Metastability in molecular dynamics and Bayesian inference methods

Florian Maire

Université de Montréal, dépt. de mathématiques et de statistique

with:

- Vio Buchete (University College Dublin, School of Physics)
- Olivier Roux (École Centrale Marseille & UCD)
- Pierre Vandekherkove (Univ. Paris-Est Marne-la-Vallée, LAMA)

Two contexts

Characterization of folded and unfolded conformations in protein dynamic model

Locally informed Markov chains

Two contexts

Context 1: Protein dynamics

Softwares allow to simulate protein dynamics (forward simulation):

$$\text{input parameters} \overset{CPU}{\rightsquigarrow} \text{a trajectory } \{z_t, t > 0\},$$

where $z_t \in \mathsf{Z} \subset \mathbb{R}^p$ ($p \gg 1$) includes distance/angle between amino acids, energy levels, relative speeds, interactions with solvent, etc. at time t.

Some issues:

- analysis difficult (too many details)
- statistical problem
- computational problem: relevant processes have different characteristic times

⬇ coarse graining the dynamics

Stochastic (typically Markov) processes used to model protein trajectories $\{z_t,\ t > 0\}$

$$\{X_t,\ t > 0\}, \quad X_t \in \mathsf{X},$$

where X is some high dimensional finite state-space.

Context 2: Computational methods is Bayesian statistics

Bayesian context: observed data Y described by a likelihood model f_X with parameter $X \in \mathsf{X} \subset \mathbb{R}^d$ and a prior p on X.

Bayesian analysis: estimate quantities

$$\pi\phi := \mathbb{E}\{\phi(X) \mid Y\} = \int_{\mathsf{X}} \pi(\mathrm{d}X \mid Y)\phi(X)\,, \quad \pi(\mathrm{d}X \mid Y) \propto p(\mathrm{d}X) f_X(Y)\,.$$

Issue: $\mathbb{E}\{\phi(X) \mid Y\}$ is often intractable.

⬇ numerical integration

Many schemes (MC, Sequential MC, Importance-Sampling, etc.), focus is on Markov chain Monte Carlo methods:

$$\{X_k,\, k \in \mathbb{N}\} \quad \text{s.t.} \quad \frac{1}{n}\sum_{k=1}^{n} \phi(X_k) \to \pi\phi \;(as)$$

A common denominator in two contexts

Analysis/interpretation of Molecular dynamics $\Leftrightarrow$ understand the Markov semi-group $\{P_t,\, t > 0\}$

$$\frac{\mathrm{d}P_t}{\mathrm{d}t} = \mathcal{L}P_t, \qquad P_t(x,\cdot) := \mathbb{P}_x(X_t \in \cdot)\,,$$

and in particular its generator $\mathcal{L}$.
$\Rightarrow$ continuous time process, discrete state space

Inferring a parametric statistical model with MCMC $\Leftrightarrow$ estimate how the Markov kernel $\{P^k,\, k \in \mathbb{N}\}$

$$P^{k+1} = P^k P\,, \qquad P^k(x,\cdot) := \mathbb{P}_x(X_k \in \cdot)\,,$$

transform recursively any measure $\mu_0 \mapsto \mu_0 P$ s.t., for a large number of applications k, $\|(\mu_0 P^k)\phi - \pi\phi\| \approx 0$, for some metric $\|\cdot\|$.
$\Rightarrow$ discrete time process, general state space

Characterization of folded and unfolded conformations in protein dynamic model

Analysis of protein dynamics

A key property for a protein dynamics model $\{X_t,\ t > 0\}$ is to identify folded and unfolded states.

