

Sparse Approximation Algorithms for High Dimensional Parametric Initial Value Problems

M. Hansen and C. Schillings and C. Schwab

Research Report No. 2013-10
March 2013

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Sparse Approximation Algorithms for High Dimensional Parametric Initial Value Problems [★]

M. Hansen, Cl. Schillings, and Ch. Schwab

Abstract We consider the efficient numerical approximation on nonlinear systems of initial value Ordinary Differential Equations (ODEs) on Banach state spaces \mathcal{S} over \mathbb{R} or \mathbb{C} . We assume the right hand side depends *in affine fashion* on a vector $y = (y_j)_{j \geq 1}$ of possibly countably many parameters, normalized such that $|y_j| \leq 1$. Such affine parameter dependence of the ODE arises, among others, in mass action models in computational biology and in stoichiometry with uncertain reaction rate constants. We review results from [19] on N -term approximation rates for the parametric solutions, i.e. summability theorems for coefficient sequences of generalized polynomial chaos (gpc) expansions of the parametric solutions $\{X(\cdot; y)\}_{y \in U}$ with respect to tensorized polynomial bases of $L^2(U)$. We give sufficient conditions on the ODEs for N -term truncations of these expansions to converge on the entire parameter space with efficiency (i.e. accuracy versus complexity) being independent of the number of parameters viz. the dimension of the parameter space U . We investigate a heuristic adaptive approach for computing sparse, approximate representations of the $\{X(t; y) : 0 \leq t \leq T\} \subset \mathcal{S}$. We increase efficiency by relating the accuracy of the adaptive initial value ODE solver to the estimated detail operator in the Smolyak formula. We also report tests

Markus Hansen

Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland
e-mail: markus.hansen@sam.math.ethz.ch.

Claudia Schillings

Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland
e-mail: claudia.schillings@sam.math.ethz.ch.

Christoph Schwab

Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland
e-mail: christoph.schwab@sam.math.ethz.ch.

[★] Research supported in part by the European Research Council (ERC) under the FP7 program AdG247277 and by the Eidgenössische Technische Hochschule (ETH) Zürich under the ETH-Fellowship Grant FEL-33 11-1

which indicate that the proposed algorithms and the analyticity results of [19] hold for more general, nonaffine analytic dependence on parameters.

Key words: Ordinary differential equations, initial value problem, parametric dependence, analyticity in infinite dimensional spaces, Taylor series, N -term approximation, adaptive sparse grid, Smolyak interpolation, Leja points.

1 Introduction

Numerous systems in engineering and life- and in social sciences are modelled by initial value ordinary differential equations (ODEs). In particular, complex systems require state spaces \mathcal{S} of high or even infinite dimension.

In recent years, in particular in connection with applications in life-sciences, climate-sciences but also in economics, particular attention has been paid to *initial value ODE models for systems with uncertainty*. We mention only stoichiometric descriptions of biochemical reaction pathways with uncertain reaction rate constants, chemical reaction cascades with uncertain reaction rate constants, mass action models with uncertain reaction rates. In complex systems, the goal of computation is in *obtaining the system characteristics on the entire parameter space in one single numerical forward simulation*. Besides the efficient *numerical forward solution* of parametric initial value ODEs by combination of *adaptive parameter collocation approaches* with *adaptive numerical initial value solvers* such as [17, 18] and the references there, additional problems consist in optimization resp. in optimal control of systems described by initial value ODEs.

Some form of *Sparsity* in the parametric dependence of the solution (resp. the control resp. the optimum) is necessary in order to allow for efficient approximations of the parametric solutions on the entire, possibly high-dimensional parameter space. Here, we present theoretical results from [19] on the sparsity of solutions of parametric ODEs and propose computational approaches which allow to exploit computationally the sparse parameter dependence of the solutions.

Unless stated otherwise, the state space \mathcal{S} is assumed to be a separable, reflexive Banach space, and will be understood over the coefficient field \mathbb{R} ; occasionally, however, we shall also work with the extension of \mathcal{S} to the coefficient field \mathbb{C} . By $\mathbb{R}^{\mathbb{N}}$ and $\mathbb{C}^{\mathbb{N}}$, we denote the countable cartesian products of \mathbb{R} and \mathbb{C} , respectively. Likewise, $U = (-1, 1)^{\mathbb{N}}$ will denote the countable product of the open interval $(-1, 1)$ and $\bar{U} = [-1, 1]^{\mathbb{N}}$. We shall denote the state of the system by $X(t) \in \mathcal{S}$ for $t \in [0, T]$. The parameter dependence of X on the parameter sequence $y \in U$ is indicated by $X(t; y)$.

On the parameter domain U , we consider *high-dimensional, parametric, deterministic ODE initial value problems (ODE IVP)*:

Given $x_0(y) \in \mathcal{S}$ and $T \in (0, \infty)$, find $X(t, x_0; y) : [t_0, T] \times \mathcal{S} \times U \rightarrow \mathcal{S}$ such that in \mathcal{S}

$$\frac{dX}{dt} = f(t, X; y), \quad X(t_0; y) = x_0(y), \quad t_0 \leq t \leq T, \quad \forall y \in U. \quad (1)$$

Here, \mathcal{S} denotes the *state space* of the parametric model (1). We shall mostly be concerned with the case of initial value ordinary differential equations (ODEs), when $\mathcal{S} = \mathbb{R}^d$, with particular attention to the case of high or even infinite dimensional state spaces, i.e. \mathbb{R}^d with large d , but [19] covers also the infinite dimensional case, when \mathcal{S} is a *separable and reflexive Banach space*.

We denote $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. We use standard multiindex notation: for a vector $y = (y_j)_{j \geq 1}$ of parameters and for a sequence $\nu \in \mathbb{N}_0^{\mathbb{N}}$ of nonnegative integers, we denote by $\mathfrak{F} = \{\nu \in \mathbb{N}_0^{\mathbb{N}} : |\nu| < \infty\}$. As any $\nu \in \mathfrak{F}$ has only finitely many nonzero entries, the definitions

$$\nu! = \prod_{j \in \mathbb{N}} \nu_j!, \quad |\nu| = \sum_{j \in \mathbb{N}} \nu_j, \quad \partial_y^\nu = \frac{\partial^{|\nu|}}{\partial y_1^{\nu_1} \partial y_2^{\nu_2} \dots}$$

for multi-factorials, the length of a multi-index ν and for the partial derivative of order ν are well-defined for $\nu \in \mathfrak{F}$.

In practice efficient solution methods in the case where the number of parameters is large are of interest. In particular, it would be highly desirable to identify methods which are *dimensionally robust*, i.e. whose efficiency (meaning accuracy versus computational cost measured in terms of the total number of floating point operations to achieve this accuracy) is *provably robust with respect to the number of parameters* which requires consideration of (1) for *parameter sequences*. In [19] we showed, analogously to earlier results for linear, elliptic partial differential equations [10, 11, 8, 16] sparsity of the parametric solutions' dependence on y .

It is well-known and classical (e.g. [23, Chap. 13]) that for parametric right hand sides $f(t, X; y)$ which are Lipschitz continuous with respect to (t, X) and which depend *analytically* on the parameters y , the solution $X(t; y)$ in turn depends analytically on the parameter vector y . In [19], we extended the proof in [23] of this (classical) result to a possibly countable number of parameters with *quantitative bounds on the size of domains of analyticity*. This allows us to establish in [19] *best N -term convergence rates for parametric expansions of the solution $X(t; y)$* under a *sparsity hypothesis* on the vector field $f(t, X; y)$. The rates of best N -term approximation are shown to be achievable with *N -term truncated Taylor expansions* of the solution $X(t, y)$ in the parameter space U which we prove to converge uniformly for all y belonging to the parameter domain U . The key mathematical principle behind these results is the fact that *sparsity in the input vector field $f(t, X; y)$ implies sparsity (in a sense to be made precise below) in the parametric solution's Taylor expansion*

$$X(t; y) = \sum_{\nu \in \mathfrak{F}} T_\nu(t) y^\nu, \quad T_\nu(t) := \frac{1}{\nu!} (\partial_y^\nu X(t; y))|_{y=0}, \quad t_0 \leq t \leq T, \quad y \in U. \quad (2)$$

In [19], similar results are also established for other polynomial expansions of the solution, such as Legendre or Chebyshev expansions.

The theoretical result on sparse parameter dependence in [19] opens the perspective of dimensionally robust, adaptive algorithms for the efficient solution of large systems of parametric ODE's on possibly infinitely dimensional parameter spaces. This requires to address the following issues: first, under the (*unrealistic*) *assumption of having available exact solutions of the ODE IVP (1) for a single instance of the parameter vector $y \in U$ at unit cost*, concrete sequences of sparse, *monotone* index sets $\mathcal{M}_N \subset \mathfrak{F}$ (to which we will also refer as “sparsity models”) for at most N “active” Taylor coefficients $T_\nu(t)$, $\nu \in \mathcal{M}_N$, can be constructed such that the corresponding, finitely truncated parametric expansions

$$X_{\mathcal{M}_N}(t; y) = \sum_{\nu \in \mathcal{M}_N} T_\nu(t) y^\nu \quad (3)$$

realize the best N -term asymptotic convergence rate.

