Quantitative approximations of evolving probability measures

Andreas Eberle

March 28, 2011

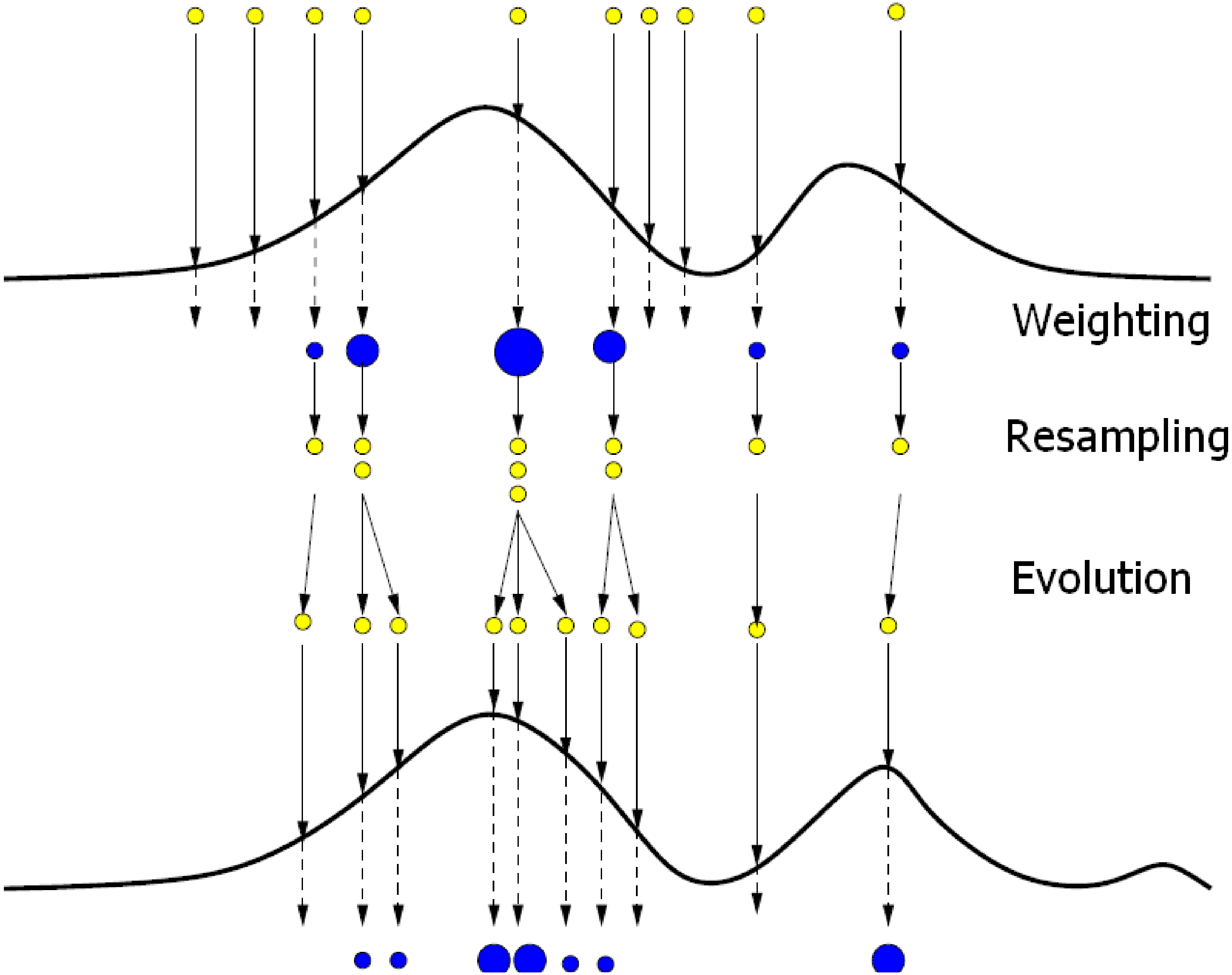
1 MOTIVATION; SEQUENTIAL MC SAMPLERS

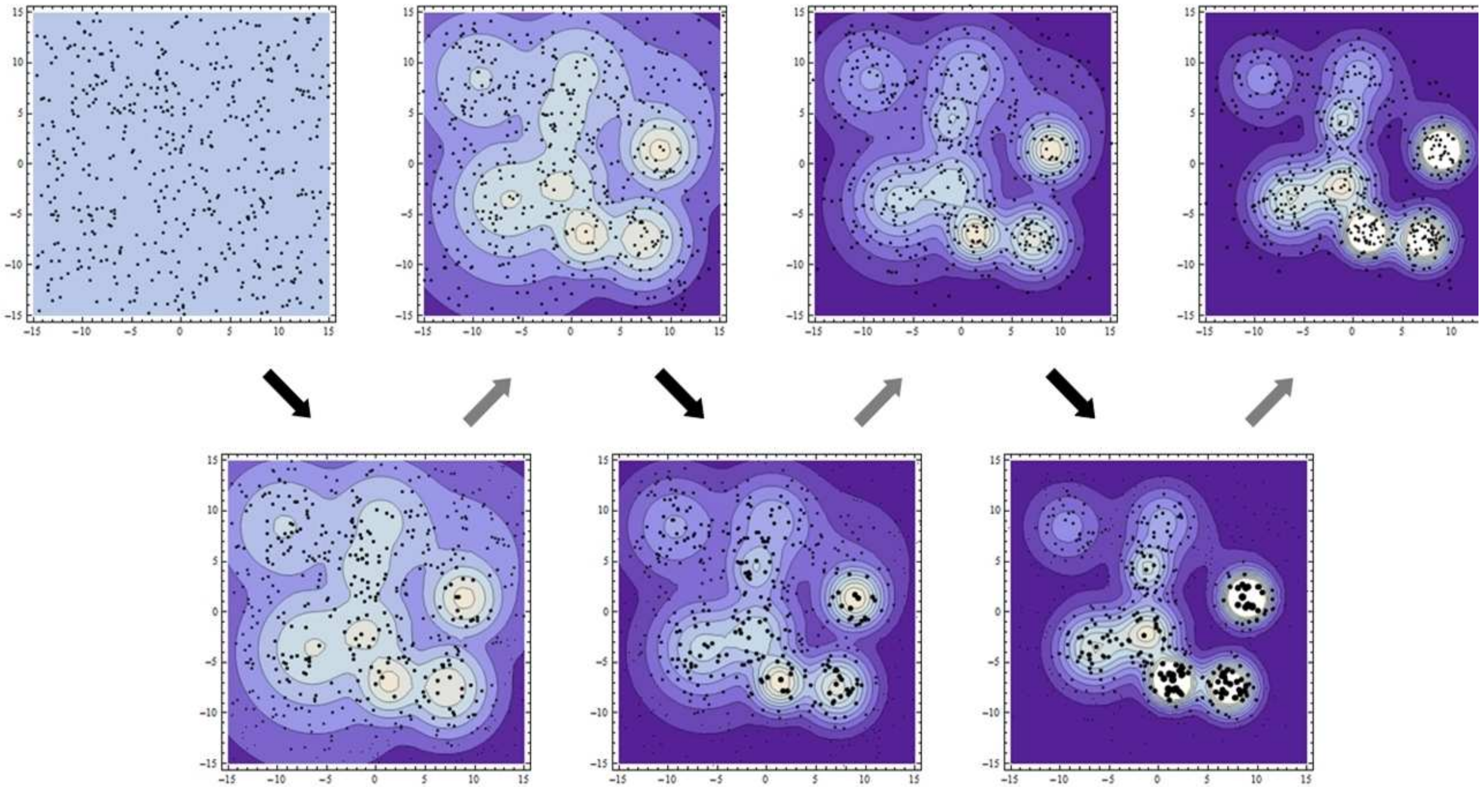
$\mu_0, \mu_1, \ldots, \mu_k$ probability measures, $\lambda_0, \lambda_1, \ldots, \lambda_k \in \mathbb{N}$,
$p_1, p_2, \ldots, p_k$ Markov kernels s.t. $\mu_t p_t = \mu_t$

ALGORITHM (SMCMC with multinomial resampling).

- Initialization:
  - Sample X_0^i $(1 \leq i \leq N)$ i.i.d. $\sim \mu_0$, set $\eta_0^N := N^{-1} \sum_{i=1}^N \delta_{X_0^i}$
- Step: For $t := 1$ to k do
  - **SIR**: Sample X_t^j i.i.d. $\sim \sum_{i=1}^N w_t^i \cdot \delta_{X_{t-1}^i}$, $w_t^i \propto d\mu_t / d\mu_{t-1}(X_{t-1}^i)$
  - **MCMC**: For $m := 1$ to λ_t do
    - $*$ Sample Y_t^i condit. indep. $\sim p_t(X_t^i, \cdot\,)$; set $X_t^i := Y_t^i$
  - Set $\eta_t^N := N^{-1} \sum_{i=1}^N \delta_{X_t^i}$

P. Del Moral, A. Doucet, A. Jasra, J. R. Statist. Soc. B **68** (2006)


Weighting
Resampling
Evolution



15
10
5
0
-5
-10
-15

