ASC Report No. 17/2007

Local Error Estimates for Moderately Smooth Problems: Part II – SDEs and SDAEs with Small Noise

Thorsten Sickenberger, Ewa Weinmüller, Renate Winkler

Institute for Analysis and Scientific Computing
Vienna University of Technology — TU Wien
www.asc.tuwien.ac.at ISBN 978-3-902627-00-1

Most recent ASC Reports

16/2007 *Thorsten Sickenberger, Ewa Weinmüller, Renate Winkler*
Local Error Estimates for Moderately Smooth ODEs and DAEs

15/2007 *Markus Brunk, Ansgar Jüngel*
Simulation of Thermal Effects in Optoelectronic Devices Using Coupled Energy-transport and Circuit Models

14/2007 *Jose Antonio Carrillo, Maria Pia Gualdani, Ansgar Jüngel*
Convergence of an Entropic Semi-discretization for Nonlinear Fokker-Planck Equations in R^d

13/2007 *Ansgar Jüngel, Stefan Krause, Paola Pietra*
A Hierarchy of Diffuse Higher-order Moment Equations for Semiconductors

12/2007 *Markus Brunk, Ansgar Jüngel*
Numerical Coupling of Electric Circuit Equations and Energy-transport Models for Semiconductors

11/2007 *Markus Brunk, Ansgar Jüngel*
Numerical Simulation of Thermal Effects in Coupled Optoelectronic Device-circuit Systems

10/2007 *Ansgar Jüngel, Ingrid Violet*
First-Order Entropies for the Derrida-Lebowitz-Speer-Spohn Equation

09/2007 *Maike Schulte, Anton Arnold*
Discrete Transparent Boundary Conditions for the Schrödinger Equation – a Compact Higher Order Scheme

08/2007 *Anton Arnold, Jose Antonio Carrillo, Corinna Klapproth*
Improved Entropy Decay Estimates for the Heat Equation

07/2007 *Christoph Erath, Samuel Ferraz-Leite, Stefan Funken, Dirk Praetorius*
Energy Norm Based A Posteriori Error Estimation for Boundary Element Methods in Two Dimensions

Institute for Analysis and Scientific Computing
Vienna University of Technology
Wiedner Hauptstraße 8–10
1040 Wien, Austria

E-Mail: admin@asc.tuwien.ac.at
WWW: http://www.asc.tuwien.ac.at
FAX: +43-1-58801-10196

ISBN 978-3-902627-00-1

ASC TU WIEN

BIT manuscript No.
(will be inserted by the editor)

Local error estimates for moderately smooth problems: Part II SDE:s and SDAE:s with small noise

Thorsten Sickenberger · **Ewa Weinmüller** · **Renate Winkler**

Received March 8 2007 / Accepted December 8 2008

Abstract The paper consists of two parts. In the first part of the paper, we proposed a procedure to estimate local errors of low order methods applied to solve initial value problems in ordinary differential equations (ODEs) and index-1 differential-algebraic equations (DAEs). Based on the idea of Defect Correction we developed local error estimates for the case when the problem data is only moderately smooth, which is typically the case in stochastic differential equations. In this second part, we will consider the estimation of local errors in context of mean-square convergent methods for stochastic differential equations (SDEs) with small noise and index-1 stochastic differential-algebraic equations (SDAEs). Numerical experiments illustrate the performance of the mesh adaptation based on the local error estimation developed in this paper.

Mathematics Subject Classification (2000) 65C30 · 60H35 · 65L06 · 65L80 · 65L50

Keywords local error estimation · step-size control · adaptive methods · stochastic differential equations · small noise · stochastic differential-algebraic equations · mean-square numerical methods

The first author acknowledges support by the BMBF-project 03RONAVN and the second author support by the Austrian Science Fund Project P17253.

T. Sickenberger
Department of Mathematics, Heriot-Watt University, Edinburgh, Scotland, United Kingdom. E-mail: t.sickenberger@hw.ac.uk

E. Weinmüller
Institut für Analysis und Scientific Computing, Technische Universität Wien, Wiedner Hauptstraße 8-10, A-1040 Wien, Austria. E-mail: e.weinmueller@tuwien.ac.at

R. Winkler
Bergische Universität Wuppertal, Fachbereich C - Mathematik, Gaußstraße 20, D-42199 Wuppertal, Germany. E-mail: winkler@math.uni-wuppertal.de

1 Introduction.

In the first part of the paper we focused our attention on the local error estimates in context of ODEs and index-1 DAEs. We were especially interested in low order linear multi-step methods applied to solve initial value problems. The error estimation procedure based on the idea of Defect Correction, proved to work dependably for the case when the problem data is only moderately smooth, which is typically the case in stochastic differential equations.

Many SDEs of interest in science and engineering feature small noise. When the noise is small, one can expect that the stochastic system, though of a completely different analytical character, has a solution that is somehow 'close' to a deterministic one. Our aim here is to show that the strategies developed in Part I of this paper for low order linear multi-step methods can be applied also in context of small noise SDEs and SDAEs for stochastic analogues of linear multi-step methods and work reliably for this class of problems. However, we will discuss to which extent this statement is correct depending on the size (smallness) of the noise in relation to the step-sizes or accuracy requirements. We concentrate on two-step schemes, since the higher numerical effort for higher deterministic order pays off only if the noise is *very* small. Our interest in stochastic multi-step methods originates from applications with small noise in circuit simulation, where especially the backward differential equation (BDF) and the trapezoidal rule have proven valuable in the deterministic case.

Several variable step-size strategies for SDEs were developed during the last few years. Most of them are based on path-wise arguments and lead to path-wise different step-size sequences. The classical paper [6] proposes a path-wise strategy by comparing results of a given integration scheme with those of a higher order method. The approaches discussed in [12] and [4] utilize the comparison of two Runge-Kutta schemes of different order. In [11] conditions are provided that imply mean-square convergence of the Euler-Maruyama scheme with path-wise different step-size sequences. A different approach was developed in [8], where the authors obtain step-size sequences that are optimal for asymptotically small step-sizes.

In contrast to the above approaches we aim at an efficient estimate of the mean-square of the local error and present a strategy for controlling the step-size in the numerical integration based on this. For moderate tolerances the use of adaptively chosen steps should provide much more accurate results than the computation with the same number of constant steps. An important prerequisite for estimating local errors in context of SDEs and SDAEs is an understanding and a precise definition of local and global errors and their relation to each other. We use the notion of mean-square local and global errors and the corresponding convergence analysis developed in [3, 18]. As in [17] for the family of Euler-Maruyama schemes we propose to estimate the mean-square of the local errors by means of a number of simultaneously computed solution paths, which leads to a step-size sequence that is identical for all paths.

The paper is organized in the following way. In Section 2 we give a brief introduction into the ideas and techniques used in the analysis of SDEs, in particular the Itô calculus. To help those who are not familiar with the concept of SDEs we will concentrate

on the main ideas and avoid technicalities. Stochastic linear multi-step methods and their convergence properties are recapitulated in Section 3. The analysis of the local errors is done in Section 4. In Section 5 we provide a reliable error estimate for systems with small noise and in Section 6 we adapt this technique for SDAEs. We present the step-size control algorithm in Section 7 and finally, report on numerical experiments illustrating the performance of the mesh adaptation in Section 8.

We will repeatedly refer to results and formulas from Part I of this paper and cite formula (s.x) from Part I by (I.s.x).

2 Itô calculus and SDEs.

SDEs formally arise when white noise of a certain intensity is added to the right hand side of an ODE. In the engineering literature the resulting system is often simply written as

$$x'(t) = f(t,x(t)) + g(t,x(t))\ \xi(t), \tag{2.1}$$

where the time t varies in a compact interval $\mathscr{J} = [0,t_{end}]$. The drift and diffusion functions are given as $f,g : \mathscr{J} \times \mathbb{R}^n \to \mathbb{R}^n$, and ξ is a scalar Gaussian white noise process. Considering a number m of independent white noise sources leads to

$$x'(t) = f(t,x(t)) + \sum_{r=1}^{m} g_r(t,x(t))\ \xi_r(t), \tag{2.2}$$

where the diffusion function is given as $G = (g_1,\ldots,g_m) : \mathscr{J} \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ and $\xi = (\xi_1,\ldots,\xi_m)^T$ is an m-dimensional Gaussian white noise process.

Gaussian white noise is a model of random fluctuations, comprising many small independent random events like thermal noise in physics or volatility in finance. It is a stochastic process depending on time t and chance ω. The latter argument is usually omitted in the notation. The noise is called white since the spectral density of the process is constant, meaning that all frequencies equally contribute to the process. In the time domain Gaussian white noise is characterized by Gaussian distributed and delta correlated random values $\xi(t)$. Realizations or paths $\xi(\cdot,\omega)$ of Gaussian white noise turn out to be highly irregular and nowhere continuous. A serious mathematical description of the problems (2.1) or (2.2) begins by introducing the Brownian motion or the Wiener Process that is caused by integrating the white noise "$W(t) = \int_0^t \xi(s)\mathrm{d}s = \int_0^t \mathrm{d}W(s)$".

The (scalar) Brownian motion W is defined as a stochastic process given on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ with a filtration $\{\mathscr{F}_t\}_{t\geq 0}$. $\mathscr{F}_t$ can be seen as the σ-algebra of those events that are observable with the information available at time t. The Brownian motion is characterized by $W(0) = 0$ and independent non-overlapping increments that are Gaussian distributed, $W(t+h) - W(t) \sim N(0,h)$. Typical realizations $W(\cdot,\omega)$, also called paths, are Hölder-continuous with exponents smaller than $\frac{1}{2}$, nowhere differentiable, not of bounded variation and with quadratic variation $\langle W\rangle_t = t$ (see, e.g., [9]). Multi-dimensional Wiener processes consist of independent scalar Wiener processes.

Problem (2.2) is then understood as a stochastic integral equation

$$X(t) = X_0 + \int_0^t f(s,X(s))\,\mathrm{d}s + \sum_{r=1}^{m} \int_0^t g_r(s,X(s))\,\mathrm{d}W_r(s)\,, \tag{2.3}$$

and is abbreviated as $\mathrm{d}X(t) = f(t,X(t))\mathrm{d}t + G(t,X(t))\mathrm{d}W(t)$ with $X(0) = X_0$, where the second integral is a stochastic integral, and W denotes an m-dimensional Wiener process (or Brownian motion).

Due to the unbounded variation of paths of the Wiener process one cannot interpret the stochastic integrals in (2.3) as ordinary Riemann-Stieltjes integrals, but need to apply a special stochastic calculus. In this paper we use the Itô calculus. First, we recall the definition of an Itô integral.

Let W be a Wiener process and Y be an adapted stochastic process on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with filtration $\{\mathscr{F}_t\}_{t\in\mathscr{J}}$. The stochastic process Y is called adapted if $Y(t)$ is $\mathscr{F}_t$-measurable. We consider partitions $0 = t_0 < t_1 < \cdots < t_N = t_{end}$ of the interval $\mathscr{J}$ and set $\mathbf{h} = \max_{0\le i\le N}(t_i - t_{i-1})$. Then the Itô integral is defined as the limit of non-anticipative Riemann sums for $\mathbf{h} \to 0$ in $L_2(\Omega,\mathbb{R})$. The integrand is evaluated at left endpoints,

$$\int_0^t Y(s)\mathrm{d}W(s) = \lim_{\mathbf{h}\to 0} \sum_i Y(t_i)[W(t_{i+1}) - W(t_i)] \quad \text{in } L_2(\Omega,\mathbb{R})\,.$$

Itô integrals satisfy the usual linearity and additivity properties of integrals, moreover, zero expectation, $\mathbb{E}\int_0^t Y(s)\mathrm{d}W(s) = 0$, and Itô isometry,

$$\mathbb{E}\Big(\int_0^t Y(s)\mathrm{d}W(s)\Big)^2 = \int_0^t \mathbb{E}(Y(s))^2\mathrm{d}s\,.$$

With this interpretation of the stochastic integrals the problem (2.3) is called an Itô SDE. Let the drift function f and the diffusion function G be continuous and Lipschitz continuous with respect to the second variable, and assume that X_0 is a given $\mathscr{F}_0$-measurable initial value, independent of the Wiener process with finite second moments, i.e., square integrable. Then there exists a path-wise unique strong solution $X : \mathscr{J} \times \Omega \to \mathbb{R}^n$ of (2.3) with square integrable values $X(t) \in L_2(\Omega,\mathbb{R}^n)$.