One particular class of sparsity models are *monotone index sets* $\Lambda \subset \mathfrak{F}$ which were introduced in [8] in the context of adaptive Taylor approximations of parametric elliptic partial differential equations. This notion is based on the following ordering of \mathfrak{F} : for any two indices $\mu, \nu \in \mathfrak{F}$, we say that $\mu \leq \nu$ if and only if $\mu_j \leq \nu_j$ for all $j \geq 1$. We will also say that $\mu < \nu$ if and only if $\mu \leq \nu$ for all $j \in \mathbb{N}$ and if $\mu_j < \nu_j$ for at least one value of j .

Definition 1. A sequence $(a_\nu)_{\nu \in \mathfrak{F}}$ of nonnegative real numbers is said to be *monotone decreasing* if and only if for all $\mu, \nu \in \mathfrak{F}$

$$\mu \leq \nu \Rightarrow a_\nu \leq a_\mu.$$

A set $\emptyset \neq \Lambda \subset \mathfrak{F}$ is called *monotone* if and only if $\nu \in \Lambda$ and $\mu \leq \nu \Rightarrow \mu \in \Lambda$.

Once concrete, monotone \mathcal{M}_N sparsity models have been selected, the evaluation of the truncations (3) requires approximation of the expansion coefficients $T_\nu(t)$ in (2) for $\nu \in \mathcal{M}_N$. Naturally, the assumption of an exact solution of the ODE IVP (1) for a single instance of the parameter vector y in $O(1)$ work and memory is not realistic. Thus to still achieve the rate of best N -term approximation also for the approximate partial sums

$$\tilde{X}_{\mathcal{M}_N}(t; y) = \sum_{\nu \in \mathcal{M}_N} \tilde{T}_\nu(t) y^\nu, \quad (4)$$

where $\tilde{T}_\nu(t) \in \mathcal{S}$ are the Taylor coefficients obtained with an approximate initial value ODE solver, the effort for computing the coefficients has to be

balanced against the respective impact for approximating $X(t; y)$ by an initial value ODE solver.

In doing so, we obtain an approximate, adaptive numerical solution of the parametric ODE IVP (1) to a prescribed accuracy ε *uniformly on the entire parameter domain* U . This ultimately enables us to approximately calculate all further relevant information about the parametric solution (e.g. statistical moments), again up to an arbitrary prescribed accuracy, by several classes of adaptive approximation algorithms based on Galerkin projection (see, e.g. [14]) or by sparse collocation as in [21, 4, 3] or by adaptive truncation (4) of the Taylor expansions (2) as in [8]. We let B denote a separable Banach space both over \mathbb{R} as well as its complexification over \mathbb{C} (i.e. an extension of B whose restriction to real valued elements coincides with the original space B). We shall need spaces of (differentiable) functions with values in B . We denote by $C(U; B) \equiv C^0(U; B)$ the space of functions from U into B which are, as B -valued functions, continuous on U (where U is equipped with the product topology). Moreover, for any $k \in \mathbb{N}$, we denote by $C^k([0, T]; B)$ the space of functions $f : [0, T] \rightarrow B$ whose k -th Fréchet derivative $\frac{d^k f}{dt^k}$ with respect to $t \in [0, T]$ belongs to $C^0([0, T]; B)$. These spaces $C^k([0, T]; B)$, equipped with the norms

$$\|f\|_{C^k([0, T]; B)} := \max_{0 \leq j \leq k} \left\{ \left\| \frac{d^j f}{dt^j} \right\|_{C^0([0, T]; B)} \right\}, \quad k \in \mathbb{N}, \quad (5)$$

are themselves Banach spaces. Similar notations are used, if the interval $[0, T]$ is itself replaced by another Banach space \mathcal{S} . Then the derivatives $\frac{df}{dx}$ have to be understood as Fréchet derivatives, i.e. $\frac{df}{dx}$ is a mapping from \mathcal{S} taking values in $\mathcal{L}(\mathcal{S}, B)$, the space of bounded linear operators from \mathcal{S} into B . Finally, spaces of locally Lipschitz continuous functions will be defined below.

2 Parametric Initial Value ODEs

For a parameter sequence $y = (y_j)_{j \geq 1} \in U$ and a Banach state space \mathcal{S} , we assume given an initial state $x_0(y) \in \mathcal{S}$ and a parametric family of vector fields $f(t, X; y) : [0, T] \times \mathcal{S} \times U \mapsto \mathcal{S}$. Then we are interested in solving (1) numerically to a prescribed tolerance *uniformly for all values* $y \in U$.

As we think of applications to large mass-action models in computational chemistry and biology, attention will be in the following on the particular case when the dependence of the vector field f in (1) on the parameter vector $y \in U$ is affine, i.e. for every $t \in [0, T]$ and every $X \in \mathcal{S}$,

$$f(t, X; y) = f_0(t, X) + \sum_{j \geq 1} y_j f_j(t, X), \quad 0 \leq t \leq T < \infty. \quad (6)$$

Here, we assume that each $f_j \in (f_j)_{j \geq 0}$ is continuous with respect to t and satisfies certain Lipschitz conditions with respect to X uniform in $t \in [0, T]$. For the non-parametric problem $\frac{dX}{dt} = g(t, X)$, $X(t_0) = x_0$, it is classical that the right-hand-side g being locally Lipschitz continuous, i.e. for every $X_0 \in \mathcal{S}$ there is a neighbourhood $U = U(X_0)$ such that

$$\forall X, X' \in U \quad \forall t \in [0, T] : \quad \|g(t, X) - g(t, X')\|_{\mathcal{S}} \leq L(X_0) \|X - X'\|_{\mathcal{S}} \quad (7)$$

for some constants $L(X_0)$, implies existence and uniqueness of local solutions, i.e. existence of unique solutions on some maximally extended subinterval $[0, \delta) \subset [0, T]$, see e.g. [12]. To obtain global, parametric solutions we imposed in [19] a local Lipschitz condition: for every $R > 0$, there exist constants $L(R) > 0$ such that for every $X, X' \in B_R = \{X \in \mathcal{S} : \|X\|_{\mathcal{S}} \leq R\}$ and for every $t \in [0, T]$ holds

$$\|g(t, X) - g(t, X')\|_{\mathcal{S}} \leq L(R) \|X - X'\|_{\mathcal{S}},$$

where

$$L(R) := \|g\|_{\ell\text{Lip}(\mathcal{S}, R)} = \sup_{t \in [0, T], X \neq X' \in B_R} \frac{\|g(t, X) - g(t, X')\|_{\mathcal{S}}}{\|X - X'\|_{\mathcal{S}}} < \infty.$$

A continuous function g belongs to $\ell\text{Lip}(\mathcal{S})$, if $L(R) < \infty$ for all $R > 0$. The subclass $\ell\text{Lip}_0(\mathcal{S})$ consists of all functions $g \in \ell\text{Lip}(\mathcal{S})$ which additionally fulfill $g(t, 0) = 0$ for all $t \in [0, T]$. Then $\ell\text{Lip}_0(\mathcal{S})$ equipped with the increasing family of norms $\|\cdot\|_{\ell\text{Lip}(\mathcal{S}, R)}$ becomes a complete locally convex vector space. Our main assumption on (6) is $f_j \in \ell\text{Lip}_0(\mathcal{S})$ for all j , i.e. for $j = 0, 1, 2, \dots$ holds

$$L_j(R) = \sup_{t \in [0, T], X \neq X' \in B_R} \frac{\|f_j(t, X) - f_j(t, X')\|_{\mathcal{S}}}{\|X - X'\|_{\mathcal{S}}} < \infty, \quad f_j(t, 0) = 0. \quad (8)$$

In order to prove results which are independent of the number of terms in the affine expansion (6), we shall further require *summability of the coefficient sequence* $(f_j)_{j \geq 1}$. Specifically, we assume the sequence of Lipschitz constants to be summable, i.e.

$$\forall R > 0 : \quad (L_j(R))_{j \geq 1} \in \ell^1(\mathbb{N}). \quad (9)$$

Under this assumption, the sum in (6) converges uniformly with respect to $y \in U$ and for all $(t, X) \in [0, T] \times \mathcal{S}$. In [19], we showed

Proposition 1. *Let the conditions (8) and (9) be satisfied. Then the sum in (6) converges absolutely and uniformly in U as a $\ell\text{Lip}_0(\mathcal{S})$ -valued mapping.*

Moreover, we may also consider \mathcal{S} to be a complex Banach space: besides being of independent interest, the proofs in [19] used analytic continuations

and complex variable techniques even for problems with real-valued parameter sequences $y \in U$. In [19] we showed

Theorem 1. *Assume (8) and (9). Moreover, suppose the initial condition $x_0 \in C(\mathcal{U}, \mathcal{S})$ satisfies*

$$\sup_{z \in \mathcal{U}} \|x_0(z)\|_{\mathcal{S}} \leq (1 - \kappa)r, \quad r = Re^{-L(R)T/\kappa} \quad (10)$$

for some $R > 0$ and $0 < \kappa < 1$, where $\mathcal{U} = \{\zeta \in \mathbb{C} : |\zeta| < 1\}^{\mathbb{N}}$.

