hausdorff CENTER FOR MATHEMATICS

universität bonn iam

Coupling approaches to high-dimensional continuous systems

Andreas Eberle

University of Bonn

February 2016

Outline

Introduction

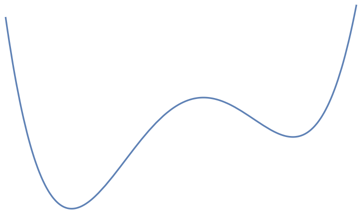
A. Overdamped Langevin dynamics

$$dX_t \;=\; b(X_t)\,dt \;+\; dB_t \qquad \text{in } \mathbb{R}^d$$

e.g. $b(x) = -\frac{1}{2}\nabla V(x) \implies$ Stationary distribution

$$\mu(x) = Z^{-1}\, e^{-V(x)}$$

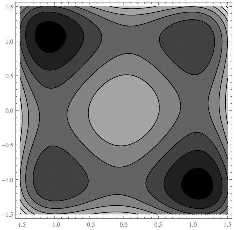
V not necessarily convex !



B. Mean-field interactions

$$dX_t^i \;=\; -\frac{1}{2}\nabla V(X_t^i)\,dt \;-\; \frac{\alpha}{n}\sum_{j\neq i}\nabla W(X_t^i - X_t^j)\,dt \;+\; dB_t^i \quad \text{in } \mathbb{R}^{d\cdot n}$$

e.g. $W(x) = |x|^2/2$, $\;V$ non-convex.



C. Euler scheme

Markov chain on $\mathbb{R}^d$ with transition step

$$X'_{\mathrm{Euler}} \;=\; x + h\,b(x) + \sqrt{h}\,Z, \quad Z \sim N(0, I),\; h > 0 \text{ stepsize},$$

e.g. $b(x) = -\frac{1}{2}\nabla V(x) \;\longrightarrow\;$ Stochastic gradient descent ,

stationary distribution $\mu_h \neq \mu$, $\mu_h \to \mu$ as $h \downarrow 0$.

D. Metropolis Adjusted Langevin Algorithm (MALA)

Assume $b(x) = -\frac{1}{2}\nabla V(x)$. Transition step:

$$X' = \begin{cases} X'_{\mathrm{Euler}} & \text{if } U \leq a(x, X'_{\mathrm{Euler}}), \\ x & \text{otherwise,} \end{cases} \quad U \sim \mathrm{Unif}(0,1) \text{ indep. of } Z$$

$$a(x, x') = \frac{\mu(x')p(x', x)}{\mu(x)p(x, x')} \wedge 1 \quad \text{Metropolis-Hastings accept. prob.}$$

Goal

Contractivity of transition kernels w.r.t. appropriate metrics on probability measures.

Consequences

- Quantitative non-asymptotic bounds for distance to equilibrium
- Variance bounds and concentration inequalities for ergodic averages (Joulin & Ollivier, AOP 2010)
- Stability under perturbations (Pillai & Smith, Rudolf & Schweizer)

Approaches

- L^2 decay, spectral gap, ...
- Decay of relative entropy, log. Sobolev inequalities, ...
- Couplings, Wasserstein distances, ...

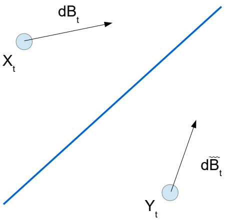
$$\begin{array}{ccc} & \text{Coupling} & \\ & \downarrow & \\ & \text{Metric on probability measures} & \leftarrow \quad \text{optimize} \\ & \downarrow & \\ & \text{Contraction rate} & \end{array}$$

Kantorovich (L^1 Wasserstein) distance based on metric ρ:

$$\mathcal{W}_\rho(\mu,\nu) \;=\; \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[\rho(X,Y)], \quad \mu,\nu \in \mathrm{Prob}(\mathbb{R}^d).$$

Couplings on $\mathbb{R}^d$

A. Reflection coupling



$$
\begin{aligned}
dX_t &= b(X_t)\,dt + dB_t \\
dY_t &= b(Y_t)\,dt + (I - 2e_t e_t^T)\,dB_t, \quad e_t = \frac{X_t - Y_t}{|X_t - Y_t|},
\end{aligned}
$$

for $t < T := \inf\{s \geq 0 : X_s = Y_s\}$, $Y_s = X_s$ for $t \geq T$.