LLN: $\int f\, d\eta_t^N \;=\; \frac{1}{N}\sum_{i=1}^N f(X_t^i) \;\to\; \int f\, d\mu_t$ as $N \to \infty$.

CLT: $\sqrt{N} \cdot (\int f\, d\eta_t^N - \int f\, d\mu_t) \;\to\; N(0, \sigma_t(f)^2)$.

REFERENCES:

- O. Cappé, E. Moulines, T. Ryden. Inference in Hidden Markov Models, Springer 2005.
- P. Del Moral. Feynman-Kac Formulae, Springer 2004.
- N. Chopin. CLT for sequential SMC methods and its application to Bayesian inference. *Annals of Statistics* **32 (6)** (2004)
- H. R. Künsch. Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis. *Annals of Statistics* **33(5):** 1983-2021, 2004.

Related methods:

- *Parallel tempering*, Geyer (1991)
- *Equi-energy sampler*, Kou, Zhou, Wong, Annals of Statistics **34** (2006)

Goal: “Feasible” non-asymptotic error bounds (N fixed)

- not too far off in simple examples (e.g. moving Gaussians)
- applicable in high dimensions (at least for product models)
- extension to simple multimodal cases

Setup here: Continuous time, discrete state space
A.E., C. Marinelli. Quantitative approximations of evolving probability measures and sequential MCMC methods. Submitted October 2010.