Then the IVP (1) (with $t_0 = 0$) admits a unique solution $X \in \mathcal{B}_{r,R}^1 \subset C^1([0, T]; C(\mathcal{U}; \mathcal{S}))$, where

$$\mathcal{B}_{r,R}^1 = \left\{ Y \in C^1([0, T]; C(\mathcal{U}; \mathcal{S})) : \sup_{(t,z) \in [0,T] \times \mathcal{U}} e^{-tL(R)/\kappa} \|Y(t,z)\|_{\mathcal{S}} \leq r \right\}$$

If, in addition, for some $k \in \mathbb{N}$

$$\forall j \geq 0 : f_j : [0, T] \times \mathcal{S} \longrightarrow \mathcal{S} \text{ is } k\text{-times continuously differentiable,} \quad (11)$$

then for every $z \in \mathcal{U}$ the unique solution $X(\cdot, x_0(z); z)$ of (1) belongs to $C^{k+1}([0, T]; \mathcal{S})$.

Moreover, the solution $X(\cdot, x_0; z)$ depends continuously on the data x_0 and parameters z . If additionally the functions f_j are analytic as $C^k([0, T]; \mathcal{S})$ -valued mappings, then X is analytic on \mathcal{U} as a $C^{k+1}([0, T]; \mathcal{S})$ -valued mapping.

3 Sparsity

It was shown in [19] that if the sequence f_j in (1) is sparse in the sense that if $(\|f_j\|_{\ell\text{Lip}_0(\mathcal{S}, R)})_{j \geq 1} \in \ell^p(\mathbb{N})$ for all $R > 0$ for some $0 < p < 1$, then the sequence $(T_\nu)_{\nu \in \mathfrak{F}}$ of Taylor coefficients of the solution is equally sparse.

Theorem 2. *Consider the parametric IVP ODE (1) for parameter vectors $y \in U = [-1, 1]^{\mathbb{N}}$. If there exist real numbers $R > 0$ and $0 < \kappa < 1$ with the following properties:*

1. In (1) the vector field f depends on the parameter vector y in the affine fashion (6) with the coefficient functions f_j satisfying for some $0 < p < 1$

$$\left(\|f_j\|_{\ell\text{Lip}_0(\mathcal{S}, R)} \right)_{j \geq 1} \in \ell^p(\mathbb{N}) \quad \text{and} \quad \left(\bar{\rho}_j \|f_j\|_{\ell\text{Lip}_0(\mathcal{S}, R)} \right)_{j \geq 1} \in \ell^1(\mathbb{N}), \quad (12)$$

where the scaling vector $\bar{\rho}$ is given by $\bar{\rho}_j = \max\left(1, \frac{\delta}{4L_j(R)}\right)$ for some arbitrary fixed $\delta > 0$, and $L_j(R) := \|f_j\|_{\ell\text{Lip}_0(\mathcal{S}, R)}$.

2. The initial data $x_0 \in C([0, T] \times U_{\bar{\rho}}; \mathcal{S})$ satisfies

$$\sup_{z \in U_{\bar{p}}} \|x_0(z)\| \leq (1 - \kappa) R e^{-TL(\bar{p}, R)/\kappa}. \quad (13)$$

Then the Taylor expansion (2) of the parametric solution $X(t; y)$ of (1) is p -sparse in the following sense: for every $N \in \mathbb{N}$, there exists a monotone set $\Lambda_N \subset \mathfrak{F}$ of indices $\nu \in \mathfrak{F}$ corresponding to N Taylor coefficients T_ν with largest norm in $C_{L(\bar{p}, R)/\kappa}([0, T]; \mathcal{S})$ such that it holds

$$\sup_{y \in U} \left\| X(\cdot; y) - \sum_{\nu \in \Lambda_N} T_\nu(t) y^\nu \right\|_{L(\bar{p}, R)/\kappa, T, \mathcal{S}} \leq CN^{-r}, \quad r = \frac{1}{p} - 1 \quad (14)$$

and where

$$\sum_{\nu \in \Lambda_N} T_\nu(t) y^\nu \in \mathbb{P}_{\Lambda_N}(U; C^1([0, T]; \mathcal{S})). \quad (15)$$

For $0 < p \leq 1$ as in (12), $(T_\nu)_{\nu \in \mathfrak{F}} \in \ell^p(\mathfrak{F}; C^1([0, T]; \mathcal{S}))$. Finally, let (11) be satisfied for some $k \geq 0$. Denote by $\Lambda_N^k \subset \mathfrak{F}$ a monotone set of N largest Taylor coefficients (measured in $C_{L(\bar{p}, R)/\kappa}^{k+1}([0, T]; \mathcal{S})$). Then

$$\sup_{y \in U} \left\| X(\cdot; y) - \sum_{\nu \in \Lambda_N^k} T_\nu(t) y^\nu \right\|_{C_{L(\bar{p}, R)/\kappa}^{k+1}([0, T]; \mathcal{S})} \leq CN^{-r}, \quad r = \frac{1}{p} - 1. \quad (16)$$

In [19], also results analogous to Theorem 2 for N -term approximations with monotone index sets for tensorized Legendre and Chebyshev systems are proved. The sparsity result Theorem 2 yields the *existence* of a family of sparse, N -term polynomial approximations of the parametric solutions $X(t; y)$. Apart from monotonicity its proof does not shed light on the structure resp. on the construction of concrete sets $\Lambda_N \subset \mathfrak{F}$ which would yield the proven convergence rate with, possibly, a suboptimal constant.

Unlike in the case of linear problems which was considered in [8], due to the strongly nonlinear nature of the problem (1), stable computation of Taylor coefficients $T_\nu(t)$ is, in general, not advisable (although in biological systems engineering schemes are developed for efficient computation of *sensitivities* $T_{e_j}(t)$).

We therefore consider *collocation approximations* of (1) using Smolyak type collocation operators which are unisolvent on monotone sets (see, e.g. [9]). In order to exploit sparsity in polynomial expansions of the parametric solutions, as provided by Theorem 2, with collocation schemes, it is important that for monotone sets $\Lambda \subset \mathfrak{F}$ of “active” polynomial coefficients we have available unisolvent, sparse polynomial interpolants.

4 Numerical Examples

In the present section, we present *heuristic adaptive algorithms* which attempt to iteratively localize a sequence $\{\Lambda_N\}_{N \in \mathbb{N}}$ of monotone sets which, although possibly not optimal in the sense of best N -term approximation, will deliver the optimal rate for given summability of the parametric inputs. We place particular attention on high-dimensional parameter spaces, but also investigate the scaling of the proposed algorithms with respect to the dimension p of the state space \mathcal{S} (always assumed here to be finite dimensional, i.e. $\mathcal{S} = \mathbb{R}^p$).

We emphasize that the examples which are presented here are illustrative model problems, and that the development of “industrial strength” numerical solvers for high-dimensional, parametric initial value ODEs is, currently, in its infancy; the present section is intended to give a first indication of scaling and performance of the proposed methods, and, in particular, also identifies specific directions for further algorithm development.

The proposed algorithm successively tries to identify the most profitable indices in a neighborhood of the monotone set Λ_N of currently active indices in terms of error and work contribution. Following [13, 2], the profit of a multiindex $\nu \in \mathfrak{F} = \mathbb{N}_0^d$ is given by

$$g_\nu = \max_{t \in \Xi} \frac{\Delta E(\nu; t)}{\Delta W(\nu)}, \quad \Xi \subset [0, T], \quad (17)$$

where $\Delta W(\nu) = \prod_{k=1}^d (m_{\nu_k} - m_{\nu_k - 1})$, $m_{-1} := 0$, denotes the work contribution and $\Delta E(\nu; t) = \sum_{j \in I} \|X(t; x_{j_1}^{\nu_1}, \dots, x_{j_d}^{\nu_d}) - A(U; \tilde{\Lambda})X(t; x_{j_1}^{\nu_1}, \dots, x_{j_d}^{\nu_d})\|_{\mathcal{S}}$ the error contribution with $\tilde{\Lambda} = \{o \in \Lambda : |o| < |\nu|\}$ and $I = \{j \in \mathbb{N}^d : j_l = 1, \dots, m_{i_l}, l = 1, \dots, d\}$. Note that the multiindices are assumed to be finite, i.e. $\mathfrak{F} = \mathbb{N}_0^d$, which obviously fits into the theoretical framework discussed above, namely by setting the remaining entries equal to zero in the infinite-dimensional case.