The value of the solution process at fixed time t is a random variable $X(t,\cdot) = X(t)$ whose argument ω is usually dropped. For a fixed sample ω representing a fixed realization of the driving Wiener process, the function $X(\cdot,\omega)$ is called a realization or a path of the solution. Due to the influence of the Gaussian white noise, typical paths of the solution are nowhere differentiable. While using the sloppy formulations (2.1), (2.2) one always should keep this in mind.

The unbounded variation of the Wiener paths leads to a modified chain rule, the famous Itô formula. Let X be a solution of (2.3). Let the function $y : \mathscr{J} \times \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable and additionally, have continuous second partial

derivatives with respect to the second variable. Then the Itô formula reads

$$y(t,X(t)) - y(0,X(0)) = \int_0^t \Big(y'_t + y'_x f + \frac{1}{2}\sum_{r=1}^{m} y''_{xx}[g_r,g_r] \Big)(s,X(s))\,\mathrm{d}s \quad (2.4)$$
$$+ \sum_{r=1}^{m} \int_0^t y'_x g_r(s,X(s))\,\mathrm{d}W_r(s), \quad t \in \mathscr{J}.$$

The term involving the second derivative of y is sometimes called Itô correction. We refer, e.g., to [1] for more details.

For further reference we state the following notations and definitions: We denote by $|\cdot|$ the Euclidian norm in $\mathbb{R}^n$ and by $\|\cdot\|$ the corresponding induced matrix norm. The mean-square norm of a vector valued square integrable random variable $Z \in L_2(\Omega,\mathbb{R}^n)$ will be denoted by $\|Z\|_{L_2} := (\mathbb{E}|Z|^2)^{1/2}$.

Let us denote by $C^{s-1,s}$ the class of all functions $y : \mathscr{J} \times \mathbb{R}^n \to \mathbb{R}^n$ having continuous partial derivatives up to order $s-1$ and continuous partial derivatives of order s with respect to the second variable. Moreover, let C^K be the class of functions $y : \mathscr{J} \times \mathbb{R}^n \to \mathbb{R}^n$ satisfying the linear growth condition of the form

$$|y(t,x)| \le K(1+|x|^2)^{\frac{1}{2}}, \qquad \forall t \in \mathscr{J},\ x \in \mathbb{R}^n\,. \quad (2.5)$$

Furthermore, we introduce the notation

$$I^{t,t+h}_{r_1,r_2,\ldots,r_j}(y) = \int_t^{t+h} \int_t^{s_1} \cdots \int_t^{s_{j-1}} y(s_j,X(s_j))\,\mathrm{d}W_{r_1}(s_j)\cdots\mathrm{d}W_{r_j}(s_1), \quad (2.6)$$

where $r_i \in \{0,1,\ldots,m\}$ and $\mathrm{d}W_0(s) = \mathrm{d}s$, for general multiple stochastic (Itô) integrals. If $y \equiv 1$ we write $I^{t,t+h}_{r_1,r_2,\ldots,r_j}$. Note that the integral $I^{t,t+h}_r$ is simply the increment $W_r(t+h) - W_r(t)$ of the scalar Wiener process W_r. To estimate the multiple integrals (2.6) we will use the following lemma (see Lemma 2.1 and 2.2 in [14]).

Lemma 2.1 *For any function $y \in C^K$ and any $t \in \mathscr{J}$, $h > 0$, such that $t+h \in \mathscr{J}$, we have*

$$\mathbb{E}(I^{t,t+h}_{r_1\ldots r_j}(y)|\mathscr{F}_t) = 0, \text{ if } r_i \neq 0 \text{ for some } i \in \{1,\ldots,j\}, \quad (2.7)$$
$$\|I^{t,t+h}_{r_1,\ldots,r_j}(y)\|_{L_2} = O(h^{l_1+l_2/2}), \quad (2.8)$$

where l_1 is the number of zero indices r_i and l_2 the number of non-zero indices r_i.

The present paper is devoted to SDEs and SDAEs with *small* noise. Following [15] we indicate the size of the noise by a small factor ε in front of the diffusion coefficient, i.e., $G = \varepsilon\widehat{G}$. Thus, we consider SDEs with small noise of the type

$$X(s)\Big|_0^t = \int_0^t f(s,X(s))\mathrm{d}s + \varepsilon\int_0^t \widehat{G}(s,X(s))\mathrm{d}W(s), \quad X(0) = X_0. \quad (2.9)$$

We stress that the presence of the small parameter ε in (2.9) is not required to design the respective numerical schemes. It is used here, because it is crucial in the discussion of the *errors* of the involved numerical schemes.

3 Stochastic linear multi-step schemes: Mean-square consistency, stability and convergence.

In the literature on numerical methods for SDEs two concepts of convergence are discussed, weak and strong convergence. Weak convergence relates to Monte-Carlo methods and is mainly concerned with statistical properties of the solution of the SDEs, such as moments of the solution process like expectation and variance. The term strong convergence is often used synonymously for the expression mean-squares convergence, i.e., convergence in the norm $\|\cdot\|_{L_2}$. It is used when one is interested in paths of solutions. The related applications originate e.g., from the computation of phase noise in circuit simulation [5,20], or pricing of American options in finance. In this paper, we investigate the mean-square convergence.

¿From the variety of stochastic linear multi-step schemes discussed in [3,18], we consider those schemes which only include information from the increments of the driving Wiener process. We are not interested in schemes that include higher stochastic integrals because they are not efficient for problems with small noise. Moreover we restrict the subsequent analysis to two-step schemes. This simplifies the notation but is not a serious restriction. Stochastic schemes with higher deterministic order of convergence than two are only efficient in case of extremely small noise.

Let

$$\Gamma := \{0 = t_0 < t_1 < \cdots < t_i < \cdots < t_N = t_{end}\} \tag{3.1}$$

be a deterministic partition of the interval $\mathscr{J}$. We denote the length of the subinterval $[t_{i-1}, t_i]$ by $h_i = t_i - t_{i-1}$, $i = 1, \ldots, N$. Let $\mathbf{h}$ be the maximal step-size of Γ, $\mathbf{h} := \max_{1 \le i \le N} h_i$ and $\kappa_i = h_i / h_{i-1}$, $i = 2, \ldots, N$, be the step-size ratio. A linear two-step Maruyama method takes the form

$$\sum_{j=0}^{2} \alpha_{j,i} X_{i-j} = h_i \sum_{j=0}^{2} \beta_{j,i} f(t_{i-j}, X_{i-j}) + \sum_{j=1}^{2} \gamma_{j,i} \sum_{r=1}^{m} g_r(t_{i-j}, X_{i-j})\, I_r^{t_{i-j}, t_{i-j+1}}, \tag{3.2}$$

where $i = 2, \ldots, N$. Generally, X_i denotes an approximation to the solution $X(t_i)$. We assume that the initial values $X_0, X_1 \in L_2(\Omega, \mathbb{R}^n)$ are given such that X_i is $\mathscr{F}_{t_i}$-measurable for $i = 0, 1$. Let the coefficients of the scheme be normalized in such a way that $\alpha_{0,i} = 1$ for all i.

As in the deterministic case, usually only $X_0 = X(0)$ is given by the initial value problem and the value X_1 needs to be computed numerically. This can be done by a suitable one-step method chosen in such a way that it provides X_1 up to the desired asymptotical accuracy. The coefficients $\alpha_{j,i}$, $\beta_{j,i}$, $\gamma_{j,i}$ depend on the ratio $\kappa_i = h_i / h_{i-1}$. We emphasize that an explicit discretization is used for the diffusion term. For $\beta_{0,i} = 0$ the method (3.2) is explicit, otherwise it is drift-implicit.

We consider mean-square convergence of the method (3.2) in the sense discussed in Milstein and others [2,3,13–15,22]. The scheme (3.2) is called *mean-square convergent with order* γ ($\gamma > 0$) if the *global error*, $X(t_i) - X_i$, satisfies

$$\max_{i=0,\ldots,N} \|X(t_i) - X_i\|_{L_2} \le C \cdot \mathbf{h}^{\gamma},$$

with a constant $C > 0$ independent of $\mathbf{h}$. Note that in the literature the term *strong* convergence is sometimes used synonymously for this property.

As in Part I and following [3,18], we define the *local error* as the defect that is obtained when the exact solution values are inserted into the numerical scheme, i.e., it is given by

$$\begin{aligned} L_i := & \sum_{j=0}^{2} \alpha_{j,i} X(t_{i-j}) - h_i \sum_{j=0}^{2} \beta_{j,i} f(t_{i-j}, X(t_{i-j})) \\ & - \sum_{j=1}^{2} \gamma_{j,i} \sum_{r=1}^{m} g_r(t_{i-j}, X(t_{i-j}))\, I_r^{t_{i-j},t_{i-j+1}}, \quad i = 2,\dots,N, \\ L_i := & X(t_i) - X_i, \qquad\qquad i = 0,1. \end{aligned} \tag{3.3}$$

Note that L_i is now a vector valued random variable as $X(t_i)$ and X_i. In order to exploit the adaptivity and independence of the stochastic terms arising on disjoint subintervals, we represent the local error in the form

$$L_i = R_i + S_i =: R_i + S_{1,i} + S_{2,i-1}, \quad i = 2,\dots,N, \tag{3.4}$$

where each $S_{j,i}$ is F_{t_i} measurable with $\mathbb{E}(S_{j,i}|\mathscr{F}_{t_{i-1}}) = 0$, cf. [3]. The representation (3.4) is not unique. One useful choice is provided by

$$R_i = \mathbb{E}(L_i|\mathscr{F}_{t_{i-2}}), \quad S_{2,i-1} = \mathbb{E}(L_i - R_i|\mathscr{F}_{t_{i-1}}), \quad S_{1,i} = L_i - R_i - S_{2,i-1}.$$

Here, in the hypothetical case that $L_i = c_0 I_r^{t_{i-1},t_i} + c_1 I_r^{t_{i-2},t_{i-1}} + c_2$ holds, we have $R_i = c_2$, $S_{2,i-1} = c_1 I_r^{t_{i-2},t_{i-1}}$ and $S_{1,i} = c_0 I_r^{t_{i-1},t_i}$.