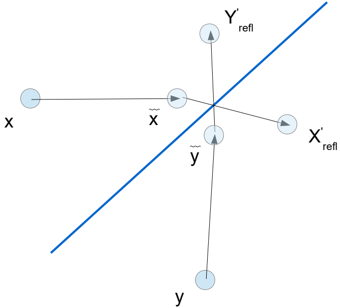
B. Componentwise reflection coupling

$$\begin{aligned} dX^i &= -\frac{1}{2}\nabla V(X^i)\,dt - \frac{\alpha}{n}\sum_{j\neq i}\nabla W(X^i - X^j)\,dt + dB^i, \\ dY^i &= -\frac{1}{2}\nabla V(X^i)\,dt - \frac{\alpha}{n}\sum_{j\neq i}\nabla W(X^i - X^j)\,dt + (I - 2\,1_{|X^i-Y^i|>\delta} e_i e_i^T)dB^i \end{aligned}$$

where $e_i = (X^i - Y^i)/|X^i - Y^i|$ and $\delta > 0$ small $(\delta \downarrow 0)$.

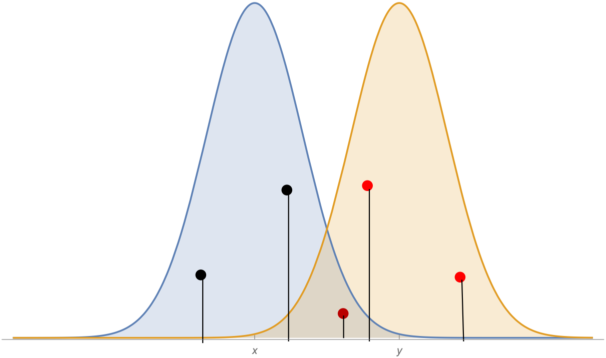
C. Optimized coupling for Euler

$$
\begin{aligned}
X' &= \hat{x} + \sqrt{h}\, Z, & \hat{x} &= x + hb(x), \\
Y'_{\mathrm{refl}} &= \hat{y} + \sqrt{h}\,(I - 2ee^T)Z, & e &= (\hat{x} - \hat{y})/|\hat{x} - \hat{y}|.
\end{aligned}
$$



Not good for small distances !

Optimized coupling for Euler



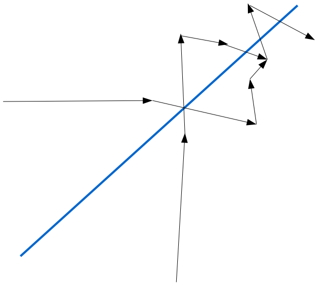
$$Y' = \begin{cases} X' & \text{if } \tilde{U} \leq \frac{\phi_{\hat{y},hI}(X')}{\phi_{\hat{x},hI}(Y')}, \\ Y'_{\mathrm{refl}} & \text{otherwise,} \end{cases} \quad \tilde{U} \sim \mathrm{Unif}(0,1) \text{ indep. of } Z.$$

Optimized coupling for Euler

- $Y' = X'$ with maximal probability.
- Optimal coupling of one step transition kernels for any Kantorovich distance based on a metric

$$\rho(x,y) = f(|x-y|) \quad \text{with } f : [0,\infty) \to [0,\infty) \text{ concave.}$$

- Correct discretization of reflection coupling !



D. Non-optimal coupling for MALA

$$X' = \begin{cases} X'_{\mathrm{Euler}} & \text{if } U \leq a(x, X'_{\mathrm{Euler}}), \\ x & \text{if } U > a(x, X'_{\mathrm{Euler}}), \end{cases}$$

$$Y' = \begin{cases} Y'_{\mathrm{Euler}} & \text{if } U \leq a(y, Y'_{\mathrm{Euler}}), \\ x & \text{if } U > a(y, Y'_{\mathrm{Euler}}), \end{cases}$$

$$U \sim \mathrm{Unif}(0,1) \text{ independent of } Z \text{ and } \tilde{U},$$

- Optimized coupling for proposals.
- Maximal coupling for Acceptance Rejection step.
- Nevertheless the combined coupling is not optimal.

