Rate of convergence of SDE with discontinuous drift

Antoine Lejay (Inria Nancy Grand-Est)

From a joint work with
Arturo Kohatsu-Higa (Ritsumeikan University)
& Kazuhiro Yasuda (Hosei University)

International Conference on Monte Carlo techniques, Paris, July 2016

Our problem

$$\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t$$

- σ uniformly elliptic, bounded, "regular enough"
- b bounded but discontinuous

How to construct an approximation of X_T with a control on the error?

The Euler-Maruyama scheme is

- Simple to set-up, whatever the dimension
- Efficient in general

Euler-Maruyama (EM) scheme

- Time horizon $T > 0$
- Time step T/n
- $t_i = iT/n$ and $\varphi(t) = t_i$ with $t_i \leqslant t < t_{i+1}$
- $\xi_i \sim \mathcal{N}(0,1)$, iid.

EM scheme: Compute recursively $\widehat{X}_0 = 0$ and

$$\widehat{X}_{i+1} = \widehat{X}_i + \sigma(t_i, \widehat{X}_i)\sqrt{\frac{T}{n}}\xi_i + b(t_i, \widehat{X}_i)\frac{T}{n}$$

For the theory: continuous EM scheme

$$\overline{X}_t = x + \int_0^t \sigma(\varphi(s), \overline{X}_{\varphi(s)})\, \mathrm{d}B_s + \int_0^t b(\varphi(s), \overline{X}_{\varphi(s)})\, \mathrm{d}s$$

so that $\overline{X}$ and X are on the same probability space and

$$(\widehat{X}_i)_{i=0,\dots,n} \overset{\text{dist.}}{=} (\overline{X}_{t_i})_{i=0,\dots,n}$$

Rate of convergence of the EM scheme

The number of steps n is the parameter to adjust.

- $\star$ strong convergence at rate δ if

$$\mathbb{E}[|X_T - \overline{X}_T|^2]^{1/2} \leqslant \frac{C}{n^\delta},$$

- $\star$ weak convergence at rate δ if

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| \leqslant \frac{C(f)}{n^\delta}$$

for $f \in \mathfrak{F}$, a class of test functions.

- Strong conv $\Longrightarrow$ weak conv with $\mathfrak{F}$ =Lipschitz functions
- Weak rate is more difficult to establish than strong rate. But
  - ☺ Gives better rate than strong rates
  - ☺ Corresponds to what is actually computed (think to option prices)

Traditional approach for the weak rate of convergence

$\mathcal{L}$ is the infinitesimal generator of X

$\Longrightarrow u(0,x) = \mathbb{E}_x[f(X_T)]$ with

$$\mathcal{L} = \frac{1}{2} a_{i,j}(t,x) \frac{\partial^2}{\partial x_i \partial x_j} + b_i(t,x) \frac{\partial}{\partial x_i}, \;\; a = \sigma\sigma^T$$

$$\begin{cases} \dfrac{\partial u(t,x)}{\partial t} + \mathcal{L}u(t,x) = 0 \\ u(T,x) = f(x) \end{cases}$$

Write

$$\begin{aligned} \mathbb{E}_x[f(\overline{X}_T)] - \mathbb{E}_x[f(X_T)] &= \mathbb{E}[f(\overline{X}_T)] - u(0, \overline{X}_0) \\ &= \sum_{i=0}^{n-1} \mathbb{E}[u(t_{i+1}, \overline{X}_{t_{i+1}}) - u(t_i, \overline{X}_{t_i})] \end{aligned}$$

Traditional approach for the weak rate of convergence

$$\mathbb{E}[f(\overline{X}_T)] - \mathbb{E}[f(X_T)]$$
$$= \sum_{i=0}^{n-1} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \frac{1}{2}(a_{i,j}(s,\overline{X}_s) - a_{i,j}(t_i,\overline{X}_{t_i}))\partial^2_{ij}u(r,\overline{X}_r)\,\mathrm{d}r\right]$$
$$+ \sum_{i=0}^{n-1} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \frac{1}{2}(b_i(s,\overline{X}_s) - b_i(t_i,\overline{X}_{t_i}))\partial_i u(r,\overline{X}_r)\,\mathrm{d}r\right]$$

Perform Taylor development up to order 4 to approximate the integrals. □

Theorem (Talay, ...) If f, a and b are $\mathcal{C}^4$, the weak rate of convergence is of order 1.

Rem. The strong rate of convergence is of order 1/2.