In what follows, we consider for a given monotone index set Λ_N the Smolyak interpolant (cf. [1])

$$\mathcal{A}(U; \Lambda_N)X(t; y) = \sum_{\nu \in \Lambda_N} (\Delta^\nu X(t; \cdot))(y),$$

with the increment $\Delta^\nu X(t; \cdot) = \bigotimes_{j=1}^d (U^{i_j} - U^{i_j - 1})X(t; \cdot)$, $\nu \in \Lambda_N$,

$$U^{-1} := 0, U^n(X(t; \cdot)) = \begin{cases} id, & \text{for } n = 0 \\ \sum_{j=1}^{m_n} X(t; x_j^n) l_j^n, & \text{for } n \geq 1 \end{cases},$$

l_j^n univariate Lagrange interpolation polynomials at

1. Clenshaw-Curtis abscissas x_j^n in $[-1, 1]$, i.e.

$$x_j^n = -\cos\left(\frac{\pi(j-1)}{m_n-1}\right), j = 1, \dots, m_n \text{ if } m_n > 1 \text{ and } n_1 = 0 \text{ if } m_n = 1$$

with $m_0 = 1$ and $m_n = 2^{n-1} + 1$, for $n \geq 1$

2. symmetrized Leja abscissas x_j^n , e.g.

$$x_1^n = 0, \text{ if } j = 1,$$

$$x_2^n = 1, \text{ if } j = 2,$$

$$x_3^n = -1, \text{ if } j = 3 \text{ and}$$

$$x_j^n = \arg \max_{x \in [-1,1]} \prod_{k=1}^{j-1} |x - x_k^n|, j = 4, \dots, m_n, \text{ if } j \text{ even,}$$

$$x_j^n = -x_{j-1}^n, j = 4, \dots, m_n, \text{ if } j \text{ odd,}$$

with $m_n = 2n + 1$, for $n \geq 0$ (cf. [5, 6, 7]).

In order to approximate the monotone index set maximizing the profit of each index with respect to the indicator (17), we consider the following algorithm, due to [13].

```

1: function ASG
2:    $d_a \leftarrow 2, \nu \leftarrow (0,0), \Lambda \leftarrow \emptyset, \mathbf{I}_{active} \leftarrow \{\nu\}$ 
3:   Compute  $\Delta^\nu(X(t;\cdot))$  and the error indicator  $g_\nu$ 
4:   while  $\max_{\nu \in \mathbf{I}_{active}} (g_\nu) > tol$  do
5:     Select  $\nu$  from  $\mathbf{I}_{active}$  with largest  $g_\nu$ 
6:     if  $\nu = \nu - e_{d_a}$  then
7:        $d_a \leftarrow d_a + 1, \mathbf{I}_{active} \leftarrow \mathbf{I}_{active} \cup \{e_{d_a}\}$ 
8:       Compute  $\Delta^{e_{d_a}}(X(t;\cdot))$  and the error indicator  $g_{e_{d_a}}$ 
9:     end if
10:     $\Lambda \leftarrow \Lambda \cup \{\nu\}$ 
11:    for  $j = 1, \dots, d_a$  do
12:       $o \leftarrow \nu + e_j$ 
13:      if  $o - e_m \in \Lambda, \forall 1 \leq m \leq d$  then
14:         $\mathbf{I}_{active} \leftarrow \mathbf{I}_{active} \cup \{o\}$ 
15:        Compute  $\Delta^o(X(t;\cdot))$  and the error indicator  $g_o$ 
16:      end if
17:    end for
18:  end while
19: end function

```

Due to the monotonicity requirement of the index set Λ , it holds $\{0\} \subset \Lambda$, so that the proposed algorithm starts with the initial index set $\Lambda_0 = 0$. Then, all feasible neighbor indices are computed, so that the monotonicity of the set Λ is preserved and the new index with the largest profit is added to the index set Λ . This procedure is repeated until the estimated profit of the remaining indices is smaller or equal than a given tolerance tol . Note that the dimension of the indices is also adaptively controlled, i.e. the dimension of the parameter space is iteratively enlarged according to the above results. The example concerns a parametric initial value problem of the following form. Given $X(0, y) = x_0 \in \mathbb{R}$, $T = 1$, $U = [-1, 1]^d$, find $X(t; y) : [0, 1] \times U \rightarrow \mathbb{R}$ such that $X(0; y) = x_0 = 1 \in \mathbb{R}$

$$\frac{dX}{dt} = f(t, X; y) = \left(f_0 + \sum_{j=1}^d f_j y_j\right)^\sigma X, \quad 0 \leq t \leq 1, \quad (18)$$

with $f_i = \left(\frac{1}{i+1}\right)^s$, $s > 1$, $i \in \mathbb{N}_0$.

In (18), the exponent $\sigma = \pm 1$ in all experiments which are considered below. We note that $\sigma = +1$ implies *affine dependence* of the right hand side in (18) on the parameter vector y , as was assumed in the theoretical setting in Sections 2 - 3 above. Therefore, the parametric family of solutions admits, indeed, sparse representations with respect to the parameter vector y . *We emphasize, however, that due to the linear character of the ODE IVP (18), this can also be verified directly from the explicit expression (19).* We note that (19) reveals that for $\sigma = +1$, *the parametric solution is a separable function of the parameters y_j* , so that very favourable approximation properties by the Smolyak interpolation can be expected. On the other hand, for $\sigma = -1$, the exact solution is not separable with respect to the parameters y_j . By direct analysis of the solution formula (19) ahead it can be verified that the parametric solution of (18) allows a representation as unconditionally convergent Taylor expansions with p -summable coefficients, even though for $\sigma = -1$ the dependence of $f(t, X; y)$ on the parameter vector y in (18) is not affine, and the abstract existence theory in [19, Sect. 2 & 3] is not applicable; the N -term approximation results in [19, Sect. 4], however, are applicable based on Taylor coefficient estimates obtained from (19) ahead (indicating, among others, that the theory in [19] could be generalized to certain types of nonaffine, analytic dependence of $f(t, X; y)$ on the parameter vector y). To study the potential of the sparse approximation with respect to the variable y , we will compare the resulting index sets with results based on the exact solution of (18) given by

$$X(t; y) = x_0 \exp\left(t\left(f_0 + \sum_{j=1}^d f_j y_j\right)^\sigma\right), \quad 0 \leq t \leq 1, \quad \forall y \in U. \quad (19)$$

4.1 Separable parametric ODE

First, we consider the case of separable solutions $\sigma = 1$ and restrict the discussion to the ten-dimensional case, i.e.

$$f(t, X; y) = \left(1 + \sum_{j=1}^{10} y_j \left(\frac{1}{j+1}\right)^s\right) X.$$

In the following figure, the adaptively constructed monotone index sets based on Clenshaw-Curtis and Leja points as well as the corresponding error

contributions, where the adaptivity indicator is chosen as

$$g_\nu = \max_{t \in \Xi} \frac{\Delta E(\nu; t)}{\Delta W(\nu)}, \quad \Xi = \{0, 0.1, 0.2, \dots, 1.0\}.$$

are shown. Results based on the exact solution using linear ansatz functions to discretize the time interval $[0, 1]$ ($\Delta t = 2^{-8}$) are compared to the numerical solution of the ODE (18) using MATLAB's ode45 (Runge-Kutta method 4(5) with variable time step and dense output, see [17, 18] for more details) with prescribed tolerance $eps = 2.22045 \cdot 10^{-14}$. In the case of Clenshaw-Curtis as well as of Leja points, it can be stated that both approaches lead to the same adaptive index sets, so that the approximation of the solution (18) does not affect the approximation of the solution with respect to the parameter sequence y , see Figure 1.