Mean-square convergence is implied by local properties of the scheme (3.2) by means of numerical stability in the mean-square sense. Numerical stability estimates the influence of any perturbations of the right-hand side of the discrete scheme on the global solution of that discrete scheme. Taking the local errors as special perturbations and applying the numerical stability estimate to them results in the following convergence theorem, which is a special case of Theorem 3.2 in [18]:

Theorem 3.1 *Let the coefficients $\alpha_{j,i} = \alpha_j(\kappa_i)$, $j = 0,1,2$ of the stochastic linear two-step scheme (3.2) be continuous in a neighborhood of $\kappa_i = 1$ and let the coefficients of the related constant step-size scheme satisfy Dahlquist's root condition. Moreover, let us assume that the coefficient functions f and g_r, $r = 1,\dots,m$, are globally Lipschitz continuous with respect to their second argument. Then there exists constants κ, K $(\kappa < 1 < K)$, $a \geq 0$, $h_{max} > 0$ and a stability constant $S > 0$ such that the following statement holds for each grid Γ having the property $\mathbf{h} := \max_{i=1,\dots,N} h_i \leq h_{max}$, $\mathbf{h} \cdot N \leq a \cdot (T - t_0)$ and $\kappa \leq h_i/h_{i-1} \leq K$ for all i:*

For all representations (3.4) of the local error L_i the following estimate of the global error holds:

$$\max_{i=0,\dots,N} \|X_i - X(t_i)\|_{L_2} \leq S \Big\{ \max_{i=0,1} \|L_i\|_{L_2} + \max_{i=2,\dots,N} \Big(\frac{\|R_i\|_{L_2}}{\mathbf{h}} + \frac{\|S_i\|_{L_2}}{\mathbf{h}^{1/2}} \Big) \Big\}. \tag{3.5}$$

Remark 3.1 The coefficients $\alpha_{j,i}$ of the stochastic analogues of the two-step Adams-Bashforth formula and the trapezoidal rule do not depend on step-size ratios and hence these schemes are stable for any step-size sequence. The stochastic variant of the two-step BDF is numerically stable (as is wellknown from the deterministic case) if and only if

$$0 < \kappa_i \leq K < 1+\sqrt{2} \quad \text{for } i \geq 2. \tag{3.6}$$

In the following we assume that the requirements specified in Theorem 3.1 are satisfied. For the proof of the convergence of order γ, it is now sufficient to find a representation (3.4) of the local error L_i such that

$$\|R_i\|_{L_2} \leq c_R \cdot h_i^{\gamma+1}, \tag{3.7}$$

and

$$\|S_i\|_{L_2} \leq c_S \cdot h_i^{\gamma+\frac{1}{2}}, \quad i = 1,\ldots,N, \tag{3.8}$$

with constants $c_R > 0$ and $c_S > 0$ independent of h_i. Together, (3.7) and (3.8) imply the estimates

$$\|\mathbb{E}(L_i|\mathscr{F}_{t_{i-2}})\|_{L_2} = O(h_i^{\gamma+1}), \qquad \|L_i\|_{L_2} = O(h_i^{\gamma+\frac{1}{2}}), \quad i = 1,\ldots,N,$$

see Lemma 2.8 in [3], known as consistency in the mean and consistency in the mean-square sense, respectively. We observe that in case of the k-step schemes the conditional mean has to be taken with respect to the σ-algebra $\mathscr{F}_{t_{i-k}}$.

4 Analysis of the local error.

We shall now analyze the local error terms by means of appropriate Itô-Taylor expansions, where we take special care in separating the multiple stochastic integrals over the different subintervals of integration. To this aim we introduce the following operators Λ_0 and Λ_r, $r = 1,\ldots,m$, defined on $C^{1,2}$ and $C^{0,1}$, respectively:

$$\Lambda_0 y = y'_t + y'_x f + \frac{1}{2}\sum_{r=1}^{m}\sum_{i,j=1}^{n} y''_{x_i x_j} g_{ri} g_{rj}, \quad \Lambda_r y = y'_x g_r, \;\; r = 1,\ldots,m, \tag{4.1}$$

and refer the reader to the notation for multiple Wiener integrals introduced in (2.6). Using these operators the Itô formula for a function y in $C^{1,2}$ and the solution X of (2.3) reads

$$y(t,X(t)) = y(t_0,X(t_0)) + I_0^{t_0,t}(\Lambda_0 y) + \sum_{r=1}^{m} I_r^{t_0,t}(\Lambda_r y), \quad t \in \mathscr{J}. \tag{4.2}$$

Before exploiting the effect of the small parameter ε in the expansions of the local error, we recall the notation $g_r = \varepsilon \hat{g}_r$ and introduce operators $\Lambda_0^f, \hat{\Lambda}_0$ and $\hat{\Lambda}_r$, $r = 1,\ldots,m$, defined on $C^{1,2}$ and $C^{0,1}$, respectively, by

$$\Lambda_0^f y := y'_t + y'_x f, \quad \hat{\Lambda}_0 y := \frac{1}{2}\sum_{r=1}^{m}\sum_{i,j=1}^{n} y''_{x_i x_j} \hat{g}_{ri}\hat{g}_{rj}, \quad \hat{\Lambda}_r y := y'_x \hat{g}_r\,. \tag{4.3}$$

In terms of the original definition (4.1) we have

$$\Lambda_0 y = \Lambda_0^f y + \varepsilon^2 \hat{\Lambda}_0 y, \qquad \Lambda_r y = \varepsilon \hat{\Lambda}_r y. \tag{4.4}$$

After straightforward but lengthy computations one arrives at the following result, which we cite from [18, L.4.4 and C.4.5]. To be consistent with the representation there we use the same notation.

Lemma 4.1 *Let us assume that the coefficients $f, \hat{g}_r, r = 1, \dots, m$, of the small noise SDE (2.9), as well as $\Lambda_0^f f = f'_x f + f'_t$ belong to the class $C^{1,2}$ with $\Lambda_0 f, \Lambda_0 \hat{g}_r, \hat{\Lambda}_r f$, $\hat{\Lambda}_q \hat{g}_r, \Lambda_0 \Lambda_0^f f$, $\hat{\Lambda}_r \Lambda_0^f f \in C^K$ for $r, q = 1, \dots, m$. Let the stochastic two-step scheme (3.2) satisfy the consistency conditions*

$$\begin{aligned} &\sum_{j=0}^{2} \alpha_{j,i} = 0, \qquad && \alpha_{0,i} + \frac{1}{\kappa_i}(\alpha_{0,i} + \alpha_{1,i}) = \sum_{j=0}^{2} \beta_{j,i}\ , \\ &\alpha_{0,i} = \gamma_{1,i}, \qquad && \alpha_{0,i} + \alpha_{1,i} = \gamma_{2,i}, \end{aligned} \tag{4.5}$$

for all i. Then the local error of the method (3.2) for the small noise SDE (2.9) allows the representation

$$L_i = R_i^{\diamond} + S_{1,i}^{\diamond} + S_{2,i-1}^{\diamond}, \quad i = 2, \dots, N, \tag{4.6}$$

where $R_i^{\diamond}, S_{j,i}^{\diamond}, j = 1, 2$ are $\mathscr{F}_{t_i}$-measurable with $\mathbb{E}(S_{j,i}^{\diamond} | \mathscr{F}_{t_{i-1}}) = 0$, and

$$\begin{aligned} R_i^{\diamond} = &\Big[(\frac{1}{\kappa_i^2} + \frac{2}{\kappa_i} + 1)\alpha_{0,i} + \frac{1}{\kappa_i^2}\alpha_{1,i} - (\frac{2}{\kappa_i} + 2)\beta_{0,i} - \frac{2}{\kappa_i}\beta_{1,i} \Big] \frac{h_i^2}{2} \Lambda_0^f f(t_{i-2}, X(t_{i-2})) \\ &+ \tilde{R}_i^{\diamond f} + \varepsilon^2 \hat{R}_i^{\circ}, \end{aligned}$$

with

$$\begin{aligned} \tilde{R}_i^{\diamond f} = &(\alpha_{0,i} - 2\beta_{0,i}) \frac{h_i^2}{2} I_0^{t_{i-2}, t_{i-1}}(\Lambda_0 \Lambda_0^f f) \\ &+ (\alpha_{0,i} - \beta_{0,i} - \beta_{1,i}) h_i I_{00}^{t_{i-2}, t_{i-1}}(\Lambda_0 \Lambda_0^f f) - \beta_{0,i} h_i I_{00}^{t_{i-1}, t_i}(\Lambda_0 \Lambda_0^f f) \\ &+ (\alpha_{0,i} + \alpha_{1,i}) I_{000}^{t_{i-2}, t_{i-1}}(\Lambda_0 \Lambda_0^f f) + \alpha_{0,i} I_{000}^{t_{i-1}, t_i}(\Lambda_0 \Lambda_0^f f), \end{aligned} \tag{4.7}$$

$$\|\hat{R}_i^{\circ}\|_{L_2} = O(h_i^2), \tag{4.8}$$

$$\|S_{1,i}^{\diamond}\|_{L_2} = O(\varepsilon^2 h_i + \varepsilon h_i^{3/2}), \qquad \|S_{2,i-1}^{\diamond}\|_{L_2} = O(\varepsilon^2 h_i + \varepsilon h_i^{3/2}). \tag{4.9}$$

Corollary 4.1 *Let the coefficients $f, \hat{g}_r, r = 1, \dots, m$, of the SDE (2.9) satisfy the assumptions of Lemma 4.1 and suppose that they are Lipschitz continuous with respect to their second variable. Let us assume that the stochastic linear two-step scheme with variable step-size (3.2) is stable, its coefficients satisfy the consistency conditions (4.5) and*

$$\left(\frac{1}{\kappa_i^2} + \frac{2}{\kappa_i} + 1\right) \alpha_{0,i} + \frac{1}{\kappa_i^2}\ \alpha_{1,i} - \left(\frac{2}{\kappa_i} + 2\right) \beta_{0,i} - \frac{2}{\kappa_i}\ \beta_{1,i} = 0 \tag{4.10}$$

holds. Then the global error of the scheme (3.2) applied to solve (2.9) shows the following asymptotic behavior:

$$\max_{i=0,\ldots,N} \|X(t_i) - X_i\|_{L_2} = O(\mathbf{h}^2 + \varepsilon\mathbf{h} + \varepsilon^2\mathbf{h}^{1/2}) + O(\max_{i=0,1} \|X(t_i) - X_i\|_{L_2}).$$

Consequently, starting from the deterministic linear two-step schemes of order two or higher and choosing the parameters $\gamma_{1,i}, \gamma_{2,i}$ according to $\gamma_{1,i} = \alpha_{0,i}$ and $\gamma_{2,i} = \alpha_{0,i} + \alpha_{1,i} = -\alpha_{2,i}$, the global error of the resulting two-step Maruyama scheme can be estimated as $O(\mathbf{h}^2 + \varepsilon\mathbf{h} + \varepsilon^2\mathbf{h}^{1/2})$. More generally, stochastic linear multi-step Maruyama schemes with deterministic order $p \geq 2$ have a global error $O(\mathbf{h}^p + \varepsilon\mathbf{h} + \varepsilon^2\mathbf{h}^{1/2})$ provided that the coefficients $f, \hat{g}_r$ are sufficiently smooth.

Example 4.1 For illustration, we present stochastic variants of three well known deterministic methods of order two. The stochastic trapezoidal rule, also known as stochastic θ-method with $\theta = \frac{1}{2}$, is the one-step scheme with the coefficients $\alpha_{0,i} = 1$, $\alpha_{1,i} = -1$, $\beta_{0,i} = \beta_{1,i} = \frac{1}{2}$, $\gamma_{1,i} = 1$, $\alpha_{2,i} = \beta_{2,i} = \gamma_{2,i} = 0$,

$$X_i - X_{i-1} = h_i \frac{1}{2}\big(f(t_i, X_i) + f(t_{i-1}, X_{i-1})\big) + \sum_{r=1}^{m} g_r(t_{i-1}, X_{i-1})\, I_r^{t_{i-1},t_i}. \tag{4.11}$$

The Adams-Bashforth-Maruyama scheme is given by

$$X_i - X_{i-1} = h_i \left(\frac{\kappa_i + 2}{2} f(t_{i-1}, X_{i-1}) - \frac{\kappa_i}{2} f(t_{i-2}, X_{i-2}) \right) + \sum_{r=1}^{m} g_r(t_{i-1}, X_{i-1})\, I_r^{t_{i-1},t_i} \tag{4.12}$$

with $\alpha_{0,i} = 1$, $\alpha_{1,i} = -1$, $\beta_{0,i} = \frac{\kappa_i+2}{2}$, $\beta_{1,i} = -\frac{\kappa_i}{2}$, $\gamma_{1,i} = 1$ and $\beta_{2,i} = \alpha_{2,i} = \gamma_{2,i} = 0$. The stochastic two-step backward differentiation formula, BDF_2, takes the form

$$\begin{aligned} X_i - \frac{(\kappa_i+1)^2}{2\kappa_i+1} X_{i-1} + \frac{\kappa_i^2}{2\kappa_i+1} X_{i-2} &= h_i \frac{\kappa_i+1}{2\kappa_i+1} f(t_i, X_i) \\ &+ \sum_{r=1}^{m} g_r(t_{i-1}, X_{i-1})\, I_r^{t_{i-1},t_i} - \frac{\kappa_i^2}{2\kappa_i+1} \sum_{r=1}^{m} g_r(t_{i-2}, X_{i-2})\, I_r^{t_{i-2},t_{i-1}}. \end{aligned} \tag{4.13}$$

Here, one has $\alpha_{0,i} = 1$, $\alpha_{1,i} = -\frac{(\kappa_i+1)^2}{2\kappa_i+1}$, $\alpha_{2,i} = \frac{\kappa_i^2}{2\kappa_i+1}$, $\beta_{0,i} = \frac{\kappa_i+1}{2\kappa_i+1}$, $\beta_{1,i} = \beta_{2,i} = 0$, and $\gamma_{1,i} = 1$, $\gamma_{2,i} = -\frac{\kappa_i^2}{2\kappa_i+1}$.

