Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Sparse adaptive approximation of high dimensional parametric initial value problems*

M. Hansen and Ch. Schwab

Research Report No. 2011-64
October 2011

Revised: May 2012

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

*Research partially supported by the European Research Council (ERC) under the FP7 programm AdG247277 and by the Eidgenössische Technische Hochschule (ETH) Zürich under the ETH-Fellowship Grant FEL-33 11-1

Sparse Adaptive Approximation of High Dimensional Parametric Initial Value Problems ⋆

M. Hansen and Ch. Schwab

Abstract We consider nonlinear systems of ordinary differential equations (ODEs) on a state space $\mathcal{S}$. We consider the general setting when $\mathcal{S}$ is a Banach space over $\mathbb{R}$ or $\mathbb{C}$. We assume the right hand side depends *affinely linear* on a vector $y = (y_j)_{j\geq 1}$ of possibly countably many parameters, normalized such that $|y_j| \leq 1$. Under suitable analyticity assumptions on the ODEs, we prove that the parametric solution $\{X(t;y) : 0 \leq t \leq T\} \subset \mathcal{S}$ of the corresponding IVP depends holomorphically on the parameter vector y, as a mapping from the infinite-dimensional parameter domain $U = (-1,1)^{\mathbb{N}}$ into a suitable function space on $[0,T] \times \mathcal{S}$. Such affine parameter dependence of the ODE arises, among others, in mass action models in computational biology (see, e.g. [18]) and in stochiometry with uncertain reaction rate constants. Using our analytic regularity result, we prove summability theorems for coefficient sequences of generalized polynomial chaos (gpc) expansions of the parametric solutions $\{X(\cdot;y)\}_{y\in U}$ with respect to tensor product orthogonal 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.

Markus Hansen
Seminar for Applied Mathematics, ETH Zürich, 8092 Zürich, Switzerland
e-mail: markus.hansen@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

Key words: Ordinary differential equations, initial value problem, parametric dependence, analyticity in infinite dimensional spaces, Taylor series, N-term approximation.

Contents

1 Introduction

Numerous phenomena in engineering and life- and in social sciences are modelled by initial value ordinary differential equations (ODEs); if, in particular, the systems of interest are complex, they are described by vectors on state spaces $\mathcal{S}$ of high or even infinite dimension, with multiple time scales. Accordingly, the *numerical solution of initial value problems for deterministic and stochastic ODEs* is a basic problem in engineering and in the sciences. For a survey of the state of the art on theory and implementation of numerical initial value solvers, we refer to the monographs [11, 12, 13] and to the references therein.

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.* As cases in point, we mention only stochiometric descriptions of biochemical reaction pathways with uncertain reaction rate constants, chemical reaction cascades with uncertain reaction rate constants, mass action models [18] with uncertain reaction rates. In mathematical models of such systems, the number of parameters subject to uncertainty is often large, and the interest in modelling consists in *obtaining the system characteristics on the entire parameter space in one single numerical forward simulation.* Besides the efficient *forward solution* of parametric initial value ODEs, additional problems consist in optimization resp. in optimal control of systems described by initial value ODEs. Here it is again of interest to obtain a parsimonious numerical representation of the parametric dependence of the control resp. the optimum on the entire, possibly high-dimensional parameter space.

The analysis of *approximability* of solutions of a class of abstract initial value problems on possibly infinite dimensional parameter spaces is the purpose of the present paper. Due to the so-called curse of dimensionality, the approximation of parametric solutions by standard tensor product interpolation methods is not feasible even for a few dozen parameters. It is the aim of the present paper to identify sufficient conditions for *sparsity* of several types of polynomial expansions for solutions of parametric initial value ODEs. Under the assumption of *affine dependence on the parameters* which arise, for example, in Karhúnen-Loève expansions of the random input data, the parametric solutions of the ODEs admit, in turn, unconditionally convergent expansions into polynomial series with respect to the possibly infinitely many parameters.

Throughout, we shall use the following *notation.* 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 $\overline{U} = [-1,1]^{\mathbb{N}}$.

1.1 The Scope of Problems

On the parameter domain U, we wish to solve the *high-dimensional, parametric, deterministic ODE initial value problem (ODE-IVP)*:
Given $x_0(y) \in \mathcal{S}$ and $T \in (t_0, \infty)$, find $X(t, x_0; y) : [t_0, T] \times \mathcal{S} \times U \to \mathcal{S}$ such that in $\mathcal{S}$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = f(t, X; y) \ , \quad X(t_0; y) = x_0(y) \ , \quad t_0 \le t \le T \ , \quad \forall\, y \in U \ . \qquad (1)$$

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 will consider also the infinite dimensional case, when $\mathcal{S}$ *is a separable and reflexive Banachspace*.

In practice efficient solution methods in the case where the number of parameters is very 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*. Hence our approach consists in directly tackling the case of an infinite (but countable) number of parameters.

It is well-known and classical (see, e.g., the text [22, 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. This *local analytic dependence of the solutions on the parameters* is, in fact, a quite generic phenomenon which appears in many systems with analytic, nondegenerate parameter dependence as a consequence of (a suitable version of) the implicit function theorem (see, e.g., [17, Theorem 2.1.2]). In the present paper, we extend the proof in [22] of this (classical) result to a possibly countable number of parameters with *quantitative bounds on the size of domains of analyticity*. This mathematical result will be used to establish *best N-term convergence rates for various parametric expansions of the solution $X(t; y)$* under a *sparsity hypothesis* on the vector field $f(t, X; y)$ in (1). Our main result is that dimension-independent rates of best N-term approximation are achieveable with *N-term truncated Taylor expansions* of the solution $X(t, y)$ in the parameter space U. In the present paper, we establish for several types of expansions the uniform and unconditional convergence for all y belonging to an infinite-dimensional parameter domain U. Moreover, we establish that *sparsity in the input vector field $f(t, X; y)$ implies, for example, sparsity (in a sense to be made precise below) in the parametric solutions' formal Taylor-expansions, i.e.*

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

We also prove analogous results also for other polynomial expansions of the solution, such as Legendre or Chebyshev expansions.

Similar results have been obtained for a number of parametric *partial differential equations* in [7, 8, 16]. We note, however, that the proofs in these references do not directly generalize to problems (1), so that we give a new, and self-contained argument here.

The theoretical result in this paper is used in [14] to design a class of dimensionally robust practical algorithms for the efficient solution of large systems of parametric ODE's on possibly infinitely dimensional parameter spaces required 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, so-called *monotone* sequences of sparse 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

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

realize the best N-term asymptotic convergence rate. Second, once such truncations have been selected, it will be necessary to solve the initial value problems by ODE solvers for approximation of the expansion coefficients $T_\nu(t)$ in (2). We will furthermore prove here that N-term approximations with sets $\mathcal{M}_N$ constrained to be monotone sets achieve the same rates as best N-term approximations by truncated Legendre or Chebyshev series as well. Since, *for monotone sparsity models* $\mathcal{M}_N$*, the polynomial space* $\mathbb{P}_{\mathcal{M}_N} = \mathrm{span}\{y^\nu : \nu \in \mathcal{M}_N\}$ *is independent of the particular choice of polynomial bases*, the results in the present paper (in particular, Theorem 11) imply that *the best* N*-term convergence rates can be realized by sparse tensor polynomial interpolation schemes on monotone sparsity models* $\mathcal{M}_N$ *independently of the choice of the univariate polynomial basis.* This allows, in particular, to choose the univariate coordinate basis functions in tensorized Smolyak interpolation schemes such as those analyzed in [1] and the references there according to other criteria, e.g. such that certain condition numbers are minimized. We refer to [14].

The application of Smolyak type interpolation schemes with respect to the parameter vector $y \in U$ in (1) amounts to the numerical solution of (1) for many instances of the parameter vector y. In the present paper, *we assume that* (1) *can be solved exactly.* In practice, however, only approximate, numerical solutions of (1) are available and efficient computational algorithms will require the *approximate solution of* (1) *via initial value ODE* *solvers* such as those in [11, 12]. In [14], we develop computational aspects of combined sparse parametric polynomial interpolation methods based on the mathematical results in the present paper. In particular, we use in [14] the independence of $\mathbb{P}_{\mathcal{M}_N}$ of the univariate polynomial basis functions to design Smolyak type interpolation algorithms based on *hierarchic sequences* of unisolvent, univariate interpolation points; in particular, *Chebychev* and *Leja* point sequences (see, e.g. [4] and the references there) are considered. In [14], we propose computational strategies which allow to solve the $|\mathcal{M}_N|$ many instances of the parametric ODE (1) in parallel. The mathematical results in the present paper will also guide the judicious choice of the tolerances in termination criteria of adaptive initial value ODE solvers.

We remark that the analyticity and sparsity results for the parametric solution families of (1) form the basis of additional mathematical developments. Among them are problems of optimization and of optimal control of high-dimensional, parametric initial value ODEs; here the controls can be expected to depend analytically on the parameter vector. We refer to [19] for a corresponding development in the context of parametric linear elliptic partial differential equations.

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

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

where $\widetilde{T}_\nu(t) \in \mathcal{S}$ are the mentioned approximate Taylor coefficients, the effort for computing the coefficients have to be balanced against the respective impact for approximating $X(t;y)$. In doing so, we obtain an adaptive approximate numerical solution of the parametric ODE-IVP (1) to a presribed 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. [10]) or by sparse collocation as in [20, 3, 2, 14] or by adaptive truncation (4) of the Taylor expansions (2) as in [6].

1.2 Notation and Function Spaces

Throughout this work, we shall use the following *standard notation*: by (a,b) we denote for $-\infty \le a < b \le +\infty$ the open interval $\{x \in \mathbb{R} : a < x < b\}$, by $[a,b] = \overline{(a,b)} = \{x \in \mathbb{R} : a \le x \le b\}$ its closure, by $(a,b)^p = (a,b) \times ... \times (a,b)$

its p-fold Cartesian product. For two sequences $a = (a_j)_{j\geq 1}$ and $b = (b_j)_{j\geq 1}$ such that $-\infty < a_j < b_j < \infty$ for all values of j, we identify the set $(a,b)^{\mathbb{N}}$ with the countable Cartesian product

$$(a,b)^{\mathbb{N}} = \prod_{j\geq 1}(a_j, b_j)\,.$$

Throughout, we assume that the time interval of evolution of the system (1) is $[0,T]$. We shall denote the state of the system by $X(t) \in \mathbb{R}^d$ for $t \in [0,T]$. The parameter dependence of X on $y \in U$ is indicated by $X(t;y)$. We shall also consider extensions of problem (1) to complex values of the parameter vector y. To this end, we denote by $\mathbb{C}^{\mathbb{N}}$ the set of all sequences with values in $\mathbb{C}$. We denote $\mathbb{N} = \{1,2,...\}$ 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

$$\mathfrak{F} = \{\nu \in \mathbb{N}_0^{\mathbb{N}} : |\nu| < \infty\}\,. \tag{5}$$

As any $\nu \in \mathfrak{F}$ has only finitely many nonzero entries, the definitions

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

for multi-factorials, the length of a multi-indices ν and for the partial derivative of order ν are well-defined for $\nu \in \mathfrak{F}$. To state and prove results on existence, regularity and numerical approximation errors of solutions, we require certain function spaces. In what follows, we let B denote a separable Banach space with norm $\|\cdot\|_B$. We shall, by abuse of notation, denote by B both the vector space 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 continuous functions $f : [0,T] \to B$ whose k-th Fréchet derivative $\frac{\mathrm{d}^k f}{\mathrm{d}t^k}$ 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

$$\|f\|_{C^k([0,T];B)} := \max_{0\leq j\leq k}\left\{\|\tfrac{\mathrm{d}^j f}{\mathrm{d}t^j} f\|_{C^0([0,T];B)}\right\}\,, \quad k \in \mathbb{N}\,, \tag{6}$$

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 $\frac{\mathrm{d}f}{\mathrm{d}x}$ have to be understood as Fréchet derivatives, i.e. $\frac{\mathrm{d}f}{\mathrm{d}x}$ 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.

To define r-integrable functions on U with values in B, we assume given on U a probability measure $\mu(\mathrm{d}y)$, which is defined on the measurable space $(U, \mathfrak{A}) = \otimes_{j\geq 1}((-1,1); \mathcal{B}^1)$ where $\mathcal{B}^1$ denotes the sigma algebra of Borel sets on $(-1,1)$. For $0 < r \leq \infty$, we denote by $L^r(U;B)$ the space of all measurable functions $f : U \to B$ which are r-summable in the sense that

$$\|f\|_{L^r(U;B)} := \left(\int_{y\in U} \|f(y)\|_B^r \mu(\mathrm{d}y) \right)^{1/r} < \infty \tag{7}$$

(with the integral replaced by the essential (w.r.to μ) supremum in the limiting case $r = \infty$). If the measure μ is clear from the context, we write $L^r(U;B)$ for brevity.

An important role in this work will be played by *analytic*, B-valued functions. To this end, we will denote for an open, nonempty set $\mathcal{U} \subset \mathcal{S}$ by $\mathcal{A}(\mathcal{U};B)$ the space of all mappings $u : \mathcal{U} \to B$ which are analytic. The definitions for analyticity and (weak and strong) holomorphy for the cases where the state space is either finite-dimensional ($\mathcal{S} = \mathbb{C}^d$), $\mathcal{S}$ is a (separable) Banach space or $\mathcal{S}$ is a locally convex vector space are provided in Section 3.1. Thus, our setting and results include in particular also parametric, nonlinear evolution PDEs. Further function spaces (spaces of locally Lipschitz continuous functions and certain weighted spaces of continuous functions) will be defined as they occur in the text.

1.3 Outline of the paper

The outline of the present paper is as follows. In the next section, we precise the parametric initial value ODE problem (1), in spaces of k-times continously differentiable functions which take values in the state space $\mathcal{S}$. We state and prove a global existence result for solutions of (1). The proof is based on applying Banach's fixed-point theorem to a Volterra integral equation reformulation of (1) in exponentially weighted spaces due to [22]. The use of weights allows to deal with the possible exponential growth of solutions and avoids tedious continuation arguments.

Section 3 is devoted to showing *analyticity* of the parameter dependence of the parametric solutions. To this end, basic notions of analyticity in Banach spaces are recapitulated, and then analytic dependence of solutions of (1) on the parameter sequence y is established. Particular attention is paid to the quantative bounds on the size of the domains of analyticity of the solution.

Section 4.1 contains statements and proofs of our main results, the proof of best N-term approximation rates of the parametric solutions $X(t;y)$ of (1). Three particular types of approximation are considered: tensorized Taylor-, Tschebyscheff and Legendre expansions of $X(\cdot;y)$ with respect to the coordinate vector y. The proof uses arguments from [7, 8], and is given in Sections

4.2 - 4.7. We emphasize, however, that the present arguments combine results from [7, 8] and allow to sharpen also some of the results obtained in these references. In Section 4.8, finally, we consider the *restricted N-term approximations* where the sparsity models $\mathcal{M}_N \subset \mathfrak{F}$ are restricted to the class of monotone index sets (as in [6], see also Definition 5.