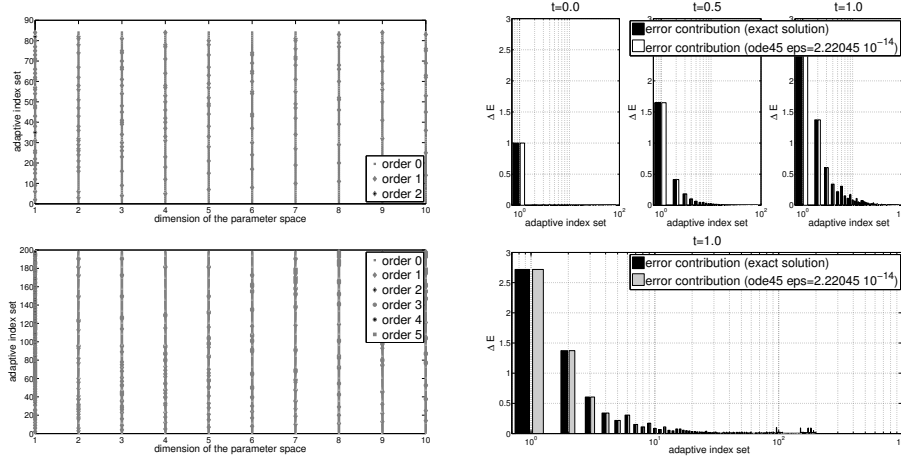


Fig. 1 Adaptively constructed index set Λ (left) and comparison of the error contribution of each index $\nu \in \Lambda$ exemplarily shown at time $t_0 = 0.0, t_5 = 0.5, t_{10} = 1.0$ based on the exact solution (black) and numerical solution (gray) ($\sigma = 1, s = 4, d = 10, tol = 10^{-4}$, exact: $\Delta t = 2^{-8}, \Xi = \{0, 0.1, 0.2, \dots, 1.0\}$, ode45: $eps = 2.22045 \cdot 10^{-14}$, Clenshaw-Curtis nodes (above), Leja nodes (below)).

We further investigate this effect by comparing the resulting grids for the case $s = 4$ (cf. Figure 2 using Clenshaw-Curtis points and Leja nodes).

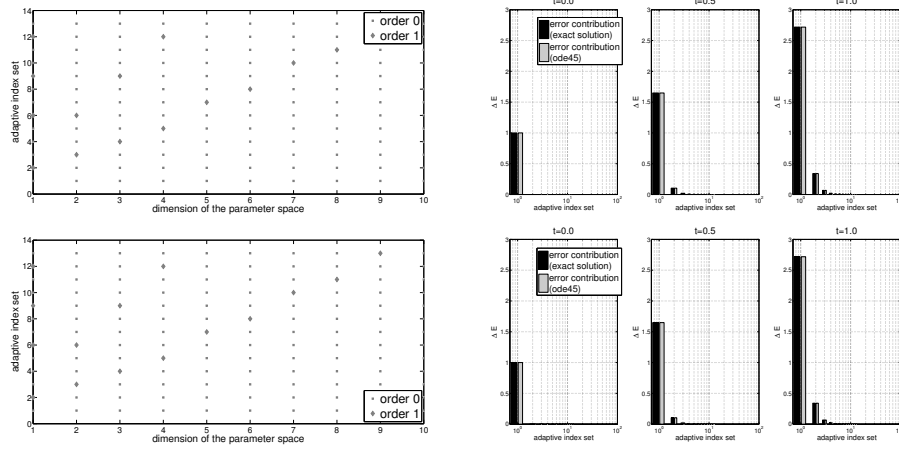


Fig. 2 Adaptively constructed index set Λ (left) and comparison of the error contribution of each index $\nu \in \Lambda$ exemplarily shown at time $t_0 = 0.0, t_5 = 0.5, t_{10} = 1.0$ based on the exact solution (black) and numerical solution (gray) ($\sigma = 1, s = 4, d = 10, tol = 10^{-4}, exact: \Delta t = 2^{-8}, \Xi = \{0, 0.1, 0.2, \dots, 1.0\}, ode45: eps = 2.22045 \cdot 10^{-14}$, Clenshaw-Curtis nodes (above), Leja nodes (below)).

There is a perfect match between the two solutions in both cases, i.e. considering Clenshaw-Curtis as well as Leja points. Finally, the sparsity of the solution $X(t; \cdot)$ with respect to y is explored by adapting the accuracy of the ODE solver according to the impact of the index on the solution. Therefore, we estimate the error contribution of a new feasible index ν by the maximum of the error contributions of all predecessors

$$\Delta \tilde{E}(\nu) = \max_{\nu - e_j \in \Lambda, \forall j \in \text{supp}(\nu)} \max_{t \in \Xi} \Delta E(\nu - e_j; t), \quad (20)$$

normalized by $\Delta \hat{E}(\nu) = \frac{\Delta \tilde{E}(\nu)}{\Delta \tilde{E}(0)}$, so that the error tolerance eps of the ODE solver is chosen as

$$eps_{adapt}(\nu) = \frac{eps}{\Delta \hat{E}(\nu)}. \quad (21)$$

The resulting adaptive index sets for the case $s = 2$ as well as $s = 4$ are practically identical to the set obtained by using the exact solution. Figure 3 illustrates the error contribution based on the proposed adaptive strategy compared to the solutions depicted in Figures 1 and 2. We can observe that the fully adaptive strategy yields the same accuracy as the reference solution while minimizing at the same time the work contribution for each index. The same effect can be explored using Leja points.

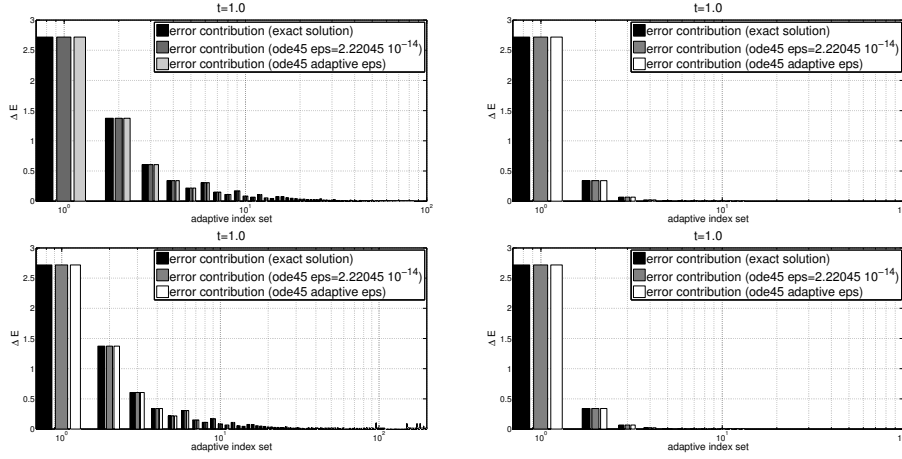


Fig. 3 Comparison of the error contribution of each index $\nu \in \Lambda$ for the case $s = 2$ (left) and $s = 4$ (right) exemplarily shown at time $t_{10} = 1.0$ based on the adaptive strategy controlling the tolerance of the ODE solver (light gray) with the exact solution (black) and numerical solution with fixed error tolerance (gray) ($\sigma = 1$, $d = 10$, $tol = 10^{-4}$, exact: $\Delta t = 2^{-8}$, $\Xi = \{0, 0.1, 0.2, \dots, 1.0\}$, ode45: $eps = 2.22045 \cdot 10^{-14}$, ode45: adaptive eps , Clenshaw-Curtis nodes (above), Leja nodes (below)).

The speedup in terms of function evaluations required for the approximation of the ODE solution and in terms of computation time are summarized in the following table.

	Clenshaw-Curtis nodes		Leja nodes	
	# feval	CPU feval	# feval	CPU feval
$s = 2$				
fixed tolerance	202575	41.28 sec.	328006	66.52 sec.
adaptive strategy	95399	20.84 sec.	138089	29.504 sec.
$s = 4$				
fixed tolerance	29984	6.36 sec.	29985	6.14 sec.
adaptive strategy	14348	3.20 sec.	14350	3.13 sec.

Table 1 Comparison of the number of function evaluations and computation time using the proposed adaptive strategy to control the error tolerance of the ODE solve and the non-adaptive approach using a fixed tolerance of $eps = 2.22045 \cdot 10^{-14}$ ($\sigma = 1$, $d = 10$, $tol = 10^{-4}$, Apple Mac Mini, 2.66 GHz Intel Core 2 Duo, 4GB).

To verify the efficiency of the proposed method in the high-dimensional case, the underlying problem is considered for $d = 100$. As in the previous example, a variation of the parameter s ($s = 2$, $s = 4$) as well as results based on the exact solution (19) and numerical solution of (18) using the Matlab ode45 solver with fixed and adaptive error tolerance considering Clenshaw-Curtis and Leja points are presented.