5 Local error estimates.

This section consists of two parts. In Section 5.1 we investigate the structure of the local error to find out under which condition the deterministic terms still dominate the behavior and therefore, order two can be observed. In Section 5.2 we attempt to apply techniques developed in Part I of this paper to find estimates for the local error in case of small noise. It turns out that this is possible only under further restrictions concerning the size of the noise in relation to the size of the step-size.

5.1 Structure of the local error.

A very crucial point in the numerics of SDEs is that the contributions of the deterministic and stochastic parts of the local error to the global error are qualitatively different. We may and we have to use the fact, that the accumulated stochastic parts of the local error contribute weaker to the global error than the accumulated deterministic parts. The sum of N independent centered random values with mean-square norm 1 has the mean-square norm $\sqrt{N}$, whereas the sum of N deterministic values with norm 1 may be equal to N. This is reflected in the stability estimate (3.5). Therefore, a natural approach consists in controlling the local error contribution

$$\eta_i := \|R_i\|_{L_2}/h_i + \|S_i\|_{L_2}/h_i^{1/2}, \tag{5.1}$$

where $L_i = R_i + S_i$. Recall that the representation $L_i = R_i + S_i$ of the local error is not unique and therefore, an important issue is to find an adequate representation that is on the one hand as simple as possible and requires only low smoothness, and on the other hand gives satisfactory estimates. We will use the representation (4.6),

$$L_i = R_i^{\diamond} + S_{1,i}^{\diamond} + S_{2,i-1}^{\diamond}, \;\; i = 2, \dots, N,$$

given in Lemma 4.1.

Let us consider a stochastic linear two-step scheme (3.2) with deterministic order two and suppose that the smoothness assumptions of Corollary 4.1 are satisfied. Then, using notation from Lemma 4.1, we have

$$\begin{aligned} R_i^{\diamond} &= \tilde{R}_i^{\diamond f} + \varepsilon^2 \hat{R}_i^{\diamond}, \\ \|\hat{R}_i^{\diamond}\|_{L_2} &= O(h_i^2), \qquad \|S_{1,i}^{\diamond} + S_{2,i-1}^{\diamond}\|_{L_2} = O(\varepsilon^2 h_i + \varepsilon h_i^{3/2}). \end{aligned}$$

This gives

$$\|R_i^{\diamond}\|_{L_2}/h_i = \|\tilde{R}_i^{\diamond f}\|_{L_2}/h_i + O(\varepsilon^2 h_i), \quad \|S_i^{\diamond}\|_{L_2}/h_i^{1/2} = O(\varepsilon^2 h_i^{1/2} + \varepsilon h_i).$$

A more detailed investigation shows that

$$\tilde{R}_i^{\diamond f} = c_i \cdot h_i^3 \Lambda_0^f \Lambda_0^f f(t_i, X(t_i)) + \text{higher order terms},$$

where c_i is the deterministic error constant. The L_2-norm of the higher order terms can be estimated by $h_i^3 \cdot O(h_i + \varepsilon h_i^{1/2} + \varepsilon^2)$ provided that $\Lambda_0^f \Lambda_0^f f$ is sufficiently smooth, or more precisely, belongs to the class $C^{1,2}$ with derivatives in C^K. Also, changing the evaluation point from $(t_i, X(t_i))$ to $(t_{i-1}, X(t_{i-1}))$ alters only higher order terms. Consequently, the local error term $c_i \cdot h_i^3 \Lambda_0^f \Lambda_0^f f(t_i, X(t_i))$ is responsible for the global error term of order $O(\mathbf{h}^2)$. Note that the expression

$$\begin{aligned} \Lambda_0^f \Lambda_0^f f(t_i, X(t_i)) &= \Lambda_0^f (f_t' + f_x' f)(t_i, X(t_i)) \\ &= \big((f_t' + f_x' f)_t' + (f_t' + f_x' f)_x' f \big)(t_i, X(t_i)) \end{aligned}$$

becomes $x'''(t_i)$ in the limit $\varepsilon \to 0$.

Now, we ask the question, for which choices of ε and $\mathbf{h}$ the $O(\mathbf{h}^2)$ term dominates the term $O(\varepsilon\mathbf{h}+\varepsilon^2\mathbf{h}^{1/2})$, and so we can still observe the $O(\mathbf{h}^2)$ behavior. Clearly, both these terms depend on the actual coefficients $f,\hat{g}_r, r=1,\ldots,m$, of the SDE (2.9) and their derivatives. Assuming moderate function values, the term $O(\mathbf{h}^2)$ dominates $O(\varepsilon\mathbf{h}+\varepsilon^2\mathbf{h}^{1/2})$, if $\mathbf{h}^2 \gg \varepsilon^2\mathbf{h}^{1/2}$, i.e., $\mathbf{h} \gg \varepsilon^{4/3}$, and $\mathbf{h}^2 \gg \varepsilon\mathbf{h}$, i.e., $\mathbf{h} \gg \varepsilon$. Obviously, the second condition is stronger, in general. Summarizing, we can expect to observe the order two behavior of the global error if $\mathbf{h} \gg \varepsilon$. An example illustrating this statement in context of constant step-sizes can be found in [3].

In the following, we assume that the step-sizes used in the computation and the noise are properly related, the noise being small enough with $\mathbf{h} \gg \varepsilon$.

5.2 Estimation of the dominant error term via defect evaluation.

We are now able to adapt the techniques developed in Part I of this paper [19] to small noise SDEs. Again, we consider two different multi-step schemes, the basic one (3.2) which we identify with the solver scheme, and an auxiliary one which is used to define the defect for the estimation of the local discretization error (3.3) of the basic scheme. As in Section 3 we restrict the exposition here to one or two-step methods as solver schemes and two-step methods as auxiliary schemes. Two-step auxiliary schemes will turn out to be suitable to provide error estimates for stochastic multi-step schemes with deterministic order two. However, the principles presented in this section are applicable also to handle more general multi-step schemes.

We consider an auxiliary scheme of the form

$$\begin{aligned}\sum_{j=0}^{2} \bar{\alpha}_{j,i}\bar{X}_{i-j} = h_i \sum_{j=0}^{2} \bar{\beta}_{j,i} f(t_{i-j},\bar{X}_{i-j}) \\ + \sum_{j=1}^{2} \bar{\gamma}_{j,i} \sum_{r=1}^{m} g_r(t_{i-j},\bar{X}_{i-j})\, I_r^{t_{i-j},t_{i-j+1}}, \quad i=2,\ldots,N,\end{aligned} \tag{5.2}$$

with given initial values $\bar{X}_0 = X_0, \bar{X}_1 \in L_2(\Omega,\mathbb{R}^n)$ such that $\bar{X}_i$ is $\mathscr{F}_{t_i}$-measurable for $i=0,1$. Let the coefficients of the scheme be normalized in such a way that $\bar{\alpha}_{0,i}=1$. The local truncation error of (5.2), $\bar{L}_i$, is again, defined as the defect that results from inserting the exact solution values into the numerical scheme (5.2).

As in Part I we consider the defect obtained by substituting the approximations X_i computed from (3.2) into the scheme (5.2),

$$\begin{aligned}D_i := \sum_{j=0}^{2} \bar{\alpha}_{j,i}X_{i-j} - h_i \sum_{j=0}^{2} \bar{\beta}_{j,i} f(t_{i-j},X_{i-j}) \\ - \sum_{j=1}^{2} \bar{\gamma}_{j,i} \sum_{r=1}^{m} g_r(t_{i-j},X_{i-j})\, I_r^{t_{i-j},t_{i-j+1}}, \quad i=2,\ldots,N.\end{aligned} \tag{5.3}$$

Let us denote the solutions of the schemes (3.2) and (5.2) obtained using exact starting values by $X_i^\star$ and $\bar{X}_i^\star$, respectively,

$$X_i^\star = -\sum_{j=1}^{2} \alpha_{j,i} X(t_{i-j}) + h_i \beta_{0,i} f(t_i, X_i^\star) + h_i \sum_{j=1}^{2} \beta_{j,i} f(t_{i-j}, X(t_{i-j})) + \sum_{j=1}^{2} \gamma_{j,i} \sum_{r=1}^{m} g_r(t_{i-j}, X(t_{i-j}))\, I_r^{t_{i-j},t_{i-j+1}}, \tag{5.4}$$

$$\bar{X}_i^\star = -\sum_{j=1}^{2} \bar{\alpha}_{j,i} X(t_{i-j}) + h_i \bar{\beta}_{0,i} f(t_i, \bar{X}_i^\star) + h_i \sum_{j=1}^{2} \bar{\beta}_{j,i} f(t_{i-j}, X(t_{i-j})) + \sum_{j=1}^{2} \bar{\gamma}_{j,i} \sum_{r=1}^{m} g_r(t_{i-j}, X(t_{i-j}))\, I_r^{t_{i-j},t_{i-j+1}}, \tag{5.5}$$

where $i = 2, \ldots, N$. For explicit schemes ($\beta_{0,i} = 0$ and $\bar{\beta}_{0,i} = 0$) we immediately have

$$L_i = X(t_i) - X_i^\star, \qquad \bar{L}_i = X(t_i) - \bar{X}_i^\star,$$

but in general,

$$\begin{aligned} L_i &= X(t_i) - X_i^\star - h_i \beta_{0,i} \big(f(t_i, X(t_i)) - f(t_i, X_i^\star) \big) \\ &= \big(I - h_i \beta_{0,i} J_i \big) (X(t_i) - X_i^\star), \end{aligned} \tag{5.6}$$

and

$$\begin{aligned} \bar{L}_i &= X(t_i) - \bar{X}_i^\star - h_i \bar{\beta}_{0,i} \big(f(t_i, X(t_i)) - f(t_i, \bar{X}_i^\star) \big) \\ &= \big(I - h_i \bar{\beta}_{0,i} \bar{J}_i \big) (X(t_i) - \bar{X}_i^\star). \end{aligned} \tag{5.7}$$

Here, $J_i = \int_0^1 f_x'(t_i, sX(t_i) + (1-s)X_i^\star)\, ds$, $\bar{J}_i = \int_0^1 f_x'(t_i, sX(t_i) + (1-s)\bar{X}_i^\star)\, ds$, and f is supposed to be differentiable with respect to its second variable x.

Fully analogously to Lemma I.3.1, one can prove the following property of the defect $D_i^\star$ defined in (5.8) and obtained by substituting $X_i^\star$ into (5.2).

Lemma 5.1 *Let $f(t,x) \in C^{0,1}$ be continuous and continuously differentiable with respect to x. Let the step-size* **h** *be sufficiently small to guarantee that the matrix $\big(I - h_i \beta_{0,i} J_i\big)$ is nonsingular. Then the defect $D_i^\star$,*

$$D_i^\star := X_i^\star + \sum_{j=1}^{2} \bar{\alpha}_{j,i} X(t_{i-j}) - h_i \bar{\beta}_{0,i} f(t_i, X_i^\star) - h_i \sum_{j=1}^{2} \bar{\beta}_{j,i} f(t_{i-j}, X(t_{i-j})) + \sum_{j=1}^{2} \bar{\gamma}_{j,i} \sum_{r=1}^{m} g_r(t_{i-j}, X(t_{i-j}))\, I_r^{t_{i-j},t_{i-j+1}}, \tag{5.8}$$

satisfies

$$D_i^\star = \bar{L}_i - L_i + h_i (\bar{\beta}_{0,i} - \beta_{0,i}) J_i \big(I - h_i \beta_{0,i} J_i \big)^{-1} L_i. \tag{5.9}$$

The structure of the defect allows a special representation of the local truncation error.

