L^2_\varrho(\Gamma)} &\leq W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \exp\left(\sum_{n=1}^{N} \log(1-\epsilon_n) \right)^{1/(e\mathcal{K})} \\
&\leq W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \exp\left(-\sum_{n=1}^{N} \epsilon_n \right)^{1/(e\mathcal{K})} \\
&\leq W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \exp\left(-\frac{N\delta}{e\mathcal{K}} \right),
\end{aligned}$$

and the final result follows by recalling the definition of $\mathcal{K}$ and using Lemma 4. □

5.4 Convergence result: non-nested case

We now focus on the case of non-nested sequences of collocation points. Observe that the profits (20) are derived by the profit definition (9) combining the work contribution (8) and Lemma 5.

Lemma 7 *Under Assumption A2, the auxiliary profits*

$$P^b(\mathbf{i}) = C_E C_{\mathbb{M}}^N \prod_{n=1}^{N} e^{-g_n m(i_n-1)}$$

are such that

$$P(\mathbf{i}) \le P^b(\mathbf{i}), \quad \forall \ \mathbf{i} \in \mathbb{N}_+^N, \tag{31}$$

where $P(\mathbf{i})$ are the profits (20). Moreover, the sequence $\{P^b(\mathbf{i})\}_{\mathbf{i}\in\mathbb{N}_+^N}$ is monotone according to Definition 1, i.e.

$$P^{b,*}(\mathbf{i}) = \max_{\mathbf{j}\ge\mathbf{i}} P^b(\mathbf{j}) = P^b(\mathbf{i}), \quad \forall \mathbf{i} \in \mathbb{N}_+^N$$

and, under Assumption A3, it satisfies the weighted τ-summability condition (14) for every $0 < \tau < 1$. In particular, there holds

$$\sum_{\mathbf{i}\in\mathbb{N}_+^N} P^b(\mathbf{i})^\tau \Delta W(\mathbf{i}) \le (C_E C_{\mathbb{M}}^N)^\tau \prod_{n=1}^{N} \left(\widehat{C}_m \frac{e^{\tau g_n}}{e^{\tau g_n}-1} + \frac{2}{\tau g_n e} \frac{e^{\tau g_n/2}}{e^{\tau g_n/2}-1} \right).$$

Proof Inequality (31) follows from Assumption A2, while the fact that $\{P^b(\mathbf{i})\}_{\mathbf{i}\in\mathbb{N}_+^N}$ is monotone is a straightforward consequence of its definition. As for the summability property, we proceed as in the proof of Lemma 6 and rewrite

$$\begin{aligned}\sum_{\mathbf{i}\in\mathbb{N}_+^N} P^b(\mathbf{i})^\tau \Delta W(\mathbf{i}) &= \sum_{\mathbf{i}\in\mathbb{N}_+^N} C_E^\tau \prod_{n=1}^{N} \left[\left(C_{\mathbb{M}} e^{-g_n m(i_n-1)} \right)^\tau m(i_n) \right] \\ &= C_E^\tau C_{\mathbb{M}}^{\tau N} \prod_{n=1}^{N} \sum_{i_n=1}^{\infty} \left(e^{-g_n m(i_n-1)} \right)^\tau m(i_n), \end{aligned} \tag{32}$$

to study the summability of

$$\mathbb{S}_n = \sum_{i_n=1}^{\infty} \left(e^{-g_n m(i_n-1)} \right)^\tau m(i_n), \quad n = 1, \dots, N.$$

We split the sum as

$$\begin{aligned}\mathbb{S}_n &= 1 + \sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} m(i_n) \\ &= 1 + \sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} d(i_n) + \sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} m(i_n-1),\end{aligned}$$

and we consider the two sums separately. The first one can be bounded as in Lemma 6, Eqs. (26)–(27),

$$\sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} d(i_n) \leq C_m \sum_{i_n=1}^{\infty} e^{-\tau g_n i_n},$$

while the second one can be bounded as

$$\sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)} m(i_n-1) \leq \frac{2}{\tau g_n e} \sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)/2}.$$

exploiting the elementary fact that for every $x > 0$ and for every $\epsilon > 0$, there holds $x \leq \frac{1}{\epsilon e} e^{\epsilon x}$. Combining the two bounds, we obtain

$$\begin{aligned}
\mathbb{S}_n &\leq 1 + C_m \sum_{i_n=1}^{\infty} e^{-\tau g_n i_n} + \frac{2}{\tau g_n e} \sum_{i_n=2}^{\infty} e^{-\tau g_n m(i_n-1)/2} \\
&\leq \max\{1, C_m\} \sum_{i_n=0}^{\infty} e^{-\tau g_n i_n} + \frac{2}{\tau g_n e} \sum_{i_n=0}^{\infty} e^{-\tau g_n i_n/2} \\
&\leq \widehat{C}_m \frac{e^{\tau g_n}}{e^{\tau g_n} - 1} + \frac{2}{\tau g_n e} \frac{e^{\tau g_n/2}}{e^{\tau g_n/2} - 1},
\end{aligned}$$

and the proof is concluded by substituting this bound into (32). □

We are now ready to give the full proof of Theorem 3.

Proof of Theorem 3 In the case of non-nested collocation points, the quasi-optimal sparse grid is built using the profits from (20). The proof is analogous to that of Theorem 2. Due to Lemma 7, there holds

$$P^*(\mathbf{i}) = \max_{\mathbf{j} \geq \mathbf{i}} P(\mathbf{i}) \leq \max_{\mathbf{j} \geq \mathbf{i}} P^b(\mathbf{i}) = P^b(\mathbf{i}),$$

such that from Theorem 1 and Lemma 7 we have

$$\begin{aligned}
&\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \\
&\quad \leq W^{1-1/\tau}_{\mathcal{I}(w),m} \left(\sum_{\mathbf{i} \in \mathbb{N}_+^N} P^b(\mathbf{i})^\tau \Delta W_j \right)^{1/\tau} \\
&\quad \leq W^{1-1/\tau}_{\mathcal{I}(w),m} \left((C_E C^N_{\mathbb{M}})^\tau \prod_{n=1}^{N} \left(\widehat{C}_m \frac{e^{\tau g_n}}{e^{\tau g_n} - 1} + \frac{2}{\tau g_n e} \frac{e^{\tau g_n/2}}{e^{\tau g_n/2} - 1} \right) \right)^{1/\tau},
\end{aligned} \tag{33}$$

to be minimized with respect to τ. Next, we suppose τ to be small, such that

$$\widehat{C}_m \frac{e^{\tau g_n}}{e^{\tau g_n}-1} + \frac{2}{\tau g_n e}\frac{e^{\tau g_n/2}}{e^{\tau g_n/2}-1} \approx \frac{\widehat{C}_m}{\tau g_n} + \frac{4}{(\tau g_n)^2 e} \approx \frac{4}{(\tau g_n)^2 e},$$

and

$$\begin{aligned}\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V\otimes L^2_\varrho(\Gamma)} &\approx W^{1-1/\tau}_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{4^N}{e^N \tau^{2N} \prod_{n=1}^N g_n^2} \right)^{1/\tau} \\ &= W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{\mathcal{K}^N}{\tau^{2N}} \right)^{1/\tau},\end{aligned}$$

with $\mathcal{K}^N = \dfrac{4^N}{e^N W_{\mathcal{I}(w),m} g_m^{2N}}$. With some calculus analogous to (29), we then obtain

$$\frac{d}{d\tau}\left(\frac{\mathcal{K}^N}{\tau^{2N}}\right)^{1/\tau} = 0 \Leftrightarrow \log\frac{\tau^2}{\mathcal{K}} = 2 \Leftrightarrow \tau = e\sqrt{\mathcal{K}}.$$

We now go back to the original bound (33) and apply Lemma 3, obtaining

$$\begin{aligned}&\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V\otimes L^2_\varrho(\Gamma)} \\ &\le W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{1}{W_{\mathcal{I}(w),m}} \prod_{n=1}^N \left(\widehat{C}_m \frac{(1-\epsilon_n)e}{\tau g_n} + \frac{2}{\tau g_n e}\frac{(1-\epsilon_n)e}{\tau g_n/2} \right) \right)^{1/\tau}, \\ &= W_{\mathcal{I}(w),m} C_E C^N_{\mathbb{M}} \left(\frac{1}{W_{\mathcal{I}(w),m}} \prod_{n=1}^N \frac{1-\epsilon_n}{\tau g_n} \left(\widehat{C}_m e + \frac{4}{\tau g_n} \right) \right)^{1/\tau}, \end{aligned} \tag{34}$$

which holds for $\tau g_n \le x_{cr}(\epsilon_n)$. This condition can be satisfied using the lower bound in Lemma 3, i.e. by choosing ϵ_n such that

$$e\sqrt{\mathcal{K}} g_n = \frac{2 g_n \sqrt{e}}{g_m \sqrt[2N]{W_{\mathcal{I}(w),m}}} = \alpha_L - \beta_L \epsilon_n \Rightarrow \epsilon_n = \left(\alpha_L - \frac{2 g_n \sqrt{e}}{g_m \sqrt[2N]{W_{\mathcal{I}(w),m}}} \right) \frac{1}{\beta_L}.$$