2 Parametric Initial Value ODEs

2.1 Problem Formulation. Existence Result

For a parameter vector $y = (y_j)_{j\geq 1} \in U$ 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 $X \in \mathcal{S}$,

$$f(t, X; y) = f_0(t, X) + \sum_{j\geq 1} y_j f_j(t, X) \ , \quad 0 \leq t \leq T < \infty \ . \tag{8}$$

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

$$\frac{dX}{dt} = g(t, X) \ , \quad X(t_0) = x_0 \ , \tag{9}$$

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

$$\forall X, X' \in U \quad \forall t \in [0, T] : \quad \|g(t, X) - g(t, X')\|_{\mathcal{S}} \leq L(X_0)\|X - X'\|_{\mathcal{S}} \tag{10}$$

for some constants $L(X_0)$, implies existence and uniqueness of local solutions of (9), i.e. existence of unique solutions on some maximally extended subinterval $[0, \delta) \subset [0, T]$. This result is classical in the scalar case $\mathcal{S} = \mathbb{R}$, and it extends immediately to the general state space $\mathcal{S}$ considered here, see e.g. [9]. In the parametric case the naïve way would be to assume this condition to hold pointwise, i.e. for every fixed $y \in U$; i.e., to assume that condition (10) holds for $f(\cdot, \cdot; y)$. This assumption then yields local existence of the parametric solution, but unfortunately the existence-interval $[0, \delta)$ might depend on the parameter, i.e. in this argument δ cannot be chosen independent of $y \in U$.

For our further arguments we would rather like to have global solutions, i.e. on the whole interval $[0,T]$. To obtain these, we impose slightly more restrictive conditions. More precisely, we suppose the condition that for every $R>0$, there exist constants $L(R)>0$ such that for every $X,X'\in B_R=\{X\in\mathcal{S}:\|X\|_{\mathcal{S}}\le R\}$ and for every $t\in[0,T]$ holds

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

where the optimal constants $L(R)$ are given by

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

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

$$L_j(R)=\sup_{t\in[0,T],X\ne X'\in B_R}\frac{\|f_j(t,X)-f_j(t,X')\|_{\mathcal{S}}}{\|X-X'\|_{\mathcal{S}}}<\infty\,,\quad f_j(t,0)=0\,. \tag{11}$$

In order to prove results which are independent of the number of terms in the affine expansion (8), we shall further require *summability of the coefficient sequence* $(f_j)_{j\ge1}$. Specifically, we assume the sequence of Lipschitz constants to be summable, i.e.

$$\forall R>0:\quad \big(L_j(R)\big)_{j\ge1}\in\ell^1(\mathbb{N})\,. \tag{12}$$

Under this assumption, the sum in (8) converges uniformly with respect to $y\in U$ and for all $(t,X)\in[0,T]\times\mathcal{S}$.

Proposition 1. *Let the conditions* (11) *and* (12) *be satisfied. Then the sum in* (8) *converges absolutely and uniformly in* U *as a* $\ell\mathrm{Lip}_0(\mathcal{S})$*-valued mapping, and it holds* $f\in C(U;\ell\mathrm{Lip}_0(\mathcal{S}))$.

Proof. We fix $0<N<M<\infty$ and denote by $S_N[f]$ the N-term partial sum of (8). Then we estimate with $|y_j|\le1$ for all $j\ge1$ and with the assumptions (11) and (12) and the triangle inequality

$$\sup_{t\in[0,T]}\sup_{y\in U}\|(S_M[f]-S_N[f])(t,\cdot;y)\|_{\ell\mathrm{Lip}(\mathcal{S},R)}\le\sum_{j=N+1}^{M}\sup_{t\in[0,T]}\|f_j(t;\cdot)\|_{\ell\mathrm{Lip}(\mathcal{S},R)}\to0$$

as $N,M\to\infty$, since by assumption (12) we have

$$\sum_{j\geq 1} L_j(R) = \sum_{j\geq 1} \sup_{t\in[0,T]} \|f_j(t;\cdot)\|_{\ell \mathrm{Lip}(\mathcal{S},R)} < \infty \quad \forall R > 0 \; .$$

This shows that the expression (8) is well-defined for every $y \in U$ and for every $(t,X) \in [0,T] \times \mathcal{S}$. Moreover, the assumption (11) for all $j \geq 0$ implies that $f(t,X;y)$ satisfies a *Lipschitz condition uniform* in $y \in U$: for every $X, X' \in B_R$ holds

$$\begin{aligned}
&\sup_{t\in[0,T]} \sup_{y\in U} \|f(t,X;y) - f(t,X';y)\|_{\mathcal{S}} \\
&\leq \sup_{t\in[0,T]} \|f_0(t,X) - f_0(t',X')\|_{\mathcal{S}} \\
&+ \sup_{t\in[0,T]} \sup_{y\in U} \left| \sum_{j\geq 1} |y_j| \|f_j(t,X) - f_j(t',X')\|_{\mathcal{S}} \right| \\
&\leq \left(\sup_{t\in[0,T]} \|f_0(t;\cdot)\|_{\ell \mathrm{Lip}(R)} + \sum_{j\geq 1} \sup_{t\in[0,T]} \|f_j(t;\cdot)\|_{\ell \mathrm{Lip}(R)} \right) \|X - X'\|_{\mathcal{S}} \\
&= L(R) \|X - X'\|_{\mathcal{S}}
\end{aligned} \tag{13}$$

where the Lipschitz constant

$$L(R) := \sum_{j\geq 0} \sup_{t\in[0,T]} \|f_j(t;\cdot)\|_{\ell \mathrm{Lip}(\mathcal{S},R)} \tag{14}$$

is finite by (12). This proves the asserted uniform Lipschitz condition for the infinite sum (8) in U.

For the proof of the continuous dependence we first observe

$$\begin{aligned}
&\|f(\cdot,\cdot;y) - f(\cdot,\cdot;y')\|_{\ell \mathrm{Lip}(\mathcal{S},R)} \\
&\qquad \leq \sum_{j\geq 1} \|f_j(\cdot,\cdot) - f_j(\cdot,\cdot)\|_{\ell \mathrm{Lip}(\mathcal{S},R)} |y_j - y'_j| \leq \sum_{j\geq 1} L_j(R) |y_j - y'_j|
\end{aligned}$$

for all $y, y' \in U$. Then since by Assumption (12) for every $\varepsilon > 0$, we can find some $N(\varepsilon)$ such that $\sum_{j>N} L_j(R) < \varepsilon$, and by choice of the product topology $y' \to y$ if, and only if, $\sum_{1\leq j\leq J} |y_j - y'_j| \to 0$ for every $J \in \mathbb{N}$, this estimate proves the continuous dependence of f on $y \in U$.

As we shall see below this local Lipschitz condition (13) together with some scaling assumption on the initial data already implies existence and uniqueness of solutions $X(t, x_0; y)$ for all $y \in U$ by standard fixed point arguments. Furthermore, it also implies additional regularity of the solution $X(t, x_0; y)$ with respect to $t \in [0,T]$.

Before we come to this we shall further extend the parametric problem to include complex parameters. More precisely, looking more closely at the proof of Proposition 1 we immediately see that the arguments carry over to the parameter domain $\mathcal{U} = \{z \in \mathbb{C}^{\mathbb{N}} : |z_j| \leq 1\,, j \in \mathbb{N}\}$, if we define

$$f(t,X;z) = f_0(t,X) + \sum_{j\geq 1} z_j f_j(t,X)\ , \quad 0 \leq t \leq T < \infty\ . \tag{15}$$

Moreover, we may also consider $\mathcal{S}$ to be a complex Banach space. The motivation for this extension lies in the observation that under certain conditions on the functions f_j the solution depends analytically on the parameters, which will be discussed in the following section.

Theorem 1. *Assume* (11) *and* (12)*. Moreover, suppose the initial condition* $x_0 \in C(\mathcal{U},\mathcal{S})$ *satisfies*

$$\sup_{z\in\mathcal{U}} \|x_0(z)\|_{\mathcal{S}} \leq (1-\kappa)r\,, \qquad r = Re^{-L(R)T/\kappa} \tag{16}$$

for some $R > 0$ *and* $0 < \kappa < 1$.

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

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

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

$$\forall j \geq 0: \quad f_j : [0,T]\times\mathcal{S} \longrightarrow \mathcal{S} \text{ is k-times continuously differentiable}\,, \tag{17}$$

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

Moreover, the solution $X(\cdot,x_0;z)$ *on the data* x_0 *and parameters* z*. More precisely: For every* $x_0 \in C(\mathcal{U};\mathcal{S})$ *satisfying* (16) *and for every* $z, z' \in \mathcal{U}$ *holds*

$$\begin{aligned}&\left\|X\left(\cdot,x_0(z);z\right) - X\left(\cdot,x_0(z');z'\right)\right\|_{C([0,T];\mathcal{S})} \\ \leq\ &\frac{\mathrm{e}^{L(R)T/\kappa}}{1-\kappa}\left(\|x_0(z)-x_0(z')\|_{\mathcal{S}} + \kappa r\|z-z'\|_{\ell^\infty(\mathbb{N})}\right)\ .\end{aligned} \tag{18}$$

Similarly, for every $x_0, x_0' \in C(\mathcal{U},\mathcal{S})$ *satisfying* (16) *and for every* $z \in \mathcal{U}$ *holds*

$$\left\|X\left(\cdot,x_0(z);z\right) - X\left(\cdot,x_0'(z);z\right)\right\|_{C([0,T];\mathcal{S})} \leq \frac{\mathrm{e}^{L(R)T/\kappa}}{1-\kappa}\|x_0(z)-x_0'(z)\|_{\mathcal{S}}\ . \tag{19}$$

2.2 Proof of the Existence result

Though, as mentioned before, this existence result is known, and the proof of Theorem 1 is very similar to the scalar case, we include it anyway in order to obtain explicit, quantitative bounds which will be of importance later on.

The proof of Theorem 1 will be based on a fixed point argument along the lines of [22]. Specifically, Theorem 1 will follow from a more general result on Volterra integral equations. To motivate it, we observe that the IVP (1) can be recast into a Volterra Integral Equation in $\mathcal{S}$, i.e.

$$X(t;z) = x_0(z) + \int_{t_0(z)}^{t} f\big(s, X(s;z); z\big)\,\mathrm{d}s\ , \quad 0 \le t \le T < \infty \tag{20}$$

or, in operator notation, into the fixed point equation

$$X = \Theta[X]\ , \quad X \in C([0,T]\times\mathcal{U};\mathcal{S}) \tag{21}$$

where the Volterra integral operator $\Theta[X]$ for $X \in C([0,T]\times\mathcal{U};\mathcal{S})$ is defined by

$$\Theta[X](t;z) := x_0(z) + \int_{t_0(z)}^{t} f\big(s, X(s;z); z\big)\,\mathrm{d}s\ . \tag{22}$$

For $f(\cdot,\cdot;z)$ satisfying a uniform local Lipschitz condition (13), Θ becomes a contraction on $C([0,T]\times\mathcal{U};\mathcal{S})$, if we equip this space with a suitable norm (depending on T and L). We shall now prove this within a slightly more general setting.

Lemma 1. *Assume that the vector field $k(t,s,X;z) : [0,T]^2\times\mathcal{S}\times\mathcal{U}\mapsto\mathcal{S}$ is continuous on its domain of definition and it satisfies a local Lipschitz condition uniformly for all $s,t\in[0,T]$ and for all $z\in\mathcal{U}$, i.e. for every $R>0$ there exists some $L(R)>0$ such that for every $X,X'\in B_R$ and for every $0\le t,s\le T<\infty$ holds*

$$\sup_{z\in\mathcal{U}} \|k(t,s,X;z) - k(t,s,X';z)\|_{\mathcal{S}} \le L(R)\|X-X'\|_{\mathcal{S}}\ . \tag{23}$$

Assume moreover that for every $X\in\mathcal{S}$, $k(t,s,X;z) : [0,T]^2\times\mathcal{U}\mapsto\mathcal{S}$ and $t_0(z) : \mathcal{U}\mapsto[0,T]$, $x_0(z) : \mathcal{U}\mapsto\mathcal{S}$ are continuous with respect to $(t,s,z)\in[0,T]^2\times\mathcal{U}$ and $z\in\mathcal{U}$, respectively.
We then consider the set $C([0,T]\times\mathcal{U};\mathcal{S})$ equipped with the weighted uniform norms $\|\cdot\|_{\alpha,T,\mathcal{S}}$, where for $\alpha\ge 0$

$$\|X\|_{\alpha,T,\mathcal{S}} := \sup_{(t,z)\in[0,T]\times\mathcal{U}} \{\mathrm{e}^{-\alpha|t-t_0(z)|}\|X(t;z)\|_{\mathcal{S}}\}\ . \tag{24}$$

We shall also use the notation $C_\alpha([0,T]\times\mathcal{U};\mathcal{S})$ to indicate this weighted setting. Let some $0<\kappa<1$ and $R>0$ be given, and put $r = Re^{-L(R)T/\kappa}$.

Then the parametric integral operator

$$\Theta[X](t;z) := x_0(z) + \int_{t_0(z)}^{t} k\big(t,s,X(s;z);z\big)\,\mathrm{d}s \tag{25}$$

is a contraction on $\mathcal{B}_{r,R} = \{X \in C([0,T]\times\mathcal{U};\mathcal{S}) : \|X\|_{L(R)/\kappa,T,\mathcal{S}} \le r\}$*. More precisely, it holds*

$$\forall X, X' \in \mathcal{B}_{r,R}: \quad \|\Theta[X] - \Theta[X']\|_{L(R)/\kappa,T,\mathcal{S}} \le \kappa \|X - X'\|_{L(R)/\kappa,T,\mathcal{S}}\,. \tag{26}$$

Proof. By the continuity assumptions on the kernel function k, the initial data x_0 and on t_0, the operator Θ defined in (25) satisfies $\Theta[C([0,T]\times\mathcal{U};\mathcal{S})] \subset C([0,T]\times\mathcal{U};\mathcal{S})$.

Concerning the contraction property, let some $R > 0$ be given. We first note that $\|X\|_{L(R)/\kappa,T,\mathcal{S}} \le r$ implies $\|X(t,z)\|_{\mathcal{S}} \le r\mathrm{e}^{L(R)T/\kappa} = R$ by choice of r. Hence we conclude from the local Lipschitz condition (23) for $X, X' \in \mathcal{B}_{r,R}$

$$\big\|(\Theta[X] - \Theta[X'])(t,z)\big\|_{\mathcal{S}} \le L(R) \left| \int_{t_0(z)}^{t} \big\|X(s;z) - X'(s;z)\big\|_{\mathcal{S}}\,\mathrm{d}s \right|\,.$$

Injecting the factor $1 = \exp\big(L(R)|t-t_0(z)|/\kappa\big)\exp\big(-L(R)|t-t_0(z)|/\kappa\big)$ into the integral on the right hand side we obtain for every $t \in [0,T]$ and for every $z \in \mathcal{U}$ the bound

$$\begin{aligned}\big\|(\Theta[X] - \Theta[X'])(t,z)\big\|_{\mathcal{S}} &\le L(R)\|X - X'\|_{L(R)/\kappa,T,\mathcal{S}} \left| \int_{t_0(z)}^{t} \mathrm{e}^{L(R)|s-t_0(z)|/\kappa}\,\mathrm{d}s \right| \\ &\le \frac{L(R)}{L(R)/\kappa}\|X - X'\|_{L(R)/\kappa,T,\mathcal{S}}\mathrm{e}^{L(R)|t-t_0(z)|/\kappa}\,.\end{aligned}$$

Dividing both sides by $\mathrm{e}^{L(R)|t-t_0(z)|/\kappa}$ and taking the supremum in the resulting bound over all $t \in [0,T]$ and all $z \in \mathcal{U}$, we find

$$\forall X, X' \in \mathcal{B}_{r,R}: \quad \big\|\Theta[X] - \Theta[X']\big\|_{L(R)/\kappa,T,\mathcal{S}} \le \kappa\|X - X'\|_{L(R)/\kappa,T,\mathcal{S}}\,.$$

This completes the proof. □.