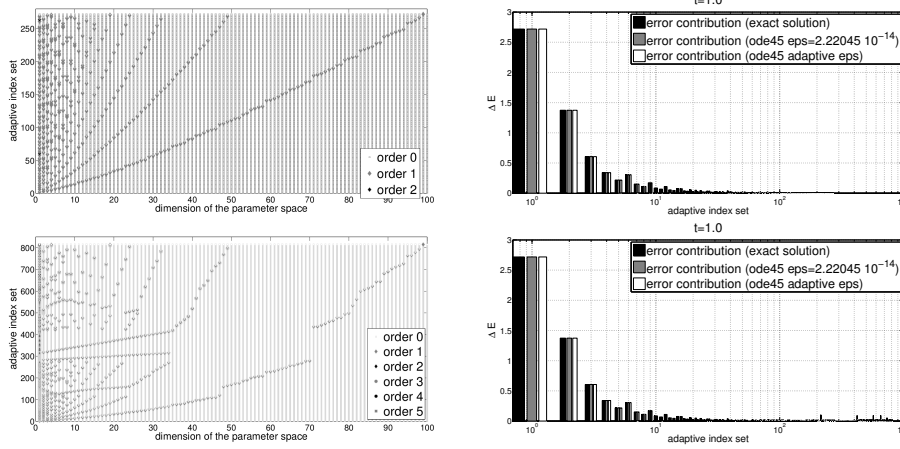


Fig. 4 Adaptively constructed index set Λ (left) and comparison of the error contribution of each index $\nu \in \Lambda$ exemplarily shown at time $t_{10} = 1.0$ based on the adaptive strategy controlling the tolerance of the ODE solver (light gray) with the exact solution (black) and numerical solution with fixed error tolerance (gray) ($\sigma = 1$, $d = 100$, $s = 2$, $tol = 10^{-4}$, exact: $\Delta t = 2^{-8}$, $\Xi = \{0, 0.1, 0.2, \dots, 1.0\}$, ode45: $eps = 2.22045 \cdot 10^{-14}$, ode45: adaptive eps , Clenshaw-Curtis nodes (above), Leja nodes (below)).

Figure 4 illustrate the case $s = 2$, where the grids based on the exact as well as on the non-adaptive and adaptive numerical solution of the ODE are identical. Increasing the parameter ($s = 4$) leads to the same monotone index set as in the ten-dimensional case (see Figure 3). The statistics for the case $d = 100$ are summarized in Table 2.

	Clenshaw-Curtis nodes		Leja nodes	
	# feval	CPU feval	# feval	CPU feval
<i>s = 2</i>				
fixed tolerance	3591145	830.50 sec.	4089927	935.77 sec.
adaptive strategy	1302439	312.56 sec.	1468805	349.63 sec.
<i>s = 4</i>				
fixed tolerance	29984	9.86 sec.	29985	9.58 sec.
adaptive strategy	14348	4.73 sec.	14350	4.76 sec.

Table 2 Comparison of the number of function evaluations and computation time using the proposed adaptive strategy to control the error tolerance of the ODE solver with the non-adaptive approach using a fixed tolerance of $eps = 2.22045 \cdot 10^{-14}$ ($\sigma = 1$, $d = 100$, $tol = 10^{-4}$, Apple Mac Mini, 2.66 GHz Intel Core 2 Duo, 4GB).

4.2 Non-separable ODE

We will now discuss the non-separable case, i.e. $\sigma = -1$ in (18). Numerical results are presented considering a variation of the parameter s in the ten-dimensional case, i.e. $d = 10$. The following figure shows the adaptive index

set and comparison of the error contribution of each index based on the adaptive strategy controlling the tolerance of the ODE solver with the exact solution and numerical solution with fixed error tolerance using Clenshaw-Curtis interpolation points.

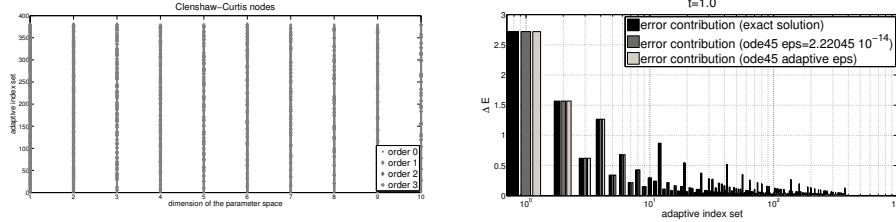


Fig. 5 Adaptively constructed index set Λ (left) and comparison of the error contribution of each index $\nu \in \Lambda$ exemplarily shown at time $t_{10} = 1.0$ based on the adaptive strategy controlling the tolerance of the ODE solver (light gray) with the exact solution (black) and numerical solution with fixed error tolerance (gray) ($\sigma = -1$, $d = 10$, $s = 2$, $tol = 10^{-4}$, $exact: \Delta t = 2^{-8}$, $\Xi = \{0, 0.1, 0.2, \dots, 1.0\}$, ode45: $eps = 2.22045 \cdot 10^{-14}$, ode45: adaptive eps , Clenshaw-Curtis nodes).

Comparing the results with the separable case (cf. Figure 1), we can state that the number of indices of the adaptive sparse grid is enlarged approximately by factor 5 to reach the given tolerance. In the case of Leja points, the prescribed tolerance cannot be reached in a reasonable computation time due to the higher number of indices (resulting from the linear growth of the interpolation nodes) and the related overhead caused by the search of new admissible indices.

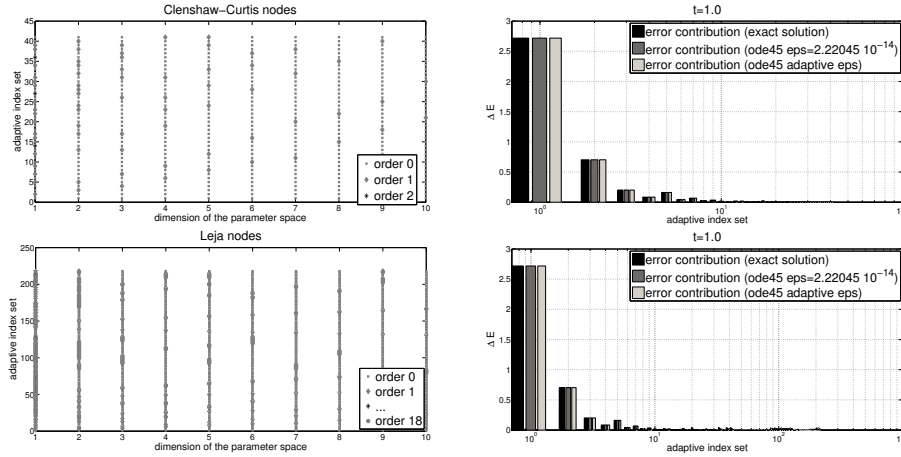


Fig. 6 Adaptively constructed index set Λ (left) and comparison of the error contribution of each index $\nu \in \Lambda$ exemplarily shown at time $t_{10} = 1.0$ based on the adaptive strategy controlling the tolerance of the ODE solver (light gray) with the exact solution (black) and numerical solution with fixed error tolerance (gray) ($\sigma = -1$, $d = 10$, $s = 3$, $tol = 10^{-4}$, $exact: \Delta t = 2^{-8}$, $\Xi = \{0, 0.1, 0.2, \dots, 1.0\}$, ode45: $eps = 2.22045 \cdot 10^{-14}$, ode45: adaptive eps , Clenshaw-Curtis nodes (above), Leja nodes (below)).

Therefore, we additionally present results for the case $s = 3$ in order investigate the influence of the choice of interpolation nodes in the non-separable case, see Figure 6. Increasing the sparsity with respect to the parameter y can be exploited by the algorithm in both cases, that means in the case of Clenshaw-Curtis and Leja points, cf. Figure 7.

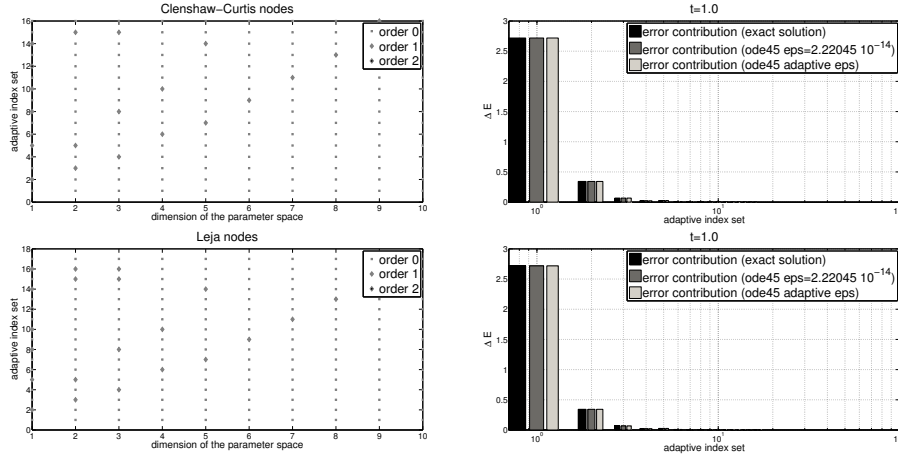


Fig. 7 Adaptively constructed index set Λ (left) and comparison of the error contribution of each index $\nu \in \Lambda$ exemplarily shown at time $t_{10} = 1.0$ based on the adaptive strategy controlling the tolerance of the ODE solver (light gray) with the exact solution (black) and numerical solution with fixed error tolerance (gray) ($\sigma = -1$, $d = 10$, $s = 4$, $tol = 10^{-4}$, exact: $\Delta t = 2^{-8}$, $\mathcal{E} = \{0, 0.1, 0.2, \dots, 1.0\}$, ode45: $eps = 2.22045 \cdot 10^{-14}$, ode45: adaptive eps , Clenshaw-Curtis nodes (above), Leja nodes (below)).

