ON THE COMPLEXITY OF PARABOLIC INITIAL VALUE PROBLEMS WITH VARIABLE DRIFT

KNUT PETRAS AND KLAUS RITTER

Abstract. We study the intrinsic difficulty of solving linear parabolic initial value problems numerically at a single point. We present a worst case analysis for deterministic as well as for randomized (or Monte Carlo) algorithms, assuming that the drift coefficients and the potential vary in given function spaces. We use fundamental solutions (parametrix method) for equations with unbounded coefficients to relate the initial value problem to multivariate integration and weighted approximation problems. Hereby we derive lower and upper bounds for the minimal errors. The upper bounds are achieved by algorithms that use Smolyak formulas and, in the randomized case, variance reduction. We apply our general results to equations with coefficients from Hölder classes, and here, in many cases, the upper and lower bounds almost coincide and our algorithms are almost optimal.

1. Introduction

Consider a linear parabolic equation

$$\Delta u + \sum_{j=1}^{d} b_j \cdot \frac{\partial u}{\partial x_j} + b_0 \cdot u = \frac{\partial u}{\partial t} \tag{1}$$

on $]0, t_0] \times \mathbb{R}^d$ with initial condition

$$u(0, \cdot) = \varphi, \tag{2}$$

and suppose that the vector $\mathbf{b} = (b_0, \ldots, b_d)$ of coefficients belongs to a function class $\mathbf{B}$. The latter is typically defined by smoothness properties and growth conditions. We wish to compute the solution u of any initial value problem with $\mathbf{b} \in \mathbf{B}$ at a single point (t_0, x_0), and we study deterministic and randomized (or Monte Carlo) algorithms that use a finite number of functions values of the coefficient vector $\mathbf{b}$. Well-known algorithms of this form are, e.g., finite difference schemes for the parabolic equation or weak Ito-Taylor schemes for the associated stochastic differential equation, see (7) and (8).

We follow a worst case approach for the analysis and comparison of algorithms. For parabolic initial value problems the approach is used for the first time by Plaskota, Wasilkowski, and Woźniakowski (2000) and then by Kwas and Li (2003), and Kwas (2004). They consider (1) and (2) with $b_1 = \cdots = b_d = 0$ but variable b_0 and φ.

For simplicity we assume that φ is fixed. The intrinsic difficulty of our computational problem is quantified by the n-th minimal errors

$$\mathrm{e}^{\mathrm{det}}(n, \mathbf{B}) = \inf_{P_n \in \mathfrak{A}_n^{\mathrm{det}}} \mathrm{e}(P_n, \mathbf{B})$$

and

$$\mathrm{e}^{\mathrm{ran}}(n, \mathbf{B}) = \inf_{P_n \in \mathfrak{A}_n^{\mathrm{ran}}} \mathrm{e}(P_n, \mathbf{B}).$$

Date: January 2005.

Dagstuhl Seminar Proceedings 04401
Algorithms and Complexity for Continuous Problems
http://drops.dagstuhl.de/opus/volltexte/2005/149

Here $\mathfrak{A}_n^{\rm det}$ and $\mathfrak{A}_n^{\rm ran}$ denote the classes of all deterministic and randomized algorithms, respectively, that use n function values of the coefficient vector $\mathbf{b}$ at adaptively chosen nodes from $[0,t_0]\times\mathbb{R}^d$. Moreover, $\mathrm{e}(P_n,\mathbf{B})$ is the worst case (or maximal) error of an algorithm P_n for initial value problems with $\mathbf{b}\in\mathbf{B}$. We are interested in the asymptotic behaviour of the minimal errors as n tends to ∞, and we derive asymptotic upper and lower bounds. Furthermore, we wish to determine algorithms from $\mathfrak{A}_n^{\rm det}$ and $\mathfrak{A}_n^{\rm ran}$ with error close to the corresponding minimal error and with computational cost close to n.

Our analysis relies on the results by Deck and Kruse (2002), who construct fundamental solutions for parabolic equations with unbounded coefficients under mild regularity assumptions. It follows that $u(t_0,x_0)$ is given as a rapidly convergent series of weighted integrals of increasing dimension, where the integrands are tensor products of the coefficients b_j.

A general technique, which is due to Bakhvalov (1959) and Novak (1988), is available to establish lower bounds for integration problems. We extend this technique to derive the same lower bounds for the non-linear problem of solving parabolic initial value problems with variable coefficients.

To provide upper bounds for the minimal errors and to construct corresponding algorithms we truncate the series representation for $u(t_0,x_0)$ and approximate the remaining tensor products of coefficients. Furthermore, our randomized algorithm use the deterministic one for variance reduction. Hereby we show that solving the initial value problem is almost as easy as L_∞-approximation of the coefficients b_j with respect to the weight function $(t,x)\mapsto\bigl(1+|x|^\delta\bigr)^{-1}$ for any $0\le\delta<1$. In fact, one may use Smolyak formulas to approximate tensor products of coefficients, so that the latter problems altogether are almost as easy as approximation of a single coefficient. The computational cost of the resulting algorithms is almost proportional to the number n of evaluations of $\mathbf{b}$. We rely on results by Wasilkowski and Woźniakowski (1995), who provide a general analysis of Smolyak formulas with emphasis on bounds that explicitly depend on the dimension.

We apply our general results to parabolic equations with coefficients from Hölder classes $C_M^{r,\alpha}$, where r is the order of differentiability and α and M denote the Hölder exponent and constant, respectively. To derive the upper bounds we also use results (partially with slight modification) on weighted approximation by Wasilkowski and Woźniakowski (2001). Put

$$\gamma=\frac{r+\alpha}{d+1},$$

and consider any $\varepsilon>0$ in the sequel.

For

$$\mathbf{B}=\bigl\{b\in C_M^{r,\alpha}:|b(0,0)|\le M'\bigr\}^{d+1}$$

with $r=0$ and $\alpha<1$ we get the lower bounds

$$\liminf_{n\to\infty}\bigl(n^{\gamma}\cdot\mathrm{e}^{\rm det}(n,\mathbf{B})\bigr)>0,\qquad \liminf_{n\to\infty}\bigl(n^{\gamma+1/2}\cdot\mathrm{e}^{\rm ran}(n,\mathbf{B})\bigr)>0 \tag{3}$$

and the upper bounds

$$\lim_{n\to\infty}\bigl(n^{\gamma-\varepsilon}\cdot\mathrm{e}^{\rm det}(n,\mathbf{B})\bigr)=0,\qquad \lim_{n\to\infty}\bigl(n^{\gamma+1/2-\varepsilon}\cdot\mathrm{e}^{\rm ran}(n,\mathbf{B})\bigr)=0. \tag{4}$$

According to Deck and Kruse (2002) we additionally have to impose a growth condition with exponent $0\le\beta<1$ for coefficients of smoothness $r+\alpha\ge1$. Hence we study the

classes

$$\mathbf{B} = \bigl\{ b \in C_M^{r,\alpha} : \sup_{(t,x)\in[0,t_0]\times\mathbb{R}^d} |b(t,x)|/(1+|x|^\beta) \le M' \bigr\}^{d+1}. \tag{5}$$

In this case (3) is valid, too, but instead of (4) we get

$$\lim_{n\to\infty} \bigl(n^{\vartheta-\varepsilon} \cdot \mathrm{e}^{\mathrm{det}}(n,\mathbf{B}) \bigr) = 0, \qquad \lim_{n\to\infty} \bigl(n^{\vartheta+1/2-\varepsilon} \cdot \mathrm{e}^{\mathrm{ran}}(n,\mathbf{B}) \bigr) = 0 \tag{6}$$

with

$$\vartheta = \gamma \cdot \frac{1-\beta}{r+\alpha-\beta}.$$

However, on subclasses of functions whose local Hölder constants decay sufficiently fast as $|x|$ tends to ∞ we get the bounds (3) and (4) again. This holds true in particular for compactly supported functions from $C_M^{r,\alpha}$.

We conclude that for coefficients with low regularity or for smooth coefficients with Hölder constants that decay sufficiently fast the minimal errors converge to zero almost like $n^{-\gamma}$ for deterministic algorithms and $n^{-(\gamma+1/2)}$ for randomized algorithms. Moreover, our algorithms are almost optimal in these cases. On the other hand, we do not have sharp bounds in the case (5) if $r+\alpha \ge 1$. In fact, the upper bounds (6) and lower bounds (3) differ significantly for large values of γ, i.e., if the smoothness $r+\alpha$ is large compared to the spatial dimension d. Note that in all these cases randomized algorithms are significantly better than deterministic ones if γ is small.

Associated with every linear parabolic equation there is a stochastic differential equation

$$dX_t = \mu(t,X_t)\,dt + \sigma(t,X_t)\,dW_t \tag{7}$$

for $t \in [0,t_0]$. The principal part of the elliptic operator is related to the diffusion coefficient σ, and in particular for (1) we have $\sigma = \sqrt{2}\cdot \mathrm{Id}_d$. Furthermore,

$$\mu(t,x) = (b_1(t_0-t,x),\ldots,b_d(t_0-t,x))^\top,$$

and W denotes a d-dimensional standard Brownian motion. The Feynman-Kac representation of u yields

$$u(t_0,x_0) = E\left(\varphi(X_{t_0}) \cdot \exp\left(\int_0^{t_0} b_0(s,X_s)\,ds \right) \right) \tag{8}$$

for parabolic equations in general, where X is the solution of (7) with initial condition

$$X_0 = x_0.$$

Consider the particular case

$$\Delta u + b_0 \cdot u = \frac{\partial u}{\partial t} \tag{9}$$

of equation (1), which corresponds to a stochastic differential equation with solution $X_t = x_0 + \sqrt{2}\,W_t$. The latter is a Gaussian process, and a Taylor expansion of the exponential function in (8) shows that $u(t_0,x_0)$ may be expressed as a series of weighted integrals. The integrands are the tensor products $\varphi \otimes b_0 \otimes \cdots \otimes b_0$, and the weight functions correspond to mixtures of multivariate normal distributions, since X is Gaussian. If $b_0 = 0$, too, then $u(t_0,x_0)$ is simply given as the integral of φ with respect to a d-dimensional normal distribution.

This is the starting point of the analysis by Plaskota, Wasilkowski, and Woźniakowski (2000), Kwas and Li (2003), and Kwas (2004). More precisely, Plaskota, Wasilkowski, and

Woźniakowski (2000) and Kwas and Li (2003) study deterministic algorithms in the cases $d = 1$ and $d \geq 1$, respectively, and Kwas (2004) analyzes randomized (as well as quantum) algorithms in the case $d \geq 1$.

The stochastic differential equation (7) corresponding to (1) can be solved explicitly only in exceptional cases and leads to a non-Gaussian process X, in general. Still, by the results from Deck and Kruse (2002), we can proceed in a similar way as Plaskota, Wasilkowski, and Woźniakowski (2000), Kwas and Li (2003), and Kwas (2004).