Remark 1. In case of a global Lipschitz condition, i.e. for all $X, X' \in \mathcal{S}$, $0 \le t, s \le T$ and $z \in \mathcal{U}$ we have

$$\|k(t,s,X,z) - k(t,s,X';z)\|_{\mathcal{S}} \le L\|X - X'\|_{\mathcal{S}}\,,$$

the operator Θ becomes a contraction on the whole set $C_{L/\kappa}([0,T]\times\mathcal{U};\mathcal{S})$ for every $0 < \kappa < 1$.

This lemma is the basis for an existence result for integral equations with the Volterra operator Θ in (25) which we will deduce next.

Lemma 2. *Assume that the vector field* $k(t,s,z;y) : [0,T]^2 \times \mathcal{S} \times \mathcal{U} \mapsto \mathcal{S}$ *is continuous on its domain of definition and it satisfies the local Lipschitz condition* (23) *uniformly for all* $s,t \in [0,T]$ *and for all* $z \in \mathcal{U}$. *Moreover, assume that for every* $X \in \mathcal{S}$, $k(t,s,X;z) : [0,T]^2 \times \mathcal{U} \mapsto \mathcal{S}$ *and* $t_0(z) : \mathcal{U} \mapsto [0,T]$, $x_0(z) : \mathcal{U} \mapsto \mathcal{S}$ *are continuous. Finally, assume the function* x_0 *satisfies*

$$\sup_{z\in\mathcal{U}} \|x_0(z)\|_{\mathcal{S}} =: r_0 \leq (1-\kappa)R\mathrm{e}^{-L(R)T/\kappa} \tag{27}$$

for some $R > 0$ *and* $0 < \kappa < 1$. *Then the parametric Volterra integral equation*

$$\textit{find} \quad X \in C([0,T]\times\mathcal{U};\mathcal{S}) : \qquad X = \Theta[X] \tag{28}$$

with the Volterra integral operator Θ *defined in* (25) *admits a unique solution* $X \in \mathcal{B}_{r,R}$.

Moreover, for every $X^{(0)} \in \mathcal{B}_{r,R}$ *the sequence*

$$X^{(k+1)} := \Theta[X^{(k)}], \quad k = 0,1,2,\ldots, \tag{29}$$

converges in $C([0,T]\times\mathcal{U};\mathcal{S})$ *and* $C_{L(R)/\kappa}([0,T]\times\mathcal{U};\mathcal{S})$ *towards* X*:*

$$\|X - X^{(k)}\|_{C([0,T]\times\mathcal{U};\mathcal{S})} \to 0 \qquad k \to \infty\ .$$

Proof. By the assumptions and by Lemma 1, the Volterra integral operator Θ is a contraction on $\mathcal{B}_{r,R} \subset C([0,T]\times\mathcal{U};\mathcal{S})$ with $r = R\mathrm{e}^{-L(R)T/\kappa}$.

The existence of a unique fixed point X of (28) in the (complete metric) space $\mathcal{B}_{r,R}$ equipped with the metric induced by the norm $\|\cdot\|_{L(R)/\kappa,T,\mathcal{S}}$ now follows from Banach's fixed point theorem, if we can determine a closed subset $\mathcal{B} \subset \mathcal{B}_{r,R}$, such that $\Theta : \mathcal{B} \longrightarrow \mathcal{B}$.

For this purpose we consider the initial data x_0 as an element of $C([0,T]\times\mathcal{U};\mathcal{S})$ (i.e. constant with respect to $t \in [0,T]$), upon which we have $\|x_0\|_{L(R)/\kappa,T,\mathcal{S}} = \sup_{z\in\mathcal{U}} \|x_0(z)\|_{\mathcal{S}} =: r_0$. Now put $\mathcal{B} = \mathcal{B}_{r_1}(x_0) = \{X \in C([0,T]\times U;\mathcal{S}) : \|X - x_0\|_{L(R)/\kappa,T,\mathcal{S}} \leq r_1\}$. Choosing $r_1 = \frac{\kappa r_0}{1-\kappa}$, it holds $\mathcal{B} \subset \mathcal{B}_{r_0/(1-\kappa),R}$, which in view of condition (27) implies $\mathcal{B} \subset \mathcal{B}_{r,R}$. We further observe $x_0 = \Theta[0]$ (due to $f_j(t,0) = 0$ for all j), hence we obtain from Lemma 1

$$\|\Theta[X] - x_0\|_{L(R)/\kappa,T,\mathcal{S}} \leq \kappa\|X\|_{L(R)/\kappa,T,\mathcal{S}} \leq \kappa(r_0 + r_1) = \frac{\kappa r_0}{1-\kappa} = r_1\ .$$

Thus we have shown that this choice of $\mathcal{B}$ indeed has the required property $\Theta[\mathcal{B}] \subset \mathcal{B}$.

Altogether we conclude that we may apply Banach's fixed point theorem to obtain the unique fixed point $X \in \mathcal{B}$. The same theorem then also yields that this fixed point may be obtained via the iteration (29), with convergence with respect to the norm $\|\cdot\|_{L(R)/\kappa,\mathcal{S},T}$ (although the fixed point theorem is applied to the set $\mathcal{B}$, the statements concerning the uniqueness of the fixed point as well as the convergence of the iteration extend to $\mathcal{B}_{r,R} \supset \mathcal{B}$, since Θ is a contraction on this larger set). However, since it holds $\|X'\|_{C([0,T]\times\mathcal{U};\mathcal{S})} \leq$

$\mathrm{e}^{\alpha T}\|X'\|_{\alpha,T,\mathcal{S}}$ for all $X' \in C([0,T]\times\mathcal{U};\mathcal{S})$ and all $\alpha \geq 0$, the iteration also converges in the unweighted uniform norm.

Remark 2. Note that in case of global Lipschitz continuity (i.e. the Lipschitz constant L does not depend on R), condition (27) is redundant (the right hand side may be chosen arbitrarily large). On the other hand, if k is not globally Lipschitz continuous (i.e. if $L(R)$ is an increasing function) then condition (27) can be interpreted as a small-data-assumption.

We can now give the *proof of Theorem 1*: We rewrite the initial value problem (1) with right hand side f as in (8) as Volterra integral equation (20) with kernel function k given by $k(t,s,X;z) = f(s,X;z)$ and $t_0(z) \equiv 0$. By assumption (12) and Proposition 1, the function f defined in (8) satisfies a local Lipschitz condition with Lipschitz constants $L(R)$ as in (14). Therefore Lemmas 1 and 2 are applicable and the existence of a unique solution $X \in \mathcal{B}_{r,R} \subset C([0,T]\times\mathcal{U};\mathcal{S})$ of (20) follows.

From the continuity of $f(t,X;y)$ with respect to $t \in [0,T]$ and $z \in \mathcal{U}$ and from (20) it immediately follows that $X \in C^1([0,T];C(\mathcal{U};\mathcal{S})) \cap C(\mathcal{U};C^1([0,T];\mathcal{S}))$. Under the additional assumption (17), it further follows from (20) that $X \in C(\mathcal{U};C^{k+1}([0,T];\mathcal{S}))$ by repeated differentiation.

Concerning the dependence on the parameter $z \in \mathcal{U}$, we shall use the fixed point equation (28). Then we find (we use the shorthand notation $X(t;z) = X(t,x_0(z);z)$)

$$
\begin{aligned}
&\|X(t;z) - X(t;z')\|_{\mathcal{S}} \\
&\quad \leq \|x_0(z) - x_0(z')\|_{\mathcal{S}} + \left\| \int_0^t \big(f(s,X(s;z);z) - f(s,X(s;z');z') \big)\,\mathrm{d}s \right\|_{\mathcal{S}} \\
&\quad \leq \|x_0(z) - x_0(z')\|_{\mathcal{S}} + \left\| \int_0^t \big(f_0(s,X(s;z)) - f_0(s,X(s;z')) \big)\,\mathrm{d}s \right\|_{\mathcal{S}} \\
&\quad + \left\| \int_0^t \sum_{j\geq 1} \Big(\big(f_j(s,X(s;z)) - f_j(s,X(s;z')) \big) z_j + f_j(s,X(s;z'))(z_j - z_j') \Big)\,\mathrm{d}s \right\|_{\mathcal{S}} \\
&\quad \leq \|x_0(z) - x_0(z')\|_{\mathcal{S}} + L_0(R)\int_0^t \big\|X(s;z) - X(s;z')\big\|_{\mathcal{S}}\,\mathrm{d}s \\
&\quad + \sum_{j\geq 1}\int_0^t \Big(L_j(R)\big\|X(s;z) - X(s;z')\big\|_{\mathcal{S}} + L_j(R)\|X(s;z')\|_{\mathcal{S}}|z_j - z_j'| \Big)\,\mathrm{d}s \\
&\quad \leq \|x_0(z) - x_0(z')\|_{\mathcal{S}} + \frac{\kappa}{L(R)}\sum_{j\geq 0} L_j(R) e^{tL(R)/\kappa} \sup_{s\in[0,T]} \mathrm{e}^{-sL(R)/\kappa}\big\|X(s;z) - X(s;z')\big\|_{\mathcal{S}} \\
&\quad + \frac{\kappa}{L(R)}\sum_{j\geq 1} L_j(R)|z_j - z_j'|\mathrm{e}^{tL(R)/\kappa} \sup_{s\in[0,T]} \mathrm{e}^{-sL(R)/\kappa}\big\|X(s;z')\big\|_{\mathcal{S}}\,.
\end{aligned}
$$

Multiplying by $\mathrm{e}^{-tL(R)/\kappa}$ and taking the supremum over $t \in [0,T]$ then results in the estimate

$$(1-\kappa) \sup_{t\in[0,T]} \mathrm{e}^{-tL(R)/\kappa} \|X(t;z) - X(t;z')\|_{\mathcal{S}}$$
$$\leq \|x_0(z) - x_0(z')\|_{\mathcal{S}} + \frac{\kappa}{L(R)} \sup_{t\in[0,T]} \mathrm{e}^{-tL(R)/\kappa} \big\|X(t;z')\big\|_{\mathcal{S}} \sum_{j\geq 1} L_j(R)|z_j - z_j'| \,. \tag{30}$$

Now recall $X \in \mathcal{B}_{r,R}$, i.e. $\|X\|_{L(R)/\kappa,T,\mathcal{S}} \leq r = R\mathrm{e}^{-L(R)T/\kappa}$. The estimate (18) then is a simplified (though slightly weaker) variant of (30).

Again by the fixed point equation (28) we find

$$\begin{aligned}
&\|X(t,x_0;z) - X(t,x_0';z)\|_{\mathcal{S}} \\
&\leq \|x_0 - x_0'\|_{\mathcal{S}} + \left\| \int_0^t \big(f(s, X(s,x_0;z);z) - f(s,X(s,x_0';z);z) \big) \, \mathrm{d}s \right\|_{\mathcal{S}} \\
&\leq \|x_0 - x_0'\|_{\mathcal{S}} + L(R) \int_0^t \big\| X(s,x_0;z) - X(s,x_0',z) \big\|_{\mathcal{S}} \, \mathrm{d}s \\
&\leq \|x_0 - x_0\|_{\mathcal{S}} + \kappa e^{tL(R)/\kappa} \sup_{s\in[0,T]} \mathrm{e}^{-sL(R)/\kappa} \big\| X(s,x_0;z) - X(s,x_0';z) \big\|_{\mathcal{S}} \,.
\end{aligned}$$

Multiplying by $\mathrm{e}^{-tL(R)/\kappa}$ and taking the supremum over $t \in [0,T]$ then proves (19). This completes the proof. □

Remark 3. Instead of the Volterra operator Θ defined in (25) we could have used parametric Volterra operators $\Theta_z : C([0,T];\mathcal{S}) \longrightarrow C([0,T];\mathcal{S})$,

$$\Theta_z[X](t) = x_0(z) + \int_{t_0(z)}^t k_z\big(t,s,X(s)\big) \, \mathrm{d}t \,,$$

i.e. considering the situation for every fixed instance of parameters $z \in \mathcal{U}$. Then the existence result follows by essentially the same arguments (dropping the supremum over $z \in \mathcal{U}$ everywhere). The estimate (30) can be derived in exactly the same way, and then the choice of the product topology on $\mathcal{U}$, Assumption (12) and assuming continuity of x_0 proves continuity of X on $\mathcal{U}$. Together with the continuity of $z \mapsto f(\cdot,\cdot;z)$ and (1) we further obtain $X \in C(\mathcal{U}; C^1([0,T];\mathcal{S}))$. Finally, we obtain the following a priori estimates.

Corollary 1. *The solution $X(t,x_0(z);z)$ of (1) satisfies*

$$\big\|X\big(t,x_0(z);z\big)\big\|_{\mathcal{S}} \leq \frac{\mathrm{e}^{tL(R)/\kappa}}{1-\kappa} \|x_0(z)\|_{\mathcal{S}} \,, \quad \|X\|_{L(R)/\kappa,T,\mathcal{S}} \leq \frac{1}{1-\kappa} \sup_{z\in\mathcal{U}} \|x_0(z)\|_{\mathcal{S}} \,.$$

Remark 4. We shall complete this section by discussing two particular examples which show the necessity of the small-data-assumption imposed in Theorem 1. Therein we simply consider the scalar non-parametric setting.

i) The problem $\dot{X} = X^n$, $n \geq 2$, $X(0) = x_0$, has the solution

$$X(t) = -\frac{1}{n-1}\left(\frac{1}{x_0^{n-1}} - (n-1)t\right)^{-\frac{1}{n-1}} ,$$

which clearly is only well-defined on the interval $[0, \frac{1}{(n-1)x_0^{n-1}})$. Hence for this interval to contain $[0, T]$ we need the initial datum x_0 to be small enough.

iia) The problem $\dot{X} = \mathrm{e}^X$, $X(0) = x_0$, has the solution

$$X(t) = -\log\left(\mathrm{e}^{-x_0} - t\right) ,$$

which is well-defined on the interval $[0, T]$ if, and only if, $x_0 < -\log T$. This example shows that in general we indeed need an additional assumption on the functions f_j, apart from the Lipschitz condition (11).

iib) The problem $\dot{X} = \mathrm{e}^{-X}$, $X(0) = x_0$, has the solution

$$X(t) = \log\left(\mathrm{e}^{x_0} + t\right) ,$$

which is well-defined for all $t > 0$, independent of the choice of x_0. Closer inspection of our arguments reveals that the proof of Theorem 1 may be reformulated to show existence of solutions on the symmetric interval $[-T, T]$, upon which all conditions necessary in example (iia) apply here as well (note that both functions yield the same Lipschitz constants $L(R)$).

3 Analytic parameter dependence

In the previous section, we showed for any instance of the parameter vector $z \in \mathcal{U}$ the existence of a unique solution in $C^1([0, T]; \mathcal{S})$ of the parametric initial value problem (1). Moreoever, the proof also showed that the mapping $\mathcal{U} \ni z \mapsto X(t; z) \in C^1([0, T]; \mathcal{S})$ is continuous.

To this end, we imposed local Lipschitz conditions (11) for the f_j and the assumption (12) of summability of the Lipschitz constants $L_j(R)$, which implied a local Lipschitz condition for the right hand side function f uniform in the parameter space $\mathcal{U}$. In the present section, we will sharpen this result in proving that the parameter dependence $\mathcal{U} \ni z \mapsto X(t; z) \in C^1([0, T]; \mathcal{S})$ (with $C([0, T]; \mathcal{S})$ either in the weighted or unweighted setting) is, in fact, *analytic*. This will be achieved by tracking analytic dependence through the fixed point iteration.