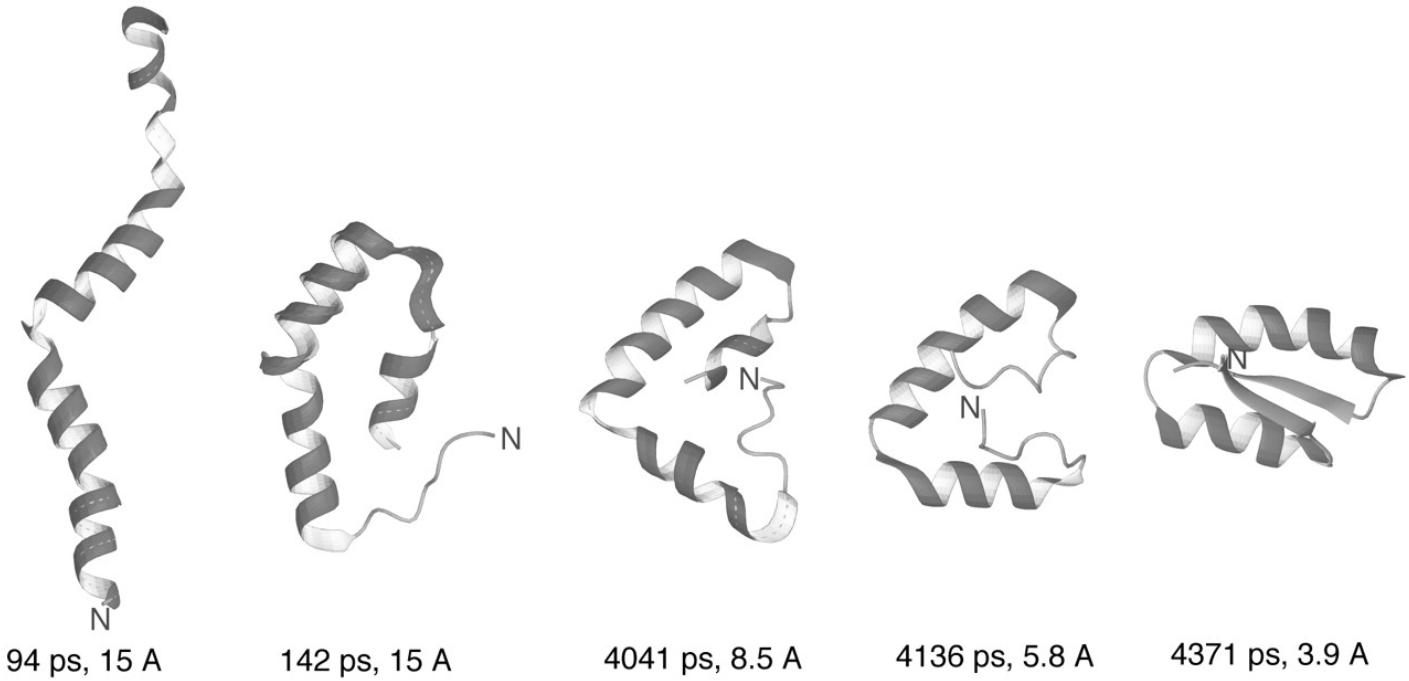


Figure: Protein folding pathway of 1E0G obtained in Langevin dynamics simulations (A. Liwo et al, PNAS, 2005)

$\Rightarrow$ X is usually large, how to identify subsets $F \subset \mathrm{X}$ (resp. $U \subset \mathrm{X}$) where the protein is folded (resp. unfolded)?

How is it done in practice?

Assumption

The process $\{X_t,\ t>0\}$ *is* μ*-reversible, i.e.* $\langle \mathcal{L}x, y\rangle_\mu = \langle x, \mathcal{L}y\rangle_\mu$.

Algorithm 1 mapping $\mathrm{X} = \{1, 2, \ldots, d\} \to \{U, F\}$

1: diagonalize $\mathcal{L}$, get $\mathrm{Sp}(\mathcal{L}) = \{\lambda_1, \lambda_2, \ldots\}$ the right eigenvectors $\Phi_1, \Phi_2, \ldots$ s.t:

$$\mathcal{L}\Phi_i = \lambda_i \Phi_i, \qquad \lambda_1 > \lambda_2 > \lambda_3 > \cdots$$

2: calculate the ratio $r = \lambda_3/\lambda_2$

3: if r is large enough (eg $r > 10$) normalize the second right eigenvector:

$$\Phi_2 \mapsto \bar{\Phi}_2 := \left\{ \frac{\Phi_2(i) - \min \Phi_2}{\max \Phi_2 - \min \Phi_2} \right\}_i$$