Note also that we will need $\epsilon_n > 0$ to ensure convergence of the estimate; namely, for any $0 < \delta < \epsilon_{max}$ we require $\epsilon_n > \delta$, which implies

$$\delta < \left(\alpha_L - \frac{2 g_n \sqrt{e}}{g_m \sqrt[2N]{W_{\mathcal{I}(w),m}}} \right) \frac{1}{\beta_L} \quad \Rightarrow \quad W_{\mathcal{I}(w),m} > \left(\frac{2\sqrt{e} g_n}{g_m(\alpha_L - \delta\beta_L)} \right)^{2N}.$$

Moreover, under the additional assumption that $\widehat{C}_m e \leq 4/(\tau g_n)$, i.e.

$$\widehat{C}_m e \leq \frac{4}{e\sqrt{\mathcal{K}} g_n} = \frac{2 g_m \sqrt[2N]{W_{\mathcal{I}(w),m}}}{g_n \sqrt{e}} \qquad \Rightarrow \qquad W_{\mathcal{I}(w),m} > \left(\frac{\widehat{C}_m e^{3/2} g_n}{2 g_m} \right)^{2N},$$

we can bound the term $(\widehat{C}_m e + 4/(\tau g_n))$ in (34) with $8/(\tau g_n)$, such that (34) can be rewritten as

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq W_{\mathcal{I}(w),m} C_E C_{\mathbb{M}}^N \left(\frac{8^N}{W_{\mathcal{I}(w),m}} \prod_{n=1}^{N} \frac{1-\epsilon_n}{\tau^2 g_n^2} \right)^{1/\tau},$$

which can be simplified further by inserting the nearly optimal value of τ that we computed previously:

$$\begin{aligned}
\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} &\leq W_{\mathcal{I}(w),m} C_E C_{\mathbb{M}}^N \left(\frac{8^N}{W_{\mathcal{I}(w),m}} \frac{1}{e^{2N} \mathcal{K}^N} \prod_{n=1}^{N} \frac{1-\epsilon_n}{g_n^2} \right)^{1/\tau} \\
&= W_{\mathcal{I}(w),m} C_E C_{\mathbb{M}}^N \left(\frac{8^N}{W_{\mathcal{I}(w),m}} \frac{e^N W_{\mathcal{I}(w),m} g_m^{2N}}{e^{2N} 4^N} \prod_{n=1}^{N} \frac{1-\epsilon_n}{g_n^2} \right)^{1/\tau} \\
&= W_{\mathcal{I}(w),m} C_E C_{\mathbb{M}}^N \left(\left(\frac{2}{e} \right)^N \prod_{n=1}^{N} (1-\epsilon_n) \right)^{1/\tau} \\
&\leq W_{\mathcal{I}(w),m} C_E C_{\mathbb{M}}^N \left(\left(\frac{2}{e} \right)^N e^{-N\delta} \right)^{1/\tau} \\
&= W_{\mathcal{I}(w),m} C_E C_{\mathbb{M}}^N \exp\left(-\frac{N(\delta + 1 - \log 2)}{e\sqrt{\mathcal{K}}} \right) \\
&= W_{\mathcal{I}(w),m} C_E C_{\mathbb{M}}^N \exp\left(-\frac{N(\delta + 1 - \log 2) g_m \sqrt[2N]{W_{\mathcal{I}(w),m}}}{2\sqrt{e}} \right).
\end{aligned}$$

The proof is then concluded by using Lemma 4. □

6 On the choice of collocation points

Before moving to the numerical testing of the quasi-optimal sparse grids, we discuss here some possible families of uni-variate collocation points that can be used in this context. More specifically, we will assess the values of the associated quantities $\overline{\mathbb{M}}^q_n$, cf. Definition 2, and we will verify whether such families satisfy Assumptions A2 and A3, that are needed for the convergence results in Theorems 2 and 3 to hold. In particular, we will consider here Gauss–Legendre, Clenshaw–Curtis, Leja, and Gauss–Patterson uni-variate collocation points that are commonly used for sparse-grid approximations of problems depending on uniform random variables, such as the test case that we will consider in Sect. 7.

Gauss–Legendre points. The classical Gauss–Legendre collocation points (see e.g. [39]) are non-nested points that we will use together with the level-to-nodes relation

$$m_{lin}(i) = i, \quad \Rightarrow d_{lin}(i) = 1. \tag{35}$$

The following result on the quantity $\mathbb{M}_n^{m(i_n)}$ holds true.

Lemma 8 *If $\mathcal{U}_n^{m(i_n)}$ is built over Gaussian abscissas, then $\mathbb{M}_n^{m(i_n)} = 1$.*

Proof Let $\mathcal{Q}_n^{m(i_n)}$ be the quadrature rule built over the same abscissas used for $\mathcal{U}_n^{m(i_n)}$,

$$\mathcal{Q}_n^{m(i_n)}[f] = \sum_{j=1}^{m(i_n)} f(t_j)\varpi_j,$$

where ϖ_j are the associated quadrature weights. Observe that since the considered abscissas are Gaussian, $\mathcal{Q}_n^{m(i_n)}$ is exact for polynomials of degree $2m(i_n) - 1$, and, in particular, $\varpi_j > 0$ for any $j = 1, \ldots, m(i_n)$ and $\sum_{j=1}^{m(i_n)} \varpi_j = 1$. Next, observe that $\left(\mathcal{U}_n^{m(i_n)}[f(t)]\right)^2$ is a polynomial of degree $2(m(i_n) - 1)$; therefore, using the fact that $\mathcal{U}_n^{m(i_n)}$ is a Lagrangian interpolant, we have

$$\begin{aligned}\left\|\mathcal{U}_n^{m(i_n)}[f]\right\|^2_{L^2_\varrho(\Gamma)} &= \int_{\Gamma_n} \left(\mathcal{U}_n^{m(i_n)}[f](t)\right)^2 \varrho_n(t)dt = \mathcal{Q}\left[\left(\mathcal{U}_n^{m(i_n)}[f]\right)^2\right] \\ &= \sum_{j=1}^{m(i_n)} \left(\mathcal{U}_n^{m(i_n)}[f](t_j)\right)^2 \varpi_j = \sum_{j=1}^{m(i_n)} f^2(t_j)\varpi_j \leq \|f\|^2_{L^\infty(\Gamma_n)},\end{aligned}$$

finishing the proof. □

This result, together with Eq. (35), implies that Assumption A2 holds with $C_{\mathbb{M}} = 1$. Assumption A3 holds with $C_m = 1$ due to Eq. (35).

Clenshaw–Curtis points. The Clenshaw–Curtis collocation points (see e.g. [40]) are the nested version of the Chebyshev collocation points, i.e.

$$y_j^i = \cos\left(\frac{(j-1)\pi}{m(i)-1}\right), 1 \leq j \leq m(i),$$

with the associated level-to-nodes relation:

$$m_{db}(i) = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } i = 1 \\ 2^{i-1} + 1, & \text{if } i > 1. \end{cases} \Rightarrow \quad d_{db}(i) = \begin{cases} 1 & \text{if } i = 1 \\ 2 & \text{if } i = 2 \\ 2^{i-2}, & \text{if } i > 2, \end{cases} \tag{36}$$

In the case of nested Clenshaw–Curtis nodes, we use the standard estimate of the "L^∞" Lebesgue constant (see e.g. [17,18]) as a bound for $\overline{\mathbb{M}}_n^q$,

$$\overline{\mathbb{M}}_n^q \leq \overline{\mathbb{M}}_{n,est}^q, \qquad \overline{\mathbb{M}}_{n,est}^q = \begin{cases} 1 & \text{for } q = 1 \\ \dfrac{2}{\pi}\log(q-1)+1 & \text{for } q \geq 2. \end{cases}$$

Combining this estimate and Eq. (36) we obtain that Assumption A2 holds with $C_{\mathbb{M}} = 1$. Assumption A3 holds with $C_m = 2$, due again to Eq. (36).

Leja points. Given a compact set X and an initial value $x_0 \in X$, Leja sequences are nested sequences of points, recursively defined as $x_k = \text{argmax}_{y\in X} |\prod_{j=0}^{k-1}(y - x_j)|$. Choosing $x_0 = 1$ and $X = [-1, 1]$ results in the so-called *standard Leja sequence*, while by choosing X as the unit disk in the complex domain together with $x_0 = 1$ and projecting the resulting sequence on $[-1, 1]$ one obtains the so-called *R-Leja sequence*, see [11,34] and references therein for details. Here we focus on the so-called *symmetrized Leja sequence* (see again [34]), which, at level i, includes $2i + 1$ points defined as

$$x_0^i = 0, \quad x_1^i = 1, \quad x_2^i = -1,$$
$$x_k^i = \begin{cases} \underset{y\in[-1,1]}{\text{argmax}} \left|\displaystyle\prod_{j=0}^{k-1}(y - x_j)\right| & \text{if } k \text{ is odd, } k \leq 2i \\ -x_{k-1} & \text{if } k \text{ is even, } k \leq 2i+1 \end{cases}$$

such that the level-to-nodes function is defined as

$$m_{SL}(i) = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } i = 1 \\ 2i-1, & \text{if } i > 1. \end{cases} \Rightarrow d_{SL}(i) = \begin{cases} 1 & \text{if } i = 1 \\ 2 & \text{if otherwise.} \end{cases}$$

Thus, Assumption A3 trivially holds with $C_m = 2$. The validity of Assumption A2 can be verified numerically, by computing lower and upper bounds for the interpolant operator norm $\mathbb{M}_n^{m_{SL}(i_n)}$. The upper bound is obtained as in the case of Clenshaw–Curtis points, by bounding $\mathbb{M}_n^{m_{SL}(i_n)}$ with the standard Lebesgue constant, which can be approximated numerically by evaluating the L^∞ norm on a very fine grid. The lower bound is instead computed by solving approximately the maximization problem appearing in the definition of $\mathbb{M}_n^{m_{SL}(i_n)}$ (see Definition 2), which can be recast into a constrained quadratic optimization problem.