To prove this, we track analyticity of the iterates and use the uniformity of the convergence in the successive approximation (29) as used in the proof of Theorem 1 to deduce analytic dependence of the solution $X(t, x_0(z); z)$ on the parameters $z \in \mathcal{U}$.

It will be necessary to impose analyticity assumptions on the dependence of the $f_j(t; x)$ on the second variable $x \in \mathcal{S}$. Since both parameter domain $\mathcal{U}$ and the domain $C^1([0, T]; \mathcal{S})$ are infinite-dimensional, this requires classical tools of Complex Analysis in infinite dimensional spaces which we will recapitulate first; our basic reference is [15].

3.1 Preliminaries on analyticity and holomorphy in infinite dimensional spaces

Our strategy to prove analytic dependence of the solutions is to track analyticity of the iterates $X^{(k)}$ in (29) and use that holomorphy is preserved under uniform convergence. It is easy to see (choose $\mathcal{S} = \mathbb{C}$), that to this end analytic dependence of the $f_j(t,x)$ in (8) on the state variable x is necessary. To formulate analytic dependence of the $f_j(t,x)$ in the presently considered general state spaces $\mathcal{S}$, we recapitulate notions of analyticity of mappings $f : \mathcal{U} \mapsto \mathcal{S}$ from [15].

Definition 1. For $m \in \mathbb{N}$ and for a set $\mathcal{U} \subset \mathbb{C}^m$ we say that a map $f : \mathcal{U} \mapsto \mathcal{S}$ is

i) *analytic*, if for each $a \in \mathcal{U}$ exists a family $\{c_\nu\}_{\nu \in \mathbb{N}_0^m} \subset \mathcal{S}$ such that the series $(c_\nu (z-a)^\nu)_{\nu \in \mathbb{N}_0^m}$ is summable and converges to $f(z)$ for z sufficiently close to a;

ii) *holomorphic* if, for any $a \in \mathcal{U}$, each first order partial derivative
$$\lim_{\xi_j \to 0} \frac{1}{\xi_j} \Big[f(a + \xi_j e_j) - f(a) \Big]$$
exists;

iii) *scalarly analytic or holomorphic* if for each $z' \in \mathcal{S}'$ the complex-valued function $z' \circ f : \mathcal{U} \mapsto \mathbb{C}$ is analytic or holomorphic on $\mathcal{U}$.

We denote the vector space of analytic maps $\mathbb{C}^m \supseteq \mathcal{U} \mapsto \mathcal{S}$ by $\mathcal{A}(\mathcal{U}, \mathcal{S})$.

Evidently, either of i) or ii) implies iii), and i), ii) and iii) are, in fact, equivalent if $\mathcal{S}$ is a Frechét space (or, even more general, a sequentially complete locally convex space). We refer to [15, Thm. 2.1.3] for details. We shall also require the notion of analyticity in the case when $\mathcal{U} \subset \mathcal{X}$ is an open set in an infinite dimensional Banachspace $\mathcal{X}$ over $\mathbb{C}$. We have (see [15, Prop. 3.1.2]):

Definition 2. Let $\mathcal{X}$ be a Banachspace over $\mathbb{C}$ and let $\emptyset \neq \mathcal{U} \subseteq \mathcal{X}$ be an open set. Then $f : \mathcal{U} \mapsto \mathcal{S}$ is *analytic*, or $f \in \mathcal{A}(\mathcal{U}, \mathcal{S})$ if for all $z' \in \mathcal{S}'$ the function $z' \circ f : \mathcal{U} \mapsto \mathbb{C}$ is analytic, i.e. $f : \mathcal{U} \mapsto \mathcal{S}$ is analytic if
$$\forall z' \in \mathcal{S}' : \quad z' \circ f \in \mathcal{A}(\mathcal{U}, \mathbb{C}) \ .$$

Analyticity of a mapping f follows from complex differentiability (see [15, Prop. 3.1.3, Cor. 3.1.4, Thm. 3.1.5]).

Proposition 2. *A map $f : \mathcal{U} \mapsto \mathcal{S}$ is analytic if and only if it is (Frechét-) differentiable at each point $a \in \mathcal{U}$, i.e. for every $a \in \mathcal{U}$ there exists a linear map $\lambda_a : \mathcal{X} \to \mathcal{S}$, such that* $\lim_{h\to 0}\big(f(a+h) - f(a) - \lambda_a(h)\big) = 0$.

For our purposes this notion of analyticity has to be further generalized.

Definition 3. Let $\mathcal{U}$ be an open subset of a locally convex complex vector space $\mathcal{X}$. Then $f : \mathcal{U} \to \mathcal{S}$ is analytic, or $f \in \mathcal{A}(\mathcal{U}, \mathcal{S})$, if it is continuous and $f\big|_{\mathcal{E}\cap\mathcal{U}}$ is analytic for every finite dimensional affine subspace $\mathcal{E} \subset \mathcal{X}$.

Using that the linear parameter dependence in the affine expansion (8) is, in fact, analytic (in the sense of Definition 3, where $\mathbb{C}^{\mathbb{N}} \supset \mathcal{U}$ is equipped with the usual locally convex topology), we can use analyticity of the iterates and the uniformity of the convergence in the successive approximation (29) that we used in the proof of Theorem 1 to deduce holomorphic dependence of the solution $X(t, x_0; y)$ on the parameters z in a complex domain.

3.2 Analyticity Assumptions

Whereas in Theorem 1 the state space $\mathcal{S}$ was allowed to be an arbitrary Banachspace over either $\mathbb{R}$ or $\mathbb{C}$, we now work under the following assumptions.

Assumption 2 1. *All Banach spaces are understood as Banach spaces over $\mathbb{C}$. In particular, we assume that all real state spaces $\mathcal{S}$ in Theorem 1 are replaced by their extensions to the coefficient field $\mathbb{C}$ (in particular $\mathcal{S} = \mathbb{R}^d$ is replaced by $\mathcal{S} = \mathbb{C}^d$). The state space $\mathcal{S}$ is always assumed to be reflexive (but does not necessarily need to be separable).*
2. *We assume $U = [-1, 1]^{\mathbb{N}}$ and $U \subset \mathcal{U} \subset \mathbb{C}^{\mathbb{N}}$.*
3. *Finally, the time parameter t is always assumed to be real.*

Assumption 3 *For every $t \in [0, T]$, the mappings $\mathcal{S} \ni X \mapsto f_j(t; X) \in \mathcal{S}$ are analytic.*

We prepare the statement and proof of our analytic dependence theorem by the observation that Assumption (12) implies with the same argument used in the proof of Proposition 1 that the N-term truncated partial sums of (8), i.e.

$$f^{(N)}(t, x; y) := f_0(t, x) + \sum_{j=1}^{N} y_j f_j(t, x) \ , \quad 0 \le t \le T < \infty \ , y \in U \ , \qquad (31)$$

are, in fact, converging uniformly for parameter vectors $z = (z_j)_{j\ge 1}$ belonging to appropriate polydiscs: for a vector $\rho = (\rho_j)_{j\ge1}$ of radii $\rho_j > 1$, we denote by

$$\mathcal{U}_\rho := \big\{z = (z_j)_{j\ge1} \in \mathbb{C}^{\mathbb{N}} : |z_j| < \rho_j,\ j \in \mathbb{N}\big\} \qquad (32)$$

the (countable) product of open discs of radii $\rho_i \geq 1$. The product of closed discs of radii ρ_i in $\mathbb{C}^{\mathbb{N}}$ will be denoted by $\overline{\mathcal{U}_\rho}$. If $\rho = (1, 1, ...)$, we write $\mathcal{U}$ for simplicity. Note that $U \subset \mathcal{U} \subset \mathcal{U}_\rho$, and that all polydiscs $\mathcal{U}_\rho$ are open subsets of $\mathbb{C}^{\mathbb{N}}$. Further note that in (31) holds $f^{(N)}(t, x; y) = f(t, x; y^{(N)})$ where, for $y \in U$, we denoted by $y^{(N)}$ the coordinate vector y anchored at zero in all but the first N coordinates, i.e. $y^{(N)} := (y_1, y_2, ..., y_N, 0, 0, ...)$.

3.3 Analyticity Result with local Lipschitz conditions

With these notations at hand, we can now state our first main result on analytic dependence of the solutions $X(t; y)$ of the parametric initial value ODEs (1) with countably many parameters.

Theorem 4. *For parameter vectors $z = y + iw \in \mathcal{U} \subset \mathbb{C}^{\mathbb{N}}$, and with $\mathcal{S}$ denoting the "complexification" of the state space $\mathcal{S}$ in (1), consider in $[0, T]$ the parametric IVP ODE*

$$\frac{\mathrm{d}X}{\mathrm{d}t}(t; z) = f(t, X; z) \quad 0 \leq t \leq T < \infty, \quad X(0; z) = x_0(z) \in \mathcal{S} \,. \tag{33}$$

Assume that in (33) the vector field f depends on the parameter vector $z = y + iw \in \mathbb{C}^{\mathbb{N}}$ in the affine fashion

$$f(t, x; z) = f_0(t, x) + \sum_{j \geq 1} z_j f_j(t, x) \,, \quad t \in [0, T], \; x \in \mathcal{S}, \; z \in \mathcal{U} \,, \tag{34}$$

with the coefficient functions $f_j(t, x)$ satisfying (17) for some $k \geq 1$ and which are, for every $t \in [0, T]$, analytic with respect to $x \in \mathcal{S}$. Assume that $f_j(t; \cdot) \in \ell\mathrm{Lip}_0(\mathcal{S})$, i.e. $f_j(t, 0) = 0$ for all $t \in [0, T]$ and the Lipschitz constants satisfy, for $j = 0, 1, 2, ...$

$$L_j(R) := \sup_{t \in [0,T]} \sup_{x \neq x' \in B_R} \frac{\|f_j(t, x) - f_j(t, x')\|_{\mathcal{S}}}{\|x - x'\|_{\mathcal{S}}} < \infty \,. \tag{35}$$

Moreover, assume that the $L_j(R)$ are such that there exists a (not necessarily bounded) polyradius $\rho = (\rho_j)_{j \geq 1}$ of radii $\rho_j > 1$ such that

$$L(\rho, R) := \sum_{j \geq 1} \rho_j L_j(R) < \infty \,, \quad \forall R > 0 \,. \tag{36}$$

Then

(i) the parametric IVP (33) admits, for every $z \in \mathcal{U}_\rho$ and every $x_0 \in \mathcal{S}$, a unique solution $X(\cdot; z) \in C^1([0, T]; \mathcal{S})$, and there holds the a-priori estimate

$$\sup_{(t,z)\in[0,T]\times\mathcal{U}_\rho} \mathrm{e}^{-tL(\rho,R)/\kappa} \|X(t,\cdot;z)\|_{\mathcal{S}} \leq \frac{1}{1-\kappa} \sup_{z\in\mathcal{U}_\rho} \|x_0(z)\|_{\mathcal{S}} < \infty. \quad (37)$$

(ii) if, for every $t \in [0,T]$, the map $\mathcal{S} \ni x \mapsto f_j(t,x) \in \mathcal{S}$ is analytic, and if $\overline{\mathcal{U}_\rho} \ni z \mapsto x_0(z) \in \mathcal{S}$ is analytic, then the solution X depends analytically on the parameter $z \in \mathcal{U}_\rho$, i.e. $X \in \mathcal{A}(\mathcal{U}_\rho; C^1([0,T];\mathcal{S}))$,

(iii) if the functions f_j in (34) satisfy $[0,T] \ni t \mapsto f_j(t,x) \in C^k([0,T];\mathcal{S})$ for every $j = 0,1,\ldots$ and for every $x \in \mathcal{S}$ for some $k \geq 1$, then $X \in \mathcal{A}(\mathcal{U}_\rho; C^{k+1}([0,T];\mathcal{S}))$.

(iv) For every $\nu \in \mathfrak{F}$, denote by $J(\nu) = \{j \in \mathbb{N} : \nu_j \neq 0\}$ the finite "support" of ν and partition $y \in U$ as $y = (y^J, y^{(J)})$, with $y^J \in [-1,1]^{|J|}$, $y^{(J)} \in U^{(J)} := [-1,1]^{\mathbb{N}\setminus J(\nu)}$. Then the parametric solution $X(\cdot;y) \in C^{k+1}([0,T];\mathcal{S})$ is, for every fixed $y^{(J)} \in U^{(J)}$, strongly holomorphic as a $C^{k+1}([0,T];\mathcal{S})$-valued function with respect to the parameters $z^J = y^J + iw^J$ in the polydisc

$$D_{\rho^J} := \{z \in \mathbb{C}^J \,\big|\, |z_j| \leq \rho_j,\ j \in J(\nu)\}, \quad (38)$$

(v) At every $z \in \mathcal{U}_\rho \subset \mathbb{C}^{\mathbb{N}}$ in (32) with the polyradius ρ as in (36), the solution X of (33) can be represented by the Taylor expansion

$$X(t;z) = \sum_{\nu\in\mathfrak{F}} T_\nu(t) z^\nu, \quad (39)$$

where

$$T_\nu(t) = \frac{1}{\nu!}\partial_y^\nu X(t;y)|_{y=0} \in C^{k+1}([0,T];\mathcal{S}), \quad \nu \in \mathfrak{F}\ .$$

The convergence of (39) is unconditional and pointwise in $[0,T]\times\mathcal{U}_\rho$, i.e. we have pointwise convergence in $\mathcal{U}_\rho$ as a mapping taking values in the space $C^{k+1}([0,T];\mathcal{S})$ (or in $C^{k+1}_{L(\rho,R)/\kappa}([0,T];\mathcal{S})$).

(vi) For real-valued arguments, i.e. $z = y \in \mathbb{R}^{\mathbb{N}}$ and for $x_0 \in \mathbb{R}^d$, $f_j(t,x) \in \mathbb{R}^d$ for $x \in \mathbb{R}^d$, the solution $X(t,x_0;z)|_{z=y}$ coincides with the solution constructed in Theorem 1; it is the unique analytic extension of this solution to complex parameter vectors $z = y + iw$.

Proof. Concerning (i) we merely note that the only modification as compared to Theorem 1 consists in replacing everywhere $\mathcal{U}$ by $\mathcal{U}_\rho$ and $L(R)$ by $L(\rho,R)$.

(ii). We use the analyticity assumptions and the fact that the class of analytic maps has the composition property (e.g. [15, Theorem 3.1.10]) to deduce that all iterates $X^i(t;z)$ in the fixed point iteration $X^{i+1} = \Theta[X^i]$ are analytic with respect to $z \in \overline{\mathcal{U}_\rho}$ for every $t \in [0,T]$. Since $\overline{\mathcal{U}_\rho}$ is, as a countable cartesian product of (closed) discs, a compact metric space by the theorem of Tychonoff, and the sequence $\{X^i\}_{i\geq 0}$ of iterates in Banach's fixed point theorem converges pointwise at every $z \in \overline{\mathcal{U}_\rho}$ to X (as a consequence of the convergence with respect to the norm $\|\cdot\|_{L(\rho)/\kappa,T,\mathcal{S}}$), it converges in fact uniformly on $\overline{\mathcal{U}_\rho}$, and hence $X \in \mathcal{A}(\overline{\mathcal{U}_\rho}; C([0,T];\mathcal{S}))$ by [15, Theorem 3.1.5c)]. Since X solves the problem (1) we obtain, once more using the composition property of analytic maps, that also $\frac{dX}{dt} \in \mathcal{A}(\overline{\mathcal{U}_\rho}; C([0,T];\mathcal{S}))$, which then yields $X \in \mathcal{A}(\overline{\mathcal{U}_\rho}; C^1([0,T];\mathcal{S}))$.