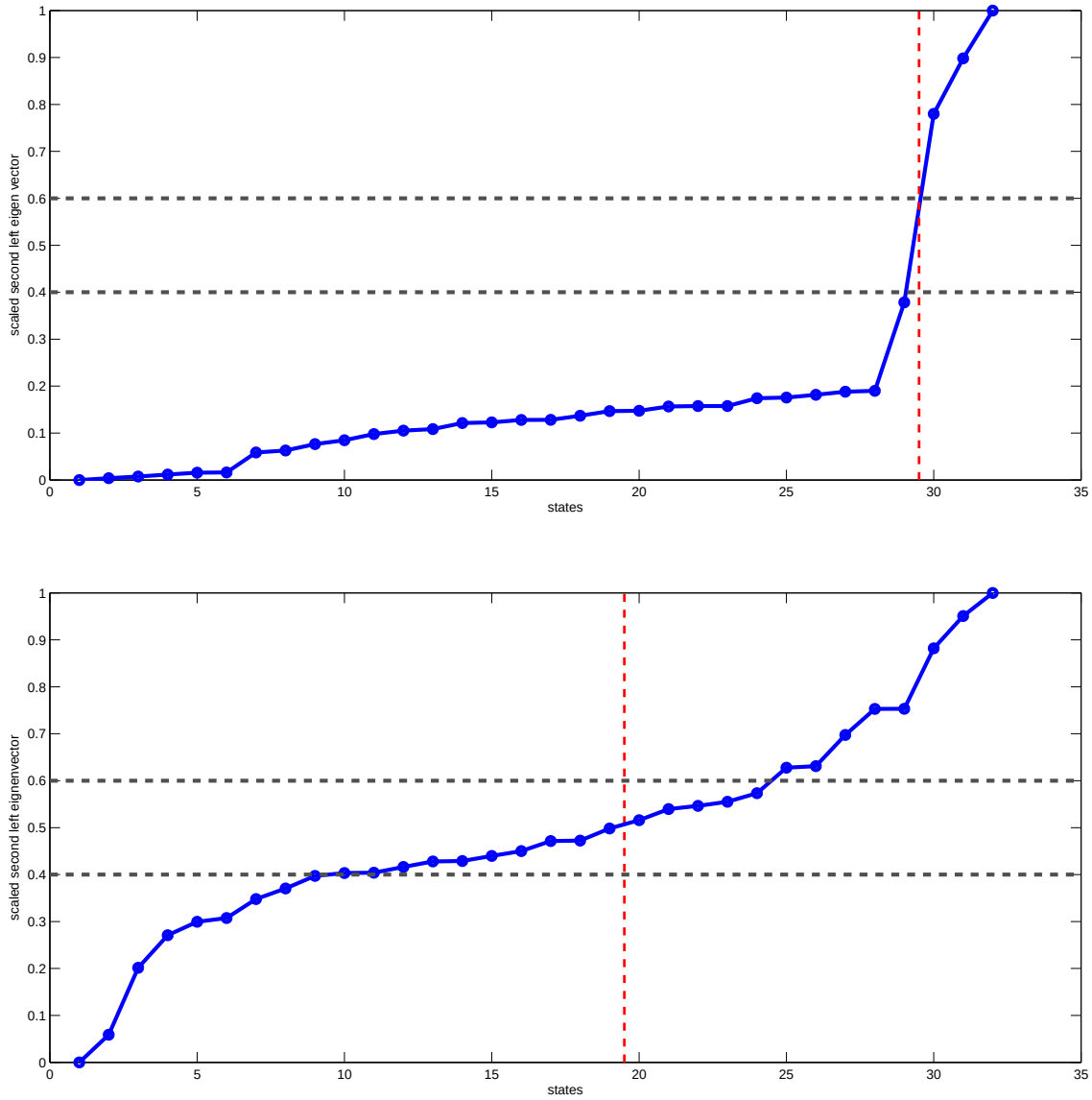
4: for each $i \in \mathrm{X}$,

- set $i \in U$ if $\Phi_2(\sigma^{-1}(i)) \ll 1/2$
- set $i \in F$ if $\Phi_2(\sigma^{-1}(i)) \gg 1/2$

---

see eg V. Buchete (2008).

Analysis of the second right eigenvectors for Ala5 peptides at 250K (top) and 350K (bottom), $X = \{1, 2, \ldots, 32\}$



Intuition behind the algorithm

Assumption

$\mathcal{L}$ is the generator of a reducible process with absorbing states U and F

Then,

- $\mathcal{L}\mathbb{1}_U = 0 \qquad \mathcal{L}\mathbb{1}_F = 0$
  (for all $h > 0$ and all $x \in \mathsf{X}$,
  $P_h\mathbb{1}_U(x) = \int P_h(x, \mathrm{d}y)\mathbb{1}_U(y) = \int_U P_h(x, \mathrm{d}y)\mathbb{1}_U(y) = \mathbb{1}_U(x)$).
- $0 \in \mathrm{Sp}(\mathcal{L})$ with multiplicity 2
- $\Phi_1 = \mathbb{1}_U$ and $\Phi_2 = \mathbb{1}_F$

Question

If $\mathcal{L}$ is not reducible but "nearly" reducible, would it allow to justify the algorithm?

Metastability

Definition

Dynamical phenomenon characterized by the existence of "sub"-processes with well-separated time scales.

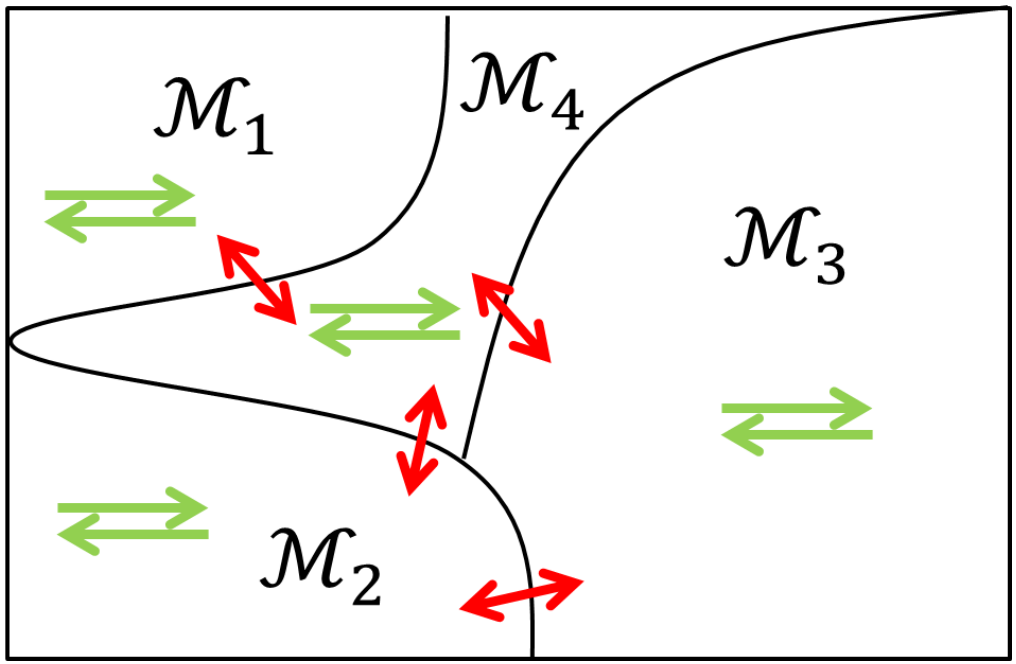


Figure: Process $\{X_t,\ t > 0\}$ with four metastable subsets $\mathcal{M}_1, \ldots, \mathcal{M}_4$.

Metastable processes: framework for analysis

- Work in Applied Maths and Stat. Phys.: M. Freidlin, A. Wendzell, A. Bovier, F. Nier, S. Meyn, M. Slowik, A. Schilchting, etc.
- Focus is on the analysis of
  - spectral properties
  - convergence

  of metastable reversible operators on discrete state space.

Different approaches:

- large deviation (path wise approach)
- potential theoretical approaches

Framework

Definition (First hitting time)

For all $A \subset \mathsf{X}$, *define* $\tau_A := \inf\{t > 0,\, X_t \in A\}$.