Similar to the separable case (cf. Table 1 and Table 2), the proposed adaptive control of the accuracy of the ODE solver can significantly reduce the overall costs of the algorithm (see Table 3). Further, the computational effort needed to construct the monotone index sets in the case of Clenshaw-Curtis and Leja nodes are comparable in the case $s = 4$ due to the low order of the grids.

	Clenshaw-Curtis nodes		Leja nodes	
	# feval	CPU feval	# feval	CPU feval
$s = 2$				
fixed tolerance	974838	195.31 sec.	-	-
adaptive strategy	528549	108.97 sec.	-	-
$s = 3$				
fixed tolerance	65433	14.28 sec.	227539	47.05 sec.
adaptive strategy	33378	7.54 sec.	88696	19.26 sec.
$s = 4$				
fixed tolerance	34596	7.60 sec.	34269	7.19 sec.
adaptive strategy	15846	3.33 sec.	15779	3.38 sec.

Table 3 Comparison of the number of function evaluations and computation time using the proposed adaptive strategy to control the error tolerance of the ODE solver with the non-adaptive approach using a fixed tolerance of $eps = 2.22045 \cdot 10^{-14}$ ($\sigma = -1$, $d = 10$, $tol = 10^{-4}$, Apple Mac Mini, 2.66 GHz Intel Core 2 Duo, 4GB).

4.3 ODE System

The separable as well as the non-separable case clearly demonstrate the speedup which can be gained by the fully adaptive strategy. The savings in CPU time become even more evident considering a high-dimensional ODE where the main part of the computational effort results from the numerical solution of the underlying differential equation. To investigate this point, we consider the following system of parametric ODEs given by $X(t; y) : [0, 1] \times U \rightarrow \mathbb{R}^p$, $T = 1$, $U = [-1, 1]^d$

$$\frac{dX}{dt} = A(y)X, \quad X(0; y) = x_0 = 1 \in \mathbb{R}^p, \quad 0 \leq t \leq 1, \quad \forall y \in U \quad (22)$$

with $A(y) = (a_{kl}(y))$, $k, l = 1, \dots, p$ given by

$$a_{kl} = \begin{cases} 1 + \sum_{j=1}^d y_j \left(\frac{1}{j+1}\right)^{s_k}, & \text{if } k = p - l + 1 \\ 0, & \text{otherwise} \end{cases}$$

and $s_1 = 1.2$, $s_k = k$, $\forall k = 2, \dots, p$. The error contribution of each index ν is estimated by the maximum error contribution of the components X_i , $i = 1, \dots, p$ with $X = (X_1, \dots, X_p)^T$. The results are summarized in Table 4.

	Clenshaw-Curtis nodes			Leja nodes		
	# feval	CPU feval	total CPU time	# feval	CPU feval	total CPU time
$d = 10, p = 10$						
fixed tolerance	223554	96.28 sec.	470.53 sec.	223142	98.21 sec.	477.26 sec.
adaptive strategy	114448	50.25 sec.	423.99 sec.	114336	50.98 sec.	429.02 sec.
$d = 10, p = 100$						
fixed tolerance	224475	450.32 sec.	3921.83 sec.	224042	445.91 sec.	3914.81 sec.
adaptive strategy	114819	227.41 sec.	3698.53 sec.	114703	228.42 sec.	3677.16 sec.

Table 4 Comparison of the number of function evaluations and computation time using the proposed adaptive strategy to control the error tolerance of the ODE solver with the non-adaptive approach using a fixed tolerance of $eps = 2.22045 \cdot 10^{-14}$ (Apple Mac Mini, 2.66 GHz Intel Core 2 Duo, 4GB).

The presented investigations show the potential of the adaptive error control of the ODE solver. In terms of function evaluations needed for the numerical solution of the underlying ODE and the corresponding CPU time, the adaptive approach is able to halve the computational effort by maintaining at the same time a comparable accuracy of the fully adaptive approximation in t and with respect to y , cf. Table 4.

The presented sparsity results open the perspective of novel, adaptive parameter collocation approaches to efficiently obtain system characteristics on the entire parameter space for numerous systems in engineering and life sciences modelled by initial value ordinary differential equations. In order to demonstrate the applicability of the proposed approach to real life applica-

tions, we will briefly discuss the following systems biology model consisting of a large, parametric system of ordinary differential equations capturing input-output behaviour of ErbB signaling pathways.

4.4 EGFR Signaling Network

In systems biology, a main issue is to understand, quantify and control the regulation of complex cellular pathways and of intercellular communication. The modelling of dynamics is essentially based on the modelling of the kinetics of the underlying biochemical reactions mainly described by mass action models. The approach proposed in this paper provides an efficient framework to quantify the influence of uncertain reaction parameters of the resulting system of ODEs. Exemplarily, we consider in the following the quantification of short term signaling by the epidermal growth factor receptor introduced by [20]. The model can be described in the following rather abstract form

$$\frac{dX}{dt} = f(t, X; y), \quad (23)$$

where $X(0; y) = x_0 \in \mathbb{R}^{23}$, $t \in [0, 120]$ and $X(t; y) : [0, 120] \times U \rightarrow \mathbb{R}^{23}$. The parameters $y \in \mathbb{R}^{51}$ are assumed to be uniformly distributed in $U = [-0.05, 0.05]^{51}$, variations of the nominal values $y_0 \in \mathbb{R}^{51}$ are given by $y_0^i \cdot 10^{y^i}$, $i = 1, \dots, 51$. Nonlinearities of the model occur due to approximations by the Michaelis-Menten kinetics. We refer to [20] for a detailed discussion of the model. The numerical experiments presented below were performed on the Brutus cluster at ETH Zurich using 32 CPUs. The quantity of interest is the dynamic behavior of state 8 (*PLC γ P*), computed at 5 time points.

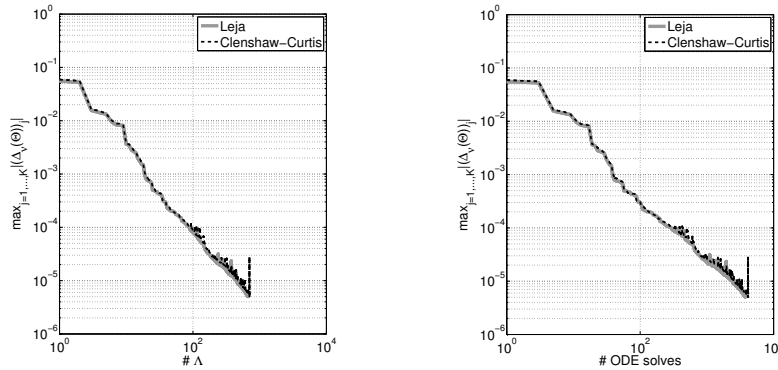


Fig. 8 L^∞ error curves of the interpolation with respect to the cardinality of the index set Λ_N based on the sequences Leja and CC (left) and with respect to the number of ODE solves needed (right), uniform distribution of the parameters in log-space, variation of $10^{\pm 0.05}$ times the nominal point.

To solve the high-dimensional system, a second-derivative integrator especially designed for applications in systems biology is used, we refer to [15] for a detailed description of the ODE solver. Figure 8 displays the estimated approximation error with respect to the cardinality of the index set as well as with respect to the number of ODE solves needed. Both interpolation formulas lead to almost identical, satisfactory results.

The following figure shows the quantity of interest for randomly chosen realisations of the parameter. In particular, we compare the reference solution with the approximated values given by the adaptive sparse grids. As Figure 9 indicates, the absolute error of the Leja and Clenshaw-Curtis based interpolation is in the predicted range, cf. Figure 8.