Implementation of the almost optimal algorithms that are constructed in the present paper, as well as in those references cited previously, requires extensive pre-computing. In fact, a straight-forward approach leads to more than n quadrature problems, which do not depend on the coefficients b_j and must be solved before implementation.

The Feynman-Kac representation is often used to derive randomized algorithms of the following form. A simulation yields samples $X^{(1)}(\omega), \ldots, X^{(m)}(\omega)$ that are approximately distributed like X, and a sample mean corresponding to the right-hand side of (8) is used as an approximation to $u(t_0, x_0)$. The simplest algorithm of this form uses the weak Euler scheme for the simulation of (7). In computational practice randomized algorithms are often preferred to deterministic ones, unless the dimension d is small. Large values of d naturally arise, e.g., in computational finance, when (7) is used to model the risk-neutral dynamics of the prices of d assets and φ denotes the discounted payoff of a European option with maturity t_0. In this case $u(t_0, x_0)$ is the value of the option at time $t = 0$.

We present a simple consequence of our lower bound for the minimal error $\mathrm{e}^{\mathrm{det}}(n, \mathbf{B})$ with Hölder classes $\mathbf{B}$. Consider the randomized algorithm that is based on the weak Euler scheme for simulation of (7) with constant step-size in time. It is well known that its bias is (at most) proportional to the step-size, if the coefficients μ and σ as well as φ and b_0 satisfy moderate smoothness assumptions and growth bounds, see, e.g., Kloeden and Platen (1999). Relating the step-size and the number of simulations in an optimal way, we get a randomized algorithm $P_n^{\mathrm{E}} \in \mathfrak{A}_n^{\mathrm{ran}}$ with error

$$\mathrm{e}(P_n^{\mathrm{E}}, \mathbf{B}) \leq c \cdot n^{-1/3}$$

for some constant $c > 0$ and with computational cost proportional to n. For $\sigma = \sqrt{2} \cdot \mathrm{Id}_d$ and $b_0 = 0$ this holds true at least if $r + \alpha > 2$. On the other hand, we have the lower bound

$$\mathrm{e}^{\mathrm{det}}(n, \mathbf{B}) \geq c' \cdot n^{-(r+\alpha)/(d+1)}$$

with some constant $c' > 0$, see (3). We thus conclude that asymptotically the simple and easily implementable algorithm P_n^{E} is better than every deterministic algorithm of the same computational cost, if

$$d > 3\,(r + \alpha) - 1.$$

For instance, if $r + \alpha$ close to 2, this superiority already holds for $d \geq 6$.

2. Problem Formulation

In the sequel we let c denote unspecified positive constants with possibly different values and we put $\mathbf{b} = (b_0, \ldots, b_d)$ as well as $D = [0, t_0] \times \mathbb{R}^d$. Moreover, we use $|\cdot|$ to denote the Euclidean norm. We always assume that

(i) every function $b_j : D \to \mathbb{R}$ is continuous and satisfies a Hölder condition

$$|b_j(t, x) - b_j(t, y)| \leq c \cdot |x - y|^{\alpha}, \qquad t \in [0, t_0],\ x, y \in K,$$

with exponent $0 < \alpha \le 1$ for every compact set $K \subseteq \mathbb{R}^d$ and a growth condition

$$\sup_{(t,x)\in D} \left(|b_j(t,x)|/(1+|x|^\beta)\right) < \infty$$

with exponent $0 \le \beta < 1$,

(ii) the function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is continuous and satisfies a growth condition

$$\sup_{x\in\mathbb{R}^d} \left(|\varphi(x)| \cdot \exp(-h \cdot |x|^2)\right) < \infty$$

for every $h > 0$.

Then the initial value problem (1) and (2) is uniquely solvable in the class of continuous functions u that satisfy

$$\sup_{(t,x)\in D} \left(|u(t,x)| \cdot \exp(-h \cdot |x|^2)\right) < \infty \tag{10}$$

for any $h > 0$, see Deck and Kruse (2002). In order to indicate the dependence of u on $\mathbf{b}$ we write $u(t,x;\mathbf{b})$.

Now we formulate the computational problem that is studied in this paper. Let $t_0 > 0$, $x_0 \in \mathbb{R}^d$, and φ with property (ii) be given. Consider a class $\mathbf{B}$ of functions $\mathbf{b} : D \to \mathbb{R}^{d+1}$ that satisfy property (i). Formally, the problem of solving the initial value problem at the point (t_0, x_0) for $\mathbf{b} \in \mathbf{B}$ is defined by the non-linear mapping $P : \mathbf{B} \to \mathbb{R}$ with

$$P[\mathbf{b}] = u(t_0, x_0; \mathbf{b}).$$

We study algorithms for the approximate computation of P that use a finite number of function values of the coefficients b_j. For simplicity we assume that a single evaluation at a node $(t,x) \in D$ already yields the values of all coefficients $b_0, \dots, b_d$ at this node. By $\mathfrak{A}_n^{\mathrm{det}}$ and $\mathfrak{A}_n^{\mathrm{ran}}$ we denote the classes of all deterministic and randomized (or Monte Carlo) algorithms, respectively, that use n evaluations for every $\mathbf{b} \in \mathbf{B}$ at adaptively chosen nodes from D. For deterministic algorithms $P_n : \mathbf{B} \to \mathbb{R}$ the worst case error on the class $\mathbf{B}$ is defined by

$$\mathrm{e}(P_n, \mathbf{B}) = \sup_{\mathbf{b}\in\mathbf{B}} |P[\mathbf{b}] - P_n[\mathbf{b}]|,$$

and for randomized algorithms P_n, which formally are random variables with values in $\mathfrak{A}_n^{\mathrm{det}}$, this quantity is defined by

$$\mathrm{e}(P_n, \mathbf{B}) = \sup_{\mathbf{b}\in\mathbf{B}} \left(E\left(\left|P[\mathbf{b}] - P_n[\mathbf{b}]\right|^2\right)\right)^{1/2}.$$

The n-th minimal errors on $\mathbf{B}$,

$$\mathrm{e}^{\mathrm{det}}(n, \mathbf{B}) = \inf_{P_n\in\mathfrak{A}_n^{\mathrm{det}}} \mathrm{e}(P_n, \mathbf{B})$$

and

$$\mathrm{e}^{\mathrm{ran}}(n, \mathbf{B}) = \inf_{P_n\in\mathfrak{A}_n^{\mathrm{ran}}} \mathrm{e}(P_n, \mathbf{B}),$$

quantify how well initial value problems with coefficients $\mathbf{b} \in \mathbf{B}$ can be solved by any deterministic or randomized algorithm that uses n values of the coefficient vector.

The number n of evaluations of $\mathbf{b}$ is a rather rough measure for the computational cost of P_n. In a detailed definition of the cost for computing $P_n[\mathbf{b}]$ we add the (expected) number of arithmetic operations (and the expected number of calls of the random number generator) to the number n. By $\mathrm{cost}(P_n, \mathbf{B})$ we then denote the worst case cost for computing $P_n[\mathbf{b}]$ on the class $\mathbf{B}$. We refer to Traub, Wasilkowski, and Woźniakowski (1988)

and Novak (1995) for the precise definition and analysis of the underlying real number model of computation.

We are interested in algorithms P_n from $\mathfrak{A}_n^{\mathrm{det}}$ or $\mathfrak{A}_n^{\mathrm{ran}}$ with error close to $\mathrm{e}^{\mathrm{det}}(n, \mathbf{B})$ or $\mathrm{e}^{\mathrm{ran}}(n, \mathbf{B})$, respectively, and with $\mathrm{cost}(P_n, \mathbf{B})$ close to n.

3. A Fundamental Solution and Basic Estimates

Deck and Kruse (2002) construct a fundamental solution for parabolic equations with unbounded coefficients under mild regularity assumptions by means of the classical parametrix method, see, e.g., Friedman (1964). Deck and Kruse consider the general case of a uniformly elliptic operator. Here we formulate and use their result in the particular case of equation (1).

Let $0 \le \tau < \tau_\mu < \cdots < \tau_1 < t \le t_0$ and $\xi, \xi_\mu, \ldots, \xi_1, x \in \mathbb{R}^d$, and put

$$Z(t, x, \tau, \xi) = \left(2\sqrt{\pi}\right)^{-d} \cdot (t-\tau)^{-d/2} \cdot \exp\left(-\frac{|x-\xi|^2}{4\,(t-\tau)}\right)$$

as well as

$$V(\tau, t) =]\tau, t[\times \mathbb{R}^d.$$

By Z_j we denote the partial derivative of the heat kernel Z with respect to the j-th component of its second argument. Moreover, we put $Z_0 = Z$. For $\mathbf{j} \in \{0, \ldots, d\}^\mu$ we define

$$\begin{aligned}&\psi^{(\mathbf{j})}(t, x, \tau_1, \xi_1, \ldots, \tau_\mu, \xi_\mu, \tau, \xi)\\ &\quad = Z(t, x, \tau_1, \xi_1) \cdot Z_{j_1}(\tau_1, \xi_1, \tau_2, \xi_2) \cdots Z_{j_{\mu-1}}(\tau_{\mu-1}, \xi_{\mu-1}, \tau_\mu, \xi_\mu) \cdot Z_{j_\mu}(\tau_\mu, \xi_\mu, \tau, \xi).\end{aligned}$$

Lemma 1 (Deck and Kruse (2002))**.** *Let*

$$0 \le \delta < 1.$$

There exists a constant $c > 0$ such that

$$f^{(\mathbf{j})}(t, x, \tau_1, \xi_1, \ldots, \tau_\mu, \xi_\mu, \tau, \xi) = \prod_{\ell=1}^{\mu} \left(1 + |\xi_\ell|^\delta\right) \cdot \left|\psi^{(\mathbf{j})}(t, x, \tau_1, \xi_1, \ldots, \tau_\mu, \xi_\mu, \tau, \xi)\right|$$

with $\mu \ge 1$ and $\mathbf{j} \in \{0, \ldots, d\}^\mu$ satisfies

$$\begin{aligned}&\int_{V(\tau,t)} \int_{V(\tau,\tau_1)} \cdots \int_{V(\tau,\tau_{\mu-1})} f^{(\mathbf{j})}(t, x, \tau_1, \xi_1, \ldots, \tau_\mu, \xi_\mu, \tau, \xi)\, d(\tau_\mu, \xi_\mu) \ldots d(\tau_1, \xi_1)\\ &\qquad \le (c\mu)^{-\mu/2} \cdot (t-\tau)^{(\mu-d-1)/2} \cdot \left(|\xi|^{\delta\cdot(\mu+1)} + \mu^{(1+\delta)/4\cdot\mu}\right) \cdot \exp\left(-\frac{c\,|x-\xi|^2}{t-\tau}\right).\end{aligned}$$