Extension to Hölder continuous coefficients

Fix $\alpha \in (0,3) \setminus \{1,2\}$.

$\mathrm{H}^{\alpha/2,\alpha}$ space of Hölder continuous functions f on $[0,T] \times \mathbb{R}^d$ which are

- $\partial_t^r \partial_x^s f$ is $(\alpha - \lfloor \alpha \rfloor)$-Hölder continuous in space for $2r + s = \lfloor \alpha \rfloor$
- $\partial_t^r \partial_x^s f$ is bounded for $2r + s \leqslant \alpha$
- $\partial_t^r \partial_x^s f$ is $(\alpha - 2r - s)/2$-Hölder continuous in time for $0 < \alpha - 2r - s < 2$

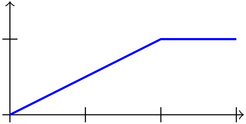
Theorem. Let $a, b \in \mathrm{H}^{\alpha/2,\alpha}$, $f \in \mathrm{H}^{2+\alpha}$, and u solution to $\partial_t u(t,x) + \mathcal{L}u(t,x) = 0$ with $u(T,x) = f(x)$.
Then $u \in \mathrm{H}^{1+\alpha/2,2+\alpha}$.

Weak rate of convergence with Hölder coefficients

Theorem (Mikulevicius & Platen)
If $a, b \in \mathrm{H}^{\alpha/2,\alpha}$ and $f \in \mathrm{H}^{2+\alpha}$, then

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]| \leqslant \frac{K}{n^{E(\alpha)}}$$

$$E(\alpha) = \begin{cases} \dfrac{\alpha}{2} & \alpha \in (0,1) \cup (1,2) \\ 1 & \alpha \in (2,3) \end{cases}$$



Our approach

- The previous approaches require a and b to be regular enough, and to share the same regularity.
- They do not apply when b is discontinuous
- We do not consider applying EM scheme to X but to an approximation of X with a regularized drift $b^\varepsilon \in \mathfrak{M}$

$$\mathrm{d}X_t^\varepsilon = \sigma(t, X_t^\varepsilon)\,\mathrm{d}B_t + b_\varepsilon(t, X_t^\varepsilon)\,\mathrm{d}t \tag{1}$$

- We consider statements of type

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T^\varepsilon)]| \leqslant \frac{C}{n^\delta}, \quad \forall f \in \mathfrak{F}$$

- When σ is constant
  - we consider $|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)]|$ using X^ε and $\overline{X}^\varepsilon$.
  - we use the fact that $\mathrm{Law}(\overline{X}^\varepsilon) \equiv \mathrm{Law}(\overline{X})$ (false when $\sigma \not\equiv$ constant)

Works on simulation of SDE with discontinuous drift

- ✎ S. Menozzi & V. Konakov (2016), densities.
- ✎ PhD thesis L. Lenôtre (2015), a and b both discont.
- ✎ P. Przybyłowicz (2013), optimal rate of convergence and adaptive algo ($d = 1$, localized discontinuities)
- ✎ PhD thesis S. Niklitschek-Soto (2013), local study around one discontinuity
- ✎ PhD thesis S. Arnold (2006), Zvonkin transform ($d = 1$)
- ✎ N. Halidias & P. Kloeden (2006): Heaviside drift
- ✎ PhD thesis L. Yan (2002), convergence but no rate
- ✎ K.S. Chan & O. Stramer (1998), discontinuity on polygons, no rate
- ✎ R. Janssen (1984), discontinuity on a surface, no rate

The difficulties

- The regularity of u is determined by the regularity of a, b and $f \in \mathfrak{F}$.
- Even if b is discontinuous, u belongs to some Sobolev space.
- Regularized drift in $\mathfrak{M}$ allows one to use "classical results" yet with exploding constants.
- The rate of convergence depends on $b_\varepsilon - b$ in a given norm, the regularity of a and $\mathfrak{F}$, the class of test functions.

How to "separate" the effects of $\mathfrak{M}$ and $\mathfrak{F}$?