Metrics and contraction rates

Kantorovich metrics

$$\mathcal{W}_\rho(\mu,\nu) \;=\; \inf_{X\sim\mu,Y\sim\nu} \mathbb{E}[\rho(X,Y)], \quad \mu,\nu \in \mathrm{Prob}(\mathbb{R}^d).$$

Kantorovich (L^1 Wasserstein) distance based on metric

$$\rho(x,y) \;=\; f(|x-y|) + \delta\,(V(x)+V(y))\,1_{x\neq y},$$

$f : \mathbb{R}_+ \to \mathbb{R}_+$ increasing, *concave*; V Lyapunov function.

A. Contraction rates for Langevin diffusions

A.E.: Reflection couplings and contraction rates for diffusions, PTRF online

$$R_t \;:=\; |X_t - Y_t|, \qquad (X_t, Y_t) \text{ reflection coupling.}$$

$$df(R) \;=\; \left(\frac{(X-Y)\cdot(b(X)-b(Y))}{R} f'(R) + 2f''(R)\right) dt \,+\, dM$$
$$\overset{!}{\leq} \; -cf(R)\,dt \,+\, dM$$

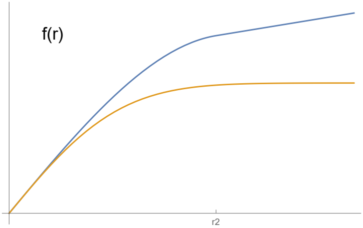
Then

$$\mathcal{W}_\rho(\mu_t, \nu_t) \;\leq\; E[f(R_t)] \;\leq\; e^{-ct} E[f(R_0)] \;=\; e^{-ct}\mathcal{W}_\rho(\mu_0, \nu_0).$$

Choose f in order to maximize the contraction rate c.

Example (One-sided Lipschitz condition)

$$(x-y)\cdot(b(x)-b(y)) \leq \begin{cases} J\cdot|x-y|^2 & \text{for } r\leq\mathcal{R}, \\ \text{const.}\cdot(1-J)\cdot|x-y|^2 & \text{for } r\geq\mathcal{R}. \end{cases}$$



- $J=0: \;\; f(r) \;=\; r-\gamma r^3 + a1_{r>0} \quad \text{for } r<r_2$
  $\Rightarrow \qquad c=\Omega(\mathcal{R}^{-2})$ independent of dimension
- $J>0: \;\; f(r) \;=\; \int_0^r \exp(-Ls^2/4)\, g(s)\, ds + a1_{r>0}$
  $\Rightarrow \qquad c=\Omega(J^{3/2}\mathcal{R}e^{-J\mathcal{R}^2/8})$ independent of dimension

Remark (Dimension dependence)

- $J = 0: \; f(r) \; = \; r - \gamma r^3 + a1_{r>0}$ for $r < r_2$

  $\Rightarrow \quad c = \Omega(\mathcal{R}^{-2})$ independent of dimension

- $J > 0: \; f(r) \; = \; \int_0^r \exp(-Ls^2/4)\, g(s)\, ds + a1_{r>0}$

  $\Rightarrow \quad c = \Omega(J^{3/2}\mathcal{R}e^{-J\mathcal{R}^2/8})$ independent of dimension

However, in applications often $\mathcal{R}$ will depend on the dimension !!!

B. Mean-field interactions

A.E.: Reflection couplings and contraction rates for diffusions, PTRF online

$$dX_t^i \;=\; -\frac{1}{2}\nabla V(X_t^i)\,dt \;-\; \frac{\alpha}{n}\sum_{j\neq i}\nabla W(X_t^i - X_t^j)\,dt \;+\; dB_t^i$$

Choice of metric:

$$\rho(x,y) \;=\; \sum_{i=1}^{n} f(|x^i - y^i|), \qquad f \text{ as above.}$$

Theorem (Contractivity for small interactions)

Suppose that $\sup\|\nabla^2 W\| < \infty$*. Then there exist constants* $\theta, c \in (0,\infty)$ *that do not depend on the dimension such that*

$$\mathcal{W}_\rho(\mu_t,\nu_t) \;\le\; e^{(\theta\alpha - c)t}\,\mathcal{W}_\rho(\mu_0,\nu_0)$$

C,D. Contraction rates for Markov chains on $\mathbb{R}^d$

A.E.: Couplings and contraction rates for Markov chains, Work in progress.