Proof. We have

$$|Z_j(t, x, \tau, \xi)| \le c_1 \cdot (t-\tau)^{-(d+1)/2} \cdot \exp\left(-\frac{|x-\xi|^2}{8\,(t-\tau)}\right)$$

for $j \ge 0$ with some constant $c_1 > 0$, cf. Deck and Kruse (2002, Eqn. (5.2)). For $0 < \delta < 1$ we now proceed as in Deck and Kruse (2002, pp. 77–79) with the particular choice $\varepsilon_\mu = 1/2 \cdot \mu^{-(1+\delta)/(2\delta)}$ (in our notation). □

We introduce the weighted L_∞-norm

$$\|g\|_{\mu,\delta} = \sup_{(\tau_\ell, \xi_\ell) \in D} \frac{|g(\tau_1, \xi_1, \ldots, \tau_\mu, \xi_\mu)|}{\prod_{\ell=1}^{\mu} \left(1 + |\xi_\ell|^\delta\right)}$$

for appropriately bounded functions $g : D^\mu \to \mathbb{R}$ and $0 \le \delta < 1$. Due to Lemma 1

$$\Gamma^{(\mathbf{j})}[g](t,x,\tau,\xi) = \int_{V(\tau,t)} \int_{V(\tau,\tau_1)} \dots \int_{V(\tau,\tau_{\mu-1})} g(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu) \cdot \psi^{(\mathbf{j})}(t,x,\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu,\tau,\xi)\, d(\tau_\mu,\xi_\mu)\dots d(\tau_1,\xi_1)$$

is well defined for $\mathbf{j} \in \{0,\dots,d\}^\mu$ and every continuous function g with $\|g\|_{\mu,\delta} < \infty$. In view of assumption (i) this applies in particular to $g = \mathbf{b}^{(\mathbf{j})}$ and $\beta \le \delta < 1$, where

$$\mathbf{b}^{(\mathbf{j})} = b_{j_1} \otimes \dots \otimes b_{j_\mu}$$

denotes the tensor product of the functions $b_{j_1},\dots,b_{j_\mu}$.

Theorem 1 (Deck and Kruse (2002))**.** *A fundamental solution for the parabolic equation* (1) *is given by*

$$\Gamma[\mathbf{b}] = Z + \sum_{\mu=1}^{\infty} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} \Gamma^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}].$$

Moreover,

$$u(t,x;\mathbf{b}) = \int_{\mathbb{R}^d} \Gamma[\mathbf{b}](t,x,0,\xi)\cdot\varphi(\xi)\,d\xi$$

holds for the continuous solution of (1) *and* (2) *that satisfies* (10)*.*

Theorem 1 yields

$$P[\mathbf{b}] = I^{(0)} + \sum_{\mu=1}^{\infty} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}] \tag{11}$$

with

$$I^{(0)} = \int_{\mathbb{R}^d} Z(t_0,x_0,0,\xi)\cdot\varphi(\xi)\,d\xi$$

and

$$I^{(\mathbf{j})}[g] = \int_{\mathbb{R}^d} \Gamma^{(\mathbf{j})}[g](t_0,x_0,0,\xi)\cdot\varphi(\xi)\,d\xi.$$

The solution $P[\mathbf{b}]$ of the initial value problem at the point (t_0,x_0) is therefore given as a series of weighted integrals with integrands being tensor products of the coefficients b_j (and the initial data φ).

We define

$$\rho^{(\mathbf{j})}(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu) = \int_{\mathbb{R}^d} \psi^{(\mathbf{j})}(t_0,x_0,\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu,0,\xi)\cdot\varphi(\xi)\,d\xi$$

if $0 < \tau_\mu < \dots < \tau_1 < t_0$ and $\rho^{(\mathbf{j})}(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu) = 0$ otherwise. Furthermore, we put

$$\lambda_\delta^{(\mathbf{j})} = \int_{\left(]0,t_0[\times\mathbb{R}^d\right)^\mu} \prod_{\ell=1}^{\mu} \left(1+|\xi_\ell|^\delta\right)\cdot\left|\rho^{(\mathbf{j})}(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu)\right| d(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu).$$

Obviously

$$I^{(\mathbf{j})}[g] = \int_{\left(]0,t_0[\times\mathbb{R}^d\right)^\mu} g(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu)\cdot\rho^{(\mathbf{j})}(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu)\,d(\tau_1,\xi_1,\dots,\tau_\mu,\xi_\mu), \tag{12}$$

and $\lambda_\delta^{(\mathbf{j})}$ is the norm of $I^{(\mathbf{j})}$ with respect to $\|\cdot\|_{\mu,\delta}$.

Lemma 2. *Let $0 \le \delta < 1$ and $\delta^* = (1-\delta)/4$. There exists a constant $c > 0$ such that*

$$\lambda_\delta^{(\mathbf{j})} \le c^\mu \cdot \mu^{-\delta^*\mu}$$

for every $\mu \ge 1$ and $\mathbf{j} \in \{0,\dots,d\}^\mu$. Moreover,

$$\sup_{\mu\in\mathbb{N}} \sup_{\mathbf{j}\in\{0,\dots,d\}^\mu} \lambda_\delta^{(\mathbf{j})} < \infty.$$

Proof. Let $c_1 > 0$. Due to Lemma 1 it suffices to verify that

$$\int_{\mathbb{R}^d} \exp(-c_1|x_0-\xi|^2)\cdot|\varphi(\xi)|\,d\xi < \infty \tag{13}$$

and that there exists a constant $c_2 > 0$ with

$$\int_{\mathbb{R}^d} |\xi|^{\delta\cdot(\mu+1)}\cdot\exp(-c_1|x_0-\xi|^2)\cdot|\varphi(\xi)|\,d\xi \le (c_2\cdot\mu)^{\delta/2\cdot\mu} \tag{14}$$

for every $\mu \ge 1$.

Because of assumption (ii) there exists a constant $c_3 > 0$ such that

$$\exp(-c_1|x_0-\xi|^2)\cdot|\varphi(\xi)| \le 1/c_3\cdot\exp(-c_3|\xi|^2).$$

Hence we have (13). For the proof of (14) we may assume $\delta\cdot(\mu+1) \ge 2$ without loss of generality. According to Ledoux and Talagrand (1991, Cor. 3.2) there exists a constant $c_4 > 0$ with

$$\int_{\mathbb{R}^d} |\xi|^p \cdot \exp\left(-|\xi|^2/2\right)\,d\xi \le (c_4\cdot p)^{p/2}$$

for every $p \ge 2$. Take $p = \delta\cdot(\mu+1)$ to complete the proof of (14). □

Put

$$\|\mathbf{b}\|_\delta = \max_{j=0,\dots,d} \|b_j\|_{1,\delta}.$$

Lemma 3. *For $\beta \le \delta < 1$ let c denote the constant from Lemma 2. Then*

$$\sum_{\mu=m+1}^{\infty} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} \left|I^{(\mathbf{j})}[b^{(\mathbf{j})}]\right| \le \left((d+1)\cdot\|\mathbf{b}\|_\delta\cdot c\right)^{m+1}\cdot m^{-\delta^* m}\cdot\left(1-(d+1)\cdot\|\mathbf{b}\|_\delta\cdot c\cdot m^{-\delta^*}\right)^{-1}$$

if $m > ((d+1)\cdot\|\mathbf{b}\|_\delta\cdot c)^{1/\delta^*}$.

Proof. By Lemma 2

$$\left|I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}]\right| \le \|\mathbf{b}^{(\mathbf{j})}\|_{\mu,\delta}\cdot c^\mu\cdot\mu^{-\delta^*\mu} \le \|\mathbf{b}\|_\delta^\mu\cdot c^\mu\cdot\mu^{-\delta^*\mu}$$

for $\mathbf{j}\in\{0,\dots,d\}^\mu$. Therefore

$$\sum_{\mu=m+1}^{\infty} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} \left|I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}]\right| \le \left((d+1)\cdot\|\mathbf{b}\|_\delta\cdot c\right)^{m+1}\cdot m^{-\delta^* m}\cdot\sum_{\mu=0}^{\infty}\left((d+1)\cdot\|\mathbf{b}\|_\delta\cdot c\cdot m^{-\delta^*}\right)^\mu,$$

and the statement follows. □

Now we truncate the series (11) at $\mu = m_n$, where m_n is chosen such that the truncation error converges to zero faster than any polynomial in $1/n$.

Lemma 4. *Let*

$$m_n = \left\lceil \ln n / \sqrt{\ln \ln n} \right\rceil$$

for $n \geq 3$. Then, for every $s > 0$, $0 \leq \delta < 1$, and $K > 0$,

$$\lim_{n\to\infty} \left(n^s \cdot \sup_{\|\mathbf{b}\|_\delta \leq K} \sum_{\mu=m_n+1}^{\infty} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} \left| I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}] \right| \right) = 0.$$

Proof. Let

$$d_n = (c/m_n)^{\delta^* m_n} \cdot n^s$$

for any constant $c > 1$ and $0 < \delta^* \leq 1/4$. According to Lemma 3 it suffices to show that d_n tends to zero. For sufficiently large n we have

$$d_n \leq \frac{c^{\ln n} \cdot n^s}{m_n^{\delta^* m_n}} = \frac{n^{\ln c + s}}{m_n^{\delta^* m_n}}.$$

Furthermore,

$$m_n^{\delta^* m_n} \geq \left(\frac{\ln n}{\sqrt{\ln \ln n}} \right)^{\frac{\delta^* \cdot \ln n}{\sqrt{\ln \ln n}}} = n^{\delta^* \cdot p_n}$$

with

$$p_n = \sqrt{\ln \ln n} - \frac{\ln \ln \ln n}{2\sqrt{\ln \ln n}}.$$

Note that $n^{\delta^* \cdot p_n}$ increases faster than any power of n, since $\lim_{n\to\infty} p_n = \infty$. □

4. Lower Bounds

We show that, under rather general assumptions on the class $\mathbf{B}$, solving the initial value problem (1) and (2) is not easier than computing an integral over a $(d+1)$-dimensional rectangle. For the latter problem lower bounds for the minimal errors are known in many cases.

The lower bounds for parabolic equations already hold if all but one of the coefficients $b_0, \dots, b_d$ vanish, i.e., if

$$\Delta u + b_j \cdot \frac{\partial u}{\partial x_j} = \frac{\partial u}{\partial t}$$

or

$$\Delta u + b_0 \cdot u = \frac{\partial u}{\partial t}.$$

Since all theses cases can be analyzed in the same way, and since the lower bounds coincide, we only present the details for the first equation with $j = 1$. Here we have $\mathbf{b} = (0, b_1, 0, \dots, 0)$, and we put

$$\rho^{(\mu)} = \rho^{(\mathbf{j})}, \qquad I^{(\mu)} = I^{(\mathbf{j})} \tag{15}$$

for

$$\mathbf{j} = (1, \dots, 1) \in \{0, \dots, d\}^\mu.$$

Hence (11) reads

$$P[(0, b_1, 0, \dots, 0)] = I^{(0)} + \sum_{\mu=1}^{\infty} I^{(\mu)}[b_1 \otimes \cdots \otimes b_1]. \tag{16}$$

Furthermore, by (12),

$$I^{(1)}[b_1] = \int_{]0,t_0[\times\mathbb{R}^d} b_1(\tau_1,\xi_1)\cdot\rho^{(1)}(\tau_1,\xi_1)\,d(\tau_1,\xi_1)$$

with

$$\rho^{(1)}(\tau_1,\xi_1) = Z(t_0,x_0,\tau_1,\xi_1)\cdot\int_{\mathbb{R}^d} Z_1(\tau_1,\xi_1,0,\xi)\cdot\varphi(\xi)\,d\xi,$$

which is a continuous function on $]0,t_0[\,\times\mathbb{R}^d$.