Extension to discrete time and continuous state spaces:
N. Schweizer. Non-asymptotic error bounds for sequential MCMC. PhD Thesis, approx. June 2011.

2 EVOLVING PROBABILITY MEASURES IN CONTINUOUS TIME

Aim : Sequential estimation / approximation of probability measures

$$\mu_t(x) \; \propto \; \exp\left(-U_t(x)\right) \; \nu(x), \qquad t \geq 0,$$

on a finite state space S.

$$H_t(x) \; := \; -\frac{\partial}{\partial t} \log \mu_t(x) \; = \; \frac{\partial}{\partial t} U_t(x) - \left\langle \frac{\partial}{\partial t} U_t, \, \mu_t \right\rangle.$$

$$\mu_t(x) \; \propto \; \exp\left(- \int_0^t H_s(x) \, ds \right) \; \nu(x)$$

Note that $\langle H_t \, , \, \mu_t \rangle \; = \; 0$ for any $t \geq 0$.

GENERAL SMC APPROACH :

1) Find evolution equation satisfied by $(\mu_t)_{t\geq 0}$ *(Fokker-Planck equation)*
2) Discretize this equation by interacting particle system.

EXAMPLE.

- $\mu_t \equiv \mu$ satisfies the Fokker-Planck equation

$$\frac{\partial}{\partial t}\mu_t \;=\; \mathcal{L}^* \mu_t$$

for any Markov generator $\mathcal{L}$ such that $\mathcal{L}^*\mu = 0$.

- Discretization $\rightsquigarrow$ MCMC

FOKKER-PLANCK EQUATION:

Let $\mathcal{L}_t,\ t \geq 0,$ be generators of a time-inhomogeneous Markov process on S satisfying

$$\mu_t(x)\mathcal{L}_t(x,y) \;=\; \mu_t(y)\mathcal{L}_t(y,x) \quad \text{(detailed balance).}$$

E.g. Metropolis-Hastings with proposal distribubtions $K_t(x,y)$:

$$\mathcal{L}_t(x,y) \;=\; K_t(x,y) \cdot \min\left(\frac{\mu_t(y)K_t(y,x)}{\mu_t(x)K_t(x,y)},\, 1\right).$$

In particular,

$$\mathcal{L}_t^* \mu_t \;=\; \mu_t \mathcal{L}_t \;=\; 0 \quad \text{(infinitesimal stationarity).}$$

FOKKER-PLANCK EQUATION:

For any continuous function $\lambda : [0, \infty) \to \mathbb{R}$, the function $t \mapsto \mu_t$ is a solution of the evolution equations

$$\frac{\partial}{\partial t}\nu_t = \lambda_t \mathcal{L}_t^* \nu_t - H_t \nu_t , \tag{1}$$

$$\frac{\partial}{\partial t}\eta_t = \lambda_t \mathcal{L}_t^* \eta_t - H_t \eta_t + \langle H_t, \eta_t \rangle \eta_t . \tag{2}$$

Proof.

- $\partial \mu_t / \partial t = -H_t \mu_t$.
- $L_t^* \mu_t = 0$ for any $t \geq 0$ by detailed balance.
- $\langle H_t, \mu_t \rangle = -\langle \partial_t \log \mu_t, \mu_t \rangle = \frac{d}{dt}\mu_t(S) = 0.$

DISCRETIZATION OF FOKKER-PLANCK EQUATION:

$X_t^N = (X_{t,1}^N, \ldots, X_{t,N}^N)$ Markov process on S^N with generator

$$\begin{aligned} \mathcal{L}_t^N \varphi(x_1, \ldots, x_N) &= \lambda_t \sum_{i=1}^N \mathcal{L}_t^{(i)} \varphi(x_1, \ldots, x_N) \\ &+ \frac{1}{N} \sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))^+ \cdot \big(\varphi(x^{i \to j}) - \varphi(x)\big) \,, \end{aligned}$$

$\mathcal{L}_t^{(i)}$ action of $\mathcal{L}_t$ on i th component.

- Independent Markov chain moves with generator $\lambda_t \cdot \mathcal{L}_t$
- $X_{t,i}^N$ replaced by $X_{t,j}^N$ with rate $\frac{1}{N}(H_t(X_{t,i}^N) - H_t(X_{t,j}^N))^+$

DISCRETIZATION OF FOKKER-PLANCK EQUATION:

$$\frac{\partial}{\partial t}\eta_t \;=\; \lambda_t \mathcal{L}_t^* \eta_t \;-\; H_t \eta_t \;+\; \langle H_t, \eta_t \rangle\, \eta_t$$

The empirical distributions

$$\eta_t^N \;=\; \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{t,i}^N} \qquad t \geq 0,$$

yield a discretization of this equation:

$$\frac{\partial}{\partial t}\mathbb{E}\left[\langle f, \eta_t^N \rangle\right] \;=\; \mathbb{E}\left[\langle f\,,\, \lambda_t \mathcal{L}_t^* \eta_t^N \;-\; H_t \eta_t^N \;+\; \langle H_t, \eta_t^N \rangle\, \eta_t^N \rangle\right]$$

LLN / Scaling limit: If $X_{0,i}^N$ $(i = 1, \ldots N)$ are i.i.d. $\sim \mu_0$, then

$$\langle f, \eta_t^N \rangle \;\approx\; \mathbb{E}[\langle f, \eta_t^N \rangle] \;\approx\; \langle f, \mu_t \rangle \qquad \text{for large } N.$$

ESTIMATORS FOR μ_t: $X_{0,i}^N$ i.i.d. $\sim \mu_0$

$$\eta_t^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{t,i}^N}\,, \qquad \nu_t^N := \exp\left(-\int_0^t \langle H_s, \eta_s^N\rangle\, ds\right)\eta_t^N\,.$$

- *Law of Large Numbers:* $\langle f, \eta_t^N\rangle,\ \langle f, \nu_t^N\rangle \to \langle f, \mu_t\rangle$ a.s. as $N \to \infty$.
- ν_t^N is an ***unbiassed*** estimator for μ_t.
- MSE w.r.t. η_t^N can be controlled by MSE w.r.t. ν_t^N.