Definition (ρ-metastability)

If there exists a subset $\mathcal{M} \subset \mathsf{X}$ *and* $\rho < 1$ *s.t.*

$$\forall x \in \mathcal{M},\, y \in \mathsf{X}\backslash\mathcal{M}\,, \quad \mathbb{P}_x\left(\tau_{\mathcal{M}\setminus\{x\}} \leq \tau_x\right) \leq \rho \mathbb{P}_y\left(\tau_{\mathcal{M}} \leq \tau_y\right)\,,$$

the process $\{X_t,\, t > 0\}$ *is said to be* ρ*-metastable w.r.t* $\mathcal{M}$.

Definition (Capacity)

$$\forall\,(A, B) \subset \mathsf{X}, \quad \mathrm{cap}(A, B) = \int_A \mathrm{d}\mu \mathcal{L} h_{A,B}$$

where $h_{A,B}$ *is the equilibrium potential,*

$$h_{A,B}(x) = \mathbb{P}_x(\tau_A < \tau_B)\mathbb{1}_{\overline{A \cup B}}(x) + \mathbb{1}_A(x)\,.$$

Definition (Dirichlet form)

For any $g, h \in L_2(\mu)$

$$\mathcal{E}(g,h) := \mu(g.\mathcal{L}h)$$
$$= (1/2) \iint \mu(\mathrm{d}x)\mathcal{L}(x,\mathrm{d}y)(h(x)-h(y))(g(x)-g(y))\,.$$

Proposition (Variational principle)

$$\mathrm{cap}(A,B) = \inf_{h \in \mathcal{H}_{A,B}} \mathcal{E}(h)$$

where $\mathcal{H}_{A,B} = \{h : \mathsf{X} \to [0,1],\ h_{|A} = 1,\ h_{|B} = 0\}$ *and* $\mathcal{E}(h) = \mathcal{E}(h,h)$.

Proposition

For all $x \in \mathsf{X}$, $A \subset \mathsf{X}\backslash\{x\}$,

$$\mathbb{P}_x\{\tau_A \leq \tau_x\} = \mathrm{cap}(x,A)/\mu(x)\,.$$

Technical results (Bovier et al., 2004)

1/ Dirichlet problem (w. boundary cdts on two sets A and B):

$$(\mathcal{L} - \lambda \mathrm{I}) h^{\lambda}_{A,B}(x) = 0\,, x \notin A \cup B, \qquad h^{\lambda}_{A,B}(x) = \mathbb{1}_A(x)\,, x \in A \cup B$$

2/ λ_0 smallest eigenvalue such that

$$(\mathcal{L} - \lambda \mathrm{I}) f = 0\,, x \notin \mathcal{M}, \qquad f_{\mathcal{M}} \neq 0$$

3/ Characterization of $\mathrm{Sp}(\mathcal{L})$

$$\lambda \in \mathrm{Sp}(\mathcal{L}) \text{ and } \lambda < \lambda_0 \iff \det \mathfrak{M}(\lambda) = 0$$

$\{\mathfrak{M}(\lambda)\}_{x,y} = \mathcal{L} h^{\lambda}_{x,\mathcal{M}\setminus\{x\}}(y)$, for all $(x, y) \in \mathcal{M}^2$

4/ There exists $\widetilde{\mathfrak{M}}$, a perturbation of $\mathfrak{M}(\lambda)$ (expending around $\zeta^{\lambda}_{x,y} = h_{x,y} - h^{\lambda}_{x,y}$),

$$\widetilde{\mathfrak{M}}_{x,y} = \mathcal{E}(h_x, h_y) / \| h_{x,\mathcal{M}\setminus\{x\}} \|_2 \| h_{y,\mathcal{M}\setminus\{y\}} \|_2$$

for all $\lambda \in \mathrm{Sp}(\mathcal{L})$, $\lambda < \lambda_0$, there is $\sigma \in \widetilde{\mathfrak{M}}$ s.t.

$$\lambda = \sigma(1 + \mathcal{O}(\rho))\,.$$

Application to the protein dynamic setup

Assumptions

- *Metastability*, $\{X_t,\ t > 0\}$ *is* ρ*-metastable wrt* $\mathcal{M} = \{U, F\}$.
- *Non-degeneracy*, *there exists* $\delta < 1$ *s.t.* $\delta := \mu(A(U))/\mu(A(F))$ *where for any* $\mathcal{M}_1 \in \mathcal{M}$, $A(\mathcal{M}_1)$ *is the attractor of* $\mathcal{M}_1$,
  $A(\mathcal{M}_1) := \{x \in \mathsf{X},\ \mathbb{P}_x\left(\tau_{\mathcal{M}_1} \leq \tau_x\right) \geq \mathbb{P}_x\left(\tau_{\mathcal{M}\setminus\mathcal{M}_1} \leq \tau_x\right)\}$.

Proposition

Under those assumptions, we have:

$$\lambda_1 = 0, \quad \lambda_2 = \frac{\mu(U)}{\mu(A(U))}\mathcal{L}(U, A(F))\left(1 + \mathcal{O}(\rho^2 + \delta)\right)$$

and the first two right eigenvectors satisfy:

$$\phi_1 = \mathbf{1}, \quad \phi_2(y) = \frac{\mathbb{P}_y\{\tau_U \leq \tau_F\}\mathbb{1}_{y\notin\{U,F\}} + \mathbb{1}_{y\in U}}{\mu(A(F))} + \mathcal{O}(\rho^2 + \delta)\,.$$

Locally informed Markov chains

Design of MCMC sampler

Suppose that n π-reversible Markov kernels are available:

$$P_1, P_2, \ldots, P_n\,.$$

We know that for any $I \subseteq \{1, \ldots, N\}$ and any probability ω on I,

$$P_\omega := \sum_{k \in I} \omega_k P_k, \quad .$$

is also π-reversible.