To do this, denote by $t_1, t_2, \ldots, t_{m(i_n)}$ the collocation points of $\mathcal{U}_n^{m(i_n)}$ and by $\ell_1, \ell_2, \ldots, \ell_{m(i_n)}$ the associated Lagrangian polynomials, and expand

$$\int_{\Gamma_n} \left(\mathcal{U}_n^{m(i_n)}[f](t)\right)^2 \varrho_n(t)dt = \int_{\Gamma_n} \left(\sum_{j=1}^{m(i_n)} f(t_j)\ell_j(t)\right)^2 \varrho_n(t)dt$$
$$= \sum_{\kappa,j}^{m(i_n)} f(t_\kappa) f(t_j) \int_{\Gamma_n} \ell_\kappa(t)\ell_j(t)\varrho_n(t)dt.$$

Observe now that, since $\ell_\kappa(y_n)\ell_j(y_n)$ is a polynomial of degree $2(m(i_n)-1)$, we can integrate it exactly with a Gaussian quadrature formula with $m(i_n)$ quadrature points $\{\widetilde{t}_1, \widetilde{t}_2, \ldots, \widetilde{t}_{m(i_n)}\}$ and associated quadrature weights $\{\varpi_1, \varpi_2, \ldots, \varpi_{m(i_n)}\}$:

$$\int_{\Gamma_n} \ell_\kappa(y_n)\ell_j(y_n)\varrho_n(y_n)dy_n = \sum_{i=1}^{m(i_n)} \ell_\kappa(\widetilde{t}_i)\ell_j(\widetilde{t}_i)\varpi_i := A_{\kappa,j}.$$

Denoting by $\boldsymbol{f}$ be the vector containing the nodal values $f_j = f(t_j)$, the computation of $\mathbb{M}_n^{m(i_n)}$ amounts then to solving the quadratic optimization problem

$$\mathcal{L} = \max_{\boldsymbol{f}\in\mathcal{R}} \boldsymbol{f}^T A \boldsymbol{f}, \quad \mathcal{R} = \{\boldsymbol{f} \in \mathbb{R}^{m(i_n)} \text{s.t. } -1 \le f_n \le 1, \quad \forall n = 1, \ldots N\},$$

and setting $\mathbb{M}_n^{m(i_n)} = \sqrt{\mathcal{L}}$. Being A positive definite, the solutions of the optimization problem are located in the corners of the feasible region $\mathcal{R}$, and multiple maxima are possible. By repeatedly running an optimization algorithm for quadratic optimization problems (here we use the active set method, see e.g. [32]) with different initial guesses, we obtain a suboptimal solution, which is however sufficient for our purposes.

Results are reported in Fig. 3, left, and they show that we can assume polynomial growth for $\overline{\mathbb{M}}_n^{m_{SL}(i_n)}$ and for $\overline{\mathbb{M}}_n^{m_{SL}(i_n)}/d_{SL}(i_n)$ (given that d_{SL} is constant for $i > 1$), which implies that Assumption A2 does not hold true; however, the theory previously developed could still be applied, cf. Remark 2.

The same conclusion can be deduced for the R-Leja sequence thanks to the results stated in [11], where polynomial growth is proved for the standard "L^∞" constant for this sequence of points.

Gauss–Patterson points. These nodes are constructed to have nested quadrature rules with maximal degree of exactness, and are tabulated (see [33]). In particular, there holds

$$m_{GP}(0) = 0, \quad m_{GP}(i_n) = 2^{i_n} - 1,$$

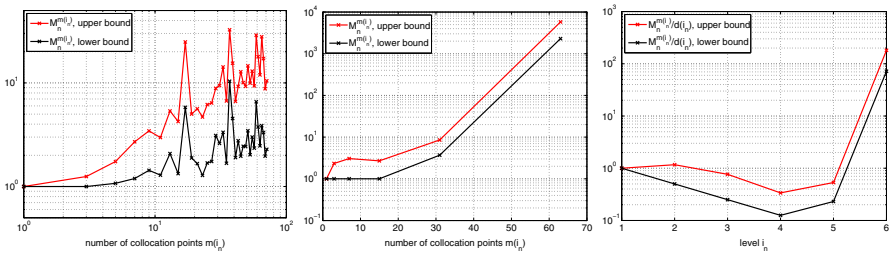

Fig. 3 *Left* the lower and upper bounds for $\mathbb{M}_n^{m_{SL}(i_n)}$ for symmetric Leja points show that $\overline{\mathbb{M}}_n^{m_{SL}(i_n)}$ grows polynomially (as does $\overline{\mathbb{M}}_n^{m_{SL}(i_n)}/d_{SL}(i_n)$). *Middle and right* $\mathbb{M}_n^{m_{GP}(i_n)}$ and $\mathbb{M}_n^{m_{GP}(i_n)}/d_{GP}(i_n)$ may asymptotically grow more than polynomially for Gauss–Patterson points instead. Note that the *left plot* is in log-log scale while the middle and the *right plots* are in semilog scale

therefore $d_{GP}(i_n) = 2^{i_n-1}$, hence Assumption A3 holds with $C_m = 2$, while the validity of Assumption A2 must be assessed again numerically.

The numerical results are shown in Fig. 3, middle/right, and they suggest that both $\overline{\mathbb{M}}_{n,est}^{m_{GP}(i_n)}$ and the ratio $\overline{\mathbb{M}}_{n}^{m_{GP}(i_n)}/d(i_n)$ may asymptotically grow more than polynomially, hence we cannot use the results obtained in the previous sections to conclude on the convergence of the quasi-optimal sparse grid approximation built using Gauss–Patterson knots.

7 Application to a diffusion problem with random inclusions

In this section, we show how the solution u of a certain class of elliptic PDEs with stochastic coefficients (namely, the so-called "inclusions problem" already examined in [4,8]) satisfies the polyellipse analyticity condition A1; we will then apply the previous Theorems 2 and 3 to establish the convergence of the quasi-optimal sparse-grid approximation of u, using both nested and non-nested points, and numerically verify such convergence results.

Let D be a convex polygonal domain in $\mathbb{R}^2$, and let $\mathbf{y}$ be an N-variate random vector whose components $y_1, \ldots, y_N$ are independent uniform random variables over $\Gamma_i = [y_{min}, y_{max}]$. The image of the random vector $\mathbf{y}$ is therefore the hypercube $\Gamma = \Gamma_1 \times \Gamma_2 \times \cdots \times \Gamma_N$, the joint probability density function of $\mathbf{y}$ is $\varrho(\mathbf{y}) = \prod_{n=1}^{N} \varrho_n(y_n) = \prod_{n=1}^{N} \frac{1}{y_{max}-y_{min}}$, and $(\Gamma, B(\Gamma), \varrho(\mathbf{y})d\mathbf{y})$ is the underlying probability space, $B(\Gamma)$ being the Borel σ-algebra on Γ. We consider the stochastic elliptic problem:

Problem 1 Find a real-valued function $u : \overline{D} \times \Gamma \to \mathbb{R}$, such that $\varrho(\mathbf{y})d\mathbf{y}$-almost everywhere there holds:

$$\begin{cases} -\operatorname{div}(a(\mathbf{x},\mathbf{y})\nabla u(\mathbf{x},\mathbf{y})) = f(\mathbf{x}) & \mathbf{x} \in D, \\ u(\mathbf{x},\mathbf{y}) = 0 & \mathbf{x} \in \partial D, \end{cases}$$

where $D = [0,1]^2$, the operators div and ∇ imply differentiation with respect to the physical coordinate only, and the diffusion coefficient is:

$$a(\mathbf{x},\mathbf{y}) = a_0 + \sum_{n=1}^{N} \gamma_n \chi_n(\mathbf{x}) y_n. \tag{37}$$

Here, $\chi_n(\mathbf{x})$ are the indicator functions of the disjoint circular sub-domains $D_n \subset D$ shown in Fig. 4, and a_0, γ_n are real coefficients such that $a(\mathbf{x},\mathbf{y})$ is strictly positive and bounded, i.e. there exist two positive constants $0 < a_{min} < a_{max} < \infty$ such that

$$0 < a_{min} \le a(\mathbf{x},\mathbf{y}) \le a_{max} < \infty, \tag{38}$$

$\varrho(\mathbf{y})d\mathbf{y}$-almost surely, $\forall\, \mathbf{x} \in D$.