Corollary 5.1 *Let us assume that Lemma 5.1 holds. Moreover, let the local errors of the schemes (3.2) and (5.2) allow representations*

$$L_i = R_i + S_i, \quad R_i = c_i h_i^{p+1} (\Lambda_0^f)^{[p]} f(t_i, X(t_i)) + R_i^\circ$$

with

$$\|R_i^\circ\|_{L_2} = O(h_i^{p+2} + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2), \quad \|S_i\|_{L_2} = O(\varepsilon h_i^{3/2} + \varepsilon^2 h_i),$$

and

$$\bar{L}_i = \bar{R}_i + \bar{S}_i, \quad \bar{R}_i = \bar{c}_i h_i^{\bar{p}+1} (\Lambda_0^f)^{[\bar{p}]} f(t_i, X(t_i)) + \bar{R}_i^\circ$$

with

$$\|\bar{R}_i^\circ\|_{L_2} = O(h_i^{\bar{p}+2} + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2), \quad \|\bar{S}_i\|_{L_2} = O(\varepsilon h_i^{3/2} + \varepsilon^2 h_i).$$

For the case $p = \bar{p}$, we additionally assume that $c_i \neq \bar{c}_i$. Then there exist $\mathscr{F}_{t_i}$-measurable variables $R_i^{\circ D^\star}$ and stochastic parts $S_i^{D^\star}$ with $\|S_i^{D^\star}\|_{L_2} = O(\varepsilon h_i^{3/2} + \varepsilon^2 h_i)$ such that

(i) for $p > \bar{p}$: $\bar{L}_i = D_i^\star + R_i^{\circ D^\star} + S_i^{D^\star}$ with $\|R_i^{\circ D^\star}\|_{L_2} = O(h_i^{\bar{p}+2} + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2)$,
(ii) for $p < \bar{p}$: $L_i = -D_i^\star + R_i^{\circ D^\star} + S_i^{D^\star}$ with $\|R_i^{\circ D^\star}\|_{L_2} = O(h_i^{p+2} + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2)$,
(iii) for $p = \bar{p}$:
$L_i = \frac{c_i}{\bar{c}_i - c_i} D_i^\star + R_i^{\circ D^\star} + S_i^{D^\star}$ with $\|R_i^{\circ D^\star}\|_{L_2} = O(h_i^{p+2} + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2)$, and
$\bar{L}_i = \frac{c_i}{\bar{c}_i - c_i} D_i^\star + R_i^{\circ D^\star} + S_i^{D^\star}$ with $\|R_i^{\circ D^\star}\|_{L_2} = O(h_i^{p+2} + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2)$.

Proof We first rewrite (5.9) by using the structure of the representations of L_i and $\bar{L}_i$ and obtain

$$D_i^* = \underbrace{\bar{L}_i}_{\bar{R}_i + \bar{S}_i} \underbrace{-L_i}_{-R_i - S_i} + \underbrace{h_i(\bar{\beta}_{0,i} - \beta_{0,i}) J_i \big(I - h_i \beta_{0,i} J_i\big)^{-1} L_i}_{\bar{\bar{R}}_i}.$$

(i) Let $p > \bar{p}$. The proof is analogous to case (ii).
(ii) Let $p < \bar{p}$. Then we have

$$\begin{aligned} D_i^\star &= -L_i + \bar{R}_i + \bar{\bar{R}}_i + \bar{S}_i, \\ L_i &= -D_i^\star + \underbrace{\bar{R}_i + \bar{\bar{R}}_i}_{R_i^{\circ D^\star}} + \underbrace{\bar{S}_i}_{S_i^{D^\star}}. \end{aligned}$$

(iii) Let $p=\bar{p}$. Then

$$D_i^* = (\bar{c}_i - c_i)h_i^{p+1}(\Lambda_0^f)^{[p]}f(t_i,X(t_i)) + \bar{R}_i^\circ - R_i^\circ + \bar{\bar{R}}_i + \bar{S}_i - S_i,$$

$$\frac{c_i}{\bar{c}_i - c_i}D_i^* = \underbrace{c_i h_i^{p+1}(\Lambda_0^f)^{[p]}f(t_i,X(t_i))}_{L_i - R_i^\circ - S_i} + \frac{c_i}{\bar{c}_i - c_i}\left(\bar{R}_i^\circ - R_i^\circ + \bar{\bar{R}}_i + \bar{S}_i - S_i\right),$$

$$\begin{aligned} L_i &= \frac{c_i}{\bar{c}_i - c_i}D_i^* + R_i^\circ + S_i - \frac{c_i}{\bar{c}_i - c_i}\left(\bar{R}_i^\circ - R_i^\circ + \bar{\bar{R}}_i + \bar{S}_i - S_i\right) \\ &= \frac{c_i}{\bar{c}_i - c_i}D_i^* + \underbrace{R_i^\circ - \frac{c_i}{\bar{c}_i - c_i}\left(\bar{R}_i^\circ - R_i^\circ + \bar{\bar{R}}_i\right)}_{R_i^{\circ D^\star}} + \underbrace{S_i - \frac{c_i}{\bar{c}_i - c_i}\left(\bar{S}_i - S_i\right)}_{S_i^{D^\star}}. \end{aligned}$$

The assertion for $\bar{L}_i$ results fully analogously.

Corollary 5.1 offers two options for designing an estimate $\mathscr{R}_i$ for the error term

$$R_i^\star = c_i h_i^{p+1}(\Lambda_0^f)^{[p]}f(t_i,X(t_i)), \tag{5.10}$$

that is assumed to dominate. According to (ii) we may choose a higher order scheme (5.2) to evaluate D_i given by (5.3) and set $\mathscr{R}_i := -D_i$. According to (iii) we may choose a scheme (5.2) with the same order $\bar{p}=p$ to evaluate D_i and set $\mathscr{R}_i := \frac{c_i}{\bar{c}_i - c_i}D_i$. In both cases we do not aim at estimating the stochastic parts of the local error, and $\mathscr{R}_i$ can be considered as an asymptotically correct estimate for $R_i^\star$ only if $D_i - D_i^\star$ is asymptotically smaller than the dominant term $R_i^\star$ itself. This means that the stochastic parts of $D_i - D_i^\star$ have to be of order $O(\varepsilon h_i^{3/2} + \varepsilon^2 h_i)$, and the remaining (deterministic or unspecified) parts of $D_i - D_i^\star$ of order $O(h_i^{p+2} + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2)$.

Let us now restrict the exposition to stochastic schemes with deterministic order $p=2$. In Part I ([19]), the defect structured as a weighted sum of f-values, proved advantageous. To obtain this structure again, we choose an auxiliary two-step scheme (5.2) with the same left-hand side as (3.2), i.e., $\bar{\alpha}_{j,i} = \alpha_{j,i}$, $j=0,1,2$. Another very important effect of this choice is that the stochastic Maruyama terms in the defect are cancelled. This yields

$$D_i = h_i \cdot \sum_{j=0}^{2}\left(\beta_{j,i} - \bar{\beta}_{j,i}\right) f(t_{i-j},X_{i-j}). \tag{5.11}$$

The freedom to choose the auxiliary scheme (5.2) now reduces to determine the coefficients $\bar{\beta}_{0,i}, \bar{\beta}_{1,i}, \bar{\beta}_{2,i}$, which additionally have to satisfy consistency conditions ensuring that the scheme (5.2) has at least the deterministic order of convergence $p=2$. Now, only one degree of freedom remains. Specifying the coefficients in such a way that the error constant of the resulting scheme satisfies $\bar{c}_i - c_i = 1$, we obtain

$$D_i = h_i \cdot \left[\frac{2\kappa_i}{\kappa_i+1}f(t_i,X_i) - 2\kappa_i f(t_{i-1},X_{i-1}) + \frac{2\kappa_i^2}{\kappa_i+1}f(t_{i-2},X_{i-2})\right], \tag{5.12}$$

in analogy to the deterministic case. As already mentioned in Section I.3.1, this structure of D_i is crucial for the property that the difference $\|D_i - D_i^\star\|$ is asymptotically smaller than $\|R_i\|$ itself. Therefore we can use $\mathscr{R}_i$ as an estimate for the dominating term in the local truncation error, which yields

$$\|L_i\|_{L_2} = \|R_i + S_i\|_{L_2} = c_i \|\mathscr{R}_i\|_{L_2} + O(h_i^4 + \varepsilon h_i^{5/2} + \varepsilon^2 h_i^2) + O(\varepsilon h_i^{3/2} + \varepsilon^2 h_i), \quad (5.13)$$

where $\mathscr{R}_i$ is of order $O(h_i^3 + \varepsilon h_i^{5/2} + \varepsilon^2 h_i)$. Unless we know the conditional expectation $\mathbb{E}(L_i|\mathscr{F}_{t_{i-1}})$, we are unfortunately not able to utilize the fact that the stochastic parts of the local error contribute less to the global error than the deterministic ones.

We now study the local behavior of (5.13) in relation of the small parameter ε and the actual step-size h_i. We assume that the local error term of order $O(h_i^3)$ in $\mathscr{R}_i$ dominates the critical stochastic error terms of order $O(\varepsilon h_i^{3/2} + \varepsilon^2 h_i)$. This is the case when h_i^3 is much larger than $\varepsilon h_i^{3/2}$, which we express by

$$\varepsilon h_i^{3/2} \ll h_i^3, \quad \text{i.e.,} \quad \varepsilon^{2/3} \ll h_i. \quad (5.14)$$

Therefore, the local truncation error L_i behaves like an order three term. Note that what we would like to control in practice are the local errors in the solutions

$$X(t_i) - X_i^\star = \left(I - h_i \beta_{0,i} J_i\right)^{-1} L_i,$$

see (5.6). They belong to the solution space, whereas the defects belong to the image space of the problem. Only for non-stiff problems, where the values of $h_i J_i$ are small compared to the identity matrix I, L_i as well as $\mathscr{R}_i$ are good approximations to $X(t_i) - X_i^\star$. For stiff problems the values of $h_i J_i$ can become considerably large and therefore $\mathscr{R}_i$ should be scaled by $\left(I - h_i \beta_{0,i} J_i\right)^{-1}$, or by a suitable approximation to this matrix. Since $\left(I - h_i \beta_{0,i} J_i\right)$ is the Jacobian of the discrete scheme (3.2), this matrix, or its good approximation, as well as its factorization are usually available. In the literature this step is also called filtering.

6 Discretization schemes for index-1 SDAEs with small noise.

We now extend the results from the previous section to SDAEs of the form

$$AX(s)|_{t_0}^t - \int_{t_0}^t f(s, X(s))\mathrm{d}s - \int_{t_0}^t G(s, X(s))\mathrm{d}W(s) = 0, \quad t \in \mathscr{J}, \quad (6.1)$$

where A is a constant singular $n \times n$ matrix. Moreover, $G = \varepsilon \widehat{G}$, where ε is a parameter characterizing the size of the noise. The following discussion for the SDAEs is similar to the respective considerations for DAEs presented in the first part of the paper, cf. Section I.4. This is due to the analogous problem structure, especially to the crucial property that the matrix A is constant. For readers convenience we will repeat the most important merits of the discussion and stress the alterations resulting from the different nature of the SDAEs.

6.1 Analysis of index-1 SDAEs.

The analytical properties of SDAEs (6.1) have been studied in [22]. Since the matrix A is singular the system (6.1) contains constraints and the solution components lying in $\ker A$ (the 'algebraic components') are not subject to differentiation. Following the definition given in [22], we say that the SDAE (6.1) is of index-1 if the following properties hold:

- The constraints are free of noise.
- The constraints are globally uniquely solvable for the algebraic variables.