4.1. **Assumptions and Preliminaries.** We assume that φ actually depends on the first component of its argument, i.e., that there exist $v,v'\in\mathbb{R}$ and $w\in\mathbb{R}^{d-1}$ such that $\varphi(v,w)\neq\varphi(v',w)$. Then the set of points (τ_1,ξ_1) with $\rho^{(1)}(\tau_1,\xi_1)\neq 0$ is dense in $]0,t_0[\times\mathbb{R}^d$.

In the sequel we consider any compact rectangle $D_0\subseteq\,]0,t_0[\,\times\mathbb{R}^d$ with non-empty interior such that

$$\inf_{(\tau_1,\xi_1)\in D_0}|\rho^{(1)}(\tau_1,\xi_1)|\geq K$$

for some constant $K>0$. Furthermore, we put

$$J[f] = \int_{D_0} f(\tau_1,\xi_1)\,d(\tau_1,\xi_1)$$

for $f\in C(D)$, and we let B_1 denote the class of all functions $b_1: D\to\mathbb{R}$ such that $(0,b_1,0,\ldots,0)\in\mathbf{B}$.

For integration on the class B_1 minimal errors are defined in the same way as for the initial value problem, i.e.,

$$\mathrm{e}^{\mathrm{det}}_{\mathrm{Int}}(n,B_1) = \inf_{J_n\in\mathfrak{A}^{\mathrm{det}}_n}\sup_{b_1\in B_1}\bigl|J[b_1]-J_n[b_1]\bigr|$$

and

$$\mathrm{e}^{\mathrm{ran}}_{\mathrm{Int}}(n,B_1) = \inf_{J_n\in\mathfrak{A}^{\mathrm{ran}}_n}\sup_{b_1\in B_1}\left(E\left(\bigl|J[b_1]-J_n[b_1]\bigr|^2\right)\right)^{1/2}.$$

We assume that there are sequences $(a_n)_{n\in\mathbb{N}}$ and $(e_n)_{n\in\mathbb{N}}$ of positive real numbers such that for every $n\in\mathbb{N}$ there exist functions $f_1,\ldots,f_{2n}: D\to\mathbb{R}$ with the following properties:

(a) $f_i\geq 0$ for $i=1,\ldots,2n$, and $\operatorname{supp} f_1,\ldots,\operatorname{supp} f_{2n}$ are pairwise disjoint and contained in D_0,

(b) for $\sigma_1,\ldots,\sigma_{2n}\in\{-1,1\}$

$$\sum_{i=1}^{2n}\sigma_i\cdot f_i\in B_1,$$

(c) for $i=1,\ldots,2n$

$$J[f_i]\geq e_n,$$

(d) for $i=1,\ldots,2n$

$$\|f_i\|^2_\infty\leq a_n.$$

Bakhvalov and Novak have used the assumptions (a)–(c) to establish lower bounds for minimal errors for (weighted) integration. We cite here their general results.

Theorem 2 (Bakhvalov (1959), Novak (1988, Sec. 2.2.4)). *Properties (a)–(c) imply*

$$\mathrm{e}_{\mathrm{Int}}^{\mathrm{det}}(n, B_1) \geq \mathrm{e}_{\mathrm{Int}}^{\mathrm{det}}\left(n, F^{(n)}\right) \geq n \cdot e_n$$

for deterministic algorithms and

$$\mathrm{e}_{\mathrm{Int}}^{\mathrm{ran}}(n, B_1) \geq \mathrm{e}_{\mathrm{Int}}^{\mathrm{ran}}\left(n, F^{(n)}\right) \geq \frac{1}{2} \cdot n^{1/2} \cdot e_n$$

for randomized algorithms. Here

$$F^{(n)} = \left\{ \sum_{i=1}^{2n} \sigma_i \cdot f_i : \sigma_1, \ldots, \sigma_{2n} \in \{-1, 1\} \right\}.$$

The same lower bounds, with an additional factor K, hold for $I^{(1)}$ instead of J.

To adopt this technique to our non-linear problem we additionally require that the unfavorable functions $f \in F^{(n)}$ are not too large, see (18) and (19) and note that

$$\sup_{f \in F^{(n)}} \|f\|_\infty^2 \leq a_n \tag{17}$$

due to (a) and (d).

Example 1. Consider a Hölder class

$$B_1 = C_M^{r,\alpha}$$

with $r \in \mathbb{N}_0$, $0 < \alpha \leq 1$, and $M > 0$. By definition, $f \in C_M^{r,\alpha}$ if and only if f has continuous partial derivatives up to order r and every r-th order partial derivative g of f satisfies

$$|g(t,x) - g(s,y)| \leq M \cdot |(t,x) - (s,y)|^\alpha, \qquad (t,x), (s,y) \in D.$$

Let $D_0 \subseteq]0, t_0[\times \mathbb{R}^d$ denote any compact rectangle with non-empty interior. Fix a function $f \in C_M^{r,\alpha}$ such that $0 \leq f \leq 1$, $\varepsilon = J[f] > 0$, and $\operatorname{supp} f$ is contained in the interior of D_0. Assume, without loss of generality, that $2n = \ell^{d+1}$ with $\ell \in \mathbb{N}$. Now rescale f as follows, $\widetilde{f}(t,x) = \ell^{-(r+\alpha)} f(\ell\, t, \ell\, x)$, and shift $\widetilde{f}$ to obtain functions $f_1, \ldots, f_{2n} \in C_M^{r,\alpha}$ with properties (a), (b), and

$$J[f_i] = \ell^{-(r+\alpha+d+1)} \cdot \varepsilon$$

as well as

$$\|f_i\|_\infty^2 \leq \ell^{-2(r+\alpha)}.$$

We may therefore take

$$e_n = (2n)^{-(r+\alpha)/(d+1)-1} \cdot \varepsilon$$

and

$$a_n = (2n)^{-2(r+\alpha)/(d+1)}$$

to satisfy properties (c) and (d).

For the integration problem we thus conclude that

$$\liminf_{n\to\infty} \frac{\mathrm{e}_{\mathrm{Int}}^{\mathrm{det}}(n, C_M^{r,\alpha})}{n^{-(r+\alpha)/(d+1)}} > 0$$

and

$$\liminf_{n\to\infty} \frac{\mathrm{e}_{\mathrm{Int}}^{\mathrm{ran}}(n, C_M^{r,\alpha})}{n^{-(r+\alpha)/(d+1)-1/2}} > 0.$$

We add that these lower bounds are sharp, i.e.,

$$\mathrm{e}_{\mathrm{Int}}^{\mathrm{det}}(n, C_M^{r,\alpha}) \asymp n^{-(r+\alpha)/(d+1)}$$

and

$$\mathrm{e}_{\mathrm{Int}}^{\mathrm{ran}}(n, C_M^{r,\alpha}) \asymp n^{-(r+\alpha)/(d+1)-1/2}.$$

See Novak (1988, Prop. 1.3.9, Prop. 2.2.9).

4.2. **Deterministic Algorithms.** Suppose that

$$\lim_{n\to\infty} \max\left(a_n, \frac{a_n}{n\cdot e_n}\right) = 0. \tag{18}$$

Then the lower bound $n \cdot e_n$ for the integration problem turns out to be a lower bound for solving the initial value problem, too, up to a constant.

If $n \cdot e_n$ tends to zero and weak equivalence $\asymp$ holds in (17), then (18) is equivalent to

$$\lim_{n\to\infty} \frac{\sup_{f\in F^{(n)}} \|f\|_\infty^2}{n\cdot e_n} = 0.$$

Theorem 3. *If* (18) *is satisfied then*

$$\liminf_{n\to\infty} \frac{\mathrm{e}^{\mathrm{det}}(n, \mathbf{B})}{n\cdot e_n} > 0.$$

Proof. Let $P_n \in \mathfrak{A}_n^{\mathrm{det}}$, and consider the deterministic algorithm

$$\widetilde{P_n} = P_n - I^{(0)} \in \mathfrak{A}_n^{\mathrm{det}}.$$

From (16) we get

$$\begin{aligned} &|P[(0, b_1, 0, \dots, 0)] - P_n[(0, b_1, 0, \dots, 0)]| \\ &\qquad \geq \left|I^{(1)}[b_1] - \widetilde{P_n}[(0, b_1, 0, \dots, 0)]\right| - \sum_{\mu=2}^{\infty} \left|I^{(\mu)}[b_1 \otimes \cdots \otimes b_1]\right|. \end{aligned}$$

By (18), a_n tends to zero, and due to (17) we thus may assume that

$$\sup_{f\in F^{(n)}} \|f\|_\infty^2 \leq (2\cdot(d+1)\cdot c)^{-1}.$$

In this case Lemma 3 with $m = 1$ implies

$$\sup_{f\in F^{(n)}} \sum_{\mu=2}^{\infty} \left|I^{(\mu)}[f \otimes \cdots \otimes f]\right| \leq 2\cdot\big((d+1)\cdot c\big)^2 \cdot a_n.$$

Furthermore,

$$\sup_{f\in F^{(n)}} \left|I^{(1)}[f] - \widetilde{P_n}[(0, f, 0, \dots, 0)]\right| \geq K\cdot n\cdot e_n,$$

see Theorem 2. Summarizing we obtain

$$\begin{aligned} \sup_{\mathbf{b}\in\mathbf{B}} |P[\mathbf{b}] - P_n[\mathbf{b}]| &\geq \sup_{f\in F^{(n)}} |P[(0, f, 0, \dots, 0)] - P_n[(0, f, 0, \dots, 0)]| \\ &\geq n\cdot e_n \cdot \left(K - 2\cdot\big((d+1)\cdot c\big)^2 \cdot \frac{a_n}{n\cdot e_n}\right). \end{aligned}$$

Use (18) to complete the proof. □

Corollary 1. *Let $r \in \mathbb{N}_0$, $0 < \alpha \le 1$, and $M > 0$, and assume that*

$$\{(0, b_1, 0, \dots, 0) : b_1 \in C_M^{r,\alpha}\} \subseteq \mathbf{B}.$$

Then

$$\liminf_{n\to\infty} \frac{\mathrm{e}^{\mathrm{det}}(n, \mathbf{B})}{n^{-(r+\alpha)/(d+1)}} > 0.$$

Proof. In view of Example 1 we can apply Theorem 3. □

4.3. **Randomized Algorithms.** In addition to (18) we now assume that

$$\lim_{n\to\infty} \frac{a_n}{(n^{1/2} \cdot e_n)^{1/m}} = 0 \tag{19}$$

and

$$\liminf_{n\to\infty} \frac{e_{n\cdot(4m-3)}}{e_n} > 0 \tag{20}$$

for some integer $m \ge 1$. Then the lower bound $1/2 \cdot n^{1/2} \cdot e_n$ for the integration problem turns out to be a lower bound for solving the initial value problem, too, up to a constant.