GOAL: Carry over approach from diffusions to Markov chains.

(X', Y') arbitrary coupling of transition steps from (x, y).

$$r = |x - y|, \qquad R' = |X' - Y'|, \qquad \Delta R = R' - r.$$

Taylor expansion yields

$$f(R') - f(r) \le \Delta R\, f'(r) + \frac{1}{2}(\Delta R\, 1_{\Delta R \in I_r})^2 \sup_{I_r} f'' - a\, 1_{R'=0}$$

where we choose

$$I_r = \begin{cases} (r - \epsilon, r) & \text{for } r > 2\epsilon, \\ (0, 3\epsilon) & \text{for } r \le 2\epsilon, \end{cases} \qquad \epsilon > 0 \text{ fixed.}$$

$$\beta(x,y) := \mathbb{E}_{x,y}[\Delta R] \quad \text{average increase of distance per step,}$$
$$\alpha(x,y) := \mathbb{E}_{x,y}[(\Delta R)^2; \Delta R \in I_r] \quad \text{fluctuations,}$$
$$\pi(x,y) := \mathbb{P}_{x,y}[R' = 0].$$

$$\mathbb{E}_{x,y}[f(R') - f(r)] \leq \beta(x,y)\, f'(r) + \frac{1}{2}\alpha(x,y) \sup_{I_r} f'' - a\,\pi(x,y) \stackrel{!}{\leq} -c\, f(r)$$

Assumptions

(A1) $\sup_{|x-y|\leq r_1} \frac{\beta(x,y)}{\alpha(x,y)} < \infty \quad \forall\, r_1 < \infty.$

(A2) $\inf_{|x-y|\leq r_1} \frac{\alpha(x,y)}{|x-y|} > 0 \quad \forall\, r_1 < \infty.$

(A3) $\limsup_{|x-y|\to\infty} \frac{\beta(x,y)}{|x-y|} < 0, \quad \limsup_{|x-y|\to\infty} \frac{\alpha(x,y)}{|x-y|} < \infty.$

Assumptions

(A1) $\sup_{|x-y|\le r_1} \frac{\beta(x,y)}{\alpha(x,y)} < \infty \quad \forall\ r_1 < \infty.$

(A2) $\inf_{|x-y|\le r_1} \frac{\alpha(x,y)}{|x-y|} > 0 \quad \forall\ r_1 < \infty.$

(A3) $\limsup_{|x-y|\to\infty} \frac{\beta(x,y)}{|x-y|} < 0, \qquad \limsup_{|x-y|\to\infty} \frac{\alpha(x,y)}{|x-y|} < \infty.$

Theorem (Contractivity)

Suppose that (A1), (A2) and (A3) hold. Then there exists an explicit concave function $f : \mathbb{R}_+ \to \mathbb{R}_+$ *and an explicit constant* $c > 0$ *depending only on the suprema and infima considered in (A1), (A2) and (A3) such that*

$$\mathbb{E}_{x,y}[\rho(X',Y')] \le (1-c)\,\rho(x,y) \quad \text{for any } x,y \in \mathbb{R}^d, \ \text{i.e.,}$$

$$\mathcal{W}_\rho(\mu p, \nu p) \le (1-c)\mathcal{W}_\rho(\mu,\nu) \quad \text{for any } \mu,\nu \in Prob(\mathbb{R}^d).$$

Consequences of Kantorovich contraction

- Quantitative replacement for classical minorization condition.
- Distance from stationary distribution.
- Non-asymptotic bounds for MSE of ergodic averages, concentration estimates (Joulin/Ollivier AAP 2010)
- Central limit theorem (Komorowski/Walczuk SPA 2011)
- Perturbation theory (!!!) (Pillai/Smith, Rudolf/Schweizer)