In order to analyze the qualitatively different solution components a theoretical decoupling is used. For constant matrices A this decoupling is compatible with the stochastic calculus and successfully carries over to SDAEs, see [22].
Again, we use constant projectors

$$Q \text{ onto } \ker A, \quad P := I - Q \text{ along } \ker A, \quad R \text{ along } \operatorname{im} A$$

to decompose the solution X into differential and algebraic components,

$$X = PX + QX =: U + V.$$

The solution X is a random vector with values in $\mathbb{R}^n$, U is a random vector with values in $\operatorname{im} P$ and V is a random vector with values in $\operatorname{im} Q$. We first decompose the original system (6.1) into a set of stochastic differential equations and stochastic constraints,

$$A(PX)(s)\Big|_{t_0}^{t} - (I-R)\int_{t_0}^{t} f(s,X(s))\mathrm{d}s - (I-R)\int_{t_0}^{t} G(s,X(s))\mathrm{d}W(s) = 0, \quad (6.2)$$

$$R\int_{t_0}^{t} f(s,X(s))\mathrm{d}s - R\int_{t_0}^{t} G(s,X(s))\mathrm{d}W(s) = 0. \quad (6.3)$$

Since we deal with an index-1 system, the constraints are free of noise, i.e., $RG = 0$. Hence, the integral $\int_{t_0}^{t} Rf(s,X(s))\mathrm{d}s$ vanishes for all $t \in \mathscr{J}$ and we can conclude that $Rf(t,U+V) = 0$ for all $t \in \mathscr{J}$. Therefore, we can solve the constraints (6.3) for the algebraic components $QX = V$,

$$Rf(t,U+V) = 0, \quad AV = 0 \Longleftrightarrow V = \hat{v}(t,U). \quad (6.4)$$

Finally, we insert $V = \hat{v}(t,U)$ into (6.2) and scale the equation by a reflexive generalized inverse A^- with $AA^- = I - R$, and $A^-A = P$. This yields the so-called *inherent regular SDE* for the differential components U,

$$U(s)\Big|_{t_0}^{t} = \int_{t_0}^{t} A^- f(s,x(s,U(s)))\mathrm{d}s + \int_{t_0}^{t} A^- G(s,x(s,U(s)))\mathrm{d}W(s), \quad (6.5)$$

where $x(s,u) = u + \hat{v}(s,u)$, and the initial condition is given by $U(t_0) = PX_0$.
We refer to [22] for the discussion of the existence and uniqueness of solutions of (6.1) based on the above theoretical decoupling.

6.2 Linear two-step Maruyama schemes.

A natural generalization of the linear two-step Maruyama schemes, cf. (3.2), for the numerical solution of SDAEs (6.1) is given by

$$\frac{1}{h_i}A\sum_{j=0}^{2}\alpha_{j,i}X_{i-j}=\sum_{j=0}^{2}\beta_{j,i}f(t_{i-j},X_{i-j})+\sum_{j=1}^{2}\gamma_{j,i}\sum_{r=1}^{m}g_r(t_{i-j},X_{i-j})\frac{I_r^{t_{i-j},t_{i-j+1}}}{h_i}. \tag{6.6}$$

However, one has to pay special attention to the discrete constraints that are implicitly included in the above scheme,

$$0=\sum_{j=0}^{2}\beta_{j,i}Rf(t_{i-j},X_{i-j})+\sum_{j=1}^{2}\gamma_{j,i}\sum_{r=1}^{m}Rg_r(t_{i-j},X_{i-j})\frac{I_r^{t_{i-j},t_{i-j+1}}}{h_i}. \tag{6.7}$$

The definition of index-1 SDAEs contains the assumption that the constraints are free of noise, i.e., $Rg_r=0$, $r=1,\ldots,m$, holds. Given this, the discrete constraints,

$$0=\sum_{j=0}^{2}\beta_{j,i}Rf(t_{i-j},X_{i-j}), \tag{6.8}$$

formally coincide with those for the deterministic case, and again, we obtain a recursion in $Rf(t_i,X_i)$. For exact initial values $Rf(t_0,X_0)=Rf(t_1,X_1)=0$, the exact solution of this recursion is trivial, $Rf(t_i,X_i)=0$. Clearly, this property can be recovered during the numerical simulations only if the recursion is stable. In Section I.4 we addressed several possibilities to realize stable recursions. Not all of the options discussed there carry over to SDAEs. Here, the situation is more complicated in general, since the discretization of the drift and the diffusion coefficient needs to be carried out differently due to their qualitatively different behavior. An implicit discretization has to be used for the drift term, whereas the diffusion term may be discretized explicitly.

The most favorable behavior can be observed for the BDF_2-Maruyama scheme. In fact, no recursion in $Rf(t_i,X_i)$ is carried out there. Applied to the SDAE (6.1) the BDF_2-Maruyama scheme takes the form

$$\begin{aligned}A\frac{X_i-\frac{(\kappa_i+1)^2}{2\kappa_i+1}X_{i-1}+\frac{\kappa_i^2}{2\kappa_i+1}X_{i-2}}{h_i}&=\frac{\kappa_i+1}{2\kappa_i+1}f(t_i,X_i)\\&+\sum_{r=1}^{m}g_r(t_{i-1},X_{i-1})\,\frac{I_r^{t_{i-1},t_i}}{h_i}-\frac{\kappa_i^2}{2\kappa_i+1}\sum_{r=1}^{m}g_r(t_{i-2},X_{i-2})\,\frac{I_r^{t_{i-2},t_{i-1}}}{h_i}.\end{aligned} \tag{6.9}$$

Other linear multi-step schemes, like the stochastic trapezoidal rule, need to be modified to guarantee a numerically stable formulation. To this end, more structural information has to be exploited. The first approach presented in Section I.4.1, formula (I.4.9), was to utilize knowledge about the constraints and to use different discretizations of the differential and algebraic components. This approach turns out to be

successful also in context of SDAEs. Using a projector R along $\mathrm{im}\,A$, a related stable scheme can be formulated as follows,

$$\frac{1}{h_i}A\sum_{j=0}^{2}\alpha_{j,i}X_{i-j} = \sum_{j=0}^{2}\beta_{j,i}(I-R)f(t_{i-j},X_{i-j}) + Rf(t_i,X_i) + \sum_{j=1}^{2}\gamma_{j,i}\sum_{r=1}^{m}g_r(t_{i-j},X_{i-j})\frac{I_r^{t_{i-j},t_{i-j+1}}}{h_i} \quad .$$

The second possibility considered in Section I.4.1, formulas (I.4.10), (I.4.11), does not work correctly in case of SDAEs. When using this approach, one cannot account for the qualitatively different discretization of drift and diffusion.

In case of SDAEs with noisy constraints, also called SDAEs with direct noise, we face even more severe difficulties. Except for the case of the BDF_2 method, the noise in the constraints accumulates to

$$Rf(t_i,X_i) = a_iRf(t_1,X_1) + b_iRf(t_0,X_0) + \sum_{j=2}^{i}c_i\sum_{r=1}^{m}Rg_r(t_{i-j},X_{i-j})\frac{I_r^{t_{i-j},t_{i-j+1}}}{h_{i-j+1}}, \quad (6.10)$$

where $i=k,\ldots,N$. The constants a_i,b_i,c_i depend on the coefficients $\alpha_{j,i}$, $\beta_{j,i}$ and $\gamma_{j,i}$ of the scheme (6.6) and do not vanish in general. A more detailed discussion of the difficulties caused by noisy constraints can be found in [21].

6.3 Local error estimate.

We define the local truncation error L_i, as before, by substituting the values of the exact solution into the scheme (6.6),

$$L_i := A\sum_{j=0}^{2}\alpha_{j,i}X(t_{i-j}) - h_i\sum_{j=0}^{2}\beta_{j,i}f(t_{i-j},X(t_{i-j})) - \sum_{j=1}^{2}\gamma_{j,i}\sum_{r=1}^{m}g_r(t_{i-j},X(t_{i-j}))\,I_r^{t_{i-j},t_{i-j+1}}, \quad i=2,\ldots,N. \quad (6.11)$$

Again, L_i satisfies the relation

$$\begin{aligned} L_i &= A(X(t_i)-X_i^\star) - h_i\beta_{0,i}\big(f(t_i,X(t_i)) - f(t_i,X_i^\star)\big) \\ &= \big(A - h_i\beta_{0,i}J_i\big)(X(t_i)-X_i^\star), \end{aligned} \quad (6.12)$$

where, $X_i^\star$ is the result of a step with exact starting values $X(t_{i-j})$, $j=1,2$ and $J_i = \int_0^1 f_x(t_i,sX_i^\star+(1-s)X(t_i))\,ds$. This representation has exactly the same form as in the deterministic case, due to the explicit discretization of the diffusion part. Also, the constraint part of L_i vanishes, i.e., $RL_i=0$, and L_i is related to the local truncation error L_i^{inh} of the discretized inherent SDE (6.5), by $L_i^{inh}=A^-L_i$ and $AL_i^{inh}=L_i$.

As before in the context of SDEs, we introduce an auxiliary linear multi-step method with coefficients $\overline{\alpha}_{j,i}$, $\overline{\beta}_{j,i}$ and $\overline{\gamma}_{j,i}$, and analyze the defect D_i of the numerical solution with respect to this second scheme.

Lemma 6.1 *Let the SDAE (6.1) be of index-1 and let $f(t,x)$ be continuous and continuously differentiable with respect to x. Let the step-size* **h** *be sufficiently small to guarantee that the matrix $\big(A-h_i\beta_{0,i}J_i\big)$ is nonsingular for all $h_i \in (0,\mathbf{h}]$. Then the defect $D_i^\star$,*

$$D_i^\star := A\Big(X_i^\star + \sum_{j=1}^{2} \bar{\alpha}_{j,i} X(t_{i-j})\Big) - h_i\bar{\beta}_{0,i} f(t_i, X_i^\star) - h_i \sum_{j=1}^{2} \bar{\beta}_{j,i} f(t_{i-j}, X(t_{i-j})) + \varepsilon \sum_{j=1}^{2} \bar{\gamma}_{j,i} \sum_{r=1}^{m} \hat{g}_r(t_{i-j}, X(t_{i-j}))\, I_r^{t_{i-j},t_{i-j+1}},$$

satisfies

$$D_i^\star = \bar{L}_i - L_i + h_i(\bar{\beta}_{0,i} - \beta_{0,i}) J_i \big(A - h_i\beta_{0,i}J_i\big)^{-1} L_i. \tag{6.13}$$

The proof is fully analogous to that of Lemma 5.1.

Since the SDAE (6.1) is of index-1, $(A-\beta h_i J_i)^{-1}(I-R) = O(1)$ holds and hence, an analogue version of Corollary 5.1 applies. Note that here L_i approximates the local error in AX_i. Depending on the available information we will monitor different quantities to satisfy accuracy requirements,

i) control $e_i := c_i \|(A-h_i\beta_{0,i}J_i)^{-1} D_i\|_{L_2} \approx \|(A-h_i\beta_{0,i}J_i)^{-1} L_i\|_{L_2}$ to match a given tolerance for X_i,
ii) control $e_i := c_i \|D_i\|_{L_2} \approx \|L_i\|_{L_2}$ to match a given tolerance for AX_i, or
iii) control $e_i := c_i \|A^- D_i\|_{L_2} \approx \|A^- L_i\|_{L_2}$ to match a given tolerance for PX_i.

7 Step-size control algorithm.

In this section we discuss algorithmic details of a step-size control which is based on the mean-square of a local error estimate. In this context, we approximate the L_2-norm $\|Z\|_{L_2}$ of a vector valued, square integrable, random variable $Z \in L_2(\Omega, \mathbb{R}^n)$ by using M samples $z^1, \ldots, z^M$ of Z,

$$\|Z\|_{L_2} \approx \left(\frac{1}{M} \sum_{\ell=1}^{M} |z^\ell|^2 \right)^{1/2}. \tag{7.1}$$

Applying this formula to the mean-square of the defect by using M simultaneously computed solution paths leads to adaptive step-size sequences that are uniform for all paths. The structure of the step-size control algorithm is almost the same as in the deterministic case, see Section I.5.2. Here, we merely comment on what modifications are necessary due to nature of stochastic differential equations.

We refer to [19] for a discussion of the initialization of the algorithm, especially for the case when the initial value problem itself does not supply enough information to start a multi-step scheme.

In the following, let us assume that two initial values X_0, $X_1 \in L_2(\Omega, \mathbb{R}^n)$ at time points t_0 and $t_1 = t_0 + h_1$, an absolute and a relative tolerance, ATOL, RTOL, and

an initial guess for the step-size h_2 are given. We consider an ensemble of M paths starting from M samples $x_0^1,\dots,x_0^M$ of the initial value X_0, and $x_1^1,\dots,x_1^M$ of the initial value X_1. Let $i := 2$ and start the step-size control algorithm:

1) Cf. Section I.5.2; the next elements $x_i^1,\dots,x_i^M$ of M paths are computed simultaneously by solving the nonlinear equations (3.2) and (6.6).

2) Cf. Section I.5.2; the L_2-norm $\|D_i\|_{L_2}$ of the defect is estimated by using the M values D_i^ℓ, $\ell = 1,\dots,M$, corresponding to the M sampled paths,

$$\begin{aligned}\widehat{D}_i &= \left(\frac{1}{M}\sum_{\ell=1}^{M}\left|h_i\left(\frac{2\kappa_i}{\kappa_i+1}f(t_i,x_i^\ell)-2\kappa_i f(t_{i-1},x_{i-1}^\ell)+\frac{2\kappa_i}{\kappa_i+1}f(t_{i-2},x_{i-2}^\ell)\right)\right|^2\right)^{1/2}\\ &\approx \|D_i\|_{L_2}.\end{aligned}$$

Depending on the problem setting, different scaling of the defect is used for the derivation of the error estimate e_i. The tolerances TOL_ν are computed componentwise with respect to the mean over all paths,

$$\mathrm{TOL}_\nu := \mathrm{ATOL} + \mathrm{RTOL}\cdot\left|\frac{1}{M}\sum_{\ell=1}^{M}x_{i,\nu}^\ell\right|.$$

3) Cf. Section I.5.2; we apply a control strategy predicting the new step-size h_{new} to match the tolerance TOL multiplied by a safety factor θ, say $\theta = 0.7$, and finally,

4) Cf. Section I.5.2; we accept the step, if $|e_{i,\nu}| \le \mathrm{TOL}_\nu$ for all components $\nu = 1,\dots,n$. In this case we set $i := i+1$, $h_i := h_{new}$ and go to 1. Otherwise we reject the step and repeat the calculations with a smaller step-size.
In case of step rejections, the already computed values of the Wiener process are stored and used to compute intermediate values according to the strategy presented in [12, 17].

8 Numerical experiments.

In this paragraph, we illustrate the performance of the step-size control strategy by means of numerical experiments for the stochastic BDF_2 applied to three test problems. We begin with a model SDE, with known exact solution, which allows accessing the exact errors of the approximation. We plot accuracy as a function of steplength to show that the classical deterministic convergence order two is retained in case of a small noise. Our second example is a MOSFET inverter circuit under the influence of thermal noise, which is modelled as a SDE. Finally, we present simulation results for a MOSFET ring oscillator.

Example 8.1 We first deal with a nonlinear scalar SDE,

$$X(t) = \int_0^t -(\alpha+\beta^2 X(s))(1-X(s)^2)ds + \int_0^t \beta(1-X(s)^2)dW(s), \qquad (8.1)$$

where $t \in [0,T]$ and W denotes a scalar Wiener process. The drift and diffusion coefficients, $f(t,x) := -(\alpha+\beta^2 x)(1-x^2)$, $G(t,x) := \beta(1-x^2)$, can be tuned by two real parameters α, β. The solution to the problem is given by (cf. [10, (4.46)])

$$X(t) = \frac{\exp(-2\alpha t + 2\beta W(t)) - 1}{\exp(-2\alpha t + 2\beta W(t)) + 1}. \tag{8.2}$$

We restrict our attention to two sets of parameters, $\alpha = -10$ and $\beta = 0.1$, and $\alpha = -10$ and $\beta = 0.01$. The tolerances are ATOL $=$ RTOL $= 10^{-2}$.

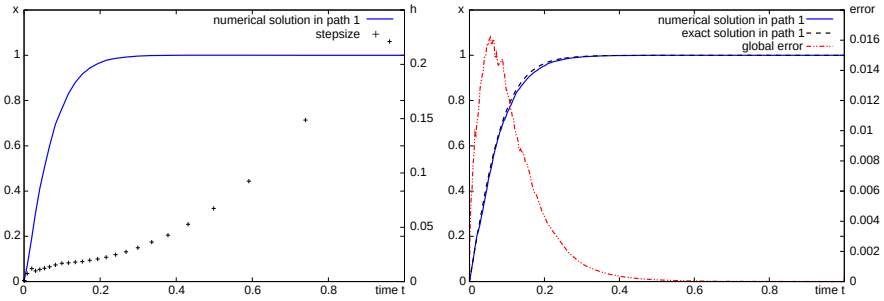

Fig. 8.1 Example 1: Numerical solution provided by BDF_2 in one path, and step-sizes (left). Exact solution, numerical solution, and its global error (right).

In the left graph of Figure 8.1 we depicted the numerical solution in the first computed path (solid line) together with the adaptively chosen step-sizes from over 100 simultaneously computed solution paths (marked by +). In the right graph the numerical solution, the exact solution (dashed line), and the observed global error (dashed dotted line) are given. The solution shows a transient behavior at the beginning of the time interval. One can observe a step-size increase towards the interval end, due to a considerable smoothness of the solution in this region.

In Figure 8.2 we present the related work precision diagrams. We plotted the tolerances ATOL $=$ RTOL (marked as $\triangle$) and the mean-square norm of the errors for adaptively chosen (marked by +) and constant (marked by $\times$) step-sizes for 100 computed paths vs. number of steps, both in logarithmic scale. The accuracy is measured using the maximum of the approximated L_2-norm of the global errors in the time interval,

$$\max_{i=1,\ldots,N} \left(\frac{1}{M} \sum_{\ell=1}^{M} |X(t_i, w_\ell) - x_i^\ell|^2 \right)^{1/2} \approx \max_{i=1,\ldots,N} \|X(t_i) - X_i\|_{L_2}, \tag{8.3}$$

where N denotes the number of steps and M the number of computed paths. Throughout, we used $M = 100$. Additional lines with slopes -2 and $-1/2$ are provided, to enable comparisons with convergence orders 2 and $1/2$, respectively.

In case of small parameter values, cf. lower graph, we observe classical convergence order $p = 2$ up to accuracies $3 \cdot 10^{-5}$. Afterwards, there is a very small region of

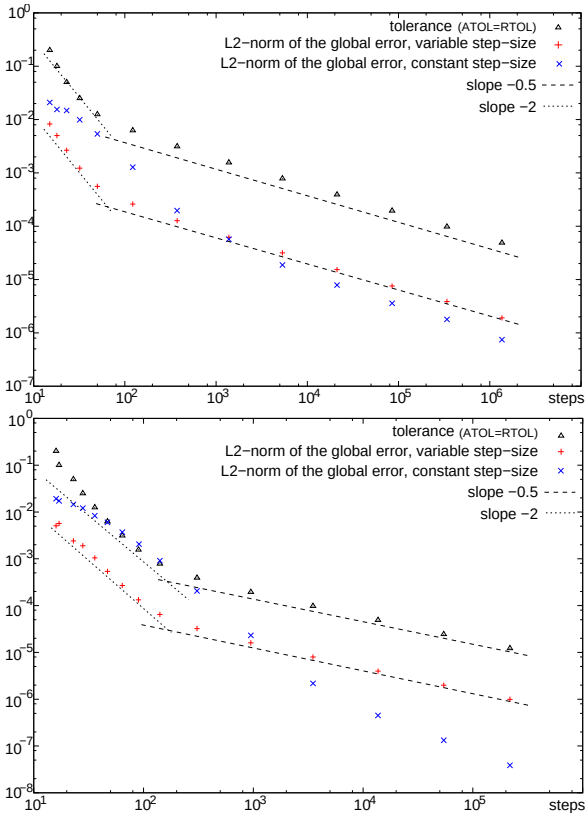

Fig. 8.2 Example 1: Tolerance and accuracy vs. steps, $\alpha = -10$, $\beta = 0.1$ (top); $\alpha = -10$, $\beta = 0.01$ (bottom).

linear convergence and thereafter the convergence order drops to only $1/2$. In the upper graph this threshold is reached at accuracies $2 \cdot 10^{-4}$. Furthermore, the use of adaptive step-sizes provides considerably more accurate results than the computation with the same number of constant steps.

The results indicate that the proposed step-size control performs well for step-sizes above the threshold and still quite reasonable for step-sizes slightly below the threshold.

Example 8.2 Now, we consider a model of an inverter circuit with a MOSFET transistor, under the influence of thermal noise. The related circuit diagram is given in Figure 8.3. The MOSFET is modelled as a current source from source to drain. The current through the MOSFET, j_D, depends on the nodal potentials at gate, e_g, drain, e_d, and source, e_s, $j_D = f_{mosfet}(e_g, e_d, e_s)$. In the model of MOSFET inverter circuit considered here, the current is controlled by the input voltage V_{in} and the nodal potential e_1 at node 1, $j_D(V_{in}, e_1) := f_{mosfet}(V_{in}, e_1, 0)$. We refer the reader to [17] for a detailed investigation of the MOSFET model.

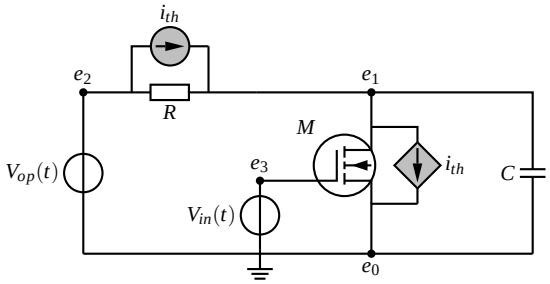

Fig. 8.3 Thermal noise sources in a MOSFET inverter circuit.

The thermal noise of the resistor and of the MOSFET is modelled by additional white noise current sources that are shunt in parallel to the original, noise-free elements. The noise intensity is given by Nyquist's rule, and hence, the associated current is modelled by additive noise,

$$i_{th} = \sigma_R \xi(t) = \sqrt{\frac{2kT}{R}}\xi(t), \tag{8.4}$$

where $\xi(t)$ is a standard Gaussian white noise process, $k = 1.38066 \cdot 10^{-23}[JK^{-1}]$ is Boltzmann's constant, T is the absolute temperature, and R is the resistance. For the thermal noise source of the MOSFET this formula is modified by considering a solution dependent resistance, see [17].

Combining Kirchhoff's current law and the element characteristics yields the following model for the output voltage e_1 at node 1:

$$C\dot{e}_1 - (V_{op} - e_1)/R + j_D(V_{in}, e_1) - \sigma_R \xi_1 + \sigma_D(V_{in}, e_1)\xi_2 = 0, \tag{8.5}$$

where ξ_1, ξ_2 are independent standard Gaussian white noise processes. We treat this system as an Itô SDE with $n = 1$, $m = 2$,

$$f(t,x) = ((V_{op} - x)/R + j_D(V_{in}, x))/C,\ g_1(t,x) = \sigma_R/C,\ g_2(t,x) = \sigma_D(V_{in}, x)/C.$$

In Figure 8.4 we present simulation results for the parameter values $C = 2 \cdot 10^{-13}[F]$, $R = 5 \cdot 10^3[\Omega]$, $U_{op} = 5[V]$, $T = 300[K]$, and the time interval $[0, 2.5 \cdot 10^{-8}][s]$. The tolerances were chosen as $\text{RTOL} = 10^{-2}$ and $\text{ATOL} = C \cdot \text{RTOL}$. To highlight the effect of the noise, we scaled the diffusion coefficients by a factor of 1000. The input voltage U_{in} and values of the output voltage e_1 are plotted vs. time. Moreover, the applied step-sizes, suitably scaled, are depicted using single crosses. We compare the results for the computation of a single path (upper graph) with those for the computation of 100 simultaneously computed solution paths (lower graph), where the thick lines additionally show the values of two different solution paths, the dashed line gives the mean of 100 paths and the thin lines the 3σ wide confidence interval for the output voltage e_1.

In Table 8.1, we compare the number of accepted and rejected steps for different numerical methods. The implicit Euler-Maruyama scheme has deterministic order $p = 1$, whereas the stochastic variants of the trapezoidal rule and the BDF_2 have

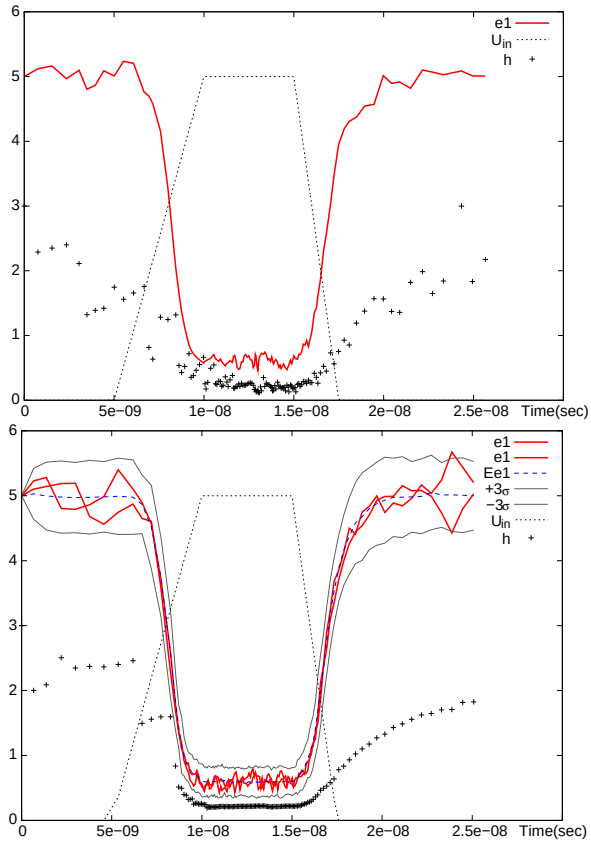

Fig. 8.4 Example 2: Simulation results for the noisy inverter circuit using BDF_2. Results for 1 path (top), and for 100 paths (bottom).

Table 8.1 Number of steps for one path and 100 paths: accepted + rejected

	RTOL $= 10^{-2}$		RTOL $= 10^{-3}$	
	1 path	100 paths	1 path	100 paths
iEul	185+36 = 221	199+5 = 204	889+193 = 1082	912+2 = 914
BDF_2	139+25 = 164	157+6 = 163	592+149 = 741	680+8 = 688
ITR	86+14 = 100	97+5 = 102	330+ 68 = 398	371+7 = 378

deterministic order $p = 2$. The ITR requires generally less steps than the BDF_2, due to the larger error constant of the latter method. Moreover, we can observe that using the information of an ensemble of simultaneously computed solution paths smoothes the step-size sequence and considerably reduces the number of rejected steps, when compared to the simulation of a single path. Also, the number of computed (accepted + rejected) steps to reach the tolerance is reduced.

Example 8.3 The final model we are dealing with is a MOSFET ring oscillator built of three coupled inverter steps with simple MOSFET models. Such an oscillator was also used for test runs in [16]. We used second order schemes for its simulation. Thermal noise in the MOSFETs and in the resistors are modelled by multiplicative and additive white noise sources. The circuit diagram is given in Figure 8.5. The corresponding noise-free circuit is a free running oscillator.

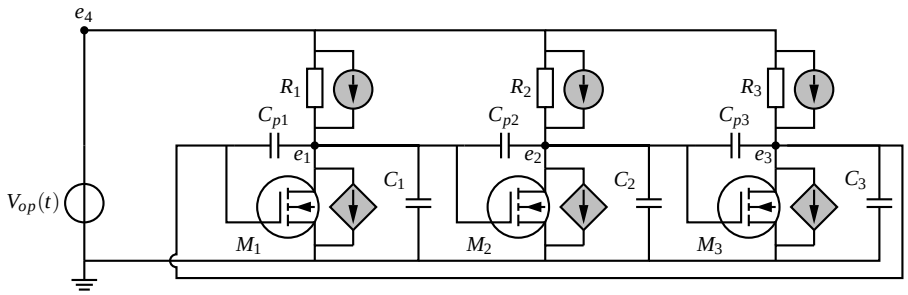

Fig. 8.5 Thermal noise sources in a MOSFET ring oscillator model.

The unknowns in the above charge oriented system are the charges for the six capacities, the four nodal potentials and the current through the voltage source. The system is of index-1, but, formally, has direct noise, because the three thermal resistance noise sources directly affect the current through the voltage source.

In Figure 8.6 we present numerical results obtained applying the stochastic trapezoidal rule ITR (upper graph) and the stochastic BDF_2 (lower graph), where 100 solution paths were computed simultaneously. We record the nodal potential at node 1. All quantities directly correspond to those computed in Example 8.2. The paths exhibit a highly visible phase noise and hence, can hardly be considered as small perturbations of the deterministic potential. The mean function appears damped and differs considerably from the noise free potential.

Using a method of second order results in a smaller number of steps compared to the drift-implicit Euler scheme. For the ITR, 766 (+4 rejected) steps were necessary, the BDF_2 preformed with 1497 (+3) steps and the implicit Euler-Maruyama scheme took 2008 (+8) steps for the simultaneous approximation of 100 solution paths.

9 Conclusions.

The theory presented in the present second part of the paper relies heavily on the techniques developed in the first part [19]. There, the Defect Correction principle has been used to derive error estimation formulas for the discretization error of the numerical solution of ODEs and DAEs. Most importantly, we were interested to construct reliable error estimates in such a way that they require only very moderate smoothness properties of the analytical solution. It turned out that a properly scaled defect may

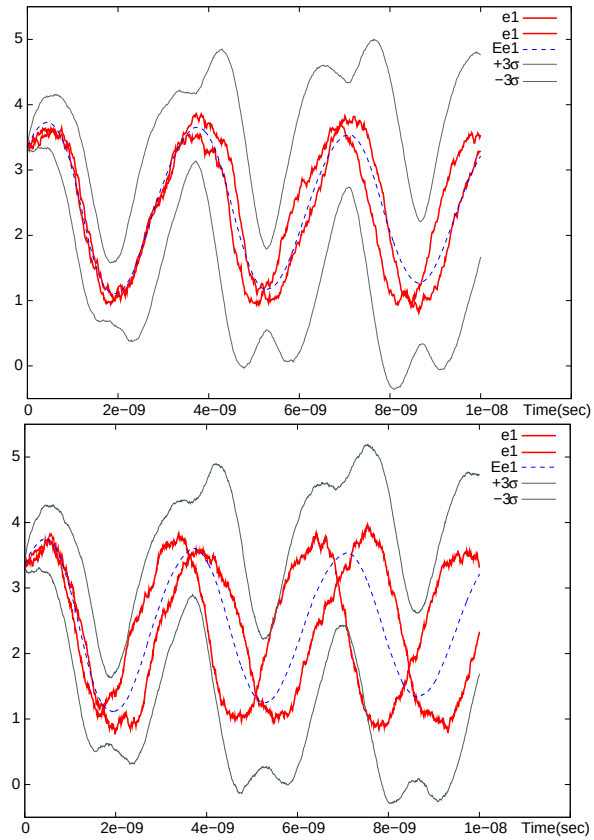

Fig. 8.6 Example 3: Two sample paths of the voltage in node 1 ($e1$), the mean over 100 sample paths (E$e1$) and the 3σ range ($\pm 3\sigma$) for the ITR (top) and for the BDF_2 (bottom).

serve as an asymptotically correct estimate of the local error. In the present investigation we were able to extend the technique and prove analogous results for SDE and SDAE with small noise.

References

1. L. Arnold. *Stochastische Differentialgleichungen.* R. Oldenbourg Verlag, München, 1973.
2. C.T.H. Baker and E. Buckwar. *Numerical analysis of explicit one-step methods for stochastic delay differential equations.* LMS J. Comput. Math. 3 (2000), pp. 315–335.
3. E. Buckwar and R. Winkler. *Multi-step methods for SDEs and their application to problems with small noise.* SIAM J. Num. Anal. 44(2) (2006), pp. 779–803.
4. P.M. Burrage and K. Burrage. *A variable stepsize implementation for stochastic differential equations.* SIAM J. Sci. Comp. 24 (2002), pp. 848–864.
5. G. Denk and R. Winkler. *Modeling and simulation of transient noise in circuit simulation.* Math. and Comp. Mod. of Dyn. Sys. (MCMDS) 13(4) (2007), pp. 383–394.
6. J.G. Gaines and T.J. Lyons. *Variable stepsize control in the numerical solution of stochastic differential equantions.* SIAM J. Appl. Math. 57 (1997), pp. 1455–1484.
7. D.J. Higham. *An algorithmic introduction to numerical simulation of stochastic differential equations.* SIAM Review 43 (2001), pp. 525–546.

8. N. Hofmann, T. Müller-Gronbach and K. Ritter. *Optimal approximation of stochastic differential equations by adaptive step-size control*. Math. Comp. 69 (2000), pp. 1017–1034.
9. I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus.* 2nd ed., Springer, Berlin, 1991.
10. P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1999.
11. H. Lamba. *An adaptive timestepping algorithm for stochastic differential equations.* J. Comput. Appl. Math. 161(2) (2003), pp. 417–430.
12. S. Mauthner. *Step size control in the numerical solution of stochastic differential equations.* J. Comput. Appl. Math. 100 (1998), pp. 93–109.
13. G.N. Milstein. *Theorem on the order of convergence for mean-square approximations of solutions of stochastic differential equations.* Theory Probab. Appl. 32(4) (1987), pp. 738–741. Translation from Teor. Veroyatn. Primen. 32(4) (1987), pp. 809–811.
14. G.N. Milstein. *Numerical Integration of Stochastic Differential Equations.* Kluwer, 1995. Translation from the Russion original of 1988.
15. G.N. Milstein and M.V. Tretyakov. *Mean-square numerical methods for stochastic differential equations with small noise.* SIAM J. Sci. Comp. 18 (1997), pp. 1067–1087.
16. C. Penski. *A new numerical method for SDEs and its application in circuit simulation.* J. Comput. Appl. Math. 115 (2000), pp. 461–470.
17. W. Römisch and R. Winkler. *Stepsize control for mean-square numerical methods for stochastic differential equations with small noise*. SIAM J. Sci. Comp. 28(2) (2006), pp. 604-625.
18. T. Sickenberger. *Mean-square convergence of stochastic multi-step methods with variable step-size*. J. Comput. Appl. Math. 212(2) (2008), pp. 300–319.
19. T. Sickenberger, E. Weinmüller and R. Winkler. *Local error estimates for moderately smooth problems: Part I - ODEs and DAEs*. BIT Numerical Mathematics 47(2) (2007), pp. 157–187.
20. T. Sickenberger and R. Winkler. *Adaptive Methods for Transient Noise Analysis*. In: G. Ciuprina, D. Ioan (Eds.), Scientific Computing in Electrical Engineering SCEE 2006, Springer, Berlin, 2007, pp. 403-410.
21. T. Sickenberger and R. Winkler. *Efficient transient noise analysis in circuit simulation*. Proceedings of the GAMM Annual Meeting 2006, Berlin, Proc. Appl. Math. Mech. 6(1) (2006), pp. 55–58.
22. R. Winkler. *Stochastic differential algebraic equations of index-1 and applications in circuit simulation.* J. Comput. Appl. Math. 157(2) (2003), pp. 477–505.