(iii). Under condition (17), we already know $X \in C^{k+1}([0,T]; C(\mathcal{U}_\rho;\mathcal{S}))$, see Theorem 1. Repeatedly differentiating the problem (1) with respect to t as well as applying the composition property of analytic maps then yields $\frac{d^jX}{dt^j} \in \mathcal{A}(\overline{\mathcal{U}_\rho}; C([0,T];\mathcal{S}))$ for $j = 1,\dots,k+1$ and hence also $X \in \mathcal{A}(\overline{\mathcal{U}_\rho}; C^j([0,T];\mathcal{S}))$, $j = 1,\dots,k+1$. We only note that for an analytic map $A : \mathcal{U} \longrightarrow \mathcal{L}(\mathcal{X},\mathcal{Y})$ (the space $\mathcal{L}(\mathcal{X},\mathcal{Y})$ being the Banach space of all bounded linear operators from the complex Banach space $\mathcal{X}$ into another complex Banach space $\mathcal{Y}$) the map $A_x : \mathcal{U} \longrightarrow \mathcal{Y}$, $z \mapsto A(z)x$, is again analytic for every $x \in \mathcal{X}$.

(iv). All preceding arguments also apply if, for fixed real parameters $y^{(J)}$, the solution X is only considered parametric with respect to the $J < \infty$ many components $z^J \in D_{\rho^J} \subset \mathbb{C}^J$ in the disc $D_{\rho^J} \subset \mathbb{C}^J$ defined in (38).

The assertion (v) is a consequence of [15, Theorem 3.1.5a)] and the statements (i), (ii) and (iii) above, and (vi) is verified directly. □

4 Sparsity

4.1 Main Result

We showed in Theorem 4 that the solution of the parametric IVP (1) depends analytically on the parameter vector $y \in \mathcal{U} \subset \mathbb{C}^{\mathbb{N}}$ provided that the summability assumptions (12) and Assumption 3 on the analyticity of the f_j hold. We show that if in addition the sequence is sparse, in the sense that if $(\|f_j\|_{\ell\mathrm{Lip}_0(\mathcal{S},\mathrm{R})})_{j\geq 1} \in \ell^p(\mathbb{N})$ 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 5. *Consider the parametric IVP ODE* (33) *for parameter vectors* $y \in U = [-1,1]^{\mathbb{N}}$. *We assume that there exist real numbers* $R > 0$ *and* $0 < \kappa < 1$ *with the following properties:*

1. *In* (33) *the vector field* f *depends on the parameter vector* z *in the affine fashion* (8) *with the coefficient functions* f_j *satisfying for some* $0 < p < 1$
$$\left(\|f_j\|_{\ell\mathrm{Lip}_0(\mathcal{S},\mathrm{R})}\right)_{j\geq 1} \in \ell^p(\mathbb{N}) \qquad \text{and} \qquad \left(\overline{\rho}_j\|f_j\|_{\ell\mathrm{Lip}_0(\mathcal{S},\mathrm{R})}\right)_{j\geq 1} \in \ell^1(\mathbb{N}) \ , \tag{40}$$
where the polyradius $\overline{\rho}$ *is given by* $\overline{\rho}_j = \frac{\delta}{4L_j(R)} + 2$ *for some arbitrary fixed* $\delta > 0$, *and where* $L_j(R) := \|f_j\|_{\ell\mathrm{Lip}_0(\mathcal{S},\mathrm{R})}$.
2. *The initial data* $x_0 \in C([0,T] \times \mathcal{U}_{\overline{\rho}}; \mathcal{S})$ *satisfies*

$$\sup_{z\in\mathcal{U}_{\overline{\rho}}} \|x_0(z)\| \le (1-\kappa)R\mathrm{e}^{-TL(\overline{\rho},R)/\kappa} \,. \tag{41}$$

Then the Taylor expansion (39) *of the parametric solution* $X(t;y)$ *of* (33) *is p-sparse in the following sense: denoting for* $N \in \mathbb{N}$ *by* $\Lambda_N \subset \mathfrak{F}$ *a set of indices* $\nu \in \mathfrak{F}$ *corresponding to* N *Taylor coefficients* T_ν *with largest norm in* $C_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})$ *it holds*

$$\sup_{z\in\mathcal{U}} \left\| X(\cdot;z) - \sum_{\nu\in\Lambda_N} T_\nu(t) y^\nu \right\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \le CN^{-r}, \quad r = \frac{1}{p} - 1 \tag{42}$$

and where

$$\sum_{\nu\in\Lambda_N} T_\nu(t) z^\nu \in \mathbb{P}_{\Lambda_N}(U; C^1([0,T];\mathcal{S}))\,. \tag{43}$$

In the real-parameter setting the parametric solution $X(\cdot;y)$, $y \in U$, *admits the expansion*

$$X(t;y) = \sum_{\nu\in\mathfrak{F}} C_\nu(t) \Xi_\nu(y) \tag{44}$$

into a series of tensorized Chebyshev polynomials Ξ_ν *with unconditional and uniform convergence on* U. *Denoting by* $\Lambda_N \subset \mathfrak{F}$ *a set of* N *Chebyshev coefficients* $C_\nu(t) \in C_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})$ *which are largest in norm, the corresponding partial sums converge at the rate*

$$\sup_{y\in U} \left\| X(\cdot;y) - \sum_{\nu\in\Lambda_N} C_\nu(\cdot) \Xi_\nu(y) \right\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \le CN^{-r}, \quad r = \frac{1}{p} - 1\,. \tag{45}$$

Likewise, the parametric solution $X(\cdot;y)$ *admits the Legendre expansion*

$$X(t;y) = \sum_{\nu\in\mathfrak{F}} X_\nu(t) P_\nu(y) \tag{46}$$

with unconditional and pointwise (for $y \in U$*) convergence and, denoting now by* $\Lambda_N \subset \mathfrak{F}$ *a set of* N *Legendre coefficients* $X_\nu(t)$ *with largest* $C_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})$ *norms, the partial Legendre sums converge at the rate*

$$\sup_{y\in U} \left\| X(\cdot;y) - \sum_{\nu\in\Lambda_N} X_\nu(t) P_\nu(y) \right\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \le CN^{-r}, \quad r = \frac{1}{p} - 1\,. \tag{47}$$

For $0 < p \le 1$ *as in* (40), *the sequences* $(T_\nu)_{\nu\in\mathfrak{F}}$, $(C_\nu)_{\nu\in\mathfrak{F}}$, $(X_\nu)_{\nu\in\mathfrak{F}}$ *of Taylor, Chebyshev and Legendre coefficients of* X *are p-summable, i.e. it holds*

$$(T_\nu)_{\nu\in\mathfrak{F}}, (C_\nu)_{\nu\in\mathfrak{F}}, (X_\nu)_{\nu\in\mathfrak{F}} \in \ell^p\left(\mathfrak{F}; C^1([0,T];\mathcal{S})\right).$$

Finally, let (17) *be satisfied for some* $k \geq 0$*. Denote by* $\Lambda_N^k \subset \mathfrak{F}$ *a set of* N *largest Taylor coefficients (measured in* $C^{k+1}_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})$*). Then it holds*

$$\sup_{z\in\mathcal{U}} \left\| X(\cdot;z) - \sum_{\nu\in\Lambda_N^k} T_\nu(t) y^\nu \right\|_{C^{k+1}_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})} \leq C N^{-r}, \quad r = \frac{1}{p} - 1\,. \tag{48}$$

Similar statements hold for the Legendre and Chebyshev expansions.

For definitions of the tensorized Legendre and Chebyshev systems as well as the respective expansion coefficients we refer to Sections 4.5 and to 4.6 ahead.

Remark 5. We emphasize that the optimal index sets Λ_N^k in general depend on k; they even need not be monotonic in k. However, quasi-optimal (monotonic in k) index sets can be obtained iteratively as follows: denote for $j = 0, 1, \ldots, k+1$ by $\Lambda_{j,N}$ the N largest Taylor coefficients of $\frac{d^j X}{dt^j}$ (in the norm of $C_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})$). Note that the Taylor coefficients for $\frac{d^j X}{dt^j}$ are exactly the t-derivatives of the Taylor coefficients for X.

Then the index set $\Lambda = \bigcup_{j=0}^{k+1} \Lambda_{j,N}$ yields a quasi-optimal $(k+2)N$-term approximation. For $k = 0$ this follows from $|\Lambda| \leq |\Lambda_{0,N}| + |\Lambda_{1,N}| \leq 2N$ and

$$\begin{aligned}
&\sup_{z\in\mathcal{U}_{\overline{\rho}}} \left\| X(\cdot;z) - \sum_{\nu\in\Lambda} T_\nu(\cdot) z^\nu \right\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})} \\
&\leq \sup_{z\in\mathcal{U}_{\overline{\rho}}} \left\| X(\cdot;z) - \sum_{\nu\in\Lambda} T_\nu(\cdot) z^\nu \right\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \\
&+ \sup_{z\in\mathcal{U}_{\overline{\rho}}} \left\| \frac{dX}{dt}(\cdot;z) - \frac{d}{dt} \sum_{\nu\in\Lambda} T_\nu(\cdot) z^\nu \right\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \\
&\leq \sum_{\nu\notin\Lambda} \|T_\nu(\cdot)\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} + \sum_{\nu\notin\Lambda} \left\| \frac{d}{dt} T_\nu(\cdot) \right\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \\
&\leq \sum_{\nu\notin\Lambda_{0,N}} \|T_\nu(\cdot)\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} + \sum_{\nu\notin\Lambda_{1,N}} \left\| \frac{d}{dt} T_\nu(\cdot) \right\|_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \\
&\leq N^{-1/p+1} \left(\sum_{\nu\in\mathfrak{F}} \|T_\nu(\cdot)\|^p_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \right)^{1/p} \\
&+ N^{-1/p+1} \left(\sum_{\nu\in\mathfrak{F}} \left\| \frac{d}{dt} T_\nu(\cdot) \right\|^p_{L(\overline{\rho},R)/\kappa,T,\mathcal{S}} \right)^{1/p} \\
&\leq c_p N^{-1/p+1} \left(\sum_{\nu\in\mathfrak{F}} \|T_\nu(\cdot)\|^p_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})} \right)^{1/p} \\
&\equiv c_p N^{-1/p+1} \left\| (T_\nu)_{\nu\in\mathfrak{F}} \right\|_{\ell^p(\mathfrak{F};C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S}))}\,.
\end{aligned}$$

For $k > 0$ one uses analogous arguments. □

4.2 Admissible poly-radii and a preliminary estimate for Taylor coefficients

We shall establish Theorem 5 in a more general setting which is also applicable to the parametric solutions considered in [7, 8, 16].

Assumption 6 *1. $\mathcal{S}$ is some Banach space over $\mathbb{C}$.*

2. *We are given a holomorphic function $u : \mathcal{U}_\rho \longrightarrow \mathcal{S}$, which satisfies an a priori estimate*

$$\sup_{z\in\mathcal{U}_\rho} \|u(z)\|_{\mathcal{S}} \leq B(\rho)\,, \tag{49}$$

where the ρ-dependent bound satisfies for every $J \in \mathbb{N}$

$$B(\rho_1, \ldots, \rho_J, 1, \ldots) \leq B_0 \prod_{j=1}^{J} \mathrm{e}^{\alpha_j \rho_j} \tag{50}$$

for some constant B_0 independent of ρ and of J for some sequence of positive real numbers α_j.

3. *The poly-radii ρ in the previous assumption shall satisfy the conditions*

$$\rho_j \geq 1\,, \quad j \in \mathbb{N}\,, \qquad \textit{and} \qquad \sum_{j\geq 1} \rho_j L_j < \infty$$

for some fixed sequence $(L_j)_{j\geq 1}$ of positive real numbers. Such a sequence ρ will be called admissible poly-radius.

4. *The given fixed sequence $(L_j)_{j\geq 1}$ shall be summable, i.e.*

$$\sum_{j\geq 1} L_j < \infty \tag{51}$$

We note that (51) is trivially satisfied in the case when the number of parameters is finite. We are now interested in estimates for the Taylor-, Legendre and Chebyshev-coefficients of u.

Proposition 3. *Let an admissible poly-radius $\rho = (\rho_j)_{j\geq 1}$ and an analytic function u as in Assumption 6 be given. Then for every $\nu \in \mathfrak{F}$ holds*

$$\|\partial^\nu u(z)\|_{\mathcal{S}} \leq \nu! B(\rho) \rho^{-\nu} = \nu! B(\rho) \prod_{j\in\operatorname{supp}\nu} \rho_j^{-\nu_j}\,, \qquad z \in \mathcal{U}_\rho\,.$$

Proof. The proof is based on the argument in [8, Lemma 2.4].

Put $E = \operatorname{supp}\nu = \{j \in \mathbb{N} : \nu_j \neq 0\}$ and set $J = |E|$. Writing $z = (z_E, z_{E^c})$, i.e. $z_E \in \mathbb{C}^J$ contains the components corresponding to indices $j \in E$, the admissibility of ρ then implies the estimate

$$\|u(z_E, 0)\|_{\mathcal{S}} \leq B(\rho) \tag{52}$$

for every $z_E \in \mathcal{U}_{\rho,E}$. W.l.o.g. we assume $E = \{1, \ldots, J\}$ (this may always be achieved by re-numerating the variables). If we further define the sequence $\widetilde{\rho}$ by

$$\widetilde{\rho}_j = \rho_j + \varepsilon\,, \quad j \in E\,, \quad \varepsilon = \frac{\delta}{\sum_{j\in E} L_j}\,, \qquad \widetilde{\rho}_j = \rho_j\,, \quad j \notin E\,,$$

for some arbitrary real number $\delta > 0$, it is easily verified that also $\widetilde{\rho}$ is an admissible poly-radius. In particular, u_E is analytic in an open neighbourhood of $\mathcal{U}_{\rho,E}$, where we are writing $u_E(z_1, \ldots, z_J) = u_E(z_E) \equiv u(z_E, 0)$.

We may thus apply Cauchy's integral formula (see, e.g., [15]) in each variable z_j, $j \in E$, to obtain

$$\begin{aligned} u_E(z_1, \ldots, z_J) &= (2\pi i)^{-J} \oint_{\Gamma_1(z_1)} \cdots \oint_{\Gamma_J(z_J)} \frac{u(z'_E, 0)}{(z'_1 - z_1)\cdots(z'_J - z_J)}\, \mathrm{d}z'_1 \cdots \mathrm{d}z'_J \\ &= (2\pi i)^{-J} \oint_{\Gamma_1} \cdots \oint_{\Gamma_J} \frac{u(z'_1 + z_1, \ldots, z'_J + z_J)}{z'_1 \cdots z'_J}\, \mathrm{d}z'_1 \cdots \mathrm{d}z'_J\,. \end{aligned}$$

where $\Gamma_j(z_j)$ denotes the circle with radius ρ_j and center z_j, and $\Gamma_j \equiv \Gamma_J(0)$. Differentiation (or directly applying the formula for derivatives) then yields

$$\partial^\nu u(0) = \frac{\partial^{|\nu|} u}{\partial z_1^{\nu_1} \cdots \partial z_J^{\nu_J}}(0) = \nu!(2\pi i)^{-J} \oint_{\Gamma_1} \cdots \oint_{\Gamma_J} \frac{u(z_1, \ldots, z_J)}{z_1^{\nu_1+1} \cdots z_J^{\nu_J+1}}\, \mathrm{d}z_1 \cdots \mathrm{d}z_J\,.$$

Eventually, together with (52) we conclude

$$\|\partial^\nu u(0)\|_{\mathcal{S}} \leq \nu! B(\rho)\rho^{-\nu} = \nu! B(\rho) \prod_{j\in E} \rho_j^{-\nu_j}\,.$$

This proves the claim. □

Note that the proposition holds for arbitrary admissible poly-radii. Hence the estimate can be improved by taking the infimum over all sequences ρ on the right hand side. Further note, that the optimal sequence ρ for such an estimate (if a minimizer exists) might depend on ν. For our purpose of proving the p-summability of the Taylor-coefficients this means that for every given $\nu \in \mathfrak{F}$ we can construct suitable admissible poly-radii $\rho(\nu)$, apply Proposition 3 and afterwards sum up the resulting estimates.