REFERENCES FOR CONTINUOUS TIME CASE:

- P. Del Moral, L. Miclo. Branching and Interacting Particle Systems Approx. of Feynman-Kac Formulae (2000)
- M. Rousset. On the control of an interacting particle approximation of Schrödinger ground states. *SIAM J. Math. An.* **38(3)** (2006)

3 QUANTITATIVE ERROR BOUNDS

$$\varepsilon_t^{N,p} := \sup \left\{ \mathbb{E}\left[\left| \langle f, \nu_s^N \rangle - \langle f, \mu_s \rangle \right|^2 \right] \,:\, s \in [0,t],\ \|f\|_{L^p(\mu_s)} \leq 1 \right\},\ p \in [2,\infty].$$

- **Aim:** Feasible bounds for $\varepsilon_t^{N,p}$ for a **fixed** number N of replicas.
- **Tool:** L^p estimates and L^p/L^q bounds for *Feynman-Kac propagators*.

FEYNMAN-KAC PROPAGATOR:

Define $q_{s,t}f$ as solution of backward equation

$$\frac{\partial}{\partial s} q_{s,t}f = -\lambda_s \mathcal{L}_s q_{s,t}f - H_s q_{s,t}f, \qquad q_{t,t}f = f,$$

Feynman-Kac representation:

$$q_{s,t}f(x) = \mathbb{E}_{s,x}\Big[e^{-\int_s^t H_r(X_r)\,dr} f(X_t)\Big],$$

where $(X_t, \mathbb{P}_{s,x})$ is Markov process with gen. $\lambda_t \mathcal{L}_t$ and init. cond. $X_s = x$.

- L^p estimates and L^p/L^q bounds for $q_{s,t}$ have been derived in:

  A.E., C. Marinelli. L^p estimates for Feynman-Kac propagators with time-dependent reference measures. *J. Math. Anal. Appl.* **365**, 2010.

AN EXPRESSION FOR THE VARIANCE:

THEOREM. For any function $f : S \to \mathbb{R}$,

$$\mathbb{E}\left[\langle f, \nu_t^N \rangle\right] = \langle f, \mu_t \rangle, \qquad \text{and}$$

$$\mathbb{E}\left[\left|\langle f, \nu_t^N \rangle - \langle f, \mu_t \rangle\right|^2\right] = \frac{1}{N}\mathrm{Var}_{\mu_t}(f) + \frac{1}{N}\int_0^t \mathbb{E}\left[V_{s,t}^N(f)\right] ds\,,$$

where

$$V_{s,t}^N(f) = -\,\langle H_s (q_{s,t} f)^2, \nu_s^N \rangle \langle 1, \nu_s^N \rangle - \langle H_s, \nu_s^N \rangle \langle q_{s,t} f^2 - (q_{s,t} f)^2, \nu_s^N \rangle$$
$$+ \frac{1}{2} \iint |H_s(z) - H_s(y)| (q_{s,t} f(z) - q_{s,t} f(y))^2 \, \nu_s^N(dy)\, \nu_s^N(dz).$$

Proof. Del Moral, Miclo 2000, EM 2010

- Define $A^f_{s,t} := \langle q_{s,t} f, \nu^N_s \rangle$.
- Then $A^f_{t,t} = \langle f, \nu^N_t \rangle$ and $A^f_{0,t} = \langle q_{0,t} f, \nu^N_0 \rangle$. Hence

$$\langle f, \nu^N_t \rangle - \langle f, \mu_t \rangle \;=\; \underbrace{A^f_{t,t} - A^f_{0,t}}_{?} + \underbrace{\langle q_{0,t} f, \nu^N_0 \rangle - \langle q_{0,t} f, \mu_0 \rangle}_{\text{standard MC error}}$$

- $s \mapsto A^f_{s,t}$ is a martingale.
- Compute $\langle A^f_{s,t} \rangle$, ...