Perturbation formula

$$\mathcal{L} = \frac{1}{2} a_{ij} \partial^2_{ij} + b_i \partial_i \text{ with semi-group } (P_t)_{t>0}$$
$$\mathcal{M} = \frac{1}{2} a_{ij} \partial^2_{ij} + c_i \partial_i \text{ with semi-group } (Q_t)_{t>0}$$

Perturbation formula

$$Q_t = P_t + \int_0^t Q_s (\mathcal{M} - \mathcal{L}) P_{t-s} \, \mathrm{d}s$$
$$= P_t + \int_0^t Q_s (b - c) \nabla P_{t-s} \, \mathrm{d}s$$

Proof.

$$Q_t - P_t = \int_0^t \mathrm{d}(Q_s P_{t-s}) = \int_0^t Q_s \mathcal{M} P_{t-s} \, \mathrm{d}s - \int_0^t Q_s \mathcal{L} P_{t-s} \, \mathrm{d}s.$$

□

Perturbation formula: stochastic version

- A stochastic version is
$$\mathbb{E}^c[f(X_t)] = \mathbb{E}^b[f(X_t)] + \mathbb{E}^b\left[\int_0^t (b-c)(s,X_s)\nabla u(s,X_s)\,\mathrm{d}s\right]$$
with
$$\partial_t u(t,x) + \mathcal{L}u(t,x) = 0, \quad u(T,x) = f(x)$$
- Another version is (Z Doléan exponential $0 \rightsquigarrow b$)
$$\mathbb{E}^c[f(X_t)] - \mathbb{E}^b[f(X_t)] = \mathbb{E}\left[\int_0^t Z_s(b-c)^\top \nabla u(s,X_s)\,\mathrm{d}s\right] + \mathbb{E}\left[(Z_T - 1)\int_0^t b^\top \nabla v(s,X_s)\,\mathrm{d}s\right]$$
with
$$\partial_t v + \frac{1}{2}a_{ij}\partial^2_{ij}v(t,x) = 0, \quad v(T,x) = f(x)$$

A control using a perturbation formula

- For a process X and $p \geqslant 1$,

$$\|g\|_{X,p} \overset{\text{def}}{=} \mathbb{E}\left[\int_0^T |g(s, X_s)|^p\right]^{1/p}$$

- $X^b \overset{\text{def}}{=}$ process generated by $\frac{1}{2}a_{ij}\partial^2_{ij} + b_i\partial_i$

- $\mathcal{C}_{\mathrm{sl}} \overset{\text{def}}{=}$ set of continuous functions with “slow growth”: $\lim_{x\to 0} |f(x)|e^{-k|x|^2} = 0$ for all $k > 0$

- a continuous, bounded, uniformly elliptic: $0 < \lambda|\xi|^2 \leqslant a\xi \cdot \xi \leqslant \Lambda|\xi|^2 \ \forall \xi \in \mathbb{R}^d$

- b bounded

Prop. If $f \in \mathcal{C}_{\text{sl}}$ and $\|\nabla v\|_{X^0,q} < +\infty$ for $1 < q \leqslant \infty$, then

$$|\mathbb{E}[f(X_T^b)] - \mathbb{E}[f(X_T^c)]| \leqslant C\|b-c\|_{X^0,p}\|\nabla v\|_{X^0,q}$$

with $p^{-1} + q^{-1} < 1$ and $2 \leqslant p$.

v sol. to the parabolic PDE without drift.

Prop'. If $f \in \mathcal{C}_{\text{sl}}$ and $\|\nabla u\|_{X,q} < +\infty$ for $1 < q \leqslant \infty$, then

$$|\mathbb{E}[f(X_T^b)] - \mathbb{E}[f(X_T^c)]| \leqslant C\|b-c\|_{X^0,p}\|\nabla u\|_{X,q}$$

with $p^{-1} + q^{-1} < 1$ and $1 \leqslant p$.

u sol. to the parabolic PDE with drift.

Proofs. Combine Girsanov and repetition of Hölder inequalities. □

Reason of this formula

- Choice of $\mathfrak{F}$ (regularity of f) $\implies \|\nabla v\|_{X^{0},q} < +\infty$ for some q.
- The distance $\|b - c\|_{X^{0},p}$ depends on $\mathfrak{F}$ through q since $p > \frac{q}{q-1}$.
- Gaussian control on the density transition of X^0 (*e.g.* if $a \in \mathrm{H}^{\alpha/2,\alpha}$) $\implies$

$$\mathbb{E}\left[\int_0^T |(b-c)(s,X_s)|^p \,\mathrm{d}s\right]^{1/p} \leqslant C\left(\int_0^T \left(\int_{\mathbb{R}^d} |(b-c)(s,x)|^q \,\mathrm{d}x\right)^{r/q} \mathrm{d}s\right)^{1/r}$$

with $\frac{d}{2q} + \frac{1}{r} < \frac{1}{p}$, or with Krylov estimates.