Next, let $V = H_0^1(D)$ be the space of square-integrable functions in D with square-integrable distributional derivatives and with zero trace on the boundary equipped

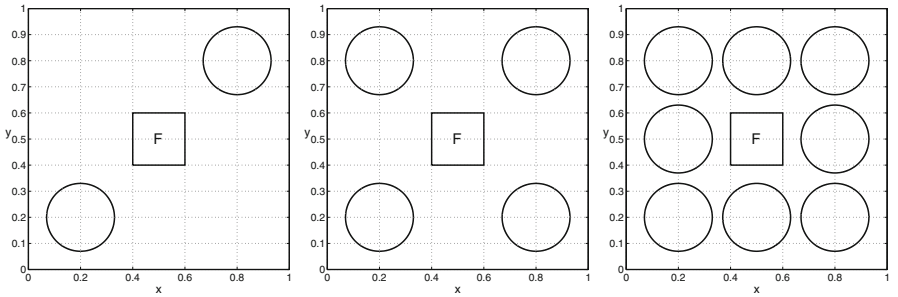

Fig. 4 Domains for the inclusions problem with 2, 4, and 8 circular inclusions

with the gradient norm $\|v\|_V = \|\nabla v\|_{L^2(D)}$. Using Lax–Milgram's Lemma, it is straightforward to show that Problem 1 is $\varrho(\mathbf{y})d\mathbf{y}$-almost everywhere well-posed in V, due to the boundedness assumption (38). Similarly, it is easy to see that $u \in L^2_\varrho(\Gamma)\otimes V$, see e.g. [4,6,8].

Remark 6 In this work, we do not address the discretization of the solution u in the physical variable $\mathbf{x}$. In this respect, all results obtained here also apply to a discrete solution u_h, obtained by introducing e.g. a finite element discretization over a triangulation $\mathcal{T}_h$ of the physical domain D and a finite element space $V_h(D) \subset H^1_0(D)$ of piecewise continuous polynomials on $\mathcal{T}_h$. Further savings may be obtained by introducing in the quasi-optimal sparse-grid procedure the possibility of approximating the PDEs associated with different points of the sparse grid on different meshes on the physical space, in the spirit of what was proposed in works [9,23,25,38,41].

We shall begin by reparametrizing the diffusion coefficient in terms of new random variables distributed over $[-1, 1]$. For the sake of notation, we will still denote the new variables as y_i, i.e. $y_i \sim \mathcal{U}(-1, 1)$. Therefore, the new diffusion coefficient will be:

$$a(\mathbf{x}, \mathbf{y}) = a_0 + \sum_{n=1}^{N} \gamma_n \chi_n(\mathbf{x}) \left(\frac{y_n + 1}{2}(y_{max} - y_{min}) + y_{min} \right). \tag{39}$$

Lemma 9 *The complex continuation u^* of the solution u corresponding to a diffusion coefficient (39) is analytic in the region*

$$\Sigma = \Sigma_1 \times \Sigma_2 \times \cdots \times \Sigma_N, \quad \Sigma_n = \{z_n \in \mathbb{C} : \mathfrak{Re}\,(z_n) \geq T_n\},$$

with $-1 \geq T_n > T_n^* = \dfrac{2a_0 + \gamma_n(y_{max} + y_{min})}{\gamma_n(y_{min} - y_{max})}$. *Moreover,* $\sup_{\mathbf{z}\in\Sigma} \|u^*\|_V \leq B_u(T_1, \ldots, T_N)$, *with*

$$B_u(T_1, \ldots, T_N) = \frac{\|f\|_{V'}}{a_0 + \min_{i=1,\ldots,N} \gamma_n \left(\frac{1-|T_n|}{2}(y_{max} - y_{min}) + y_{min} \right)}.$$

Proof See [8, Lemma 23]. □

Table 1 Values of the coefficients γ_n for the anisotropic settings

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8
Test $N = 2$	1	0.0035						
Test $N = 4$	1	0.06	0.0035	0.0002				
Test $N = 8$	1	0.25	0.06	0.015	0.0035	0.0009	0.0002	0.00005

Inclusions are numbered counterclockwise, starting from the bottom-left (south-west) corner

Corollary 1 *The solution u corresponding to a diffusion coefficient (39) satisfies Assumption A1 with $\zeta_n^* = |T_n^*| + \sqrt{|T_n^*|^2 - 1}$, and with T_n^* as in Lemma 9.*

Proof We only need to compute the parameter ζ_n^* corresponding to the largest Bernstein ellipse contained in the analyticity region Σ. This can be done by enforcing $(\zeta_n^* + \zeta_n^{*-1})/2 = |T_n^*|$. □

7.1 Numerical results

In this section, we consider three different "inclusions" geometries, with $N = 2, 4$, and 8 inclusions (see Fig. 4), where $a_0 = 1$, $y_{min} = -0.99$, and $y_{max} = 0.99$, respectively. We set homogeneous Dirichlet boundary conditions and use a constant forcing term defined on the square subdomain F located in the center of the physical domain D (see again Fig. 4), i.e. $f(\mathbf{x}) = 100\chi_F(\mathbf{x})$. Each geometry is considered in both an isotropic setting (i.e., γ_n in (37) are set to 1 for all $n = 1, \ldots, N$) and an anisotropic setting (see Table 1 for the values of γ_n).

As already mentioned, we will test the performances of two different versions of the quasi-optimal sparse grid proposed in the previous section: one using nested sequences of Clenshaw–Curtis points and the profit estimates (17) and one with non-nested Gauss–Legendre points and the profit estimates (20); from here on, we will denote these two grids as "OPT-N" and "OPT-NN". In particular, we will verify the sharpness of the convergence ansatz

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq \mathcal{A}_n \exp\left(-\mathcal{B}_n N \sqrt[N]{W_{\mathcal{I}(w),m}} \right) \tag{40}$$

for nested points and

$$\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\|_{V \otimes L^2_\varrho(\Gamma)} \leq \mathcal{A}_{nn} \exp\left(-\mathcal{B}_{nn} N \sqrt[2N]{W_{\mathcal{I}(w),m}} \right) \tag{41}$$

for non-nested points, where $\mathcal{A}_n, \mathcal{B}_n, \mathcal{A}_{nn}, \mathcal{B}_{nn}$ are numerical values. In practice, we introduce the bounded linear functional $\Theta : V \to \mathbb{R}$,

$$\Theta(u) = \int_F u(\mathbf{x}) d\mathbf{x}$$

Table 2 Absolute and normalized values of the anisotropy rates g_n for the anisotropic settings

	Test $N = 2$	Test $N = 4$	Test $N = 8$
y_1	1.41 (1)	1.41 (1)	1.41 (1)
y_2	7.16 (5)	4.31 (3)	2.88 (2)
y_3		7.16 (5)	4.31 (3)
y_4		10.01 (7)	5.70 (4)
y_5			7.16 (5)
y_6			8.51 (6)
y_7			10.01 (7)
y_8			11.40 (8)

Normalized values are shown in parenthesis, and are defined as g_n/g_1. Inclusions are numbered anticlockwise, starting from the bottom-left (south-west) corner. Finally, in the isotropic setting, the value of g is $g = 1.41$ for all variables

and we monitor the convergence of the quantity

$$\varepsilon = \sqrt{\mathbb{E}\left[\left(\Theta(\mathcal{S}^m_{\mathcal{I}(w)}[u]) - \Theta(u)\right)^2\right]}, \tag{42}$$

with respect to number of sparse-grid points, that will converge with the same rate as the full error $\left\|u - \mathcal{S}^m_{\mathcal{I}(w)}[u]\right\|_{V\otimes L^2_\varrho(\Gamma)}$ given the linearity of Θ.

From a practical point of view, we have observed in [4,6,8] that estimating the rates $g_n = \log(\zeta_n^*)$ using the a priori estimates of ζ_n^* given in Lemma 9 and Corollary 1 in the construction of the sparse grids (cf. Eqs. (17) and (20)) leads to poorly performing sparse grids; such rates are therefore better estimated with the numerical procedure described in [4,6]. The resulting values are shown in Table 2; we will address the robustness of the performance of the quasi-optimal sparse grids with respect to such values at the end of this section.

Remark 7 In our numerical experiments, the set of the w largest profits is always downward closed, i.e. we never have to explicitly enforce the admissibility condition (2), both in the nested and non-nested case. By closely following the argument in [36, Chapter 6, Lemma 19], it is actually easy to show that, regardless of the values of $g_1, \ldots, g_N$, the set of the w largest profits is necessarily downward closed when considering Gauss–Legendre points. Conversely, the set of the w largest profits is downward closed when considering Clenshaw–Curtis points under the assumption that the rates g_n are sufficiently large; however, this condition is very mild ($g_n \geq \bar{g} \approx 0.13$) and satisfied by the values in Table 2.