Question

Is there some choice of (ω, I) "better" than other?

Examples

- *Gibbs: full cdt's of π are samplable and $P_i(x_i, \cdot) \equiv \pi(\cdot \,|\, x_i)$*
- *Metropolis-within-Gibbs*

Related works

- Geometric ergodicity and hybrid Markov chains (Roberts and Rosenthal, 1997)
- Adaptive Gibbs sampler and related methods (Latuszyńky et al., 2013)
- On random-and systematic-scan samplers (Andrieu, 2016).
- etc.

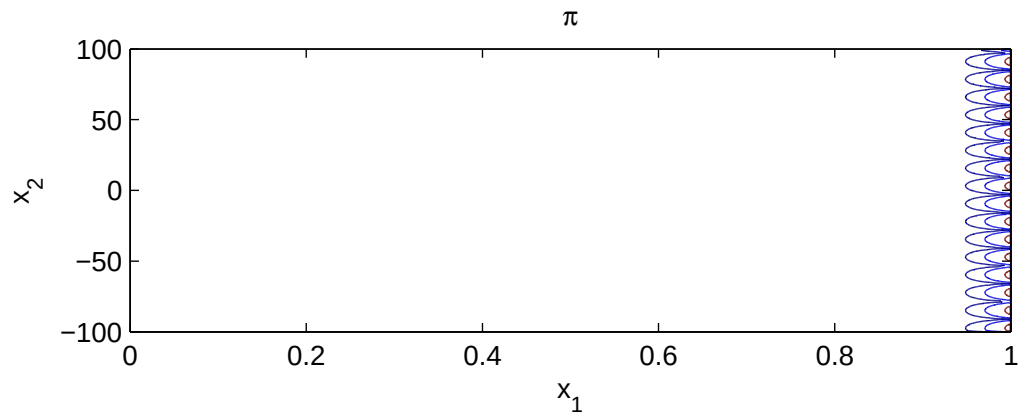
The following question has been unexplored: does it make sense to consider a state-dependant probability on I, ie

$$P_\omega(x, \cdot) = \sum_{k \in I} \omega_k(x) P_k(x, \cdot) \, ?$$

Illustration with $d = 2$

Consider the following distribution on $\mathsf{X} = [0\,,\,1] \times [-100\,,\,100]$:

$$\pi(x_1, x_2) \propto x_1^{100}\,(1 + x_1 \sin(x_2/2))\ .$$



Assume a MwG sampler is used to sample from π:

$$P_\omega = \omega P_1 + (1 - \omega) P_2\,,$$

i.e. P moves X_1 through P_1 with proba. ω (resp. for X_2).

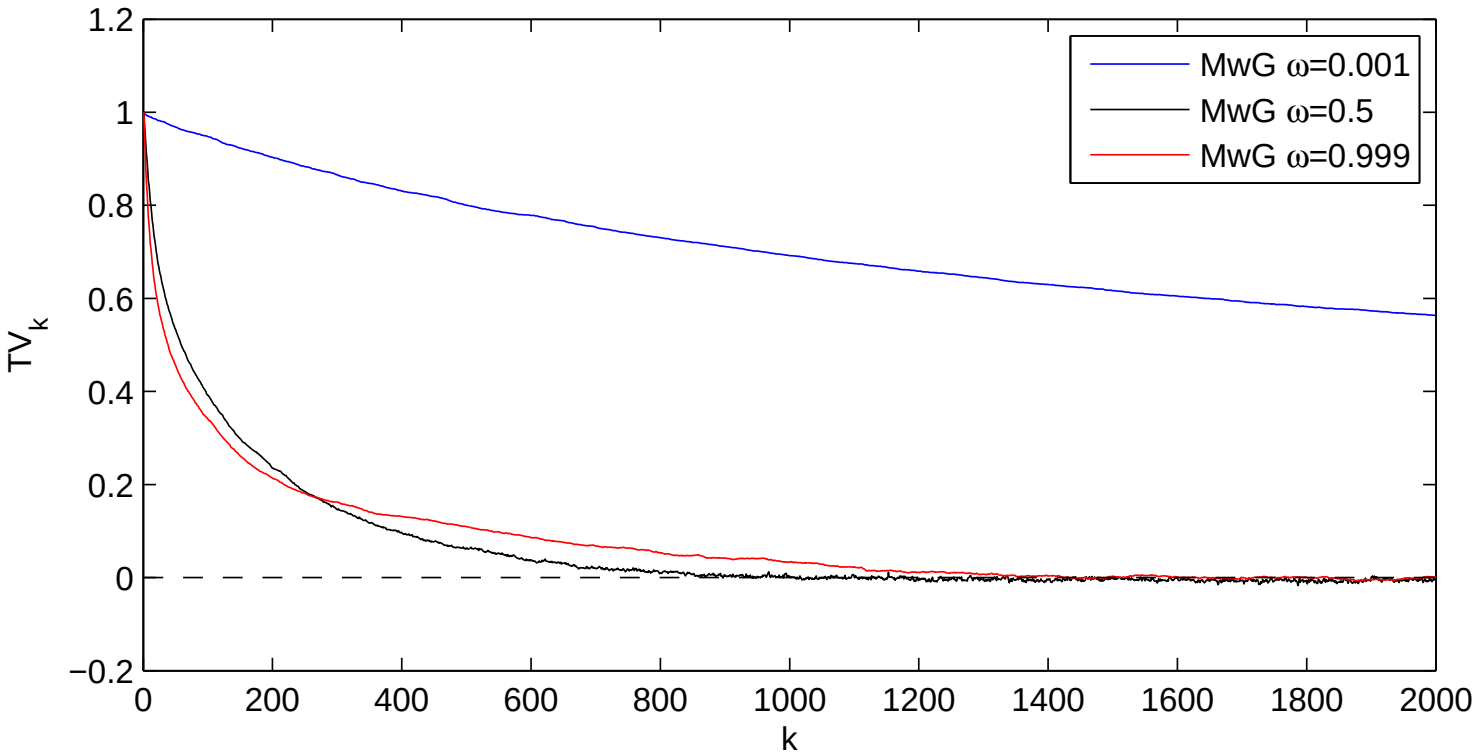
Question

How to choose ω?