BOUNDING THE MSE:

$$\mathbb{E}\left[\left|\langle f,\nu_t^N\rangle-\langle f,\mu_t\rangle\right|^2\right] \;=\; N^{-1}\mathrm{Var}_{\mu_t}(f)+N^{-1}\int_0^t \mathbb{E}\left[V_{s,t}^N(f)\right]\,ds\,,$$

$$V_{s,t}^N(f) \;=\; -\langle H_s(q_{s,t}f)^2,\nu_s^N\rangle\langle 1,\nu_s^N\rangle \;+\;\cdots$$

Decomposing $\nu_s^N=\mu_s+(\nu_s^N-\mu_s)$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[V_{s,t}^N(f)\right]\\
&\quad= \quad \underbrace{-\langle H_s(q_{s,t}f)^2,\mu_s\rangle}_{\text{asymptotic variance}} \;-\underbrace{\mathbb{E}\left[\langle H_s(q_{s,t}f)^2,\nu_s^N-\mu_s\rangle\langle 1,\nu_s^N-\mu_s\rangle\right]}_{\text{non-asymptotic correction}}+\cdots\\
&\quad\le \quad -\langle H_s(q_{s,t}f)^2,\mu_s\rangle+\varepsilon_t^{N,p}\cdot\left\|H_s(q_{s,t}f)^2\right\|_{L^p(\mu_s)}\cdot\|1\|_{L^p(\mu_s)}\;.
\end{aligned}$$

BOUNDING THE MSE:

- In order to obtain a closed estimate for $\varepsilon_t^{N,p}$, we have to bound the right hand side in terms of $\|f\|_{L^p(\mu_t)}$.
- Therefore, we need estimates of the form

$$
\begin{aligned}
-\langle H_s(q_{s,t}f)^2, \mu_s\rangle &\leq const. \cdot \|f\|^2_{L^p(\mu_t)} && (\rightsquigarrow \text{ estimate for asympt. var.})\\
\left\|H_s(q_{s,t}f)^2\right\|_{L^p(\mu_s)} &\leq const. \cdot \|f\|^2_{L^p(\mu_t)} && (\rightsquigarrow \text{ estimate for correction})
\end{aligned}
$$

A first attempt: $p = \infty$

- For $p = \infty$ we need an estimate of the form

$$\sup |q_{s,t}f| \;\leq\; C(\infty,\infty) \cdot \sup |f| \qquad \forall\ f$$

- The optimal constant in this estimate is

$$C(\infty,\infty) \;=\; \sup q_{s,t}1 \;=\; \sup_x \sum_y q_{s,t}(x,y)\ .$$

- For Markov kernels, $C(\infty,\infty)=C(1,1)=1$, but for Feynman-Kac propagators this is not the case.

- In some applications, feasible estimates for $C(\infty,\infty)$ exist. In general, however, it is not clear if and how $C(\infty,\infty)$ can be bounded efficiently.

A second attempt: $p\in(2,\infty)$

- For $p\in(2,\infty)$ we need in particular an estimate of the form

$$\|q_{s,t}f\|_{L^{2p}(\mu_s)} \;\leq\; C(p,2p)\cdot\|f\|_{L^p(\mu_t)}$$

  to control the correction term.

- Such estimates with $C(p,2p)\;\leq\;1$ are closely related to *logarithmic Sobolev inequalities (LSI)*.

- However, the estimate may hold with a reasonably sized constant $C(p,2p)>1$ even when an LSI does not hold globally !

4 NON-ASYMPTOTIC BOUNDS UNDER GLOBAL CONDITIONS

Fix $t_0 \in (0, \infty)$ (length of time interval), $p \in (6, \infty)$, $q \in (p, \infty)$, and let

$$\omega = \sup_{t \in [0,t_0]} \operatorname{osc}(H_t) \; ; \quad K_t = \int_0^t \|H_s\|_{L^q(\mu_s)} \, ds$$

$$C_t = \sup_{\langle f, \mu_t \rangle = 0} \frac{\int f^2 \, d\mu_t}{\mathcal{E}_t(f, f)}$$ Poincaré constant (inverse spectral gap)

$$\gamma_t = \sup_{\langle f^2, \mu_t \rangle = 1} \frac{\int f^2 \, \log |f| \, d\mu_t}{\mathcal{E}_t(f, f)}$$ Log-Sobolev constant

where $\mathcal{E}_t(f, f)$ is the *Dirichlet form*

$$\mathcal{E}_t(f, f) = -(f, \mathcal{L}_t f)_{L^2(\mu_t)} = \frac{1}{2} \sum_{x,y} (f(y) - f(x))^2 \, \mu_t(x) \, \mathcal{L}_t(x, y).$$

THEOREM. Suppose that

$$N \geq 40 \cdot \max\left(K_{t_0},\ 1\right), \qquad \text{and}$$

$$\lambda_t \geq \omega \cdot \max\left(\frac{p}{4} \cdot \left(1 + t \cdot \frac{p+3}{4}\right) \cdot C_t\ ,\ a(p,q) \cdot \gamma_t\right) \qquad \forall t \in [0, t_0].$$

Then

$$\varepsilon_t^{N,p} \leq \frac{2 + 8\,K_t}{N} \cdot \left(1 + \frac{16\,K_t}{N}\right) \qquad \forall\ t \in [0, t_0],$$

and, in particular,

$$\mathrm{Var}\ \left(\langle f, \nu_t^N \rangle\right) \leq \left(\mathrm{Var}_{\mu_t}(f) + \int_0^t V_{s,t}(f) + \|f\|^2_{L^p(\mu_t)}\right) N^{-1} + R_t \cdot \|f\|^2_{L^p(\mu_t)}\ N^{-2}$$

with explicit constants R_t.

$$V_{s,t}(f) \;=\; -\langle H_s(q_{s,t}f)^2,\, \mu_s\rangle \,+\, \int |H_s(x)|\left(q_{s,t}f(y) - q_{s,t}f(x)\right)^2 \mu_s(dx)\mu_s(dy)$$

COROLLARY. Under the conditions from the theorem,

$$\begin{aligned} E\ & \left[\left|\langle f, \eta_t^N\rangle - \langle f, \mu_t\rangle\right|\right] \\ \le\ & \left(\mathrm{Var}_{\mu_t}(f) + \int_0^t V_{s,t}(f) + \|f - \langle f, \mu_t\rangle\|^2_{L^p(\mu_t)}\right)^{1/2} N^{-1/2} \\ & + \tilde{R}_t \cdot \|f - \langle f, \mu_t\rangle\|_{sup}\, N^{-1}. \end{aligned}$$

with explicit constants $\tilde{R}_t$.