General heuristic

① Choose $\mathfrak{F}$, the space of terminal conditions
$\Longrightarrow$ choice of q $\Longrightarrow$ choice of p

② Choose $\mathfrak{M}$, the space of regularized drift, so that

- For some $\gamma > 0$,

$$\|b - b_\varepsilon\|_{X^{0,p}} \leqslant \mathrm{O}(\varepsilon^\gamma)$$

- For $b_\varepsilon \in \mathfrak{M}$ and $f \in \mathfrak{F}$.

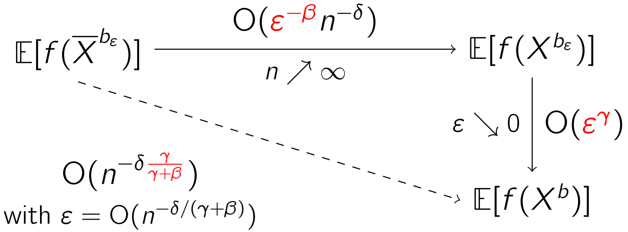
$$|\mathbb{E}[f(X_T^{b_\varepsilon})] - \mathbb{E}[f(\overline{X}_T^{b_\varepsilon})]| \leqslant \frac{C}{\varepsilon^\beta n^\delta}$$

③ Optimize over the choice of ε

$$\Longrightarrow \varepsilon = \mathrm{O}(n^{-\delta/(\gamma+\beta)})$$

$$\Longrightarrow |\mathbb{E}[f(X_T^{b})] - \mathbb{E}[f(\overline{X}_T^{b_\varepsilon})]| = \mathrm{O}(n^{-\delta\gamma/(\gamma+\beta)}).$$

General heuristic



Examples of terminal condition

☆ If $f \in \mathcal{C}_{\mathrm{sl}}$ then
$$\|\nabla v\|_{X^{0},2} \leqslant C\sqrt{\operatorname{Var} f(X^0_T)}$$

☆ If $d = 1$, $f \in \mathcal{C}_{\mathrm{sl}} \cap \mathcal{C}^1$, ∇f bounded then
$$\|\nabla u\|_{X^{b},\infty} \leqslant C\|\nabla f\|_\infty.$$

☆ Using the notion of fractional derivative (Geiss & Gobet), one may consider various values of q, even for f discontinuous.

Rate of convergence with smooth coefficients

Theorem. If $f \in \mathcal{C}^3$ with polynomial growth, $\sigma, b \in \mathcal{C}^{1,3}$ then

$$\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)] \leqslant \frac{C}{n}$$

C depends polynomially on the sup-norm of the derivatives of b (up to degree 4),

Rem. b bounded and $b_\varepsilon = b \star \rho_\varepsilon$ (mollifiers) $\implies \|\nabla^k b_\varepsilon\|_\infty \leqslant K\varepsilon^{-k}$.

We need to keep track of the dependence in the derivatives of b (gives the $\varepsilon^{-\beta}$ in the rate of conv. of the EM scheme with b_ε)

The proof relies on some idea introduced in *E. Clément, A. Kohatsu-Higa, D. Lamberton (2006)*.

Rate of convergence with smooth coefficients

Central idea of the proof.

Without drift terms, the idea is to write

$$\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)] = \mathbb{E}\left[\nabla f(\theta X_T + (1-\theta)\overline{X}_T)E_T\right]$$

where θ is a uniform in $[0,1]$ and

$$E_T = \int_0^t \nabla\left(\int_0^1 \sigma(s, \tau X_s + (1-\tau)\overline{X}_s)\,\mathrm{d}\tau\right) E_s\,\mathrm{d}W_s + \int_0^t (\sigma(s,\overline{X}_s) - \sigma(\varphi(s), \overline{X}_{\varphi(s)}))\,\mathrm{d}s.$$

Then use repeatedly the duality formula of Malliavin calculus to transform

$$\mathbb{E}\left[H \int_0^t u_s\,\mathrm{d}W_s\right] = \mathbb{E}\left[\int_0^t D_s H \cdot u_s\,\mathrm{d}s\right]$$

and get the desired control (long computations).