Illustration with $d = 2$: convergence time to π

Let π_k the distribution of X_k i.e. $\pi_k = \mu P_\omega^k$ where $X_0 \sim \mu := \text{unif(X)}$.
Define the total variation distance

$$TV_k = \|\pi - \pi_k\| = (1/2) \int_{\mathrm{X}} |\pi(x) - \pi_k(x)| \mathrm{d}x \,.$$



Locally informed MCMC

Proposition

If ω is state-dependent, P_ω is in, general[1]*, not π-invariant.*

Let $\{\bar{X}_k,\, k \in \mathbb{N}\}$, whose transition $\bar{X}_k \to \bar{X}_{k+1}$ is given by

- draw $I \sim \omega(\bar{X}_k)$,
- draw $\tilde{X} \sim P_I(\bar{X}_k, \cdot)$,
- set $\bar{X}_{k+1}$ w.p. $1 \wedge \omega_I(\tilde{X})/\omega_I(\bar{X}_k)$ and $\bar{X}_{k+1} = \bar{X}_k$ otherwise,

Proposition

The transition kernel

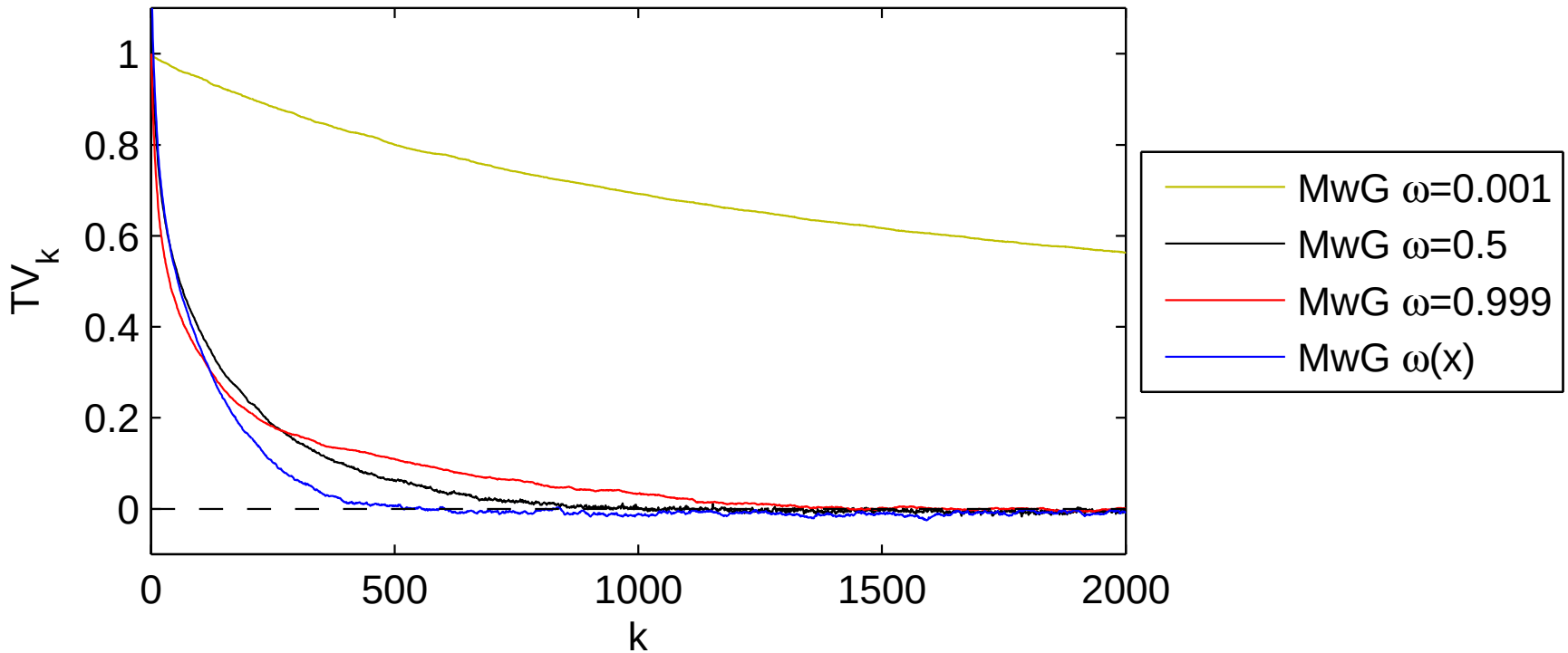
$$\bar{P}_\omega(x, \cdot) = \sum_{i=1}^{n} \omega_i(x) P_i(x, \cdot)\{1 \wedge \omega_I(\cdot)/\omega_I(x)\}$$

is π-invariant.

[1] it is iff $\mathbb{E}_\pi\{\sum_{k\in I} \omega_k(x) P_k(x, A)\} = \pi(A)$, for all $A \in \mathcal{X}$

Illustration with $d = 2$: convergence time to π

Let π_k the distribution of X_k i.e. $\pi_k = \mu P_\omega^k$ where $X_0 \sim \mu := \text{unif}(\mathsf{X})$.



with the local weight function set as

$$\omega(x) := \sqrt{1 - x_1}\,.$$

Some "poor" asymptotic properties of $\bar{P}$ (1/2)

Proposition

Let:

- $P_1, P_2, \ldots$ *be absolutely continuous kernels,*
- ω_{loc} *be state-independent and* ω_{unif} *be state-dependent proba. on* $\{1, \ldots n\}$
- $f \in L_2(\pi)$ *satisfying* $\sum_i |cov(f(X_0), f(X_i))| < \infty$

Then,

$$v(f, P_{\omega_{unif}}) \leq v(f, P_{\omega_{loc}}),$$

where $v(f, P) = \lim_{n\to\infty}(1/n)var\{\sum_{k=1}^{n} f(X_k)\}$, $X_0 \sim \pi$ *and* $X_{k+1} \sim P(X_k, \cdot)$.

Some "poor" asymptotic properties of $\bar{P}$ (2/2)

Let $\mathsf{X} = \{1, 2, 3\}$ and consider:

- $\pi_\epsilon = [(1-\epsilon)/2 \;\; (1-\epsilon)/2 \;\; \epsilon]$
- $\{\omega_{unif}(x)\}_j \propto 1$ and $\{\omega_{loc}(x)\}_j \propto \pi(j)\mathbb{1}_{x \neq j}$

Proposition

If $\epsilon < 1/3$, the spectral gap γ of the two kernels satisfy:

$$\gamma(P_{loc}, \epsilon) = \frac{1-2p}{1-p} \qquad \textit{and} \qquad \gamma(P_{unif}, \epsilon) = \epsilon \frac{3-5\epsilon}{1-\epsilon^2}\,.$$

Corollary

The speed to convergence for $\epsilon \ll 1$ is contrasting for the two methods:

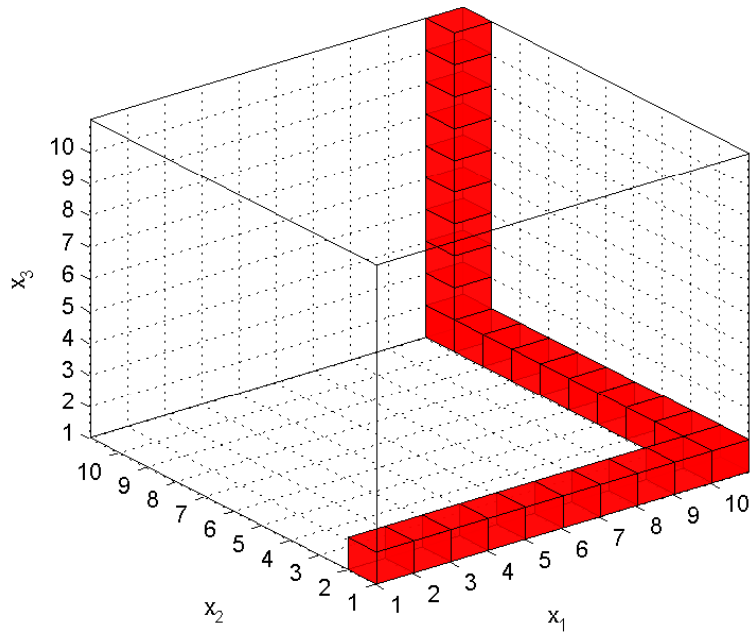
$$\sup_{\mu \in \mathcal{M}_1(\mathsf{X})} \|\mu P^n - \pi\| \approx \begin{cases} Ce^{n \log \epsilon / 1-\epsilon} & \text{for } P = P_{unif} \\ Ce^{-n\epsilon + o(\epsilon^2)} & \text{for } P = P_{loc} \end{cases}$$

A "positive" limit example

Consider the distribution π_ϵ on the hypercube $\mathsf{X} = \{1, \ldots, n\}^d$ s.t.

$$\pi_\epsilon(x) = \epsilon + (1-\epsilon)\mathbb{1}_{x \in \mathcal{F}}\,,$$

where $\mathcal{F}$ is a path on the hypercube edges:



Proposition

Define τ_{loc} (resp. τ_{unif}) the coupling time of P_{loc} (resp. P_{unif}), then when $\epsilon = 0$,

$$\mathbb{E}_{x_1}(\tau_{loc}) \leq (d/2)\mathbb{E}_{x_1}(\tau_{unif})\,, \qquad x_1 = \mathbf{1}\,.$$