If $n \cdot e_n$ tends to zero and weak equivalence $\asymp$ holds in (17), then (19) is equivalent to

$$\lim_{n\to\infty} \frac{\sup_{f\in F^{(n)}} \|f\|_\infty^2}{(n^{1/2} \cdot e_n)^{1/m}} = 0.$$

By (20) we require that the lower bound for integration decreases at most by a multiplicative factor if the number of nodes increases by the fixed multiplicative factor $4m-3$.

Theorem 4. *If* (18), (19), *and* (20) *for some integer $m \ge 1$ are satisfied then*

$$\liminf_{n\to\infty} \frac{\mathrm{e}^{\mathrm{ran}}(n, \mathbf{B})}{n^{1/2} \cdot e_n} > 0.$$

Proof. Recall that $I^{(\mu)}$ is the integral with weight function $\rho^{(\mu)}$, see (12) and (15), and let $\lambda^{(\mu)}$ denote the L_1-norm of $\rho^{(\mu)}$. Let $I^{(\mu,+)}$, $I^{(\mu,-)}$, $\lambda^{(\mu,+)}$, and $\lambda^{(\mu,-)}$ denote the corresponding mappings and L_1-norms for the positive part $\rho^{(\mu,+)}$ and the negative part $\rho^{(\mu,-)}$ of $\rho^{(\mu)}$.

Consider the classical Monte Carlo algorithm $\widehat{I}_n^{(\mu,+)} \in \mathfrak{A}_n^{\mathrm{ran}}$ for approximation of $I^{(\mu,+)}$ that is based on importance sampling according to $\rho^{(\mu,+)}$, i.e.,

$$\widehat{I}_n^{(\mu,+)}[g] = \frac{\lambda^{(\mu,+)}}{n} \cdot \sum_{i=1}^{n} g(X_i),$$

where $X_1, \dots, X_n$ are i.i.d. with distribution having Lebesgue density $\rho^{(\mu,+)}/\lambda^{(\mu,+)}$. The upper bound

$$E\left(\left|I^{(\mu,+)}[g] - \widehat{I}_n^{(\mu,+)}[g]\right|^2\right) \le \lambda^{(\mu,+)} \cdot n^{-1} \cdot I^{(\mu,+)}[g^2]$$

is well known and easy to verify. Using (17) we conclude that

$$\begin{aligned}\sup_{f\in F^{(n)}} \left(E\left(\left|I^{(\mu,+)}[f\otimes\cdots\otimes f] - \widehat{I}_n^{(\mu,+)}[f\otimes\cdots\otimes f]\right|^2\right)\right)^{1/2} &\le \lambda^{(\mu,+)} \cdot n^{-1/2} \cdot \sup_{f\in F^{(n)}} \|f\|_\infty^\mu \\ &\le \lambda^{(\mu,+)} \cdot n^{-1/2} \cdot a_n^{\mu/2}.\end{aligned}$$

In the same way we approximate $I^{(\mu,-)}$ to obtain a randomized algorithm $\widehat{I}_{2n}^{(\mu)} \in \mathfrak{A}_{2n}^{\rm ran}$ for approximation of $I^{(\mu)} = I^{(\mu,+)} - I^{(\mu,-)}$ that satisfies

$$\sup_{f\in F^{(n)}} \left(E\left(\left| I^{(\mu)}[f\otimes\cdots\otimes f] - \widehat{I}_{2n}^{(\mu)}[f\otimes\cdots\otimes f]\right|^2\right)\right)^{1/2} \le \lambda^{(\mu)}\cdot n^{-1/2}\cdot a_n^{\mu/2}. \tag{21}$$

Let $P_n \in \mathfrak{A}_n^{\rm ran}$, and consider the randomized algorithm

$$\widetilde{P_n} = P_n - I^{(0)} - \sum_{\mu=2}^{2m-1} \widehat{I}_{2n}^{(\mu)} \in \mathfrak{A}_{n\cdot(4m-3)}^{\rm ran}$$

(with $P_n,\ \widehat{I}_{2n}^{(2)},\dots \widehat{I}_{2n}^{(2m-1)}$ being independent, say). From (16) we get

$$\begin{aligned}
&|P[(0,b_1,0,\dots,0)] - P_n[(0,b_1,0,\dots,0)]| \\
&\quad \ge \left| I^{(1)}[b_1] - \widetilde{P_n}[(0,b_1,0,\dots,0)]\right| - \sum_{\mu=2}^{2m-1} \left| I^{(\mu)}[b_1\otimes\cdots\otimes b_1] - \widehat{I}_{2n}^{(\mu)}[b_1\otimes\cdots\otimes b_1]\right| \\
&\qquad - \sum_{\mu=2m}^{\infty} \left| I^{(\mu)}[b_1]\right|
\end{aligned}$$

for every realization of $\widetilde{P_n}$. As in the proof of Theorem 3 we obtain

$$\sup_{f\in F^{(n)}} \sum_{\mu=2m}^{\infty} \left| I^{(\mu)}[f\otimes\cdots\otimes f]\right| \le a_n^m\cdot c$$

with some constant $c>0$ if n is sufficiently large. Moreover,

$$\sup_{f\in F^{(n)}} \left(E\left(\left| I^{(\mu)}[f\otimes\cdots\otimes f] - \widehat{I}_{n}^{(\mu)}[f\otimes\cdots\otimes f]\right|^2\right)\right)^{1/2} \le \lambda^{(\mu)}\cdot a_n\cdot n^{-1/2}$$

for every $\mu\ge 2$ if n is sufficiently large, see (21). Finally,

$$\sup_{f\in F^{(n)}} \left(E\left(\left| I^{(1)}[f] - \widetilde{P_n}[(0,f,0,\dots,0)]\right|^2\right)\right)^{1/2} \ge K/2\cdot(4m-3)^{1/2}\cdot n^{1/2}\cdot e_{n\cdot(4m-3)},$$

see Theorem 2. Summarizing we obtain

$$\begin{aligned}
\sup_{\mathbf{b}\in\mathbf{B}} \left(E\left(|P[\mathbf{b}] - P_n[\mathbf{b}]|^2\right)\right)^{1/2} &\ge \sup_{f\in F^{(n)}} \left(E\left(|P[(0,f,0,\dots,0)] - P_n[(0,f,0,\dots,0)]|^2\right)\right)^{1/2} \\
&\ge n^{1/2}\cdot e_n\cdot\left(K/2\cdot(4m-3)^{1/2}\cdot\frac{e_{n\cdot(4m-3)}}{e_n} - \sum_{\mu=2}^{2m-1}\lambda^{(\mu)}\cdot\frac{a_n}{n\cdot e_n} - c\cdot\frac{a_n^m}{n^{1/2}\cdot e_n}\right).
\end{aligned}$$

Use (18), (19), and (20) to complete the proof. □

Corollary 2. *Let $r\in\mathbb{N}_0$, $0<\alpha\le 1$, and $M>0$, and assume that*

$$\{(0,b_1,0,\dots,0) : b_1\in C_M^{r,\alpha}\}\subseteq\mathbf{B}.$$

Then

$$\liminf_{n\to\infty} \frac{{\rm e}^{\rm ran}(n,\mathbf{B})}{n^{-(r+\alpha)/(d+1)-1/2}} > 0.$$

Proof. Consider the sequences $(a_n)_{n\in\mathbb{N}}$ and $(e_n)_{n\in\mathbb{N}}$ from Example 1. We already know that (18) is satisfied, and (20) obviously holds for every integer $m \geq 2$. It remains to verify (19) for some integer $m \geq 2$ and then to apply Theorem 4. Since

$$a_n \cdot e_n^{-2/m} \cdot n^{-1/m} \asymp n^{-(r+\alpha)/(d+1)\cdot(2-2/m)+1/m},$$

we have equivalence of

$$m > 1 + \frac{d+1}{2(r+\alpha)}.$$

and (19). □

We see that in the previous proof a low degree of smoothness or a high dimension requires to take a large value of m.

5. Upper Bounds and Almost Optimal Algorithms

We construct and analyze algorithms for the approximate computation of the solution $P[\mathbf{b}]$ of the initial value problem at the point (t_0, x_0). The algorithms are based on values of $\mathbf{b} = (b_0, \ldots, b_d) \in \mathbf{B}$ at finitely many points. The series representation (11) suggests a deterministic algorithm of the following form:

(I) approximate the tensor products $\mathbf{b}^{(\mathbf{j})} = b_{j_1} \otimes \cdots \otimes b_{j_\mu}$ by $\widetilde{\mathbf{b}^{(\mathbf{j})}}$, say, for $\mathbf{j} \in \{0, \ldots, d\}^\mu$ and $\mu = 1, \ldots, m$,

(II) approximate $P[\mathbf{b}]$ by

$$I^{(0)} + \sum_{\mu=1}^{m} \sum_{\mathbf{j}\in\{0,\ldots,d\}^\mu} I^{(\mathbf{j})}\left[\widetilde{\mathbf{b}^{(\mathbf{j})}}\right].$$

For the randomized algorithm we additionally use a variance reduction, and therefore we

(III) add Monte Carlo approximations to integrals $I^{(\mathbf{j})}\left[\mathbf{b}^{(\mathbf{j})} - \widetilde{\mathbf{b}^{(\mathbf{j})}}\right]$.

In (I) we consider approximation with respect to a weighted L_∞-norm $\|\cdot\|_{\mu,\delta}$ for

$$0 \leq \delta < 1,$$

which is motivated by the results from Section 3. Note that

$$\left\|\mathbf{b}^{(\mathbf{j})}\right\|_{\mu,\delta} = \|b_{j_1}\|_{1,\delta} \cdots \|b_{j_\mu}\|_{1,\delta} < \infty,$$

if a growth condition with exponent $\beta \leq \delta$ holds for the coefficients b_j, see Section 2. We aim at algorithms P_n with error close to $\mathrm{e}^{\mathrm{det}}(n, \mathbf{B})$ or $\mathrm{e}^{\mathrm{ran}}(n, \mathbf{B})$ and with $\mathrm{cost}(P_n, \mathbf{B})$ close to n. In view of this goal tensor products $\widetilde{\mathbf{b}^{(\mathbf{j})}} = \widetilde{b_{j_1}} \otimes \cdots \otimes \widetilde{b_{j_\mu}}$ with suitable approximations $\widetilde{b_{j_\ell}}$ for b_{j_ℓ} should not be used in (I), since then the number of arithmetic operations is too large. Instead, we use Smolyak formulas.