We will furthermore compare the performances of the quasi-optimal sparse grids "OPT-N" and "OPT-NN", with those of a number of different sparse-grid schemes proposed in the literature. In particular, we will consider:

1. A standard sparse grid (labeled "SM") built with the classical Clenshaw–Curtis abscissas together with the level-nodes relation $m(i_n) = m_{db}(i)$, see Eq. (36), and

using the multi-index set

$$\mathcal{I}_{SM}(w) = \left\{ \mathbf{i} \in \mathbb{N}_+^N : \sum_{n=1}^{N} (i_n - 1) \leq w \right\},$$

and its anisotropic counterpart ("aSM")

$$\mathcal{I}_{aSM}(w) = \left\{ \mathbf{i} \in \mathbb{N}_+^N : \sum_{n=1}^{N} g_n(i_n - 1) \leq w \right\},$$

as is proposed in [4,31]; the rates g_n used here are the same as those listed in Table 2.

2. The isotropic and anisotropic Total Degree sparse grids (labeled respectively "TD" and "aTD") proposed in [4] with Gauss–Legendre points, $m(i_n) = m_{lin}(i)$, see Eq. (35), and

$$\mathcal{I}_{TD}(w) = \left\{ \mathbf{i} \in \mathbb{N}_+^N : \sum_{n=1}^{N} (i_n - 1) \leq w \right\},$$

$$\mathcal{I}_{aTD}(w) = \left\{ \mathbf{i} \in \mathbb{N}_+^N : \sum_{n=1}^{N} g_n(i_n - 1) \leq w \right\};$$

again, the g_n rates used here are those listed in Table 2.

3. The adaptive strategy proposed in [19], in the implementation provided by [26] and available at http://www.ians.uni-stuttgart.de/spinterp (labeled "AD"). As mentioned in the Introduction and in Sect. 3, this is an adaptive algorithm that explores the set of admissible hierarchical surpluses and adds to the sparse-grid approximation the most "profitable" ones, according to suitable a posteriori estimates. The implementation considered here has a tunable parameter $\widetilde{\omega}$ that allows one to move continuously from the standard sparse grid just described ($\widetilde{\omega} = 0$) to the fully adaptive algorithm ($\widetilde{\omega} = 1$). Following [26], in the present work we have set $\widetilde{\omega} = 0.9$, which has been proven numerically to be a well performing choice. This strategy is bounded to using nested (i.e. Clenshaw–Curtis) points, and (at least in the implementation considered here) only works on problems with a finite number of dimensions. We will measure the convergence of this algorithm in terms of the *total* number of points, i.e. including also those necessary to explore the set of hierarchical surpluses.
4. Two "brute force" approximations of the quasi-optimal sparse grid (one for nested points and one for non-nested points cases, labeled "BF-N" and "BF-NN", respectively) that we obtain by actually numerically computing the profits of all hierarchical surpluses in a sufficiently large "universe" $\mathbb{U} \subset \mathbb{N}_+^N$ (see Table 3; observe that this operation requires evaluating the quantities $\delta E(\mathbf{i})$ defined in Equation (4) using in turn a suitable quadrature strategy—in this case Monte Carlo) and then sorting the corresponding profits in decreasing order. Whenever such ordering does not satisfy the admissibility condition (2), all the hierarchical surpluses needed are added (this approach is equivalent to modifying the profits according to (11)).

Table 3 Universe $\mathbb{U}$ and sizes of Monte Carlo samples considered in each computational test. Due to the different level-to-nodes relations used, we use two different sets $\mathbb{U}$ for the nested and non-nested cases

Test case	$\mathbb{U}$-nested	$\mathbb{U}$-non nested	MC samples
Iso2D	TD(8)	TD(8)	6000
Iso4D	TD(8)	TD(8)	25,000
Iso8D	TD(10)	TD(10)	50,000
Aniso2D	TD(6)	TD(10)	5000
Aniso4D	TD(6)	TD(13)	15,000
Aniso8D	TD(10)	TD(13)	25,000

We calculated error (42) with a Monte Carlo sampling, see Table 3, and we emphasize that the number of Monte Carlo samples used was verified as being sufficient for our purposes. The same sampling strategy has also been employed for the computation of the profits needed to build the brute force sparse grids "BF-N" and "BF-NN". All schemes except the adaptive strategy "AD" have been implemented using the Matlab package [37], available for download.

We show the results for the isotropic and anisotropic settings in Figs. 5 and 6, respectively; from the analysis of the numerical results, several conclusions can be drawn.

First, the proposed profit estimates are quite sharp, both in nested and non-nested cases, since the convergence curves for the "brute force" sparse grids "BF-N"/"BF-NN" and their estimated counterparts "OPT-N"/"OPT-NN" are very close in every test. Observe that while this was expected for "BF-N" and "OPT-N", given the corresponding results presented in previous works [6,7], this was not obvious for "BF-NN" and "OPT-NN", given the pessimistic approach that was adopted to estimate the work contribution of each hierarchical surplus. The non-monotone convergence curve for the "BF-NN" scheme can be explained by the fact that increasing the number of multi-indices does not necessarily lead to an increase of the number of total points in a sparse grid when using non-nested points, as previously mentioned (cf. Fig. 2 and Example 1).

Second, "OPT-N" is found to be more efficient than "OPT-NN" as was expected given convergence Theorems 2 and 3; furthermore, "OPT-N" is found to be competitive with the a posteriori sparse-grid construction ("AD"), again in agreement with previous work [6].

Third, comparing the performance of "OPT-N" and "OPT-NN" with that of non-optimized sparse grids, like Smolyak and Total Degree ("SM"/"aSM" and "TD"/"aTD"), we see that the convergence behavior of "TD"/"aTD" closely resembles that of "OPT-NN", while the same does not hold true for the corresponding nested grids, i.e., "SM"/"aSM" versus "OPT-N". Indeed, if on the one hand in the isotropic setting "SM" is competitive with the nested quasi-optimal grid (although asymptotically less efficient, at least in the cases $N = 2, 4$), on the other hand "aSM" is instead quite less efficient than "OPT-N". Observe also the significant loss of efficiency caused by using isotropic approximations in the anisotropic setting as the number of variables increases, highlighting the need for anisotropic approximation schemes.

Next, we verify the sharpness of the theoretical bounds provided in Theorems 2 and 3, and in particular of the convergence ansatz (40)–(41). To this end, we plot, in

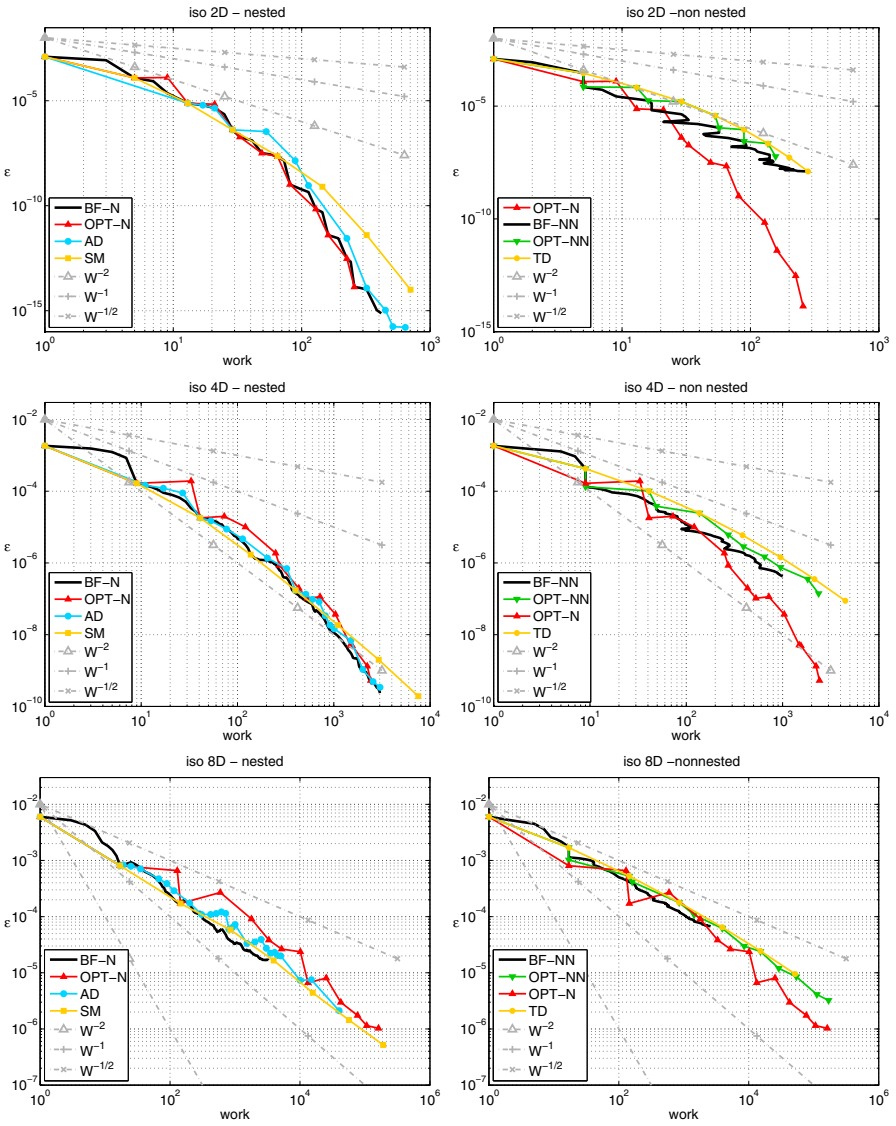

Fig. 5 Results for the **isotropic** setting. *Top row* case $N = 2$; *middle row* case $N = 4$; *bottom row* case $N = 8$. *Left column* sparse grids with nested points; *right column* sparse grids with non-nested points

semi-logarithmic scale, the quantities $N \sqrt[N]{W_{\mathcal{I}(w),m}}$, $N \sqrt[2N]{W_{\mathcal{I}(w),m}}$ versus the sparse-grid error for "OPT-N" and "OPT-NN" sparse grids. Results are shown in Fig. 7: the straight lines that we obtain indicate that the ansatz can be considered quite effective.