Rate of convergence with smooth coefficients

With a drift term, use Girsanov formula to write

$$\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X}_T)] = \mathbb{E}[\exp(L_T) f(X_T)] - \mathbb{E}[\exp(\overline{L}_T) f(\overline{X}_T)]$$

with

$$L_t = \int_0^t b^\top \sigma^{-1}(s, X_s)\, \mathrm{d}W_s - \frac{1}{2}\int_0^t b^\top a^{-1} b(s, X_s)\, \mathrm{d}s$$

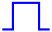
$$\overline{L}_t = \int_0^t b^\top \sigma^{-1}(\varphi(s), \overline{X}_{\varphi(s)})\, \mathrm{d}W_s - \frac{1}{2}\int_0^t b^\top a^{-1} b(\varphi(s), X_{\varphi(s)})\, \mathrm{d}s$$

and apply the same kind of computations.

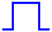
□

Examples

☆ $\sigma \in \mathfrak{M} = \mathcal{C}^{1,3}_{\mathrm{b}}$, $\mathfrak{F} = \mathcal{C}^{3}_{\mathrm{p}}$
$\Longrightarrow$ rate at most $n^{-\gamma/(\gamma+4)}$ when $\|b - b_\varepsilon\|_{X^0,p} \leqslant C\varepsilon^{-\gamma}$.

☆ $d = 1$, $b =$ , $\sigma \in \mathfrak{M} = \mathcal{C}^{1,3}_{\mathrm{b}}$, $\mathfrak{F} = \mathcal{C}^{3}_{\mathrm{p}}$
$\Longrightarrow$ rate at most $n^{-1/5+\varepsilon}$.

☆ $\sigma \in \mathfrak{M} = H^{\alpha/2,\alpha}$, $\mathfrak{F} = H^{2+\alpha}$
$\Longrightarrow$ rate at most $n^{-E(\alpha)\gamma/(\alpha+\gamma)}$ when $\|b - b_\varepsilon\|_{X^0,p} \leqslant C\varepsilon^{-\gamma}$.

☆ $d = 1$, $b =$ , $\sigma \in \mathfrak{M} = H^{\alpha/2,\alpha}$, $\mathfrak{F} = H^{2+\alpha}$
$\Longrightarrow$ rate at most $n^{-E(\alpha)/(\alpha+1)+\varepsilon}$

Case of constant diffusivity

$$X_t^b = x + B_t + \int_0^t b(X_s^b)\,\mathrm{d}s.$$

Thanks to Girsanov theorem,
the distributions of X^b, X^{b_ε}, $\overline{X}^b$ and $\overline{X}^{b_\varepsilon}$
are absolutely continuous wrt Wiener measure.

The perturbation formula may be adapted to

$$|\mathbb{E}[f(\overline{X}_T^b)] - \mathbb{E}[f(\overline{X}_T^{b_\varepsilon})]| \leqslant C(f)\|b - b_\varepsilon\|_{\mathrm{L}^p},\ \ p > d \vee 2.$$

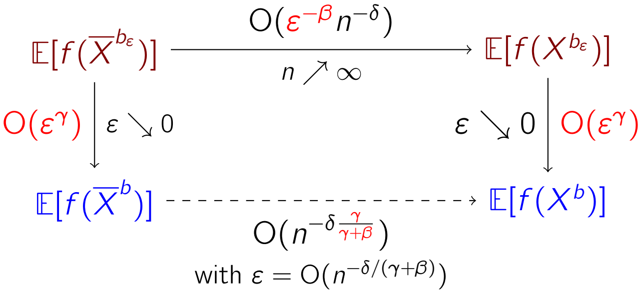
Various results may be given on

$$|\mathbb{E}[f(X_T^b)] - \mathbb{E}[f(\overline{X}_T^{b})]|$$

and not only on

$$|\mathbb{E}[f(X_T^b)] - \mathbb{E}[f(\overline{X}_T^{b_\varepsilon})]|$$

Constant diffusivity



Yet, this approach is sub-optimal

- A weak rate of order 1 could be achieved.

$$dX_t = dB_t + \begin{cases} -\theta & \text{if } X_t > 0 \\ 0 & \text{if } X_t = 0 \\ \theta & \text{if } X_t < 0 \end{cases}$$

$$|\mathbb{E}_0[f(X_T)] - \mathbb{E}_0[f(\overline{X}_T)]| \leqslant \frac{C}{n}$$

Proof. A lot of Taylor expansions and long computations. □

Conclusions

- Our approach relies on a perturbation formula and is then a "global" approach ($\neq$ local analysis around the discontinuity)
- It is flexible and allows to combine various results
- Allows to "separate" the effects of $\mathfrak{F}$ and $\mathfrak{M}$
- Mixes stochastic analysis and PDE arguments
- But provides only sub-optimal rates
- Still a lot of works to perform...