5.1. **Assumptions.** For simplicity we assume that all coefficients b_j belong to the same class B of functions, i.e.,

$$\mathbf{B} = B^{d+1}.$$

Furthermore, we require that every $b \in B$ is continuous and satisfies a Hölder condition with exponent $0 < \alpha \leq 1$, see Section 2.

Our key assumption deals with approximation of functions $b \in B$ with respect to the norm $\|\cdot\|_{1,\delta}$ for some exponent $0 \leq \delta < 1$. First of all we assume that

$$\sup_{b\in B} \|b\|_{1,\delta} \leq K_1 \tag{22}$$

with a constant $K_1 > 0$. Hence, in particular, the growth condition from Section 2 is satisfied with $\beta = \delta$. For approximation we consider linear methods $A^{(k)}$ that are based on a finite number of function values. Every such method $A^{(k)} : C(D) \to C(D)$ is of the form

$$A^{(k)}[b] = \sum_{\nu=1}^{n_k} b\big(y_\nu^{(k)}\big) \cdot h_\nu^{(k)}$$

with pairwise different nodes $y_\nu^{(k)} \in D$ and basis functions $h_\nu^{(k)} \in C(D)$. Let

$$Y^{(k)} = \big\{y_1^{(k)}, \ldots, y_{n_k}^{(k)}\big\}$$

denote the nodes that are used by $A^{(k)}$, and define $T^{(k)}(\sigma) : D \to \mathbb{R}$ by

$$T^{(k)}(\sigma)[z] = \sum_{\nu=1}^{n_k} \sigma_\nu \cdot h_\nu^{(k)}(z)$$

for $\sigma \in \mathbb{R}^{n_k}$ and $z \in D$. We assume that there exists an exponent

$$\vartheta > 0$$

and a sequence of methods $A^{(k)}$ with the following properties for every $k \in \mathbb{N}$:

(a) the nodes are nested, i.e.,

$$Y^{(k)} \subseteq Y^{(k+1)},$$

(b) their number is given by

$$n_k = 2^k - 1,$$

(c) an error bound

$$\sup_{b \in B} \|b - A^{(k)}[b]\|_{1,\delta} \le K_2 \cdot n_k^{-\vartheta}$$

holds with a constant $K_2 > 0$,

(d) and a cost bound

$$\sup_{\sigma \in \mathbb{R}^{n_k}} \operatorname{cost}(T^{(k)}(\sigma)) \le K_3 \cdot k$$

holds with a constant $K_3 > 0$.

Property (d) refers to the cost of evaluation of an approximation $A^{(k)}[b]$ at an arbitrarily chosen point $z \in D$, and it holds if the basis functions $h_\nu^{(k)}$ have small and simply shaped supports and are easy to evaluate. We add that property (d) is only used in our construction and analysis of a randomized algorithm for the initial value problem.

We present examples that involve Hölder classes $C^{r,\alpha}$, see Example 1. We rely on the approach and results (sometimes with slight modifications) that are due to Wasilkowski and Woźniakowski (2001). The respective algorithms $A^{(k)}$ use piecewise polynomial interpolation of degree r.

Example 2. Assume that

$$B = \{b \in C_M^{0,\alpha} : |b(0,0)| \le M'\}$$

with $M, M' > 0$ and

$$0 < \alpha < 1.$$

Then (22) is satisfied for every $\delta \ge \alpha$. For weighted approximation with exponent $\alpha < \delta < 1$ one can achieve the order

$$\vartheta = \frac{\alpha}{d+1}$$

as in the unweighted case on compact domains.

For $r+\alpha \geq 1$ we consider subclasses $C_M^{r,\alpha,\kappa}$ of $C_M^{r,\alpha}$. By definition $f \in C_M^{r,\alpha}$ belongs to $C_M^{r,\alpha,\kappa}$ if and only if every r-th order partial derivative g of f satisfies

$$|g(t,x)-g(s,y)| \leq \frac{M}{\max(R^\kappa,1)} \cdot |(t,x)-(s,y)|^\alpha$$

for every $R \geq 0$ and all $(t,x),(s,y) \in D$ with $|x|,|y| \geq R$. Thus, in particular $C_M^{r,\alpha,0} = C_M^{r,\alpha}$.

Example 3. Assume that

$$B = \{b \in C_M^{r,\alpha,\kappa} : \|b\|_{1,\beta} \leq M'\} \tag{23}$$

with $M, M' > 0$ and

$$r+\alpha \geq 1, \qquad 0 \leq \beta < 1, \qquad \kappa > r+\alpha-1.$$

Then (22) is satisfied for every $\delta \geq \beta$. Moreover, for every exponent $\max(\beta, r+\alpha-\kappa) < \delta < 1$, one can achieve the order

$$\vartheta = \frac{r+\alpha}{d+1}$$

for weighted approximation. We add that this upper bound holds in particular for functions with compact support. In fact, consider a compact set $D_0 \subseteq D$ with non-empty interior. Then we have

$$\{b \in C_{M''}^{r,\alpha} : \operatorname{supp} b \subseteq D_0\} \subseteq B$$

for every β and every κ, if M, M' are sufficiently large.

Example 4. Let B be given by (23) with

$$r+\alpha \geq 1, \qquad 0 \leq \beta < 1, \qquad \kappa \leq r+\alpha-1.$$

Then one can achieve the order

$$\vartheta = \frac{r+\alpha}{d+1} \cdot \frac{\delta-\beta}{r+\alpha-\kappa-\beta}$$

for weighted approximation with exponent $\beta < \delta < 1$.

5.2. **The Smolyak Algorithm.** Smolyak's construction yields approximation methods $A^{(\mu,k)} : C(D^\mu) \to C(D^\mu)$ in dimensions $\mu \geq 2$ as follows,

$$A^{(\mu,k)}[f] = \sum_{k \leq |\mathbf{i}| \leq \mu+k-1} (-1)^{\mu+k-1-|\mathbf{i}|} \cdot \binom{\mu-1}{|\mathbf{i}|-k} \cdot A^{(i_1)} \otimes \cdots \otimes A^{(i_\mu)}[f]$$

for $f \in C(D^\mu)$ and $k \in \mathbb{N}$, where

$$\mathbf{i} = (i_1, \ldots, i_\mu) \in \mathbb{N}^\mu, \qquad |\mathbf{i}| = \sum_{\ell=1}^{\mu} i_\ell.$$

For $\mu = 1$ we have $A^{(1,k)} = A^{(k)}$.

A detailed general analysis of Smolyak formulas for tensor product problems is given in Wasilkowski and Woźniakowski (1995). We apply their results to our approximation problem for the functions $\mathbf{b}^{(\mathbf{j})}$.

Because of (a) the function $A^{(\mu,k)}[f]$ depends on f via its values at the nodes from the so-called sparse grid

$$H^{(\mu,k)} = \bigcup_{|\mathbf{i}|=\mu+k-1} Y^{(i_1)} \times \cdots \times Y^{(i_\mu)}.$$

Therefore

$$A^{(\mu,k)}[f] = \sum_{\nu=1}^{n_{\mu,k}} f\left(\mathbf{y}_\nu^{(\mu,k)}\right) \cdot g_\nu^{(\mu,k)} \tag{24}$$

with suitable functions $g_\nu^{(\mu,k)} \in C(D^\mu)$, where $n_{\mu,k}$ is the number of nodes $\mathbf{y}_\nu^{(\mu,k)}$ in the sparse grid $H^{(\mu,k)}$.

Remark 1. In particular,

$$A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}] = \sum_{k \le |\mathbf{i}| \le \mu+k-1} (-1)^{\mu+k-1-|\mathbf{i}|} \cdot \binom{\mu-1}{|\mathbf{i}|-k} \cdot A^{(i_1)}[b_{j_1}] \otimes \cdots \otimes A^{(i_\mu)}[b_{j_\mu}]$$

for $\mathbf{j} \in \{0,\dots,d\}^\mu$. Hence n_k values of the coefficient vector $\mathbf{b}$, namely $(b_0(y),\dots,b_d(y))$ with $y \in Y^{(k)}$, suffice to determine $A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]$ for every dimension $\mu \ge 1$ and every $\mathbf{j} \in \{0,\dots,d\}^\mu$.

We have the following error bound.

Lemma 5 (Wasilkowski and Woźniakowski (1995))**.** *Assume that*

$$\mu - 1 \le \lambda \cdot k \tag{25}$$

for some constant $\lambda > 0$. Then there exists a constant $c > 0$ such that

$$\sup_{\mathbf{j}\in\{0,\dots,d\}^\mu} \sup_{\mathbf{b}\in\mathbf{B}} \|\mathbf{b}^{(\mathbf{j})} - A^{(\mu,k)}[b^{(\mathbf{j})}]\|_{\mu,\delta} \le (c/\mu)^\mu \cdot (\ln n_k)^{\mu-1} \cdot n_k^{-\vartheta}.$$

Proof. Let $\mu \ge 2$, and assume that $K_2 \ge 2^{-\vartheta} K_1$ without loss of generality. Lemma 2 in Wasilkowski and Woźniakowski (1995), applied[1] with $B = K_1$, $C = K_2\, 2^\vartheta$, $D = 2^{-\vartheta}$, and $E = C\,(1+2^\vartheta)$, shows that (b) and (c) yield

$$\|\mathbf{b}^{(\mathbf{j})} - A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]\|_{\mu,\delta} \le K_2^\mu \cdot \left(1+2^\vartheta\right)^{\mu-1} \cdot \binom{\mu-1+k}{k} \cdot 2^{-\vartheta(k-1)}.$$

Let $\zeta \in \mathbb{N}$. Induction readily yields $\zeta! \ge (\zeta/e)^\zeta$. Hence,

$$\binom{\zeta+k}{k} \le \frac{\exp(\zeta)\cdot(k+\zeta)^\zeta}{\zeta^\zeta} = \exp(\zeta) \cdot \left(\frac{k}{\zeta}\right)^\zeta \cdot \left(1+\frac{\zeta}{k}\right)^{(k/\zeta)\cdot(\zeta^2/k)}.$$

Therefore

$$\binom{\zeta+k}{k} \le \exp(\zeta+\zeta^2/k) \cdot \left(\frac{k}{\zeta}\right)^\zeta. \tag{26}$$

Together with (25) this yields

$$\binom{\mu-1+k}{k} \le \exp((\mu-1)\cdot(1+\lambda)) \cdot \left(\frac{k}{\mu-1}\right)^{\mu-1}.$$

Use $\ln n_k \asymp k$ to complete the proof. □

The number of nodes in the sparse grid $H^{(\mu,k)}$ is bounded as follows.

[1] Wasilkowski and Woźniakowski (1995) consider a Hilbert space setting. However, their proof is applicable here, too. See Li (2002) for an analysis in a Banach space setting.