We also investigate the convergence of the expected value of $\Theta(u)$, $\varepsilon_{\mathbb{E}[\Theta]} = |\mathbb{E}[\mathcal{S}^m_{\mathcal{I}(w)}[\Theta(u)] - \mathbb{E}[\Theta(u)]|$; see Figs. 8 and 9. As expected, the convergence in this case is faster than the convergence of the $L^2_\varrho(\Gamma)$ norm inspected previously. More-

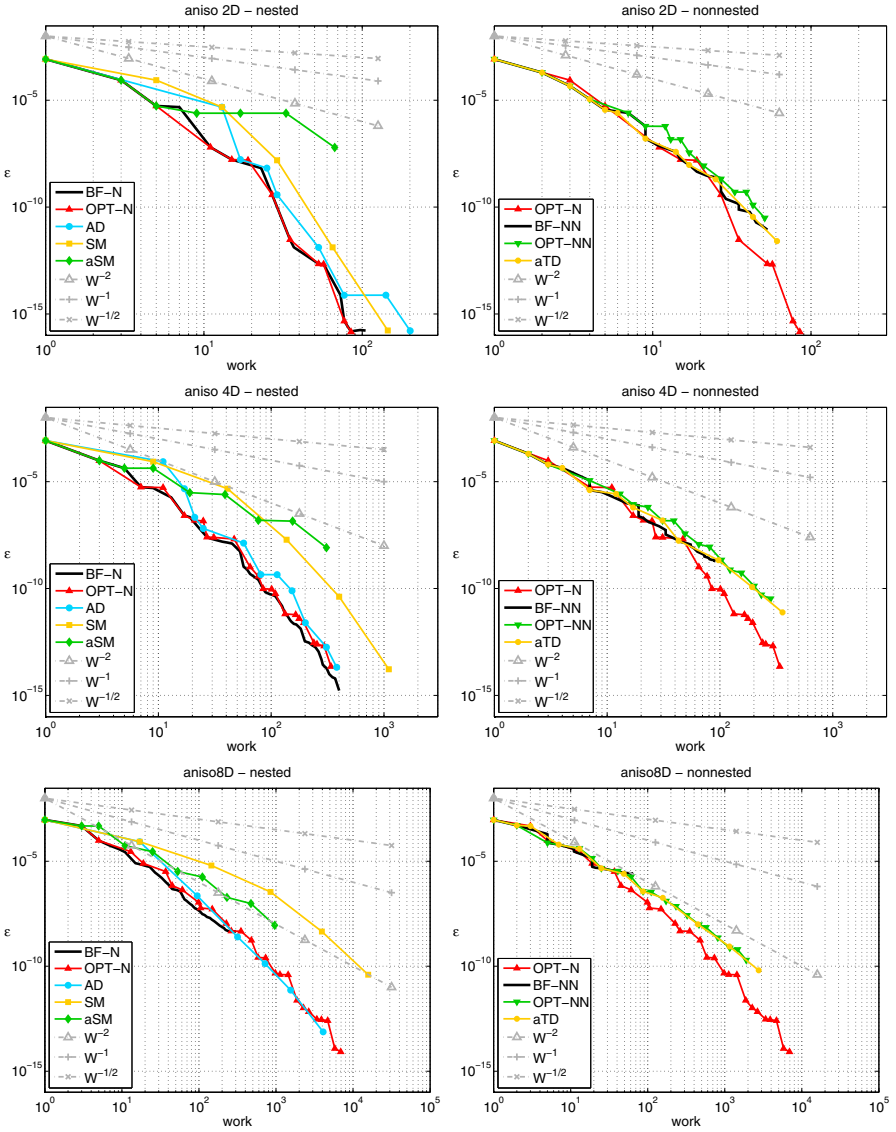

Fig. 6 Results for the **anisotropic** setting. *Top row* case $N = 2$; *middle row* case $N = 4$; *bottom row* case $N = 8$. *Left column* sparse grids with nested points; *right column* sparse grids with non-nested points

over, cancellations among hierarchical surpluses cause convergence to be significantly less smooth than the previous case. Note also that in the anisotropic setting non-nested grids are surprisingly competitive with nested schemes. A possible explanation for this is that in this test we are actually assessing the quadrature capabilities of the sparse grids rather than those of interpolation, for which Gaussian collocation points are particularly suitable. Other than this aspect, the other observations on the performances of sparse-grid schemes done in the previous analysis also apply to this case.

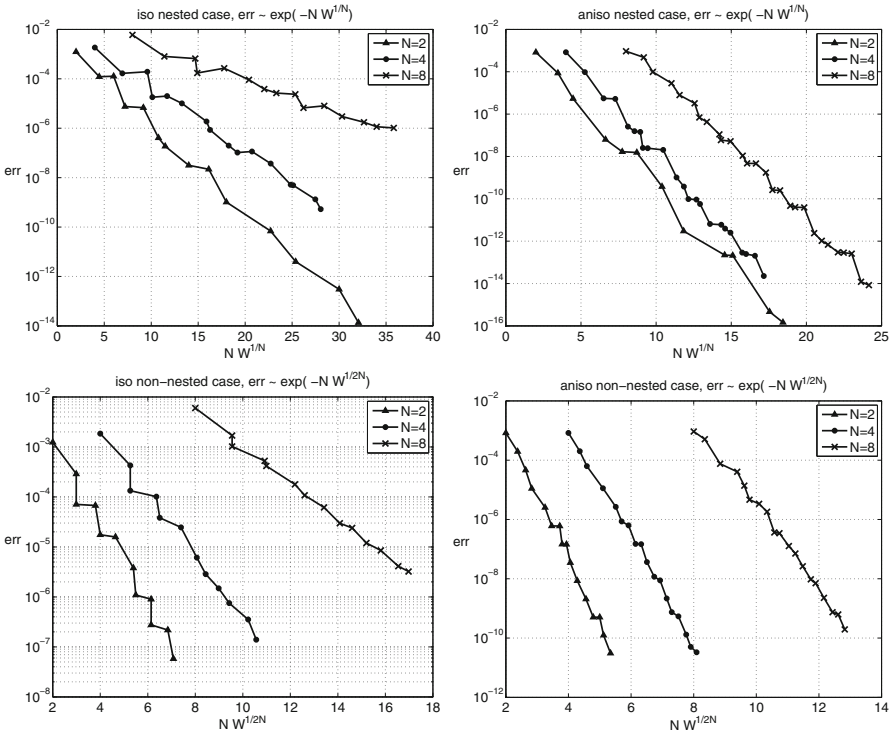

Fig. 7 Verification of the quasi-optimal sparse-grid convergence estimates. *Top row* nested case; *bottom row* non-nested case. *Left column* isotropic setting; *right column* anisotropic setting

Additional numerical results obtained on the same test for the quasi-optimal sparse grids built on Leja points can be found in the follow-up work [30]; these results show that the performance of quasi-optimal sparse grids with Leja points is essentially comparable to that of Clenshaw–Curtis points, at least when applied to problems with a moderate number of random parameters.

We conclude this section by numerically assessing the sensitivity of the performance of the quasi-optimal sparse grids with respect to the values of the coefficients g_i, appearing in Eqs. (17) and (20). To this end, we add some noise to the values of g_i in Table 2, and we compare the performance of the quasi-optimal grid thus constructed with the performance of the grids based on the "unperturbed" values of g_i. More precisely, we modify the value of each g_i by a fixed percentage; the sign of this perturbation was chosen at random. We performed three such tests with increasing levels of noise, 10, 25, and 50 %, generating three sets of randomly perturbed coefficients for each of the three noise levels.

Results obtained in the case of four isotropic inclusions are shown in Fig. 10 (similar results have also been obtained for test cases $N = 2$ and $N = 8$ inclusions, though they are not reported here). In particular, one can see that 10 % perturbations on rates do not significantly affect the resulting grids (top row), and to a certain extent even 25 % perturbations do not excessively alter the convergence of quasi-optimal sparse grids.