Study of limiting examples i.e. $\epsilon \to 0$

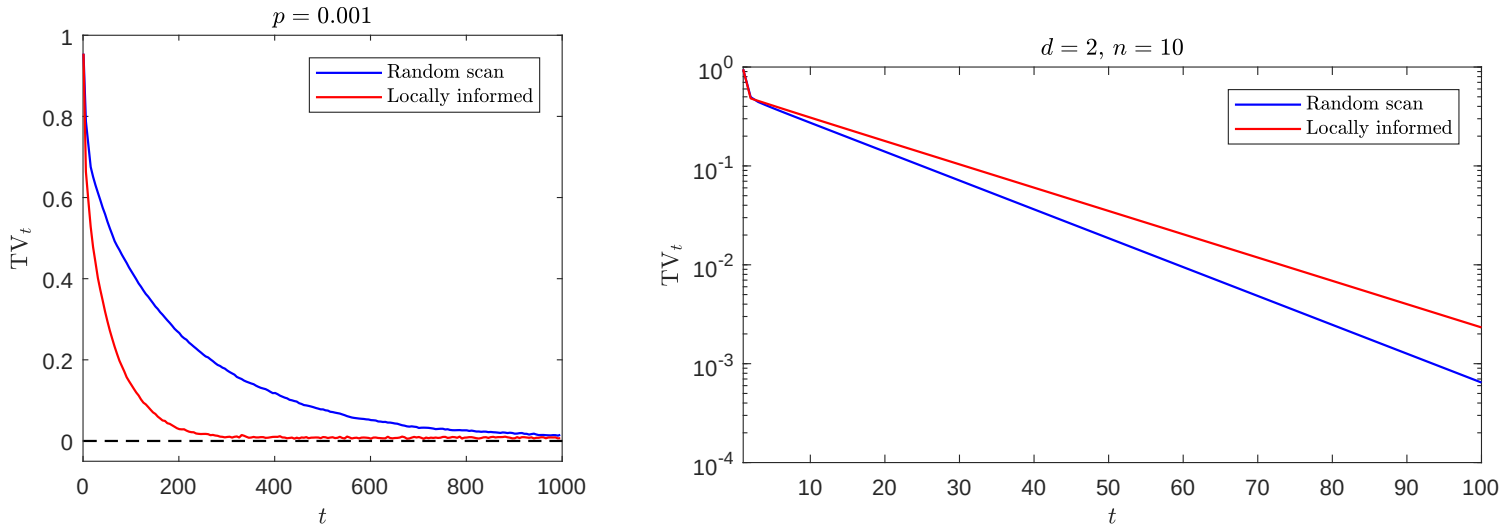


Figure: Left: empirical, $d = 7$, $p = 10^{-3}$ – Right: theoretical, $d = 2$, $p = 10^{-1}$.

+ many other examples in Maire and Vankerkhoven (2018, arXiv) showing the same convergence patterns for "filamentary distributions".

Conjecture

In similar scenarios, P_{loc} converges initially (much) faster than P_{unif} **before** *reaching a very slow asymptotic rate.*

Metastable behaviour of P_{loc}

If $x \in \mathcal{F}$, $P_{loc}(x, \bar{\mathcal{F}}) \ll 1$ and $x \notin \mathcal{F}$, $P_{loc}(x, \mathcal{F}) \ll 1$.

Definition (Poincaré inequality)

P satisfies a Poincarré inequality with constant κ if:

$$\kappa var_\pi(f) \leq \mathcal{E}(f)\,.$$

Proposition (Schlichting and Slowik (2017))

For a ρ-metastable, π-reversible Markov chain on a discrete state space, with $\mathcal{M} = \mathcal{M}_1, \mathcal{M}_2$, the optimal Poincarré constant is

$$\kappa^* = \frac{\mathrm{cap}(\mathcal{M}_1, \mathcal{M}_2)}{\pi(A(\mathcal{M}_1))\pi(A(\mathcal{M}_2))}(1 + \mathcal{O}(\rho))\,.$$

Corollary

If X is discrete, π-reversible and ρ-metastable Markov chain, the Markov kernel P contracts $L_2^0(\pi)$ such that:

$$\|P_k f\|_2 \leq e^{-2k/\kappa}\|f\|_2\,.$$

Continue to seek...

The current framework to analyse metastability is not adequate:

- it would tell us only how "bad" our locally informed chain is
- most of the theory is concerned with discrete state space
- potential wells are supposed

$\Rightarrow$ For locally informed MCMC, the metastability arises from the reversible dynamic of P and is not a by-product of a multimodale stationary distribution.

Research alternatives:

- Decompose the L1 distance, say $\text{supp}(\mu) = \mathcal{F}$:

$$\|\mu\bar{P}^k - \pi\| \leq \|\mu\bar{P}^k_{red} - \pi_{|\mathcal{F}}\| + \|\pi - \pi_{|\mathcal{F}}\| + \|\mu\bar{P}^k_{red} - \mu\bar{P}^k\|$$

  where $\bar{P}^k_{red}$ is the reducible version of $\bar{P}$. This would lead to:

$$\|\mu\bar{P}^k - \pi\| \leq Ce^{-t/\tau} + 2\epsilon + \|\mu\bar{P}^k_{red} - \mu\bar{P}^k\|, \qquad \epsilon = \pi(\mathcal{F})\,.$$

- bounding the last term from application of the Markov perturbation theory (Johndrow and Mattingly, 2018, Medina-Aguayo et al., 2018) under unif. ergodicity in V-norm of $\bar{P}$ and drift cdt of $\bar{P}$ (with same function V):

$$\|\mu\bar{P}^k_{red} - \mu\bar{P}^k\| \leq \frac{33C(L+1)\kappa}{1-\alpha}\log R/R$$

Concluding observations

A rigorous characterization of metastability in dynamical systems allows one to use the rich literature on this subject.

- ρ-metastability, potential theoretic approach: variational approach $\Rightarrow$ bounds for ρ
- most results limited to cases where X is a finite state space
- improve the precision in the clustering algorithm to folding/unfolding in protein dynamics.

For the analysis of MCMC algorithms:

- ρ-metastability may be used to show how inefficient a Markov chain is (Poincaré inequality)
- some proof technics based on splitting the dynamics according the metastable sets may be useful (Slowik and Schlichting)
- perturbation theory of V-ergodic Markov chains (Medina-Aguayo et al., Rudolf et al., Johndrow et al., etc) seems more readily applicable