EXAMPLE 1: Moving Gaussians, $d = 1$

$S = \{a, a+1, \ldots, a+\Delta-1\} \subseteq \mathbb{Z}$,

$$\mu_t(x) \;\propto\; \exp\left(-\frac{|x-m_t|^2}{2\sigma_t^2}\right) \;,$$

$$\mathcal{L}_t(x,y) \;=\; \frac{1}{2}\min\left(\frac{\mu_t(y)}{\mu_t(x)}, 1\right) \;\; \text{if } |y-x| = 1, \;\; 0 \text{ otherwise.}$$

Here the following estimates hold:

$$\begin{aligned} C_t &\;\leq\; 30\;((\sigma_t \wedge \Delta) \vee 2)^2 \\ \gamma_t &\;\leq\; 300\,\frac{\Delta^2}{(\sigma_t \wedge 1)^2} + 300\;((\sigma_t \wedge \Delta) \vee 2)^2\;\log\Delta \\ \operatorname{osc}(H_t) &\;\leq\; \left(2\frac{|\sigma_t'|}{\sigma_t} + \frac{|m_t'|}{\Delta}\right)\cdot\frac{\Delta^2}{\sigma_t^2} \end{aligned}$$

EXAMPLE 1: Moving Gaussians, $d = 1$

$S = \{a, a+1, \ldots, a+\Delta-1\} \subseteq \mathbb{Z}$,

$$\mu_t(x) \;\propto\; \exp\left(-\frac{|x - m_t|^2}{2\sigma_t^2}\right) \; .$$

- We have $\omega \le 1$ if

$$2\frac{|\sigma_t'|}{\sigma_t} + \frac{|m_t'|}{\Delta} \;\le\; \left(\frac{\sigma_t}{\Delta}\right)^2 \; .$$

- The Theorem above yields feasible bounds if

$$\frac{\sigma_t \wedge 1}{\Delta}$$ is not too small.

EXAMPLE 2: Product measures, dependence on dimension

$$S = \prod_{k=1}^{d} S_k \,, \qquad \mu_t = \bigotimes_{k=1}^{d} \mu_t^{(k)}$$

$$\Rightarrow \quad H_t(x) = -\frac{d}{dt} \log \mu_t(x) = \sum_{k=1}^{d} H_t^{(k)}(x_k)$$

$$\Rightarrow \quad \omega = \sup_{t,x,y} |H_t(x) - H_t(y)| \le \sum_{k=1}^{d} \omega^{(k)} \,.$$

$$\mathcal{L}_t(x,y) = \sum_{k=1}^{d} \mathcal{L}_t^{(k)}(x,y)$$ product dynamics

$$\Rightarrow \quad C_t = \max_k C_t^{(k)}, \quad \gamma_t = \max_k \gamma_t^{(k)} \,.$$

EXAMPLE 2: Product measures, dependence on dimension

$$S \;=\prod_{k=1}^{d} S_k\,, \qquad\qquad \mu_t \;=\; \bigotimes_{k=1}^{d} \mu_t^{(k)}$$

Assumption:

$$\omega^{(k)} \le 1 \;\;\forall\, k, \qquad\qquad C_t^{(k)},\; \gamma_t^{(k)} \text{independent of } k.$$