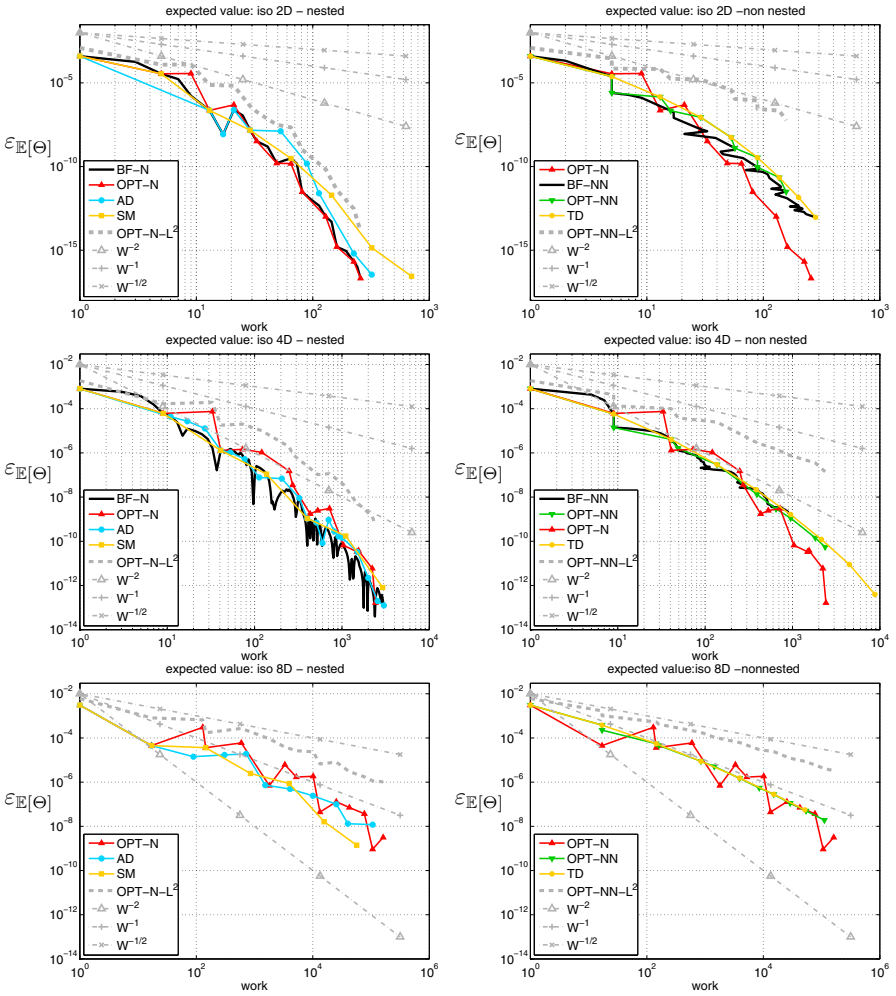

Fig. 8 Results for the **isotropic** setting, convergence of the expected value of $\Theta(u)$. *Top row* case $N = 2$; *middle row* case $N = 4$; *bottom row* case $N = 8$. *Left column* sparse grids with nested points; *right column* sparse grid with non-nested points

Observe that in some cases some of the "noisy grids" locally perform slightly better than the "unperturbed ones", possibly due to some cancellations among contributions ΔE (which we recall are not orthogonal).

8 Conclusions

In this work we proved an error estimate for stochastic collocation based on quasi-optimal sparse grids constructed by choosing the w most profitable hierarchical surpluses, that is the w surpluses with the highest error-reduction/cost ratio. The

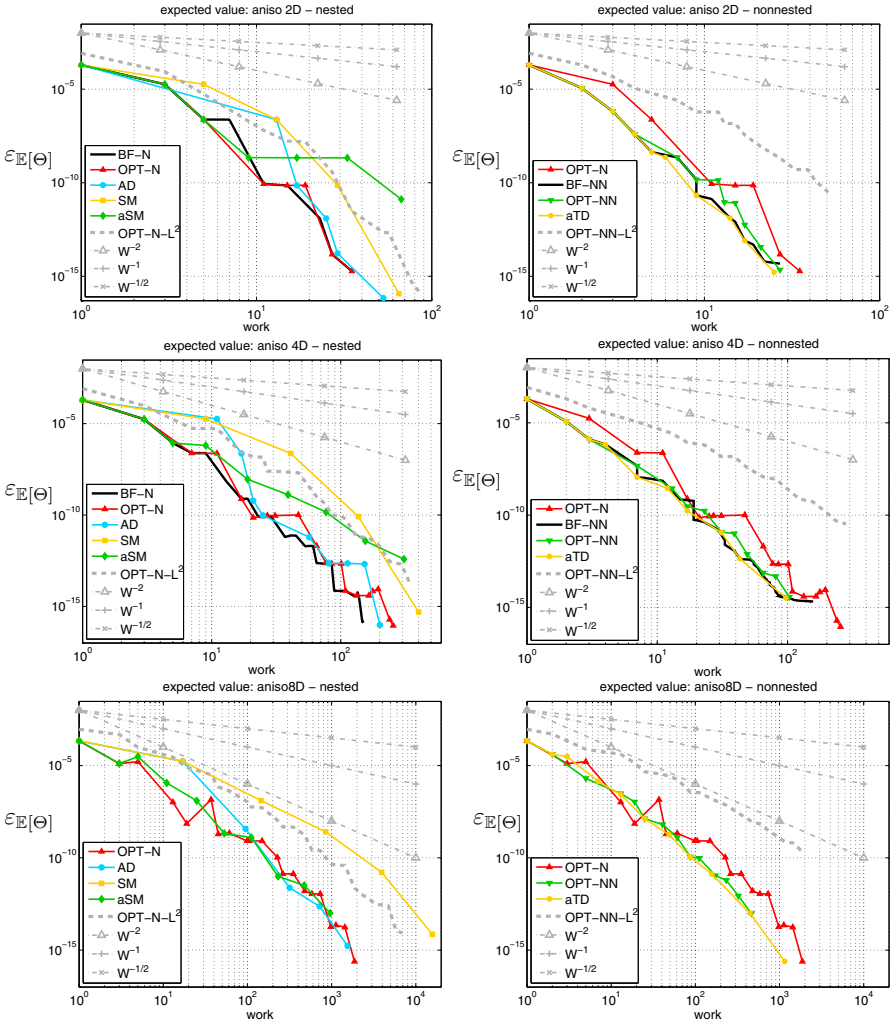

Fig. 9 Results for the **anisotropic** setting, convergence of the expected value of $\Theta(u)$. *Top row* case $N = 2$; *middle row* case $N = 4$; *bottom row* case $N = 8$. *Left column* sparse grids with nested points; *right column* sparse grid with non-nested points

convergence of these grids is proved in terms of weighted τ-summability of the profits: as the true profits are unknown, we propose to build the quasi-optimal sparse grid introducing a priori/a posteriori estimates on the decay of the profits. Next, we considered the application of such quasi-optimal sparse grid to Hilbert-space-valued functions which are analytic in certain polyellipses. We have considered two variations of the scheme, one using nested collocation points and the other using non-nested points; in both cases we were able to derive profit bounds and prove the corresponding τ-summability. After having verified that the solution of the so-called "inclusion problem" satisfied the above-mentioned analyticity Assumption, we used some numerical tests to show that the profit estimates are quite sharp and that the convergence results

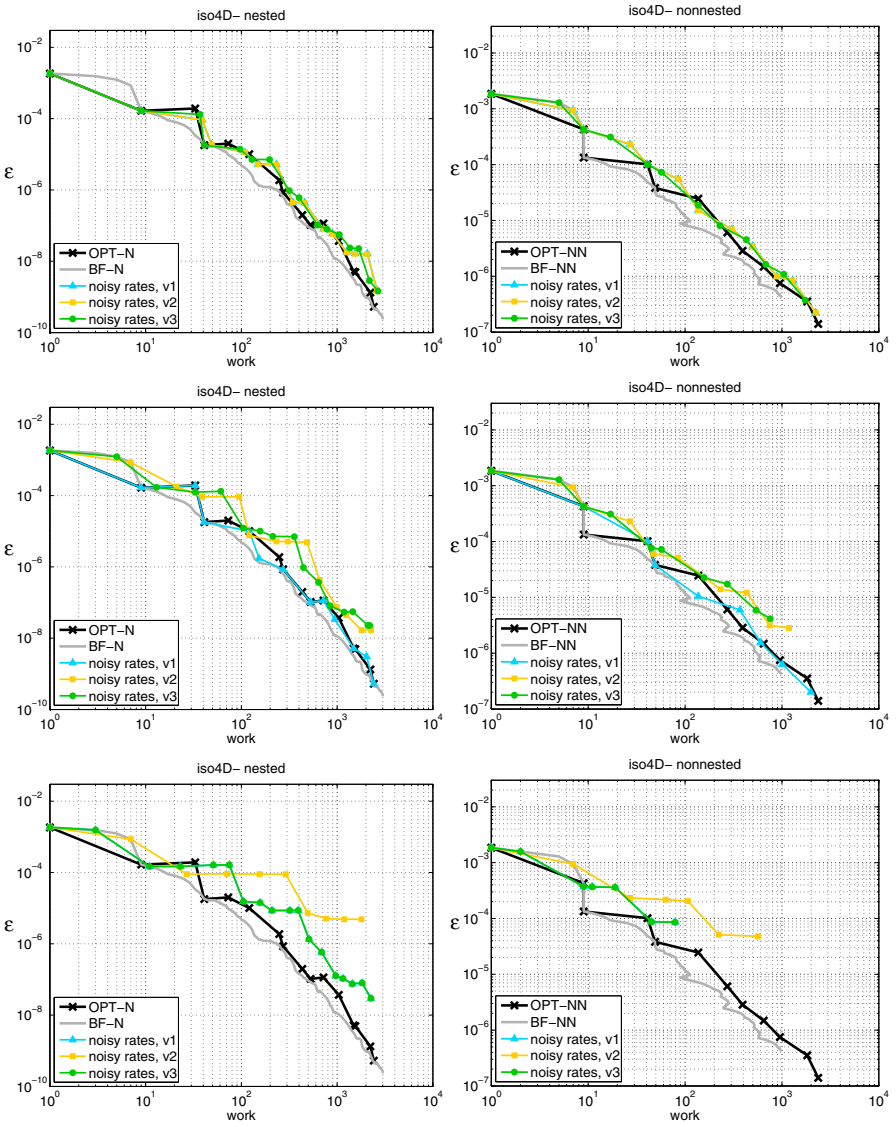

Fig. 10 Iso 4D problem with nested (*left column*) and non-nested (*right column*) sparse grids with noisy rates. From *top* to *bottom row* 10, 25, 50 % noise on rates. Three different random perturbations on the rates are shown in each plot