Within these considerations it will be necessary to have sharper bounds on the constants $B(\rho)$ which leads to the following distinction of admissible poly-radii.

Definition 4. An admissible poly-radius $\rho = (\rho_j)_{j\geq 1}$ will be called (b,δ)-admissible, if

$$\sum_{j\geq 1} \rho_j \alpha_j \leq b - \delta\ .$$

As a particular consequence, if we fix some (sufficiently large) constant b, then for every $\delta > 0$ every (b,δ)-admissible sequence ρ will satisfy the (uniform!) estimate $b(\rho) \leq b_0 e^b$.

4.3 A construction of $(b, \delta/2)$-admissible poly-radii

The basis for all the ensuing considerations is the following assumption.

Assumption 7 *The constants α_j, $j \geq 1$, shall be summable, $\sum_{j\geq 1} \alpha_j < \infty$; in other words, there exist a (sufficiently large) constant b and some $\delta > 0$ such that the sequence $\rho = (1, 1, \dots)$ is a (b,δ)-admissible poly-radius.*

Our construction is essentially based on analogous arguments in [8]. To begin with, we fix a multiindex $\nu \in \mathfrak{F}$ and choose $M \in \mathbb{N}$ such that

$$\sum_{j>M} \max(\alpha_j, L_j) \leq \frac{\delta}{12}\ .$$

This is feasible due to the Assumption (51), Assumption 7 and $\max(\alpha_j, L_j) \leq \alpha_j + L_j$. We shall use the abbreviation $\gamma_j = \max(\alpha_j, L_j)$ Without loss of generality, we assume that the indexing of the parameters z_j is chosen such that the sequence $\left(\gamma_j\right)_{j\geq 1}$ is non-increasing. Then we partition $\mathbb{N}$ into two sets $E = \{1, \dots, M\}$ and $F = \mathbb{N} \setminus E$. We further choose $\kappa > 1$ such that

$$(\kappa - 1) \sum_{j\leq M} \gamma_j \leq \frac{\delta}{4}\ .$$

Finally, we define our sequence ρ by

$$\rho_j = \kappa\,, \quad j \in E\,; \qquad \rho_j = \max\Big(1, \frac{\delta \nu_j}{4|\nu_F|\gamma_j}\Big)\,, \quad j \in F\,,$$

where ν_E denotes the restriction of ν to the index set E, and $|\nu_F|$ denotes the ℓ_1-norm of the multiindex restricted to F, i.e. $|\nu_F| = \sum_{j>M} \nu_j$ (with the convention $\frac{\nu_j}{|\nu_F|} = 0$ if $|\nu_F| = 0$).

Now we first verify that this sequence ρ is indeed $(b, \delta/2)$-admissible. To do so, we estimate

$$\begin{aligned}\sum_{j\geq 1}\rho_j\alpha_j &= \kappa\sum_{j\leq J}\alpha_j + \sum_{j>J}\max\Big(1, \frac{\delta\nu_j}{4|\nu_F|\gamma_j}\Big)\alpha_j\\ &\leq (\kappa-1)\sum_{j\leq J}\gamma_j + \sum_{j\leq J}\alpha_j + \sum_{j>J}\Big(1+\frac{\delta\nu_j}{4|\nu_F|\gamma_j}\Big)\alpha_j\\ &\leq \frac{\delta}{4} + \sum_{j\geq 1}\alpha_j + \frac{\delta}{4} \leq b - \frac{\delta}{2}\,.\end{aligned}$$

Similarly, it follows from assumption (51) that

$$\sum_{j\geq 1}\rho_j L_j \leq \frac{\delta}{2} + \sum_{j\geq 1} L_j < \infty\,.$$

By Proposition 3 we bound $t_\nu := (\nu!)^{-1}(\partial_y^\nu u)(0)$ by

$$\|t_\nu\|_{\mathcal{S}} \leq B(\rho)\rho^{-\nu} \leq b_0 e^b \Big(\prod_{j\in E}\eta^{\nu_j}\Big)\Big(\prod_{j\in F}\Big(\frac{|\nu_F|d_j}{\nu_j}\Big)^{\nu_j}\Big)\,, \tag{53}$$

where $\eta = \frac{1}{\kappa} < 1$ and $d_j = \frac{4\gamma_j}{\delta}$. Moreover, factors with exponent $\nu_j = 0$ are understood to be 1. Finally, we note that the choice of M implies

$$\|d\|_{\ell_1(F)} = \sum_{j>M} d_j \leq \frac{1}{3}\,. \tag{54}$$

□

4.4 Summability of Taylor coefficients

The constructions in the previous section are the basis for the following theorem. We shall follow closely the argument given in [8, Section 4.4]. Before stating the main theorem, we mention the following basic result of [7].

Proposition 4. *Given $0 < p < 1$, it holds $\big(\frac{|\nu|!}{\nu!}b^\nu\big)_{\nu\in\mathfrak{F}} \in \ell_p(\mathfrak{F})$ if, and only if, $\sum_{j\in\mathbb{N}} b_j < 1$ and $(b_j)_{j\in\mathbb{N}} \in \ell_p(\mathbb{N})$.*

With the help of this proposition we are finally able to prove the desired summability result.

Theorem 8. *Let the Assumptions 6 and 7 be satisfied. Moreover, suppose*

$$\big(\alpha_j\big)_{j\in\mathbb{N}}, \big(L_j\big)_{j\in\mathbb{N}} \in \ell_p(\mathbb{N})\,, \qquad i.e. \qquad \big(\gamma_j\big)_{j\in\mathbb{N}} \in \ell_p(\mathbb{N})\,. \tag{55}$$

Then, for the same value of p, the sequence $\{t_\nu\}_{\nu\in\mathcal{F}}$ of Taylor coefficients $t_\nu = \frac{1}{\nu!}\partial^\nu u(0) \in \mathcal{S}$ of the given function u is p-summable in the sense that $\big(\|t_\nu\|_{\mathcal{S}}\big)_{\nu\in\mathfrak{F}} \in \ell_p(\mathfrak{F})$.

Proof. We conclude from (53)

$$\begin{aligned}\sum_{\nu\in\mathfrak{F}}\big(\|t_\nu\|_{\mathcal{S}}\big)^p &\le b_0^p e^{bp}\sum_{\nu\in\mathfrak{F}}\Big(\prod_{j\in E}\eta^{\nu_j}\Big)^p\Big(\prod_{j\in F}\Big(\frac{|\nu_F|d_j}{\nu_j}\Big)^{\nu_j}\Big)^p\\ &\equiv b_0^p e^{bp}\Big(\sum_{\nu\in\mathfrak{F}_E}\beta(\nu)^p\Big)\Big(\sum_{\nu\in\mathfrak{F}_F}\widetilde{\beta}(\nu)^p\Big) \equiv b_0^p e^{bp}A_E A_F\,,\end{aligned}$$

where $\mathfrak{F}_E = \{\nu\in\mathfrak{F} : \operatorname{supp}\nu\subset E\}$ and $\mathfrak{F}_F = \{\nu\in\mathfrak{F} : \operatorname{supp}\nu\subset F\}$. Then it further follows

$$A_E = \Big(\sum_{\nu\in\mathfrak{F}_E}\beta(\nu)^p\Big) = \sum_{\nu\in\mathfrak{F}_E}\prod_{j\in E}\eta^{\nu_j p} = \prod_{j\in E}\sum_{n=0}^{\infty}\eta^{np} = \Big(\frac{1}{1-\eta^p}\Big)^M,$$

recall $\eta < 1$ (since E is finite and $|E| = M$ the index set $\mathfrak{F}_E$ may be identified with $\mathbb{N}_0^M$). Now we turn to showing $A_F < \infty$. Using as before the convention $0^0 = 1$ and $d^{\nu_F} = \prod_{j\in F} d_j^{\nu_j}$ we find

$$\widetilde{\beta}(\nu) = \prod_{j\in F}\Big(\frac{|\nu_F|d_j}{\nu_j}\Big)^{\nu_j} = \frac{|\nu_F|^{|\nu_F|}}{\prod_{j\in F}\nu_j^{\nu_j}}d^{\nu_F}\,, \qquad \nu\in\mathfrak{F}_F\,. \tag{56}$$

Applying the Stirling inequalities

$$\frac{n!\mathrm{e}^n}{\mathrm{e}\sqrt{n}} \le n^n \le \frac{n!\mathrm{e}^n}{\sqrt{2\pi n}}\,, \qquad n\ge 1\,,$$

we can further estimate the numerator and denominator in (56),

$$|\nu_F|^{|\nu_F|} \le |\nu_F|!\mathrm{e}^{|\nu_F|} \qquad \text{and} \qquad \prod_{j\in F}\nu_j^{\nu_j} \ge \frac{\nu_F!\mathrm{e}^{|\nu_F|}}{\prod_{j\in F}\max(1,\mathrm{e}\sqrt{\nu_j})} \ge \frac{\nu_F!\mathrm{e}^{|\nu_F|}}{\prod_{j\in F}\mathrm{e}^{\nu_j}}\,,$$

where at the end we used the bound $\max(1,\mathrm{e}\sqrt{n}) \le \mathrm{e}^n$. Altogether we then obtain from (56)

$$\widetilde{\beta}(\nu) \le \frac{|\nu_F|!}{\nu_F!}e^{|\nu_F|}d^{\nu_F} = \frac{|\nu_F|!}{\nu_F!}\prod_{j\in F}(ed_j)^{\nu_j}\,.$$

We next apply Proposition 4 to the sequence $(\mathrm{e}d_j)_{j\in\mathbb{N}}$. The assumptions of Proposition 4 are satisfied due to (54), $\mathrm{e} < 3$, and due to condition (55). Eventually, this yields $\sum_{\nu\in\mathfrak{F}_F}\widetilde{\beta}(\nu)^p < \infty$, and thus the asserted summability of the Taylor coefficients. □

4.5 Legendre Approximation

In this section we shall consider expansions of the mapping $z \mapsto u(z)$ into series of tensorized Legendre polynomials. We show that N-term truncated Legendre expansions yield optimal L^2 convergence rates. Their orthogonality with respect to the uniform Lebesgue measure on $(-1,1)$ renders them particularly suitable as bases in so-called *stochastic Galerkin Methods* (see, e.g., [10] and the references there). To fix notations, and to facilitate the (countable) tensor product construction, normalization is crucial. In the univariate case, we define the system $(P_n)_{n\geq 0}$ to be Legendre polynomials with the L_∞-normalization $\|P_n\|_{L_\infty([-1,1])} = P_n(1) = 1$ and we denote by $L_n = \sqrt{2n+1}P_n$ their L_2-normalized version, i.e.

$$\int_{-1}^{1} |L_n(s)|^2 \frac{\mathrm{d}s}{2} = 1\,, \quad n \geq 0\,.$$

Note that in particular $P_0 \equiv L_0 \equiv 1$. For $\nu \in \mathfrak{F}$, we define two families of tensorized Legendre polynomials,

$$P_\nu(z) = \prod_{j\geq 1} P_{\nu_j}(z_j) \qquad \text{and} \qquad L_\nu(z) = \prod_{j\geq 1} L_{\nu_j}(z_j)\,.$$

The property $P_0 \equiv L_0 \equiv 1$ renders $P_\nu(z)$ and $L_\nu(z)$ well-defined for all $\nu \in \mathfrak{F}$.

As a direct consequence we further note that $(L_\nu)_{\nu\in\mathfrak{F}}$ is an orthonormal basis in $L_2(U, \mathrm{d}\mu)$, where $d\mu$ is the countable product of the probability measures $\frac{dy_j}{2}$ on $[-1,1]$. The vector-valued spaces $L_2(U, \mathrm{d}\mu; \mathcal{S})$ as well as the Bochner spaces $L_p(U, \mathrm{d}\mu; \mathcal{S})$, $0 < p \leq \infty$ are to be understood similarly.

From conditions (49) and (50) we immediately conclude $u \in L_\infty(U, \mathrm{d}\mu; \mathcal{S}) \hookrightarrow L_2(U, \mathrm{d}\mu; \mathcal{S})$, thus we obtain the unique expansions

$$u(y) = \sum_{\nu\in\mathfrak{F}} u_\nu P_\nu(y) = \sum_{\nu\in\mathfrak{F}} v_\nu L_\nu(y) \tag{57}$$

with convergence in $L_2(U, \mathrm{d}\mu; \mathcal{S})$, where the $\mathcal{S}$-valued coefficients u_ν and v_ν are given by

$$v_\nu = \int_U u(y) L_\nu(y)\, \mathrm{d}\mu(y) \qquad \text{and} \quad u_\nu = \left(\prod_{j\geq 1}(1+2\nu_j)\right)^{1/2} v_\nu\,.$$

We then find the following analog of Theorem 8 for tensorized Legendre expansions.

Theorem 9. *Let the Assumptions 6 and 7 and condition* (55) *be satisfied. Then the Legendre coefficients u_ν and v_ν of the given function u are p-summable, i.e.* $\left(\|u_\nu\|_{\mathcal{S}}\right)_{\nu\in\mathfrak{F}}, \left(\|v_\nu\|_{\mathcal{S}}\right)_{\nu\in\mathfrak{F}} \in \ell_p(\mathfrak{F})$.

A proof may be obtained following along the lines of the one given in [8], but we prefer to give an alternative argument leading to sharper estimates for the Legendre coefficients. This second proof is an adaption of (real variable) estimates presented in Sec. 6 of [7].