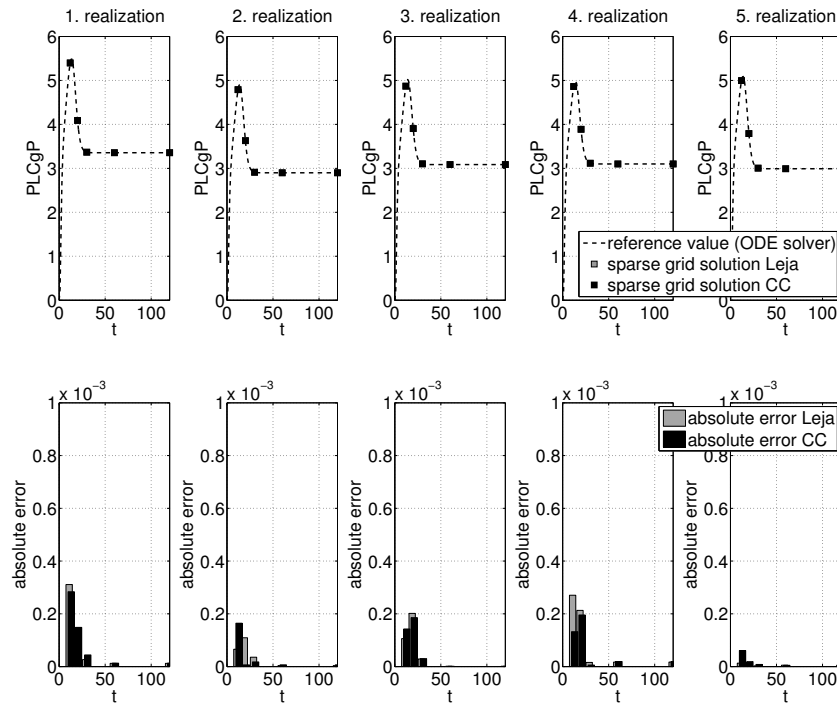


Fig. 9 Comparison of the states approximated by the adaptive sparse grid based on Leja and CC interpolation nodes and the solution of the underlying ODE at randomly chosen realizations of the parameter.

The quantity of interest can be approximated by low order Leja and Clenshaw-Curtis interpolation formulas, so that the results of both sequences are almost identical. Further, we can observe an excellent agreement with the reference solutions, which are obtained by solving the underlying system at the randomly chosen realisations.

5 Conclusions and Perspectives

We have presented recent theoretical results on sparsity of initial value ODEs on high-dimensional, parametric systems which indicate that the parametric solutions of these problems generically admit sparse expansions into tensor products of polynomials with respect to the (possibly countably many) parameters, uniformly with respect to the dimension of the state space. The sparsity result takes the form of N -term approximation rates which are, again, independent of the dimension of the parameter space, and depend only on sparsity in the system models' parameter dependence.

We then propose a numerical algorithm based on Smolyak type interpolation which is steered by adaptively probing the parameter space. At each Smolyak point, the initial value problem is solved numerically, by a stepsize - adaptive ODE initial value solver. The proposed interpolation algorithms being nonintrusive, no modification of the initial value ODE solver is necessary for the numerical solution; moreover, the collocation algorithm allows easy incorporation of parallel numerical solution of initial value ODEs corresponding to different parameter values.

The tolerance parameter of the adaptive ODE IVP solver was, moreover, controlled by the expected upper bound for the contribution of each Smolyak increment to the overall interpolant in the Smolyak formula.

Numerical experiments for model parametric ODEs indicate linear scaling of the algorithm with respect to the number of parameters as well as with respect to the dimension of the system's state space, up to dimensions of 100. Applications from parametric ODE IVP models which arise in biochemical reaction pathways indicate that the approach also allows for efficient numerical solution of complex models with several hundred parameters and state variables. The new, efficient forward simulation methods allow the efficient Bayesian inversion of such models, for large sets of noisy data with the methodology in [22]. Technical details of this aspect of our work, as well as further scalability results of our algorithm for very large models and for a large number of processors will be reported in a forthcoming publication.

References

1. V. BARTHELMANN, E. NOVAK, AND K. RITTER, *High dimensional polynomial interpolation on sparse grids*, Advances in Computational Mathematics, 12 (2000), pp. 273–288. [10.1023/A:1018977404843](https://doi.org/10.1023/A:1018977404843).
2. J. BECK, F. NOBILE, L. TAMELLINI, AND R. TEMPONE, *On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods*, Mathematical Models and Methods in Applied Sciences (M3AS), 22 (2012), pp. 1250023.1–1250023.33.
3. M. BIERI, *A sparse composite collocation finite element method for elliptic SPDEs*, SIAM J. Numer. Anal., 49 (2011), pp. 2277–2301.

4. M. BIERI, R. ANDREEV, AND CH. SCHWAB, *Sparse tensor discretization of elliptic SPDEs*, SIAM J. Sci. Comput., 31 (2009/10), pp. 4281–4304.
5. M. CALIARI, M. VIANELLO, AND L. BERGAMASCHI, *Interpolating discrete advection-diffusion propagators at Leja sequences*, J. Comput. Appl. Math., 172 (2004), pp. 79–99.
6. J.-P. CALVI AND M. PHUNG VAN, *On the Lebesgue constant of Leja sequences for the unit disk and its applications to multivariate interpolation*, Journal of Approximation Theory, 163 (2011), pp. 608 – 622.
7. ———, *Lagrange interpolation at real projections of Leja sequences for the unit disk*, Proceedings of the American Mathematical Society, 140 (2012), pp. 4271–4284.
8. A. CHKIFA, A. COHEN, R. DEVORE, AND CH. SCHWAB, *Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs*, ESAIM: Math. Mod. and Num. Anal., 47 (2013), pp. 253–280.
9. A. CHKIFA, A. COHEN, AND CH. SCHWAB, *High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs*, Tech. Report 2012-22, Seminar for Applied Mathematics, ETH Zürich, 2012.
10. A. COHEN, R. DEVORE, AND CH. SCHWAB, *Convergence rates of best N -term approximations for a class of elliptic sPDEs*, Journ. Found. Comp. Math., 10 (2010), pp. 615–646.
11. A. COHEN, R. A. DEVORE, AND CH. SCHWAB, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs*, Analysis and Applications, 9 (2011), pp. 11–47.
12. K. DEIMLING, *Nonlinear Ordinary Differential Equations in Banach Spaces*, vol. 596 of Springer Lecture Notes in Mathematics, Springer Verlag, New York, 1977.
13. T. GERSTNER AND M. GRIEBEL, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87.
14. C. J. GITTELSON, *Adaptive stochastic Galerkin methods: Beyond the elliptic case*, Tech. Report 2011-12, Seminar for Applied Mathematics, ETH Zürich, 2011.
15. P. GONNET, S. DIMOPOULOS, L. WIDMER, AND J. STELLING, *A specialized ODE integrator for the efficient computation of parameter sensitivities*, BMC Systems Biology, 6 (2012), p. 46.
16. V. HA-HOANG AND CH. SCHWAB, *Analytic regularity and polynomial approximation of stochastic, parametric elliptic multiscale PDEs*, Analysis and Applications (Singapore), 11 (2013), pp. 1350001–1–1350001–50.
17. E. HAIRER, S. P. NØRSETT, AND G. WANNER, *Solving ordinary differential equations. I*, vol. 8 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1987. Nonstiff problems.
18. E. HAIRER AND G. WANNER, *Solving ordinary differential equations. II*, vol. 14 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, second ed., 1996. Stiff and differential-algebraic problems.
19. MARKUS HANSEN AND CHRISTOPH SCHWAB, *Sparse adaptive approximation of high dimensional parametric initial value problems*, Vietnam Journal of Mathematics, (2013), pp. 1–35.
20. B. N. KHOLODENKO, O. V. DEMIN, G. MOEHREN, AND J. B. HOEK, *Quantification of short term signaling by the epidermal growth factor receptor*, J. Biol. Chem., 274 (1999), pp. 30169–30181.
21. F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2411–2442.
22. C. SCHILLINGS AND CH. SCHWAB, *Sparse, adaptive Smolyak algorithms for Bayesian inverse problems*, Tech. Report 2012-37, Seminar for Applied Mathematics, ETH Zürich, 2012.
23. W. WALTER, *Ordinary differential equations*, vol. 182 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1998. Translated from the sixth German (1996) edition by Russell Thompson, Readings in Mathematics.

Recent Research Reports

Nr.	Authors/Title
2012-43	J. Waldvogel Jost Bürgi and the discovery of the logarithms
2013-01	M. Eigel and C. Gittelsohn and C. Schwab and E. Zander Adaptive stochastic Galerkin FEM
2013-02	R. Hiptmair and M. Lopez-Fernandez and A. Paganini Fast Convolution Quadrature Based Impedance Boundary Conditions
2013-03	X. Claeys and R. Hiptmair Integral Equations on Multi-Screens
2013-04	V. Kazeev and M. Khammash and M. Nip and C. Schwab Direct Solution of the Chemical Master Equation using Quantized Tensor Trains
2013-05	R. Kaeppli and S. Mishra Well-balanced schemes for the Euler equations with gravitation
2013-06	C. Schillings A Note on Sparse, Adaptive Smolyak Quadratures for Bayesian Inverse Problems
2013-07	A. Paganini and M. López-Fernández Efficient convolution based impedance boundary condition
2013-08	R. Hiptmair and C. Jerez-Hanckes and J. Lee and Z. Peng Domain Decomposition for Boundary Integral Equations via Local Multi-Trace Formulations
2013-09	C. Gittelsohn and R. Andreev and C. Schwab Optimality of Adaptive Galerkin methods for random parabolic partial differential equations