provide the correct ansatz for the error decay, although with constants fitted numerically. The proposed method is therefore competitive with the a posteriori adaptive scheme [26] and possibly outperforms the previously proposed anisotropic sparse grids. Moreover, the results obtained appear quite robust both with respect to the estimate of the constants g_n to be fitted and to the choice of univariate collocation points. Finally, as already mentioned, the dependence on N of the estimates proposed in The-

orems 2 and 3 is a consequence of having made no assumption on the growth of the sequence $\{\zeta_n\}_{n=1}^N$ with respect to n. Extensions to the infinite-dimensional case can be obtained following the lines of [14].

References

1. Babenko, K.I.: Approximation by trigonometric polynomials in a certain class of periodic functions of several variables. Soviet Math. Dokl. **1**, 672–675 (1960)
2. Babuška, I., Tempone, R., Zouraris, G.E.: Galerkin finite element approximations of stochastic elliptic partial differential equations. SIAM J. Numer. Anal. **42**(2), 800–825 (2004)
3. Babuška, I., Nobile, F., Tempone, R.: A stochastic collocation method for elliptic partial differential equations with random input data. SIAM Rev. **52**(2), 317–355 (2010)
4. Bäck, J., Nobile, F., Tamellini, L., Tempone, R.: Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison. In: Hesthaven, J.S., Ronquist, E.M. (eds.) Spectral and High Order Methods for Partial Differential Equations, Lecture Notes in Computational Science and Engineering, vol. 76, pp. 43–62. Springer, Berlin (2011) (selected papers from the ICOSAHOM '09 conference, June 22–26, Trondheim, Norway)
5. Barthelmann, V., Novak, E., Ritter, K.: High dimensional polynomial interpolation on sparse grids. Adv. Comput. Math. **12**(4), 273–288 (2000)
6. Beck, J., Nobile, F., Tamellini, L., Tempone, R.: On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. Math. Models Methods Appl. Sci. **22**(09) (2012)
7. Beck, J., Nobile, F., Tamellini, L., Tempone, R.: A quasi-optimal sparse grids procedure for groundwater flows. In: Azaiez, M., El Fekih, H., Hesthaven, J.S. (eds.) Spectral and High Order Methods for Partial Differential Equations, Lecture Notes in Computational Science and Engineering. Springer, Berlin (2012) (selected papers from the ICOSAHOM '12 conference)
8. Beck, J., Nobile, F., Tamellini, L., Tempone, R.: Convergence of quasi-optimal Stochastic Galerkin methods for a class of PDEs with random coefficients. Comput. Math. Appl. **67**(4), 732–751 (2014)
9. Bieri, M., Andreev, R., Schwab, C.: Sparse tensor discretization of elliptic SPDEs. SIAM J. Sci. Comput. **31**(6), 4281–4304 (2009/2010)
10. Bungartz, H.J., Griebel, M.: Sparse grids. Acta Numer. **13**, 147–269 (2004)
11. Chkifa, A.: On the lebesgue constant of leja sequences for the complex unit disk and of their real projection. J. Approx. Theory **166**, 176–200 (2013)
12. Chkifa, A., Cohen, A., Devore, R., Schwab, C.: Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs. ESAIM Math. Model. Numer. Anal. **47**(1), 253–280 (2013)
13. Chkifa, A., Cohen, A., Schwab, C.: High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. Foundations of Computational Mathematics, pp. 1–33 (2013)
14. Cohen, A., Devore, R., Schwab, C.: Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE'S. Anal. Appl. (Singap.) **9**(1), 11–47 (2011)
15. Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order, 2nd edn. Cambridge University Press, New York (2002)
16. DeVore, R.A., Lorentz, G.G.: Constructive Approximation. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer, Berlin (1993)
17. Dzjadik, V.K., Ivanov, V.V.: On asymptotics and estimates for the uniform norms of the Lagrange interpolation polynomials corresponding to the Chebyshev nodal points. Anal. Math. **9**(2), 85–97 (1983)
18. Ehlich, H., Zeller, K.: Auswertung der Normen von Interpolationsoperatoren. Math. Ann. **164**, 105–112 (1966)
19. Gerstner, T., Griebel, M.: Dimension-adaptive tensor-product quadrature. Computing **71**(1), 65–87 (2003)
20. Ghanem, R.G., Spanos, P.D.: Stochastic Finite Elements: A Spectral Approach. Springer, New York (1991)
21. Griebel, M., Knapek, S.: Optimized general sparse grid approximation spaces for operator equations. Math. Comput. **78**(268), 2223–2257 (2009)
22. Gui, W., Babuka, I.: The h, p and h-p versions of the finite element method in 1 dimension—part I. The error analysis of the p-version. Numer. Math. **49**(6), 577–612 (1986)

23. Haji-Ali, A.-L., Nobile, F., Tamellini, L., Tempone, R.: Multi-index stochastic collocation for random PDEs. arXiv:1508.07467 (2015, e-print)
24. Haji-Ali, A.-L., Nobile, F., Tempone, R.: Multi-Index Monte Carlo: when sparsity meets sampling. Numer. Math., 1–40 (2015)
25. Harbrecht, H., Peters, M., Siebenmorgen, M.: On multilevel quadrature for elliptic stochastic partial differential equations. In: Sparse Grids and Applications, Lecture Notes in Computational Science and Engineering, vol. 88, pp. 161–179. Springer, Berlin (2013)
26. Klimke, A.: Uncertainty modeling using fuzzy arithmetic and sparse grids. PhD thesis, Universität Stuttgart, Shaker Verlag, Aachen (2006)
27. Le Maître, O.P., Knio, O.M.: Spectral methods for uncertainty quantification. Scientific Computation. Springer, New York (2010) (with applications to computational fluid dynamics)
28. Lubich, C.: From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis, Lectures in Advanced Mathematics. European Mathematical Society, Zurich (2008)
29. Martello, S., Toth, P.: Knapsack Problems: Algorithms and Computer Implementations. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, New York (1990)
30. Nobile, F., Tamellini, L., Tempone, R.: Comparison of Clenshaw–Curtis and Leja quasi-optimal sparse grids for the approximation of random PDEs. In: Spectral and High Order Methods for Partial Differential Equations—ICOSAHOM '14, Lecture Notes in Computational Science and Engineering, vol. 106. Springer, Berlin (2015, to appear) (also available as MATHICSE report 41/2014)
31. Nobile, F., Tempone, R., Webster, C.G.: An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal. **46**(5), 2411–2442 (2008)
32. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, Berlin (1999)
33. Patterson, T.N.L.: The optimum addition of points to quadrature formulae. Math. Comput. **22**, 847–856 (1968) [addendum, Math. Comput. **22**(104), C1–C11 (1968)]
34. Schillings, C., Schwab, C.: Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. Inverse Probl. **29**(6) (2013)
35. Shen, Jie, Wang, Li-Lian: Sparse spectral approximations of high-dimensional problems based on hyperbolic cross. SIAM J. Numer. Anal. **48**(3), 1087–1109 (2010)
36. Tamellini, L.: Polynomial approximation of PDEs with stochastic coefficients. PhD thesis, Politecnico di Milano (2012)
37. Tamellini, L., Nobile, F.: Sparse Grids Matlab kit v. 15-8. http://csqi.epfl.ch (2011–2015)
38. Teckentrup, A.L., Jantsch, P., Webster, C.G., Gunzburger, M.: A multilevel stochastic collocation method for partial differential equations with random input data. arXiv:1404.2647 (2014, e-print)
39. Trefethen, L.N.: Is Gauss quadrature better than Clenshaw-Curtis? SIAM Rev. **50**(1), 67–87 (2008)
40. Trefethen, L.N.: Approximation Theory and Approximation Practice. Society for Industrial and Applied Mathematics (2013)
41. van Wyk, H.W.: Multilevel sparse grid methods for elliptic partial differential equations with random coefficients. arXiv:1404.0963 (2014, e-print)
42. Wasilkowski, G.W., Wozniakowski, H.: Explicit cost bounds of algorithms for multivariate tensor product problems. J. Complex. **11**(1), 1–56 (1995)