Alternative Assumptions

(A1') $\sup_{|x-y|\leq r_1} \beta(x,y) < \infty \quad \forall\, r_1 < \infty.$

(A2') $\exists\, r_0 : \inf_{r_0 \leq |x-y| \leq r_1} \alpha(x,y) > 0$ and $\inf_{|x-y|<r_0} \pi(x,y) > 0.$

(A3') $\exists\, C, \lambda \in (0,\infty), V : \mathbb{R}^d \to \mathbb{R}_+ : \; pV \leq C + (1-\lambda)V.$

(Lyapunov condition)

Theorem (Contractivity)

Suppose that (A1'), (A2') and (A3') hold. If V is "growing rapidly enough as $|x| \to \infty$" then a similar result as above holds with

$$\rho(x,y) = f(|x-y|) + \delta \cdot (V(x) + V(y))\, 1_{x \neq y}.$$

See also [A.E., A. Guillin, R. Zimmer] in continuous time.

Application to Euler scheme

Euler discretization of $dX = b(X)\,dt + dB$:

$$x \;\to\; \hat{x} = x + hb(x) \;\to\; X' = \hat{x} + \sqrt{h}\,Z, \quad h > 0,\ Z \sim N(0, I).$$

Sychronuous coupling: $Y' = \hat{y} + \sqrt{h}Z$

$$R' \;=\; |X' - Y'| \;=\; |\hat{x} - \hat{y}| \;=\; \hat{r}.$$

$$\begin{aligned} \beta(x,y) &= \mathbb{E}_{x,y}[R' - r] \;=\; \hat{r} - r \;=\; \mathrm{Lip}(b)\,h\,r, \\ \alpha(x,y) &\leq \mathbb{E}_{x,y}[(R' - r)^2] \;=\; (\hat{r} - r)^2 \end{aligned}$$

$$\frac{\beta(x,y)}{\alpha(x,y)} \;\geq\; \frac{1}{\hat{r} - r} \;=\; \Omega\left(\frac{1}{hr}\right)$$

Not good !

Euler discretization of $dX = b(X)\,dt + dB$:

$$x \;\to\; \hat{x} = x + hb(x) \;\to\; X' = \hat{x} + \sqrt{h}\,Z, \quad h > 0,\; Z \sim N(0, I).$$

Reflection coupling: $Y' = \hat{y} + \sqrt{h}\,(I - 2ee^T)Z$

$$R' \;=\; |X' - Y'| \;=\; |\hat{x} - \hat{y} + 2\sqrt{h}ee^T Z|.$$

$$\beta(x,y) \;=\; \mathbb{E}_{x,y}[R' - r] \;=\; \Omega(\sqrt{h}) \qquad \text{for small } r,$$
$$\alpha(x,y) \;\leq\; \mathbb{E}_{x,y}[(R' - r)^2] \;=\; O(h) \qquad \text{for small } r,$$

$$\frac{\beta(x,y)}{\alpha(x,y)} \;=\; \Omega\left(h^{-1/2}\right)$$

Not good for small r!

Euler discretization of $dX = b(X)\,dt + dB$:

$$x \;\to\; \hat{x} = x + hb(x) \;\to\; X' = \hat{x} + \sqrt{h}\,Z, \qquad h > 0,\; Z \sim N(0, I).$$

Optimized coupling:

$$\begin{aligned} \beta(x,y) &\leq \beta_{\text{synchr.}}(x,y) \;\leq\; Lhr, \\ \alpha(x,y) &= \Omega(hr) \qquad \text{(since } \mathbb{P}[\text{reflection}] = \Omega(r)\text{)}, \end{aligned}$$

$$\frac{\beta(x,y)}{\alpha(x,y)} = O(1) \qquad \text{Good !}$$

Assumptions

There exist constants $\mathcal{R}, J, K, L \in [0, \infty)$ such that

(A1) $(x - y) \cdot (b(x) - b(y)) \;\leq\; J\,|x - y|^2$ for any $x, y \in \mathbb{R}^d$.

(A2) $(x - y) \cdot (b(x) - b(y)) \;\leq\; -K\,|x - y|^2$ for $|x - y| \geq \mathcal{R}$.

(A3) $|b(x) - b(y)| \;\leq\; L\,|x - y|$ for any $x, y \in \mathbb{R}^d$.