Lemma 6 (Wasilkowski and Woźniakowski (1995))**.**

$$n_{\mu,k} \le 2^k \cdot \binom{\mu+k-2}{k-1}.$$

Proof. This is Lemma 7 in Wasilkowski and Woźniakowski (1995) with parameters $F_0 = 1$ and $F = 2$. □

In order to estimate the cost of evaluating a Smolyak approximation $A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]$ at a point $z \in D^\mu$ we define $T^{(\mu,k)}(\sigma) : D^\mu \to \mathbb{R}$ by

$$T^{(\mu,k)}(\sigma)[z] = \sum_{k \le |\mathbf{i}| \le \mu+k-1} (-1)^{\mu+k-1-|\mathbf{i}|} \cdot \binom{\mu-1}{|\mathbf{i}|-k} \cdot T^{(i_1)}(\sigma^{(1)})[z_1] \cdots T^{(i_\mu)}(\sigma^{(\mu)})[z_\mu] \tag{27}$$

for $z \in D^\mu$ and $\sigma = (\sigma^{(1)}, \ldots, \sigma^{(\mu)})$ with $\sigma^{(\ell)} \in \mathbb{R}^{n_{i_\ell}}$. Moreover, we define

$$m_{n_k} = \left\lceil \ln n_k / \sqrt{\ln \ln n_k} \right\rceil, \qquad k \ge 2, \tag{28}$$

cf. Lemma 4. Clearly

$$n_k \asymp 2^k \tag{29}$$

and

$$m_{n_k} \asymp \frac{k}{\sqrt{\ln k}}. \tag{30}$$

Lemma 7. *For every* $\varepsilon > 0$

$$\lim_{k\to\infty} \left(n_k^{-\varepsilon} \cdot \binom{m_{n_k}+k}{k} \right) = 0.$$

Proof. By (26)

$$\binom{\zeta+k}{k} \le \exp\left(\zeta + \zeta^2/k + \zeta \cdot \ln(k/\zeta)\right),$$

and for $\zeta = m_{n_k}$ we have

$$\zeta + \zeta^2/k + \zeta \cdot \ln(k/\zeta) \asymp k \cdot \frac{\ln\sqrt{\ln k}}{\sqrt{\ln k}},$$

see (30). On the other hand we have (29). □

Lemma 8. *For every* $\varepsilon > 0$

$$\lim_{k\to\infty} \left(n_k^{-\varepsilon} \cdot \sup_{\mu=1,\ldots,m_{n_k}} \sup_{\sigma} \operatorname{cost}(T^{(\mu,k)}(\sigma)) \right) = 0.$$

Proof. Due to assumption (d), each summand in (27) can be computed at cost proportional to $|\mathbf{i}| \le \mu + k - 1$, and the number of summands is given by

$$\#\{\mathbf{i} \in \mathbb{N}^\mu : k \le |\mathbf{i}| \le \mu + k - 1\} \le \sum_{\ell=\mu}^{\mu+k-1} \binom{\ell-1}{\mu-1} = \binom{\mu+k-1}{k-1}.$$

For $\mu \le m_{n_k}$ we therefore get

$$\operatorname{cost}(T^{(\mu,k)}(\sigma)) \le (m_{n_k}+k-1) \cdot \binom{m_{n_k}+k}{k} \asymp k \cdot \binom{m_{n_k}+k}{k}.$$

It remains to apply (29) and Lemma 7. □

5.3. **A Deterministic Algorithm for the Initial Value Problem.** Let $\mu, k \in \mathbb{N}$ and $\mathbf{j} \in \{0, \dots, d\}^\mu$. The Smolyak algorithm $A^{(\mu,k)}$ leads to the approximation

$$I^{(\mathbf{j},k)}[\mathbf{b}^{(\mathbf{j})}] = I^{(\mathbf{j})}\big[A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]\big]$$

of $I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}]$, i.e., $\widetilde{\mathbf{b}^{(\mathbf{j})}} = A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]$ in (II). Note that

$$I^{(\mathbf{j},k)}[\mathbf{b}^{(\mathbf{j})}] = \sum_{\nu=1}^{n_{\mu,k}} \mathbf{b}^{(\mathbf{j})}\big(\mathbf{y}_\nu^{(\mu,k)}\big) \cdot u_\nu^{(\mathbf{j},k)},$$

where

$$u_\nu^{(\mathbf{j},k)} = I^{(\mathbf{j})}\big[g_\nu^{(\mu,k)}\big],$$

see (24). The coefficients $u_\nu^{(\mathbf{j},k)} \in \mathbb{R}$ do not depend on $\mathbf{b}$ and can therefore be pre-computed.

According to (II) a reasonable approximation to $P[\mathbf{b}]$ is given by

$$P^{(m,k)}[\mathbf{b}] = I^{(0)} + \sum_{\mu=1}^{m} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} I^{(\mathbf{j},k)}[\mathbf{b}^{(\mathbf{j})}],$$

where $m \geq 1$. The constant $I^{(0)}$ can be pre-computed, too. We take

$$m = m_{n_k}$$

for $k \geq 2$, see (28), and this choice defines a sequence of algorithms

$$P^{(k)} = P^{(m_{n_k},k)} \in \mathfrak{A}_{n_k}^{\det},$$

see Remark 1.

We present an asymptotic error bound for $P^{(k)}$ in terms of n_k and of the approximation order ϑ, see (c).

Theorem 5. *For every $\varepsilon > 0$*

$$\lim_{k\to\infty} \big(n_k^{\vartheta-\varepsilon} \cdot \mathrm{e}(P^{(k)}, \mathbf{B})\big) = 0.$$

Proof. For every $m \geq 1$

$$\big|P[\mathbf{b}] - P^{(m,k)}[\mathbf{b}]\big| \leq \sum_{\mu=m+1}^{\infty} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} \big|I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}]\big| + \sum_{\mu=1}^{m} \sum_{\mathbf{j}\in\{0,\dots,d\}^\mu} \big|I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}] - I^{(\mathbf{j},k)}[\mathbf{b}^{(\mathbf{j})}]\big|.$$

Let $\mu \leq m_{n_k}$. From (30), Lemma 2, and Lemma 5 we get

$$\begin{aligned}\big|I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}] - I^{(\mathbf{j},k)}[\mathbf{b}^{(\mathbf{j})}]\big| &\leq \lambda_\delta^{(\mathbf{j})} \cdot \big\|\mathbf{b}^{(\mathbf{j})} - A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]\big\|_{\mu,\delta} \\ &\leq \left(\frac{c \cdot \ln n_k}{\mu^{1+\delta^*}}\right)^\mu \cdot (\ln n_k)^{-1} \cdot n_k^{-\vartheta}\end{aligned}$$

for every $\mathbf{b} \in \mathbf{B}$ and $\mathbf{j} \in \{0, \dots, d\}^\mu$ with some constant $c > 0$.

For every $z > 0$

$$\sup_{\mu\in\mathbb{N}} \big(z/\mu^{1+\delta^*}\big)^\mu \leq \exp\big((1+\delta^*)/e \cdot z^{1/(1+\delta^*)}\big),$$

and therefore

$$\sum_{\mu=1}^{m_{n_k}} \big(c \cdot d \cdot \ln n_k/\mu^{1+\delta^*}\big)^\mu \leq (\ln n_k + 1) \cdot \exp\Big((1+\delta^*)/e \cdot (c \cdot d \cdot \ln n_k)^{1/(1+\delta^*)}\Big).$$

Since $\delta^* > 0$, we obtain

$$\lim_{k\to\infty}\left(n_k^{\vartheta-\varepsilon}\cdot\sup_{\mathbf{b}\in\mathbf{B}}\sum_{\mu=1}^{m_{n_k}}\sum_{\mathbf{j}\in\{0,\dots,d\}^\mu}\left|I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}]-I^{(\mathbf{j},k)}[\mathbf{b}^{(\mathbf{j})}]\right|\right)=0.$$

From Lemma 4 we already know that

$$\lim_{k\to\infty}\left(n_k^{\vartheta}\cdot\sup_{\mathbf{b}\in\mathbf{B}}\sum_{\mu=m_{n_k}+1}^{\infty}\sum_{\mathbf{j}\in\{0,\dots,d\}^\mu}\left|I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}]\right|\right)=0.$$

□

Theorem 5 immediately yields upper bounds for the error of $P^{(k)}$ on the Hölder classes from Examples 2–4. In particular, for functions with low regularity or for smooth functions whose Hölder constants decay sufficiently fast as $|x|$ tends to ∞, the algorithms $P^{(k)}$ turn out to be almost optimal in the power scale.

Corollary 3. *Let B denote a Hölder class from Example 2 or Example 3. Then, for $\mathbf{B}=B^{d+1}$ and every $\varepsilon>0$,*

$$\lim_{k\to\infty}\frac{\mathrm{e}(P^{(k)},\mathbf{B})}{\mathrm{e}^{\mathrm{det}}(n_k,\mathbf{B})^{1-\varepsilon}}=0.$$

Proof. Apply Corollary 1 and Theorem 5. □

Remark 2. Suppose that $\mathbf{B}=B^{d+1}$ with a Hölder class B from Example 4, and put

$$\gamma=\frac{r+\alpha}{d+1}$$

as well as

$$\vartheta=\gamma\cdot\frac{1-\beta}{r+\alpha-\kappa-\beta}.$$

We get the upper bound

$$\lim_{k\to\infty}\left(n_k^{\vartheta-\varepsilon}\cdot\mathrm{e}(P^{(k)},\mathbf{B})\right)=0$$

for every $\varepsilon>0$ by taking δ close to one in Example 4 and the lower bound

$$\liminf_{k\to\infty}\left(n_k^{\gamma}\cdot\mathrm{e}^{\mathrm{det}}(n_k,\mathbf{B})\right)>0.$$

In the extremal case $\kappa=0$ we have

$$\frac{r+\alpha-1}{d+1}\le\gamma-\vartheta\le\frac{r+\alpha}{d+1},$$

so that here our upper and lower bounds differ in particular if the smoothness is large, compared to the dimension.

Remark 3. Suppose that we use the tensor product algorithm $A^{(k)}\otimes\cdots\otimes A^{(k)}$ in the definition of $I^{(\mathbf{j},k)}$ instead of the Smolyak algorithm $A^{(\mu,k)}$. In both cases the resulting algorithm $P^{(k)}$ depends on the same data from $\mathbf{b}$, and in the tensor product case we have an error bound

$$\sup_{\mathbf{b}\in B}\|\mathbf{b}^{(\mathbf{j})}-A^{(k)}[b_{j_1}]\otimes\cdots\otimes A^{(k)}[b_{j_\mu}]\|_{\mu,\delta}\le\mu\cdot(K_1+K_2)^{\mu-1}\cdot K_2\cdot n_k^{-\vartheta},$$

cf. Lemma 5. Therefore the error bound from Theorem 5 and the optimality result from Corollary 3 are valid in the tensor product case, too. See, however, Remark 4.