We start by recalling Rodrigues' formula,

$$L_n(s) = \frac{(-1)^n\sqrt{2n+1}}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}s^n}(1-t^2)^n\,, \quad n \in \mathbb{N}\,.$$

If $g : [-1,1] \longrightarrow \mathcal{S}$ is some given C^∞-function of one variable, we then can use partial integration to estimate the corresponding Legendre-coefficients:

$$g_n = \frac{(-1)^n\sqrt{2n+1}}{2^n n!}\int_{-1}^{1} g(s)P_n(s)\,\mathrm{d}s = \frac{\sqrt{2n+1}}{2^n n!}\int_{-1}^{1} g^{(n)}(s)(1-s^2)^n\,\mathrm{d}s\,,$$

and it follows

$$\|g_n\|_{\mathcal{S}} \le \frac{\sqrt{2n+1}}{2^n n!} I_n \|g^{(n)}\|_{L_\infty([-1,1];\mathcal{S})}\,,$$

where $I_n = \int_{-1}^{1}(1-s^2)^n\,\mathrm{d}s$. Partial integration and the obvious relation $I_n = I_{n-1} - \int_{-1}^{1} s^2(1-s^2)^{n-1}ds$ reveal the recursion $I_n = \frac{2n}{2n+1}I_{n-1}$. Thus we have

$$I_n = \prod_{k=1}^{n}\frac{2k}{2k+1} = \frac{2^n n!}{(2n+1)(2n-1)!!} \quad \text{for} \quad n \ge 1\,, \quad I_0 = 1\,.$$

Combining these formulas gives

$$\|g_n\|_{\mathcal{S}} \le \frac{1}{\sqrt{2n+1}(2n-1)!!}\|g^{(n)}\|_{L_\infty([-1,1];\mathcal{S})} \le \frac{1}{\sqrt{2n+1}n!}\|g^{(n)}\|_{L_\infty([-1,1];\mathcal{S})}\,.$$

This one-variable estimate may now be applied to our analytic function u with respect to every parameter y_j. In this way we can estimate the Legendre coefficients u_ν by the L_∞-norm of the partial derivative $\partial^\nu u$:

$$\|u_\nu\|_{\mathcal{S}} \le \|\partial^\nu u\|_{L_\infty(U;\mathcal{S})}\prod_{j\in\operatorname{supp}\nu}\sqrt{2\nu_j+1}\cdot\frac{1}{\sqrt{2\nu_j+1}\nu_j!} = \frac{1}{\nu!}\|\partial^\nu u\|_{L_\infty(U;\mathcal{S})}\,.$$

Combining this with Proposition 3 and recalling $U \subset \mathcal{U}_\rho$ we have shown

Proposition 5. *Let the Assumption 6 be satisfied. Further let ρ be an arbitrary admissible sequence. Then we have the estimate*

$$\|u_\nu\|_{\mathcal{S}} \le B(\rho)\prod_{j\in\operatorname{supp}\nu}\rho_j^{-\nu_j} \equiv B(\rho)\rho^{-\nu}\,.$$

In other words, the Legendre coefficients u_ν may be estimated in exactly the same way as the Taylor coefficients t_ν, thus exactly the same arguments as for the proof of Theorem 8 prove also Theorem 9. We only note that obviously $\|v_\nu\|_{\mathcal{S}} \le \|u_\nu\|_{\mathcal{S}}$, hence the summability of $(\|u_\nu\|_{\mathcal{S}})_{\nu\in\mathfrak{F}}$ immediately implies the result for $(\|v_\nu\|_{\mathcal{S}})_{\nu\in\mathfrak{F}}$.

4.6 Chebyshev Approximation

As a further example of orthogonal polynomials defined on the interval $[-1,1]$ we consider the Chebyshev polynomials $\Xi_n(t) = \cos(n \arccos(t))$, $n \in \mathbb{N}$, and $\Xi_0(t) \equiv 1$. Whereas Legendre polynomials are, due to their orthogonality properties with respect to the uniform probability measure, well suited for stochastic Galerkin discretizations, the Chebyshev family is particularly well-suited for *collocation*; this has been exploited for several decades in connection with spectral methods [5]. Chebyshev polynomials satisfy

$$\|\Xi_n\|_{L_\infty([-1,1])} = 1\,, \quad n \ge 0\,, \quad \text{and} \quad \int_{[-1,1]} |\Xi_n(t)|^2 \frac{\mathrm{d}t}{\sqrt{1-t^2}} = \frac{\pi}{2}\,, \quad n \in \mathbb{N}\,.$$

As before we consider the set U equipped with the Borel σ-algebra $\mathcal{B}(U)$ and the product measure

$$\mathrm{d}\eta = \bigotimes_{j\in\mathbb{N}} \frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}\,.$$

As in the Legendre-case conditions (49) and (50) imply $u \in L_\infty(U, \mathrm{d}\eta; \mathcal{S}) \hookrightarrow L_2(U, \mathrm{d}\eta; \mathcal{S})$. Moreover, the system of tensorized polynomials $(\Xi_\nu)_{\nu\in\mathfrak{F}}$, where for $\nu \in \mathfrak{F}$ we put $\Xi_\nu(y) = \prod_{j\in\mathbb{N}} \Xi_{\nu_j}(y_j)$, constitutes an orthogonal basis. Note that they are not orthonormal with respect to the measure η, but there holds

$$\int_U |\Xi_\nu(y)|^2 \,\mathrm{d}\eta(y) = \prod_{j\in\operatorname{supp}\nu} \frac{1}{2} = 2^{-|\operatorname{supp}\nu|}\,.$$

Nevertheless, this yields the unique Chebyshev expansions

$$u(y) = \sum_{\nu\in\mathfrak{F}} w_\nu \Xi_\nu(y)\,, \quad y \in U\,, \qquad w_\nu = 2^{|\operatorname{supp}\nu|} \int_U u(y)\Xi_\nu(y)\mathrm{d}\eta(y)\,, \quad (58)$$

with convergence in $L_2(U, \mathrm{d}\eta; \mathcal{S})$.

We again aim at a summability result as in Theorems 8 and 9, and once more we start with an estimate in terms of arbitrary admissible sequences.

Proposition 6. *Under the assumptions of Proposition 5 it holds*

$$\|w_\nu\|_{\mathcal{S}} \le B(\rho) \prod_{j\in\operatorname{supp}\nu} 2\rho_j^{-\nu_j} \equiv B(\rho) 2^{|\operatorname{supp}\nu|} \rho^{-\nu}\,.$$

Proof. W.l.o.g we discuss the case $\nu = ne^1$, $n \in \mathbb{N}$, the general case being a straightforward modification by applying the single-variable case to every variable z_j with $j \in \operatorname{supp} \nu$.

Similar to the proof of Proposition 3 we write $z = (z_1, z') \in \mathcal{U}_\rho$ and put $u_1(z_1) = u(z_1, z')$ for some arbitrary $z' \in \mathcal{U}' = \prod_{j \geq 2} \{\zeta \in \mathbb{C} : |\zeta| \leq 1\}$ (note that due to the assumption on ρ we have $\mathcal{U} \subset \mathcal{U}_\rho$). Then we have $\|u_1(z_1)\|_{\mathcal{S}} \leq B(\rho)$, and we further find

$$\int_U \Xi_\nu(y) u(y)\, d\eta(y) = \int_{U'} \int_{[-1,1]} \Xi_n(t) u(t, y') \frac{\mathrm{d}t}{\pi\sqrt{1-t^2}} \mathrm{d}\eta(y') \equiv \int_{U'} \mathcal{I}_n(y')\, \mathrm{d}\eta(y')\,.$$

It clearly suffices to bound $\|\mathcal{I}_n(y')\|_{\mathcal{S}}$ independently of y'. At first we find

$$\begin{aligned}
\pi \mathcal{I}_n(y') &= \int_0^\pi \cos(n\theta) u_1(\cos\theta)\, \mathrm{d}\theta = \frac{1}{2} \int_{-\pi}^{\pi} u_1(\cos\theta) \cos(n\theta)\, \mathrm{d}\theta \\
&= \frac{1}{2i} \int_{|\zeta|=1} u_1\left(\frac{\zeta + \zeta^{-1}}{2}\right)\left(\frac{\zeta^n + \zeta^{-n}}{2}\right) \frac{\mathrm{d}\zeta}{\zeta} \equiv \frac{1}{2i} \int_{|\zeta|=1} u_1\big(\mathcal{J}(\zeta)\big)\big(\mathcal{J}(\zeta^n)\big) \frac{\mathrm{d}\zeta}{\zeta}\,.
\end{aligned}$$

The last step is verified by substituting the standard parametrization $\zeta(\theta) = \mathrm{e}^{i\theta}$ and by the Joukowsky-transform $\mathcal{J}(\zeta) = (\zeta + \zeta^{-1})/2$. Vice versa, it is well-known that $\mathcal{J}$ maps the unit circle onto the interval $[-1, 1]$ (traversed twice), since with $|\zeta| = 1$ it follows $\mathcal{J}(\zeta) = \Re(\zeta)$. Therefore we can estimate

$$\begin{aligned}
4\pi \|\mathcal{I}_n(y')\|_{\mathcal{S}} &\leq \left\| \int_{|\zeta|=1} u_1\big(\mathcal{J}(\zeta)\big) \zeta^n \frac{\mathrm{d}\zeta}{\zeta} \right\|_{\mathcal{S}} + \left\| \int_{|\zeta|=1} u_1\big(\mathcal{J}(\zeta)\big) \zeta^{-n} \frac{\mathrm{d}\zeta}{\zeta} \right\|_{\mathcal{S}} \\
&= \left\| \int_{|\zeta|=\rho_1} u_1\big(\mathcal{J}(\zeta)\big) \zeta^{n-1} \mathrm{d}\zeta \right\|_{\mathcal{S}} + \left\| \int_{|\zeta|=\rho_1^{-1}} u_1\big(\mathcal{J}(\zeta)\big) \zeta^{-n-1} \mathrm{d}\zeta \right\|_{\mathcal{S}} \\
&\leq B(\rho) \rho_1^{n-1} \cdot 2\pi\rho_1 + B(\rho) \rho_1^{-n-1} \cdot 2\pi\rho_1^{-1} = 4\pi B(\rho) \rho_1^{-n}\,.
\end{aligned}$$

Here we used Cauchy's Theorem, since $|\mathcal{J}(\zeta)| \leq \rho_1$ on $|\zeta| = \rho_1$ or $|\zeta| = \rho_1^{-1}$ and the fact that u_1 is analytic in an open neighbourhood of the disc $\{\zeta \in \mathbb{C} : |\zeta| \leq \rho_1\}$, see the proof of Proposition 3. □

The above, classical argument can essentially be found in Section 3 of [21].

By a modification of the construction and summation argument from the proof of Theorem 8 similar to the Legendre-case in [8], we then obtain an analogous summability result for coefficients in the Chebyshev expansion.

Theorem 10. *Theorem 9 remains valid upon replacing the Legendre polynomials L_ν and the the Legendre coefficients v_ν in the expansion* (57) *by the Chebyshev polynomials Ξ_ν and the Chebyshev coefficients w_ν in* (58)*.*

For later reference we present the corresponding construction of $(b, \delta/2)$-admissible sequences. As in Sec. 4.3 we start by choosing some parameter $\kappa > 1$ such that

$$(\kappa - 1)\sum_{j\geq 1}\gamma_j \leq \frac{\delta}{8}\,.$$

Further we fix an integer M such that

$$\sum_{j>M}\gamma_j \leq \frac{\delta}{24}\,,$$

and we put once more $E = \{1,\dots,M\}$ and $F = \mathbb{N}\setminus E$. Finally, we define our sequence ρ by

$$\rho_j = \kappa\,, \quad j \in E\,; \qquad \rho_j = \frac{\delta\nu_j}{4|\nu_F|\gamma_j} + 2\,, \quad j \in F\,.$$

Again we have to check first that this sequence ρ is indeed $(b,\delta/2)$-admissible. To do so, we estimate

$$\begin{aligned}\sum_{j\geq 1}\rho_j\alpha_j &= \kappa\sum_{j\leq M}\alpha_j + \sum_{j>M}\Big(2 + \frac{\delta\nu_j}{4|\nu_F|\gamma_j}\Big)\alpha_j \\ &\leq (\kappa-1)\sum_{j\leq M}\gamma_j + \sum_{j\leq M}\alpha_j + \frac{\delta}{4} + 2\sum_{j>M}\alpha_j \\ &\leq \frac{3\delta}{8} + \sum_{j\geq 1}\alpha_j + \sum_{j>M}\gamma_j \leq \frac{\delta}{2} + \sum_{j\geq 1}\alpha_j \leq b - \frac{\delta}{2}\,.\end{aligned}$$

Similarly, it follows from assumption (51) that

$$\sum_{j\geq 1}\rho_j L_j \leq \frac{\delta}{2} + \sum_{j\geq 1} L_j < \infty\,.$$

Ultimately, by applying Proposition 6 we obtain the estimate

$$\begin{aligned}\|t_\nu\|_{\mathcal{S}} &\leq 2^{|\operatorname{supp}\nu|}B(\rho)\rho^{-\nu} \\ &\leq b_0 e^b 2^M\Big(\prod_{j\in E}\eta^{\nu_j}\Big)2^{|(\operatorname{supp}\nu)\cap F|}\Big(\prod_{j\in F}\Big(\frac{|\nu_F|d_j}{\nu_j}\Big)^{\nu_j}\Big) \\ &\leq b_0 e^b 2^M\Big(\prod_{j\in E}\eta^{\nu_j}\Big)\Big(\prod_{j\in F}\Big(\frac{|\nu_F|\widetilde{d}_j}{\nu_j}\Big)^{\nu_j}\Big)\,, \qquad (59)\end{aligned}$$

where $\eta = \frac{1}{\kappa} < 1$ and $\widetilde{d}_j = \frac{8\gamma_j}{\delta}$, and the choice of M implies

$$\|\widetilde{d}\|_{\ell_1(F)} = \sum_{j>M}\widetilde{d}_j \leq \frac{1}{3}\,.$$

We shall point out two particular aspects of this construction, as they become important in Sec. 4.8:

- The first part of the sequence ρ with indices in E is independent of ν (particularly, also the partition $\mathbb{N} = E \cup F$ does not depend on ν).
- For the second part of all these sequences ρ we have $\rho_j \geq 2$ for all $j \in F$.

For the proof of Theorem 10 we finally note that the estimate (59) is of exactly the same form as the estimate (53), hence the arguments from the proof of Theorem 8 apply here as well.

4.7 Proof of Theorem 5

In the concrete situation of the parametric problem (1) we know from Theorem 4 that under the given assumptions we have $X \in \mathcal{A}(\mathcal{U}_\rho; C^1([0,T]; \mathcal{S}))$, and if additionally condition (17) holds it follows $X \in \mathcal{A}(\mathcal{U}_\rho; C^{k+1}([0,T]; \mathcal{S}))$. Together with the a priori estimates (37) we are almost in the situation of the previous summability theorems. Since we always have to fulfill the small-data-assumption (27), fulfilling the conditions (49) and (50) would imply x_0 to be bounded with respect to every component z_j, $j \in \mathbb{N}$, and thus (by Liouville's Theorem) to be constant. Eventually, this would yield the rather trivial result of affine dependence of the solution $X(\cdot; z)$ on the parameters.

However, to derive the above theorems we only need to consider the various constructed sequences $\rho(\nu)$, see Section 4.3. Then it becomes clear that we have $\rho_j(\nu) \leq \overline{\rho}_j = \max(1, \frac{\delta}{4\gamma_j})$, where $\delta > 0$ can be chosen arbitrarily (which in turn of course affects all the other constants involved). Hence we only need to satisfy the estimates (49) and (50) for the sequence $\overline{\rho}$, which in turn is ensured by the small-data-assumption (41) as well as the a priori estimate (37). □

Remark 6. As before in case of global Lipschitz-continuity of f the small-data-assumption (41) can be dropped, and we can even allow for an exponential growth of x_0 as described by the conditions (49) and (50).

Remark 7. Estimates for the initial data, as long as these satisfy the small-data-assumption, enter the summability result via the a-priori-estimates (37). In turn the small-data-assumptions are influenced via the Lipschitz-constants $L(\overline{\rho}, R)$, i.e. the second part of assumption (40).

The constant in (42) is exactly the $\ell^p(\mathfrak{F})$-quasi-norm of the sequence of Taylor coefficients $(T_\nu)_{\nu\in\mathfrak{F}}$, which depends on the a-priori-estimate for X (as a common factor in all the estimates in Proposition 3, the same factor appears in this quasi-norm) and on the $\ell^p(\mathbb{N})$-quasi-norm in the left part of (40).

Finally, higher regularity for the functions f_j does not affect the small-data-assumption or the domain of analyticity (i.e. the polydisc $\mathcal{U}_{\overline{\rho}}$), but only the a-priori-estimate. Moreover, such higher regularity of course affects the norm, in which the Taylor-coefficients are measured and hence their ℓ^p-quasi-norm as well as the index set of largest coefficients.

The same considerations apply to the Legendre and Chebyshev series of the parametric solution.