Theorem (Contractivity of Euler)

Suppose that (B1), (B2) and (B3) hold. Then there exists an explicit concave function $f : \mathbb{R}_+ \to \mathbb{R}_+$ *and explicit constants* $c, h_0 > 0$ *depending only on* $\mathcal{R}, J, K, L$ *auch that for* $h < h_0$*,*

$$\mathbb{E}_{x,y}[\rho(X', Y')] \;\leq\; (1 - ch)\,\rho(x, y) \quad \text{for any } x, y \in \mathbb{R}^d, \text{ i.e.,}$$

$$\mathcal{W}_\rho(\mu p, \nu p) \;\leq\; (1 - ch)\mathcal{W}_\rho(\mu, \nu) \quad \text{for any } \mu, \nu \in Prob(\mathbb{R}^d).$$

Choice of metric $\rho(x,y) = f(|x-y|)$:

$$f(r) \;=\; \int_0^r g(s)\, e^{-c_0 J s^2 - O(\sqrt{h})}\, ds \qquad \text{for } r \le r_1,$$

where $g : \mathbb{R}_+ \to \mathbb{R}_+$ with $1/2 \le g(s) \le 1$ for any s.

Remarks (Dimension dependence)

- Everything is explicit. The function f and the constant c depend only on $\mathcal{R}, J, K$ and L.
- Dimension-free bound provided dissipative for $|x-y| \ge \mathcal{R}$.
- However, often $\mathcal{R}$ depends on the dimension !
- Sometimes, there exists a weaker norm $\|\cdot\|_- \le |\cdot|$ such that dissipativity holds for $\|x-y\|_- \ge \mathcal{R}$ with $\mathcal{R}$ independent of the dimension.

  $\to$ Two-scale approach [R. Zimmer, work in progress]

Consequences

- Non-asymptotic bounds and concentration inequalities for ergodic averages of Euler chain.
- Relatively precise quantification of distance of Euler chain from its stationary distribution. This is different from stationary distribution of SDE !
- Approach can be combined with Lyapunov functions.

Metropolis-adjusted Langevin algorithm

Idea

- MALA as perturbation of Euler
- Contractivity of Euler
  $\Rightarrow$ Contractivity of MALA w.r.t. modified metric
- Ansatz: $f(r) = f_0(r) + a\,1_{r>0}$

$$
\begin{aligned}
\mathbb{E}_{x,y}\left[f(R'_{\mathrm{MALA}}) - f(r)\right] \;=\;& \mathbb{E}_{x,y}\left[f_0(R'_{\mathrm{MALA}}) - f_0(R'_{\mathrm{Euler}})\right] \\
&+ \mathbb{E}_{x,y}\left[f_0(R'_{\mathrm{Euler}}) - f_0(r)\right] \\
&- a \cdot \mathbb{P}_{x,y}\left[R'_{\mathrm{MALA}} = 0\right].
\end{aligned}
$$

$$\begin{aligned}\mathbb{E}_{x,y}\left[f(R'_{\mathrm{MALA}})-f(r)\right] &= \mathbb{E}_{x,y}\left[f_0(R'_{\mathrm{MALA}})-f_0(R'_{\mathrm{Euler}})\right]\\ &\quad+\mathbb{E}_{x,y}\left[f_0(R'_{\mathrm{Euler}})-f_0(r)\right]\\ &\quad-a\cdot\mathbb{P}_{x,y}\left[R'_{\mathrm{MALA}}=0\right].\end{aligned}$$

Bound for Euler

$$\mathbb{E}_{x,y}\left[f_0(R'_{\mathrm{Euler}})-f_0(r)\right] \;\le\; -c\,h\,f_0(r).$$

Bound for MALA rejection probability

[Bou Rabee/van den Eijnden CPAM 2010, AE AAP 2014]

$$E_{x,y}\left[f_0(R'_{\mathrm{MALA}})-f_0(R'_{\mathrm{Euler}})\right] \;\le\; E_{x,y}\left[\left|R'_{\mathrm{MALA}}-R'_{\mathrm{Euler}}\right|\right] \;=\; O(d\,h^{3/2})$$