Now we study the computational cost of the algorithm $P^{(k)}$. Note that $P^{(k)}$ uses the same number of arithmetic operations and function evaluations for every $\mathbf{b} \in \mathbf{B}$. We show that $\mathrm{cost}(P^{(k)}, \mathbf{B})$ is almost proportional to the number n_k of function values used.

Theorem 6. *For every $\varepsilon > 0$*

$$\lim_{k\to\infty} \left(n_k^{-1-\varepsilon} \cdot \mathrm{cost}(P^{(k)}, \mathbf{B}) \right) = 0.$$

Proof. Suppose that $\mathbf{b}$ is evaluated at the nodes from $Y^{(k)}$ and that $I^{(0)}$ as well as the coefficients $u_\nu^{(\mathbf{j},k)}$ are available for $\mu = 1, \ldots, m_{n_k}$, $\mathbf{j} \in \{0, \ldots, d\}^\mu$, and $\nu = 1, \ldots, n_{\mu,k}$. To compute $\mathbf{b}^{(\mathbf{j})}$ at the nodes from the sparse grid $H^{(\mu,k)}$ we need $n_{\mu,k} \cdot (\mu - 1)$ multiplications. Thereafter we need $2 \cdot n_{\mu,k} - 1$ multiplications and additions to compute $I^{(\mathbf{j},k)}[\mathbf{b}^{(\mathbf{j})}]$. Accumulation of these numbers yields $P^{(k)}[b]$. The total number of arithmetic operations is proportional to q_k with

$$q_k \le \sum_{\mu=1}^{m_{n_k}} \sum_{\mathbf{j}\in\{0,\ldots,d\}^\mu} n_{\mu,k} \cdot (\mu + 1) \le (d+1)^{m_{n_k}} \cdot \widetilde{q}_k,$$

where

$$\widetilde{q}_k = \sum_{\mu=1}^{m_{n_k}} n_{\mu,k} \cdot (\mu + 1).$$

Lemma 6 yields

$$q_k \le (d+1)^{m_{n_k}} \cdot 2^k \cdot \sum_{\mu=1}^{m_{n_k}} \binom{\mu + k - 2}{k-1} \cdot (\mu + 1) \le (d+1)^{m_{n_k}} \cdot 2^{k+1} \cdot m_{n_k}^2 \cdot \binom{m_{n_k} + k}{k}.$$

Thus, by (29), (30), and Lemma 7,

$$\lim_{k\to\infty} \left(n_k^{-1-\varepsilon} \cdot q_k \right) = 0.$$

It remains to use $\mathrm{cost}(P^{(k)}, \mathbf{B}) \le n_k + q_k$. □

Remark 4. Suppose that we replace the Smolyak approximations by tensor product approximations as in Remark 3. Then the computational cost of the resulting algorithm $P^{(k)}$ increases faster than every polynomial in n_k.

5.4. **A Randomized Algorithm for the Initial Value Problem.** We employ a variance reduction and proceed as in the proof of Theorem 4. Let $I^{(\mathbf{j},\pm)}[g]$ denote the integrals of g with the positive or negative part $\rho^{(\mathbf{j},\pm)}$ of $\rho^{(\mathbf{j})}$ as weight functions, see (12). Moreover, put

$$\omega_{\mu,\delta}(\tau_1, \xi_1, \ldots, \tau_\mu, \xi_\mu) = \prod_{\ell=1}^{\mu} \left(1 + |\xi_\ell|^\delta\right).$$

The classical Monte Carlo algorithms for approximating $I^{(\mathbf{j},\pm)}$ that are based on importance sampling according to $\rho^{(\mathbf{j},\pm)} \cdot \omega_{\mu,\delta}$ yield Monte Carlo algorithms $\widehat{I}^{(\mathbf{j},k)} \in \mathfrak{A}_{2n_k}^{\mathrm{ran}}$ that satisfy

$$\left(E\left(\left| I^{(\mathbf{j})}[g] - \widehat{I}^{(\mathbf{j},k)}[g] \right|^2 \right) \right)^{1/2} \le \lambda_\delta^{(\mathbf{j})} \cdot \|g\|_{\mu,\delta} \cdot n_k^{-1/2} \tag{31}$$

for $\mu \ge 1$, $\mathbf{j} \in \{0, \ldots, d\}^\mu$, and $g \in C(D^\mu)$ with $\|g\|_{\mu,\delta} < \infty$, cf. (21).

According to (III) we combine the deterministic algorithm $P^{(k)}$ from Section 5.3 with Monte Carlo approximations $\widehat{I}^{(\mathbf{j},k)}\big[\mathbf{b}^{(\mathbf{j})} - A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]\big]$ of $\widehat{I}\big[\mathbf{b}^{(\mathbf{j})} - A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]\big]$. Hereby we get the randomized algorithm $\widehat{P}^{(k)}$,

$$\widehat{P}^{(k)}[\mathbf{b}] = P^{(k)}[\mathbf{b}] + \sum_{\mu=1}^{m_{n_k}} \sum_{\mathbf{j}\in\{0,\ldots,d\}^\mu} \widehat{I}^{(\mathbf{j},k)}\big[\mathbf{b}^{(\mathbf{j})} - A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}]\big]$$

(with $\widehat{I}^{(\mathbf{j},k)}$ being independent for $\mathbf{j} \in \{0,\ldots,d\}^\mu$ and $\mu = 1,\ldots,m_{n_k}$, say).

Theorem 7. *For every $\varepsilon > 0$*

$$\lim_{k\to\infty} \Big(n_k^{\vartheta+1/2-\varepsilon} \cdot \mathrm{e}(\widehat{P}^{(k)}, \mathbf{B}) \Big) = 0.$$

Proof. Let $\mathbf{b} \in \mathbf{B}$ and put

$$f^{(\mathbf{j},k)} = \mathbf{b}^{(\mathbf{j})} - A^{(\mu,k)}[\mathbf{b}^{(\mathbf{j})}].$$

For every realization of $\widehat{P}^{(k)}$

$$P[\mathbf{b}] - \widehat{P}^{(k)}[\mathbf{b}] = \sum_{\mu=m_{n_k+1}}^{\infty} \sum_{\mathbf{j}\in\{0,\ldots,d\}^\mu} I^{(\mathbf{j})}[\mathbf{b}^{(\mathbf{j})}] + \sum_{\mu=1}^{m_{n_k}} \sum_{\mathbf{j}\in\{0,\ldots,d\}^\mu} \Big(I^{(\mathbf{j})}\big[f^{(\mathbf{j},k)}\big] - \widehat{I}^{(\mathbf{j},k)}\big[f^{(\mathbf{j},k)}\big] \Big).$$

By Lemma 2, Lemma 5, and (31)

$$\left(E\left(I^{(\mathbf{j})}\big[f^{(\mathbf{j},k)}\big] - \widehat{I}^{(\mathbf{j},k)}\big[f^{(\mathbf{j},k)}\big] \right)^2 \right)^{1/2} \le (c/\mu)^\mu \cdot (\ln n_k)^{\mu-1} \cdot n_k^{-\vartheta-1/2}$$

for some constant $c > 0$. It remains to use the arguments from the proof of Theorem 5. □

On the Hölder classes from Examples 2–4 the algorithms $\widehat{P}^{(k)}$ enjoy the same optimality properties as their deterministic counterparts $P^{(k)}$.

Corollary 4. *Let B denote a Hölder class from Example 2 or Example 3. Then, for $\mathbf{B} = B^{d+1}$ and every $\varepsilon > 0$,*

$$\lim_{k\to\infty} \frac{\mathrm{e}(\widehat{P}^{(k)}, \mathbf{B})}{\mathrm{e}^{\mathrm{ran}}(n_k, \mathbf{B})^{1-\varepsilon}} = 0.$$

Proof. Apply Corollary 2 and Theorem 7. □

Remark 5. The conclusion from Remark 2 remains valid for the randomized algorithms $\widehat{P}^{(k)}$, too, with γ and ϑ both increased by 1/2.

The computational cost of the algorithm $\widehat{P}^{(k)}$ turns out to be almost proportional to the number of function values used.

Theorem 8. *For every $\varepsilon > 0$*

$$\lim_{k\to\infty} \Big(n_k^{-1-\varepsilon} \cdot \mathrm{cost}(\widehat{P}^{(k)}, \mathbf{B}) \Big) = 0.$$

Proof. Use Lemma 8, $\mathrm{cost}\big(\widehat{I}(\mathbf{j},k), \mathbf{B}\big) \le c \cdot n_k$, and Theorem 6. □

Acknowledgment. We thank Jakob Creutzig for valuable discussions.

References

Bakhvalov, N. S. (1959), On approximate computation of integrals, Vestnik MGV, Ser. Math. Mech. Astron. Phys. Chem. **4**, 3–18.

Deck, T., and Kruse, S. (2002), Parabolic differential equations with unbounded coefficients - a generalization of the parametrix method, Acta Appl. Math. **74**, 71–91.

Friedman, A. (1964), Partial Differential Equations of Parabolic Type, Prentice-Hall, Englewood Cliffs.

Kloeden, P. E., and Platen, E. (1999), Numerical Solution of Stochastic Differential Equations, Springer, Berlin.

Kwas, M. (2004), Complexity of multivariate Feynman-Kac path integration in randomized and quantum settings, Preprint, Department of Computer Science, Columbia Univ., New York.

Kwas, M., and Li, Youming (2003), Worst case complexity of multivariate Feynman-Kac path integration, J. Complexity **19**, 730–743.

Ledoux, M., and Talagrand, M. (1991), Probability in Banach Spaces, Springer, Berlin.

Li, Youming (2002), Applicability of Smolyak's algorithm to certain Banach spaces of multivariate functions, J. Complexity **18**, 792–814.

Novak, E. (1988), Deterministic and Stochastic Error Bounds in Numerical Analysis, Lect. Notes in Math. **1349**, Springer, Berlin.

Novak, E. (1995), The real number model in numerical analysis, J. Complexity **11**, 57–73.

Plaskota, L., Wasilkowski, G. W., and Woźniakowski, H. (2000), A new algorithm and worst case complexity for Feynman-Kac path integration, J. Comput. Phys. **164**, 335–353.

Traub, J. F., Wasilkowski, G. W., and Woźniakowski, H. (1988), Information-based Complexity, Academic Press, New York.

Wasilkowski, G. W., and Woźniakowski, H. (1995), Explicit cost bounds of algorithms for multivariate tensor product problems, J. Complexity **11**, 1–56.

Wasilkowski, G. W., and Woźniakowski, H. (2001), Complexity of weighted approximation over $\mathbb{R}^d$, J. Complexity **17**, 722–740.

Institut für Angewandte Mathematik, Technische Universität Braunschweig, Pockelsstrasse 14, 38106 Braunschweig, Germany

E-mail address: k.petras@tu-bs.de

Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstrasse 7, 64289 Darmstadt, Germany

E-mail address: ritter@mathematik.tu-darmstadt.de