$\Rightarrow \quad \omega = O(d),\;\; C_t = O(1),\; \gamma_t = O(1)$

$\Rightarrow \quad$ need $N = O(d)$ and $\lambda_s = O(d)$

$\Rightarrow \quad$ total effort for a given precision is $O(d^3)$

5 NON-ASYMPTOTIC BOUNDS FROM LOCAL ESTIMATES

Example: Annealing

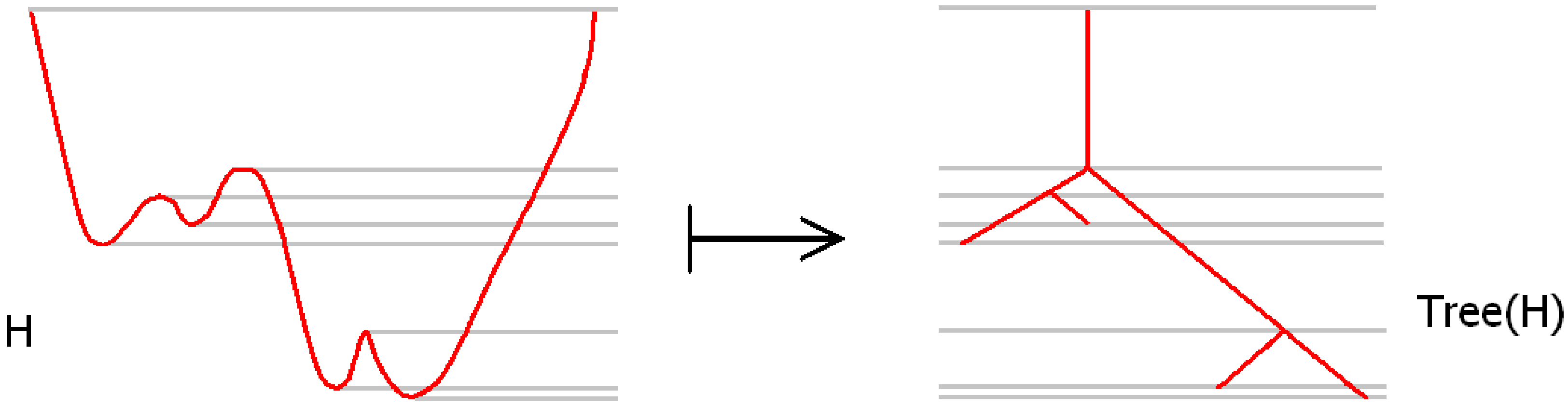
$$\mu_t(x) \;=\; Z_t^{-1} \exp(-\beta_t U(x)) \,, \qquad t \geq 0.$$

Metastability Problem :

- Local energy minima $\rightsquigarrow$ metastable states $\rightsquigarrow$ traps for Markov chain
- *Logarithmic cooling schedule:* Cool down so slowly that Markov chain escapes traps.

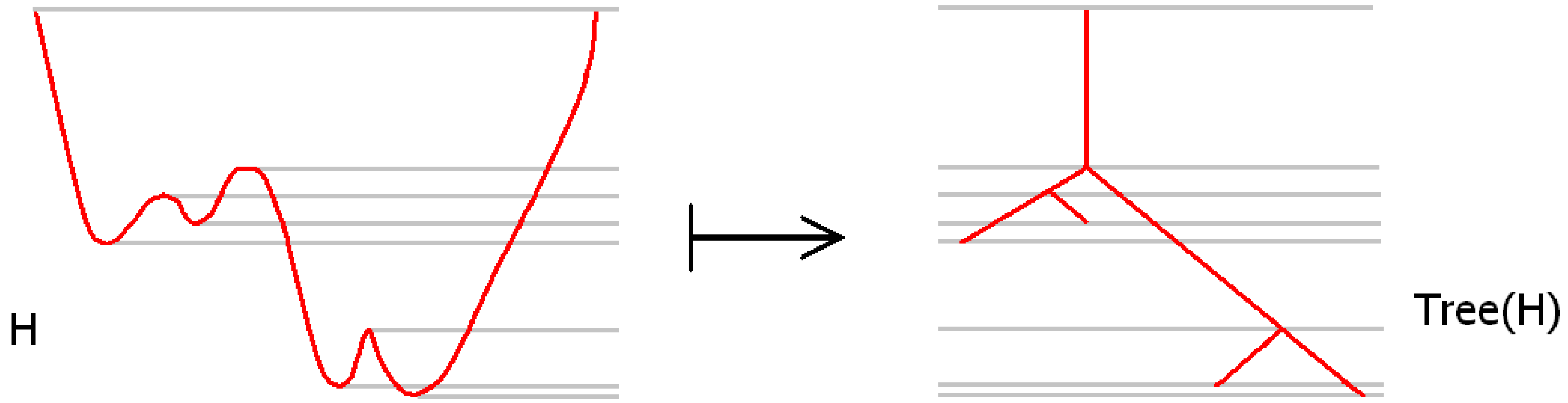
  $\rightsquigarrow$ not feasible in practice !
- *Realistic approach:* Cool down much faster.

  $\rightsquigarrow$ Markov chain eventually gets trappped

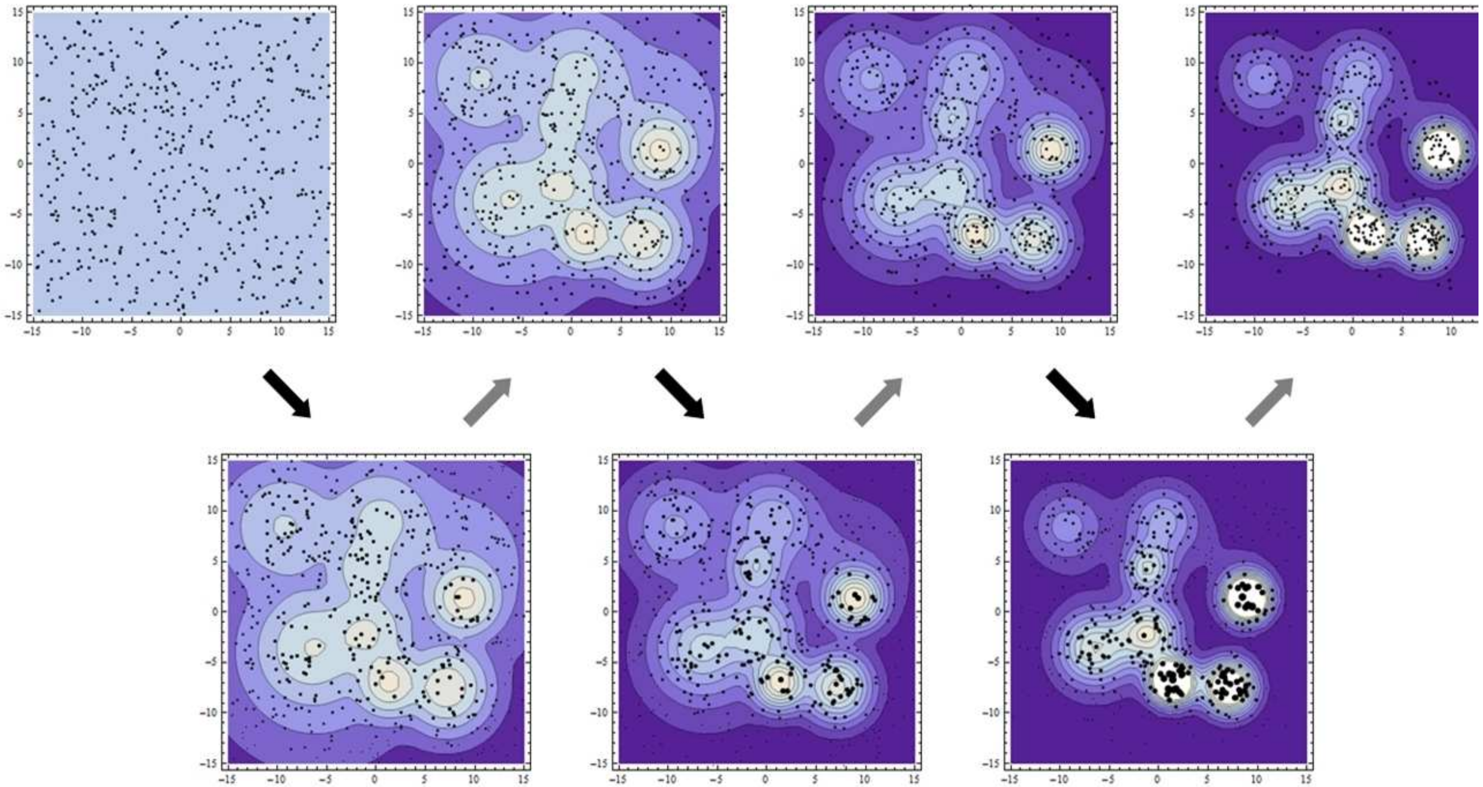


DISCONNECTIVITY TREE OF ENERGY FUNCTION:

$$\begin{array}{rcl} S & \longrightarrow & \text{Disconnectivity tree } \mathbb{T} \\ \text{Energy function } U : S \to \mathbb{R}_+ & \longrightarrow & \text{Height function } h : \mathbb{T} \to \mathbb{R}_+ \\ \text{Reference measure} \qquad \nu & \longrightarrow & \text{Density of states } \Omega(dx) \text{ on } \mathbb{T} \\ \mu_t & \longrightarrow & \bar{\mu}_t(dx) \; \propto \; e^{-\beta_t h(x)} \, \Omega(dx) \end{array}$$



- As t increases, the Markov chain gets trapped in deeper branches of the tree.
- The state space effectively splits into an increasing number of components (metastable states)



NON-ASYMPTOTIC BOUNDS FROM LOCAL ESTIMATES

$S = \bigcup S_i$ disjoint decomposition of state space. Suppose that

$$\begin{aligned} \mathcal{L}_t(x,y) &= 0 \;\forall\, t \geq 0,\; x \in S_i,\; y \in S_j \;(i \neq j), \text{ and let} \\ \mu_t^i &= \mu_t(\,\cdot\,|S_i), \qquad \|f\|^{\sim}_{L^p(\mu_t)} := \max_i \|f\|_{L^p(\mu_t^i)}\;, \\ \tilde{\varepsilon}_t^{N,p} &:= \sup\left\{\mathbb{E}\left[\left|\langle f, \nu_s^N\rangle - \langle f, \mu_s\rangle\right|^2\right] \;:\; s \in [0,t],\; \|f\|^{\sim}_{L^p(\mu_s)} \leq 1\right\}. \end{aligned}$$

THEOREM. Suppose conditions as above hold with C_t, γ_t replaced by

$$\tilde{C}_t = \max_i C_t^i\,, \qquad \tilde{\gamma}_t = \max_i \gamma_t^i\,.$$

Then

$$\tilde{\varepsilon}_t^{N,p} \leq \frac{2 + 8\,K_t\,\tilde{M}_t^2}{N} \cdot \left(1 + \frac{16\,\tilde{K}_t\,\tilde{M}_t^2}{N}\right)$$

where

$$\tilde{M}_t = \max_i \sup_{0 \leq r \leq s \leq t} \frac{\mu_s(S_i)}{\mu_r(S_i)}\,.$$

“Recipes” for applications

- Try to guarantee osc$(H_t) \leq 1$, or at least osc$(H_t^-) \leq 1$
- Use enough MCMC steps such that there is sufficient mixing in each metastable state
- Quality of error estimates depends (among other things) on structure of disconnectivity tree

6 OUTLOOK

OPEN PROBLEMS:

- Generalization to discrete time and continuous space

  see PhD thesis of Nikolaus Schweizer

- Non-asymptotic bounds under **local** mixing conditions.

  First step: Asymptotic bounds:

  A.E., C. Marinelli. Stability of nonlinear flows of probability measures related to sequential MCMC methods.

  Second step: Non-asymptotic analysis on trees:

  PhD thesis of Nikolaus Schweizer