4.8 Monotone N-term approximations

The sparsity result Theorem 5 yields the *existence* of a family of sparse, N-term truncated Taylor series of the parametric solutions $X(t;y)$. Its proof, however, 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. In [14], we test algorithms towards the end of efficient constructions of concrete sequences $\{\Lambda_N\}_{N\geq 0}$.

Due to the strongly nonlinear nature of the problem (1), these methods will be based on *collocation approximations* of (1). In order to exploit sparsity in polynomial expansions of the parametric solutions, as provided by Theorem 5, with collocation schemes, it is important that *for optimal sets* $\Lambda \subset \mathfrak{F}$ *of "active" polynomial coefficients* we have available unisolvent polynomial interpolants. For arbitrary sets $\Lambda \subset \mathfrak{F}$, it is in general difficult (if not impossibe) to design unisolvent polynomial interpolation based on N points where N is equal to the cardinality of Λ.

One particular class of index sets $\Lambda \subset \mathfrak{F}$ for which this is possible are the so-called *monotone index sets*. This class of index sets was introduced in [6] in the context of adaptive Taylor approximations of parametric elliptic partial differential equations. we now strenghten the N-term approximation properties of the Taylor series by introducing the notion of *monotonicity* of index sets $\Lambda \subset \mathfrak{F}$.

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 5. 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 non empty set $\Lambda \subset \mathfrak{F}$ is called *monotone* if and only if $\nu \in \Lambda$ and $\mu \leq \nu \Rightarrow \mu \in \Lambda$. For a monotone set $\Lambda \subset \mathfrak{F}$, we define its margin $\mathcal{M} = \mathcal{M}(\Lambda)$ as follows:

$$\mathcal{M}(\Lambda) := \{\nu \notin \Lambda \ ; \ \exists j > 0 \ : \nu - e_j \in \Lambda\} \ , \tag{60}$$

where $e_j \in \mathfrak{F}$ is the Kronecker sequence: $(e_j)_i = \delta_{ij}$ for $i, j \in \mathbb{N}$.

Notice that the margin $\mathcal{M}(\Lambda)$ is an infinite set even when Λ is finite since there are infinitely many variables. In the finite dimensional setting $d < \infty$, the margin is a finite set. Any nonempty monotone set contains the null index $(0, 0, \cdots)$, which we will denote in what follows with slight abuse of notation by 0. Intersections and unions of monotone sets are also monotone. Also, note that $\Lambda \cup \mathcal{M}(\Lambda)$ is a monotone set.

Recall $|\nu| := \sum_{i \geq 1} \nu_i$ for $\nu \in \mathfrak{F}$. We say that ν is maximal in a set $\Lambda \subset \mathfrak{F}$ if and only if there exists no $\mu > \nu$ in Λ. If $\Lambda \subset \mathfrak{F}$ satisfies $N := N(\Lambda) := \max_{\nu \in \Lambda} |\nu| < \infty$, then any $\nu \in \Lambda$ for which $|\nu| = N$ is a maximal element. In particular, any finite set Λ has at least one maximal element. If Λ is monotone and if ν is maximal in Λ, then $\Lambda - \{\nu\}$ is monotone.

The *monotone majorant of a bounded sequence* $(a_\nu)_{\nu \in \mathfrak{F}}$ is the sequence

$$\mathbf{a}_\nu := \max_{\mu \geq \nu} |a_\mu|, \quad \nu \in \mathfrak{F} \,.$$

We define $\ell^p_m(\mathfrak{F})$ as the set of all sequences which have their monotone majorant in $\ell^p(\mathfrak{F})$. Clearly, $\ell^p_m(\mathfrak{F})$ is a linear space with respect to addition of sequences and scalar multiplication. We equip this space with the norm

$$\|(a_\nu)\|_{\ell^p_m(\mathfrak{F})} := \|(\mathbf{a}_\nu)\|_{\ell^p(\mathfrak{F})},$$

Now, if $(a_\nu)_{\nu \in \mathfrak{F}} \in \ell^p_m(\mathfrak{F})$, $0 < p < 1$, and Λ_k is any monotone realization of $\Lambda^*_k((\mathbf{a}_\nu)_{\nu \in \mathfrak{F}})$, then these sets Λ_k satisfy

$$\sum_{\nu \notin \Lambda_k} |a_\nu| \leq \sum_{\nu \notin \Lambda_k} \mathbf{a}_\nu \leq \|(a_\nu)\|_{\ell^p_m(\mathfrak{F})} (k+1)^{-s}, \quad s := \frac{1}{p} - 1. \tag{61}$$

Theorem 11. *Under the assumptions of Theorem 5, each of the sequences*

$$\left(\|T_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})}\right)_{\nu \in \mathfrak{F}}, \left(\|C_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})}\right)_{\nu \in \mathfrak{F}}, \left(\|X_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})}\right)_{\nu \in \mathfrak{F}}$$

belongs to $\ell^p_m(\mathfrak{F})$.

Proof: The result will follow from the coefficient estimates obtained in Proposition 3 that we used in the proof of Theorem 5, and on arguments in [6]. We denote by $\mathcal{A}_{b,\delta}$ the set of all (b, δ)-admissible sequences ρ for which $\rho_j \geq 1$, for all j. It was shown in Proposition 3 that for any $\delta > 0$

$$\|T_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})} \leq \inf_{\rho \in \mathcal{A}_{b,\delta}} B(\rho) \rho^{-\nu} \leq b_0 e^b \inf_{\rho \in \mathcal{A}_{b,\delta}} \rho^{-\nu} \,. \tag{62}$$

Now fix $\delta > 0$ arbitrary. Then we have

$$\|T_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})} \leq b_0 e^b \inf_{\rho \in \mathcal{A}_{b,\delta/2}} \rho^{-\nu} =: b_\nu \,. \tag{63}$$

In the proof of Theorem 5 it was shown that $(b_\nu)_{\nu \in \mathfrak{F}} \in \ell^p(\mathfrak{F})$. We now observe that the sequence b_ν defined in (63) is monotone because for any $\rho \in \mathcal{A}_{b,\frac{\delta}{2}}$

$$\mu \leq \nu \Rightarrow \rho^{-\nu} \leq \rho^{-\mu},$$

and thus

$$\mu \leq \nu \Rightarrow b_\nu \leq b_\mu.$$

Therefore, if $(\mathbf{a}_\nu)$ denotes the monotone majorant of the coefficient sequence $(\|T_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})})_{\nu\in\mathfrak{F}}$, we also find that

$$\mathbf{a}_\nu \leq b_\nu.$$

It follows that $\left\|\|T_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})}\right\|_{\ell^p_m(\mathfrak{F})} \leq \|(b_\nu)\|_{\ell^p(\mathfrak{F})} < \infty$. In view of Proposition 5 exactly the same arguments apply to the sequence X_ν of Legendre coefficients. To prove that

$$\left(\|C_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})}\right)_{\nu\in\mathfrak{F}} \in \ell^p_m(\mathfrak{F})$$

we shall make use of Proposition 6 and the special construction of $(b, \delta/2)$-admissible sequences in Sec. 4.6. We find

$$\begin{aligned}\|C_\nu\|_{C^1_{L(\overline{\rho},R)/\kappa}([0,T];\mathcal{S})} &\leq \inf_{\delta>0} \inf_{\rho\in\mathcal{A}_{b,\delta}} b_0 e^b 2^{|\operatorname{supp}\nu|}\rho^{-\nu} \\ &\leq b_0 e^b \inf_{\substack{\rho\in\mathcal{A}_{b,\delta/2}, \rho_j=\kappa,\, j\in E, \\ \rho_j\geq 2,\, j\in F}} 2^{|\operatorname{supp}\nu|}\rho^{-\nu} =: \widetilde{b}_\nu\end{aligned}$$

with κ, E and F as in Sec. 4.6. Unfortunately, the sequence $(\widetilde{b}_\nu)_{\nu\in\mathfrak{F}}$ itself is not monotone. However, it holds

$$\mu \leq \nu \Rightarrow \widetilde{b}_\nu \leq 2^{|E\cap\operatorname{supp}(\nu-\mu)|}\widetilde{b}_\mu \leq 2^M s_\mu\,,$$

since due to $\rho_j \geq 2$ for $j \in F$ we have

$$\prod_{j\in\operatorname{supp}\nu\cap F} 2\rho_j^{-\nu_j} \leq \prod_{j\in\operatorname{supp}\mu\cap F} 2\rho_j^{-\mu_j}\,.$$

Thus it follows

$$\|w_\nu\|_{\mathcal{S}} \leq \widetilde{b}_\nu \leq \sup_{\mu\geq\nu} \widetilde{b}_\nu \leq 2^M \widetilde{b}_\nu\,.$$

In other words, $(\widetilde{b}_\nu)_{\nu\in\mathfrak{F}}$ is quasi-monotone, and it belonging to $\ell^p(\mathfrak{F})$ was already proven in Theorem 10. We conclude $(\widetilde{b}_\nu)_{\nu\in\mathfrak{F}} \in \ell^p_m(\mathfrak{F})$, and hence we finally find $\left(\|w_\nu\|_{\mathcal{S}}\right)_{\nu\in\mathfrak{F}}$ to belong to $\ell^p_m(\mathfrak{F})$. □

5 Conclusions

We have presented a general theory of parametric initial value ODEs on high- and possibly infinite dimensional parameter and state spaces.

Under the uniform validity of a Lipschitz condition for the parametric vector fields and under the assumption of *affine parameter dependence* on the parameters, which typically occurs in applications from stochiometry and from Karhunen-Loeve type representations in PCAs, we showed existence, uniqueness of analytic contiuations of the parametric solutions to certain polydiscs in $\mathbb{C}^{\mathbb{N}}$, with precise bounds on their size. This, in turn, was used to establish a-priori bounds on the Taylor-, Legendre- and Chebysev coefficients in the corresponding expansions of the parametric solutions, and to establish p-summability of the corresponding coefficient sequences measured in the norm $\|\circ\|_{\mathcal{S}}$ of the corresponding state spaces $\mathcal{S}$. This p-summability allowed to deduce, via a classical argument due to Stechkin, the rate of convergence of nonlinear, best N-term approximations of partial sums of such sequences. The convergence rates of best N-term approximations of the parametric solutions were found to depend only on the summability of the parametric inputs and to be in particular independent of the dimension of the parameter space U. We showed, in particular, that these rates were attained even by best N-term approximations on index sets $\Lambda_N \subset \mathcal{F}$ which were *monotone*.

Spans of tensor product polynomials with degree combinations that belong to monotone sets are independent of the choice polynomial basis in the coordinate direction. This observation allows the design and optimization of sparse, adaptive collocation approximations of Smolyak type for monotone polynomial degree combinations and rather general choices of interpolation points in the coordinate directions.

References

1. Volker Barthelmann, Erich Novak, and Klaus Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12:273–288, 2000. 10.1023/A:1018977404843.
2. Marcel Bieri. A sparse composite collocation finite element method for elliptic sPDEs. *SIAM J. Numer. Anal.*, 49:2277–2301, 2011.
3. Marcel Bieri, Roman Andreev, and Christoph Schwab. Sparse tensor discretization of elliptic SPDEs. *SIAM J. Sci. Comput.*, 31(6):4281–4304, 2009/10.
4. Jean-Paul Calvi and Manh Phung Van. On the Lebesgue constant of Leja sequences for the unit disk and its applications to multivariate interpolation. *Journal of Approximation Theory*, 163(5):608 – 622, 2011.
5. C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral methods*. Scientific Computation. Springer, Berlin, 2007. Evolution to complex geometries and applications to fluid dynamics.
6. A. Chkifa, A. Cohen, R. DeVore, and Ch. Schwab. Adaptive algorithms for sparse polynomial approximation of parametric and stochastic elliptic pdes. Technical report,

Report 2011-44, Seminar for Applied Mathematics, ETH Zürich (to appear in M2AN 2012), 2011.

7. 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(6):615–646, 2010.
8. A. Cohen, R. A. DeVore, and Christoph Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic pdes. *Analysis and Applications*, 9(1):11–47, 2011.
9. Klaus Deimling. *Nonlinear Ordinary Differential Equations in Banach Spaces*, volume 596 of *Springer Lecture Notes in Mathematics*. Springer Verlag, New York, 1977.
10. C.J. Gittelson. Adaptive stochastic Galerkin methods: Beyond the elliptic case. Technical report, Report 2011-12, Seminar for Applied Mathematics, ETH Zürich, 2011.
11. E. Hairer, S. P. Nørsett, and G. Wanner. *Solving ordinary differential equations. I*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1987. Nonstiff problems.
12. E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1996. Stiff and differential-algebraic problems.
13. Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006. Structure-preserving algorithms for ordinary differential equations.
14. Markus Hansen, Claudia Schillings, and Christoph Schwab. Sparse approximation algorithms for high dimensional, parametric initial value problems. Technical Report 2012/XX, Seminar for Applied Mathematics, ETH Zürich, 2012.
15. Michel Hervé. *Analyticity in Infinite Dimensional Spaces*, volume 10 of *De Gruyter studies in mathematics*. Walter de Gruyter, 1989.
16. V. H. Hoang and Christoph Schwab. Analytic regularity and polynomial approximation of stochastic, parametric elliptic multiscale PDEs. Technical Report 2011/07, Seminar for Applied Mathematics, ETH Zürich, 2011.
17. Lars Hörmander. *An introduction to complex analysis in several variables*, volume 7 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, third edition, 1990.
18. F. Horn and R. Jackson. General mass action kinetics. *Arch. Rational Mech. Anal.*, 47:81–116, 1972.
19. A. Kunoth and Christoph Schwab. Analytic regularity and gpc approximation for control problems constrained by linear parametric elliptic and parabolic PDEs. Technical Report 2011/54, Seminar for Applied Mathematics, ETH Zürich, 2011.
20. 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(5):2411–2442, 2008.
21. Theodore Rivlin. *The Chebyshev Polynomials*. Ser. Pure & Applied Mathematics. Wiley Interscience, New York, 1974.
22. Wolfgang Walter. *Ordinary differential equations*, volume 182 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998. Translated from the sixth German (1996) edition by Russell Thompson, Readings in Mathematics.

Research Reports

No.	Authors/Title
11-64	*M. Hansen and Ch. Schwab* Sparse adaptive approximation of high dimensional parametric initial value problems
11-63	*R. Hiptmair, G. Phillips and G. Sinha* Multiple point evaluation on combined tensor product supports
11-62	*J. Li, M. Li and S. Mao* Convergence analysis of an adaptive finite element method for distributed flux reconstruction
11-61	*J. Li, M. Li and S. Mao* A priori error estimates of a finite element method for distributed flux reconstruction
11-60	*H. Heumann and R. Hiptmair* Refined convergence theory for semi-Lagrangian schemes for pure advection
11-59	*V.A. Hoang and Ch. Schwab* N-term Galerkin Wiener chaos approximations of elliptic PDEs with lognormal Gaussian random inputs
11-58	*X. Claeys and R. Hiptmair* Electromagnetic scattering at composite objects: A novel multi-trace boundary integral formulation
11-57	*S. Mishra and L.V. Spinolo* Accurate numerical schemes for approximating intitial-boundary value problems for systems of conservation law
11-56	*P. Grohs* Finite elements of arbitrary order and quasiinterpolation for data in Riemannian manifolds
11-55	*P. Grohs* Bandlimited shearlet-type frames with nice duals
11-54	*A. Kunoth and Ch. Schwab* Analytic regularity and GPC approximation for control problems constrained by linear parametric elliptic and parabolic PDEs
11-53	*Ch. Schwab and S.A. Tokareva* High order approximation of probabilistic shock profiles in hyperbolic conservation laws with uncertain initial data
11-52	*F.Y. Kuo, Ch. Schwab and I.H. Sloan* Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients