On the Complexity of Solving Stochastic Differential Equations

Klaus Ritter

TU Darmstadt

Introduction

Stochastic differential equation

$$dX(t) = a(t, X(t))\,dt + b(t, X(t))\,dW(t),$$
$$X(0) = x_0.$$

Computational problem, e.g.,

- Input: a, b, x_0, W. Output: X.
  Strong/pathwise approximation.
- Input: a, b, x_0, f. Output: $\mathrm{E}(f(X(T)))$.
  (Related to) weak approximation.

Minimal errors

$$e_{\text{min}}(n) = \inf\{\text{error}(A) : A \text{ algorithm with } \text{cost}(A) \le n\},$$

the intrinsic difficulty of the computational problem.

Requires definition of

- Class of all algorithms:
  partial information about a, b, W, or f.
  Exmp.: finitely many function values.
- Error and cost of an algorithm.

Minimal errors

$$e_{\text{min}}(n) = \inf\{\text{error}(A) : A \text{ algorithm with } \text{cost}(A) \le n\}.$$

Typical result

$$e_{\text{min}}(n) \asymp n^{-\alpha},$$

consists of

- **upper bound**: existence (construction) of algorithms A_n such that

$$\text{cost}(A_n) \le n \quad \wedge \quad \text{error}(A_n) = O\left(n^{-\alpha}\right),$$

- **lower bound**: $\exists\, c > 0 \;\; \forall$ algorithm A :

$$\text{cost}(A) \le n \quad \Rightarrow \quad \text{error}(A) \ge c \cdot n^{-\alpha}.$$

  Key: partial information.

Information-based Complexity

$$\text{comp}(\varepsilon) = \inf\{\text{cost}(A) : A \text{ algorithm with } \text{error}(A) \le \varepsilon\}.$$

Traub, Wasilkowski, Woźniakowski (1988), *Novak* (1988), *Werschulz* (1991), *Plaskota* (1996), *Traub, Werschulz* (1998), *Ritter* (2000), *Müller-Gronbach* (2002).

Strong Approximation

1. The Setting.
2. Minimal Error for Scalar Equations and L_2-Norm.
3. Further Results.

The Setting

$$dX(t) = a(t, X(t))\, dt + b(t, X(t))\, dW(t),$$
$$X(0) = x_0.$$

Approximate $X(t)$ for all $t \in [0,1]$.

Algorithms

- Partial information about W: values at sequentially (adaptively) chosen knots t_k. Cost one per evaluation.
- Complete information about a, b, and x_0.

Average error of an algorithm $\widehat{X}$

$$e(\widehat{X}) = \left(\mathrm{E}\, \|X - \widehat{X}\|_2^2\right)^{1/2},$$

where

$$\|X(\omega) - \widehat{X}(\omega)\|_2 = \left(\int_0^1 |X(t,\omega) - \widehat{X}(t,\omega)|^2\, dt\right)^{1/2}.$$

Average cost of an algorithm $\widehat{X}$ (at least)

$n(\widehat{X}) =$ expected number of evaluations of W.

Formal definition of an algorithm $\widehat{X}$ via Borel-measurable mappings (here: scalar case)

$$\begin{aligned}\psi_k &: ([0,1]\times\mathbb{R})^{k-1} \to [0,1],\\ \chi_k &: ([0,1]\times\mathbb{R})^{k} \quad \to \{\text{STOP}, \text{GO}\},\\ \phi_k &: ([0,1]\times\mathbb{R})^{k} \quad \to L_2([0,1]).\end{aligned}$$

Sequential evaluation of W

- first step: $t_1 = \psi_1$ yields $y_1(\omega) = W(t_1, \omega)$,
- k-th step:

$$t_k(\omega) = \psi_k(t_1, y_1(\omega), \ldots, t_{k-1}(\omega), y_{k-1}(\omega))$$

yields $y_k(\omega) = W(t_k(\omega), \omega)$.

Termination criterion and output

If $\quad \chi_k(t_1, y_1(\omega), \ldots, t_k(\omega), y_k(\omega)) = \text{STOP}$,

then $\quad \widehat{X}(\omega) = \phi_k(t_1, y_1(\omega), \ldots, t_k(\omega), y_k(\omega))$.

Thus

$$n(\widehat{X}) = \mathrm{E}(\min\{k : \chi_k = \text{STOP}\}).$$

Scalar Equations and L_2-Norm

Assumption For $f = a$ and $f = b$
$\exists\, K > 0 \;\; \forall\, s, t \in [0,1] \;\; \forall\, x, y \in \mathbb{R}$

$$|f(t,x) - f(t,y)| \leq K \cdot |x - y|,$$

$$|f(s,x) - f(t,x)| \leq K \cdot (1 + |x|) \cdot |s - t|,$$

$$|f^{(0,1)}(t,x) - f^{(0,1)}(t,y)| \leq K \cdot |x - y|.$$

Recall

$$e_{\text{min}}(n) = \inf\{e(\widehat{X}) : n(\widehat{X}) \le n\}.$$

Theorem *Hofmann, Müller-Gronbach, R.* (2001)

$$e_{\text{min}}(n) \approx \frac{1}{\sqrt{6}} \cdot c(a, b, x_0) \cdot n^{-1/2}$$

with

$$c(a, b, x_0) = \int_0^1 \mathrm{E}(|b|(t, X(t)))\, dt.$$

Optimal algorithm: Milstein plus Euler with adaptive step-size control.

Remark Strong approximation related to L_2-approximation of W w.r.t. random weight function $t \mapsto |b|(t, X(t))$.

Further Results

Scalar Equations

- **Error at discrete points**, in particular

$$e(\widehat{X}) = \left(\mathrm{E} \, |X(1) - \widehat{X}(1)|^2 \right)^{1/2}.$$

See *Clark, Cameron* (1980), *Rümelin* (1982), *Cambanis, Hu* (1996), and *Müller-Gronbach* (2002).

$\rightsquigarrow$ Integration of stoch. proc. w.r.t. random weight.

- **L_p-norm, $1 \le p \le \infty$,**

$$e(\widehat{X}) = \left(\mathrm{E} \, \|X - \widehat{X}\|_p^q \right)^{1/q}$$

with $1 \le q < \infty$. See *Hofmann, Müller-Gronbach, R.* (2000), *Müller-Gronbach* (2002).

$\rightsquigarrow$ Approximation of stoch. proc. w.r.t. random weight.

- **More general information**: multiple stochastic integrals. Used in Ito-Taylor schemes, due to Milstein, Wagner, Platen, Kloeden.

  Conjecture: For L_2-norm

$$e_{\text{min}}(n) \approx \text{const}(\text{info}) \cdot c(a,b,x_0) \cdot n^{-1/2},$$

  where

$$c(a,b,x_0) = \int_0^1 \mathrm{E}(|b|(t,X(t)))\,dt.$$

  Known: $\quad$ const(point eval.) $= 1/\sqrt{6}$,
  $\qquad\qquad$ const(lin. funct.) $= 1/\pi$.

  See Hofmann, Müller-Gronbach, R. (2002), and Hofmann, Müller-Gronbach (2003).

- **Key quantity**: local mean-square smoothness of X,

$$\mathrm{E}\big((X(t+h) - X(t))^2 \mid X(t) = x\big)$$
$$= b^2(t,x) \cdot h + o(h).$$

Systems of SDEs

- **Key quantities**: local smoothness and ‘amount of non-commutativity’. See *Müller-Gronbach* (2002).

Stochastic Integration

See *Wasilkowski, Woźniakowski* (2001), *Hertling* (2001).

Equations Driven by Fractional Brownian Motion

See *Neuenkirch* (2005).

Stochastic Partial Differential Equations

See *Davie, Gaines* (2001), *Müller-Gronbach, R.* (2005a, b).

Weak Approximation

1. The Setting.

2. Minimal Errors for Hölder Classes.

3. Further Results.

The Setting

$$dX(t) = a(t, X(t))\, dt + b(t, X(t))\, dW(t),$$
$$X(0) = x_0.$$

Consider a class $\mathcal{A}$ of drift coefficients

$$a : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$$

and the particular case

$$b = I_d \in \mathbb{R}^{d \times d}.$$

For fixed $f : \mathbb{R}^d \to \mathbb{R}$, $T > 0$, and x_0

$$S(a) = \mathrm{E}(f(X(T))).$$

Approximate $S : \mathcal{A} \to \mathbb{R}$. Note that

$$S(a) = v(0, x_0),$$

where

$$\tfrac{1}{2}\Delta v + \sum_{i=1}^{d} a_i \cdot v_{x_i} = -v_t, \qquad v(T, \cdot) = f.$$

Algorithms: deterministic or randomized/Monte-Carlo

- Partial information about a: values at sequentially (adaptively) chosen knots $(t_k, x_k) \in [0, T] \times \mathbb{R}^d$.

Approximate $S : \mathcal{A} \to \mathbb{R}$, where

$$S(a) = \mathrm{E}(f(X(T))).$$

Worst case error and cost

- of a deterministic algorithm $\widetilde{S}$

$$e(\widetilde{S}) = \sup_{a \in \mathcal{A}} |S(a) - \widetilde{S}(a)|,$$

$$c(\widetilde{S}) = \sup_{a \in \mathcal{A}} \big(\# \text{ evaluations of } a + \# \text{ arithm. op.} \big),$$

- of a randomized algorithm $\widetilde{S}$

$$e(\widetilde{S}) = \sup_{a \in \mathcal{A}} \left(\mathrm{E}\, |S(a) - \widetilde{S}(a)|^2 \right)^{1/2},$$

$$c(\widetilde{S}) = \sup_{a \in \mathcal{A}} \mathrm{E} \big(\# \text{ evaluations of } a + \# \text{ arithm. op.} + \# \text{ calls of random number gen.} \big).$$

Minimal errors

$$e^{\text{det}}_{\text{min}}(n) = \inf\{e(\widetilde{S}) : \widetilde{S} \text{ det. algorithm with } c(\widetilde{S}) \leq n\},$$

$$e^{\text{ran}}_{\text{min}}(n) = \inf\{e(\widetilde{S}) : \widetilde{S} \text{ ran. algorithm with } c(\widetilde{S}) \leq n\}.$$

Question: $e^{\text{ran}}_{\text{min}}(n) \ll e^{\text{det}}_{\text{min}}(n)$?

Minimal Errors for Hölder Classes

Let $r \in \mathbb{N}_0$ and $\alpha \in \,]0,1]$, put

$$\mathcal{A}^{r,\alpha} = \text{unit ball in Hölder class } C^{r,\alpha},$$

$$\mathcal{A} = \mathcal{A}^{r,\alpha}_M = \{a \in \mathcal{A}^{r,\alpha} : \mathrm{supp}(a) \subseteq [0,T] \times [-M,M]^d\}.$$

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

Theorem *Petras, R. (2004)*

$\exists\, c > 0 \quad \forall\, n \in \mathbb{N}:$

$$e^{\text{det}}_{\text{min}}(n) \geq c \cdot n^{-\gamma},$$

$$e^{\text{ran}}_{\text{min}}(n) \geq c \cdot n^{-(\gamma+1/2)}.$$

A simple consequence

Euler scheme $\widetilde{S}$, step-size T/m, and $k = m^2$ simulations:

$\mathrm{bias}(\widetilde{S}) = O(m^{-1})$ if $r + \alpha > 2$,

cost $c(\widetilde{S}) = O(m \cdot k) = O(m^3) =: n$,

worst case error $e(\widetilde{S}) = O(n^{-1/3})$.

Thus Euler is better than every deterministic method if

$$r + \alpha > 2 \quad \wedge \quad d > 3(r+\alpha) - 1.$$

Proof of the lower bound for $e^{\text{det}}_{\text{min}}$ (here $d = 1$)

Series representation

$$S(a) = I^{(0)} + \sum_{\mu=1}^{\infty} I^{(\mu)}(\underbrace{a \otimes \cdots \otimes a}_{\mu\text{-times}})$$

with linear functionals

$$I^{(\mu)} : C\big(([0,T] \times \mathbb{R})^{\mu}\big) \to \mathbb{R}$$

being weighted integrals.

If $\|a\|_{\infty}$ is 'small' then

$$S(a) \simeq I^{(0)} + I^{(1)}(a).$$

In $\mathcal{A}^{r,\alpha}$ 'unfavorable' functions for weighted integration are 'small'.

Thus computation of $S(a)$ is not easier than computation of $I^{(1)}(a)$.

Use the lower bounds for weighted integration due to *Bakhvalov* (1959) and *Novak* (1988).

In this way: a general theorem.

Upper bounds

As previously, $\gamma = (r+\alpha)/(d+1)$ and

$$\mathcal{A} = \mathcal{A}_M^{r,\alpha} = \{a \in \mathcal{A}^{r,\alpha} : \operatorname{supp}(a) \subseteq [0,T] \times [-M,M]^d\}.$$

Theorem *Petras, R. (2004)*

For every $\varepsilon > 0$

$$e_{\text{min}}^{\text{det}}(n) = o\left(n^{-(\gamma-\varepsilon)}\right),$$
$$e_{\text{min}}^{\text{ran}}(n) = o\left(n^{-(\gamma+1/2-\varepsilon)}\right).$$

Remark

- Upper and lower bounds arbitrarily close in the power scale.
- Analogous result for unbounded coefficients under suitable growths conditions.

Proof of the upper bound for $e^{\text{det}}_{\text{min}}$ and construction of an algorithm (here $d = 1$)

Series representation

$$S(a) = I^{(0)} + \sum_{\mu=1}^{\infty} I^{(\mu)}(\underbrace{a \otimes \cdots \otimes a}_{\mu\text{-times}}).$$

Algorithm

$$\widetilde{S}(a) = I^{(0)} + \sum_{\mu=1}^{m} I^{(\mu)}\big(\widetilde{L}^{(\mu)}(a \otimes \cdots \otimes a)\big),$$

where $\widetilde{L}^{(\mu)}$ is a Smolyak algorithm (sparse grid) for approximation of functions

$$\big([0,T] \times [-M,M]\big)^{\mu} \to \mathbb{R}.$$

(Tensor product algorithms (full grids) need too many arith. operations.) Requires precomputing.

Based on

- *Wasilkowski, Woźniakowski* (1995), *Li* (2002),
- *Wasilkowski, Woźniakowski* (2001),
- *Deck, Kruse* (2002).

Further Results

Feynman-Kac Path Integration

See *Plaskota, Wasilkowski, Woźniakowski* (2000), *Kwas, Li* (2003), *Kwas* (2004) for the Gaussian case.

Variable and path-dependent functionals

$$S(f) = \mathrm{E}(f(X))$$

with $f : C([0,T]) \to \mathbb{R}$ and fixed drift and diffusion coefficients a and b, resp.

See *Dereich, Müller-Gronbach, R.* (2005).

Summary

Partial information, lower bounds, minimal errors:
concepts from Information-based Complexity.

Apply results and ideas for linear problems to computational problems for SDEs

- strong approximation: integration and approximation of stochastic processes,
- weak approximation: integration of multivariate functions.

Yields

- sharp bounds for the minimal error,
- and almost optimal algorithms.

References

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

Cambanis, S., Hu, Y. (1996). Exact convergence rate of the Euler-Maruyama scheme, with application to sampling design. Stochastics Stochastics Rep. **59**, 211–240.

Clark, J. M. C., Cameron, R. J. (1980). The maximum rate of convergence of discrete approximations. In *Stochastic Differential Systems* (B. Grigelionis, ed.) 162–171. Lect. Notes Control Inf. Sci. **25**. Springer, Berlin.

Davie, A. M., Gaines, J. G. (2001), Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations, Math. Comp. **70**, 121–134.

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

Hertling, P. (2000). Nonlinear Lebesgue and Ito integration problems of high complexity. J. Complexity **17**, 366–387.

Hofmann, N., Müller-Gronbach, T. (2004), On the global error of Ito-Taylor schemes for strong approximation of scalar stochastic differential equations, J. Complexity **20**, 732–752.

Hofmann, N., Müller-Gronbach, T., Ritter, K. (2000), Optimal approximation of stochastic differential equations by adaptive step-size control, Math. Comp. **69**, 1017–1034.

Hofmann, N., Müller-Gronbach, T., Ritter, K. (2000), Step-size control for the uniform approximation of systems of stochastic differential equations with additive noise, Ann. Appl. Prob. **10**, 616–633.

Hofmann, N., Müller-Gronbach, T., Ritter, K. (2001), The optimal discretization of stochastic differential equations, J. Complexity **17**, 117–153.

Hofmann, N., Müller-Gronbach, T., Ritter, K. (2002), Linear vs. standard information for scalar stochastic differential equations, J. Complexity **18**, 394–414

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.

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-Verlag, Berlin.

Müller-Gronbach, T. (2002), The optimal uniform approximation of systems of stochastic differential equations, Ann. Appl. Probab. **12**, 664-690.

Müller-Gronbach, T. (2004), Optimal pointwise approximation of SDEs based on Brownian motion at discrete points, Ann. Appl. Probab. **14**, 1605–1642.

Müller-Gronbach, T., Ritter, K. (2006), Lower bounds and non-uniform time discretization for approximation of stochastic heat equations, Found. Comp. Math., to appear.

Neuenkirch, A. (2006), Approximation of stochastic differential equations driven by fractional Brownian noise, Ph.D. Dissertation, TU Darmstadt.

Petras, K., Ritter, K. (2006), On the complexity of parabolic initial value problems with variable drift, J. Complexity to appear.

Plaskota, L. (1996) Noisy Information and Computational Complexity, Cambrigde Univ. Press, Cambridge.

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.

Ritter, K. (2000), Average-Case Analysis of Numerical Problems, Lect. Notes in Math. **1733**, Springer-Verlag, Berlin.

Ruemelin, W. (1982), Numerical treatment of stochastic differential equations, SIAM J. Numer. Anal. **19**, 604–613.

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

Traub, J. F., Werschulz, A. G. (1998), Complexity and Information, Cambrigde Univ. Press, Cambridge.

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., Woźniakowski, H. (2001). On the complexity of stochastic integration. Math. Comp. **70**, 685–698.

Werschulz, A. G. (1991), The Computational Complexity of differential and integral equations, Oxford Univ. Press, Oxford.