# Swapping Methods for Fleming-Viot Estimators of Quasi-Stationary Distributions

Itamar Fiorino
Paul Dupuis – Advisor
Kavita Ramanan – Second Reader

Brown University – Department of Applied Mathematics

# Setup

Take a finite set $|S| = d$ and linear operator (matrix) $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}^d$ so that, as a matrix $\mathcal{L} = Q$, its entries are non-negative. Let $\mathcal{L}^*$ be its adjoint (matrix transpose). Say we want to solve the finite-dimensional eigenvalue problem

$$\mathcal{L}^* \phi = \alpha \phi$$

for some maximal $\alpha > 0$ such that $||\phi||_1 = 1$.

# Setup

Take a finite set $|S| = d$ and linear operator (matrix) $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}^d$ so that, as a matrix $\mathcal{L} = Q$, its entries are non-negative. Let $\mathcal{L}^*$ be its adjoint (matrix transpose). Say we want to solve the finite-dimensional eigenvalue problem

$$\mathcal{L}^* \phi = \alpha \phi$$

for some maximal $\alpha > 0$ such that $||\phi||_1 = 1$.

These problems are amenable to computation from probabilistic methods, in particular through **Markov Chain Monte Carlo (MCMC)** methods.

# Continuous Time Markov Chains

A **Continuous Time Markov Chain (CTMC)** is (heuristically) a random process $(X_t)_{t>0}$ whose distribution in the future depends only on the present.

# Continuous Time Markov Chains

A **Continuous Time Markov Chain (CTMC)** is (heuristically) a random process $(X_t)_{t>0}$ whose distribution in the future depends only on the present.

A non-example would be the position of a particle shot out of a cannon, where the position and trajectory of the cannon are random.

# Continuous Time Markov Chains

A **Continuous Time Markov Chain (CTMC)** is (heuristically) a random process $(X_t)_{t>0}$ whose distribution in the future depends only on the present.

A non-example would be the position of a particle shot out of a cannon, where the position and trajectory of the cannon are random.

$$\mathbb{P}(X_{t+s} = y | \mathcal{F}_t) = \mathbb{P}(X_{t+s} = y | X_t) = P^s(X_t, y) \qquad \text{(time-homogeneous)}$$

Where $\mathcal{F}_t$ is the information in the system up to time $t$ (so that $(X_t)$ is a $(\mathcal{F}_t)$-adapted process).

# Continuous Time Markov Chains

CTMCs are normally further required to have some path-wise continuity properties, and are usually constructed via. rate-matrices $Q$, representing the rate that a particle jumps between states. $Q$ satisfies

- Off-diagonal elements are nonnegative
- Rows sum to 0

$$P^{\varepsilon} = I + \varepsilon Q + o(\varepsilon)$$

# Continuous Time Markov Chains

$$P^{\varepsilon} = I + \varepsilon Q + o(\varepsilon)$$

The **generator** of the process is

$$\mathcal{L}f(x) = \frac{\mathbb{E}\left[f(X_\delta) - f(X_0) | X_0 = x\right]}{\delta}$$

which corresponds directly with $Q$.

# Markov Chain Monte Carlo

If $\mathcal{L}$ is a rate matrix, then the empirical measures of sample paths of chains with generator $\mathcal{L}$ approach the solution to

$$\mathcal{L}^* \phi = 0$$

These eigenvectors are exactly stationary distributions that satisfy.

$$\mathbb{P}_\pi(X_t \in A) = \pi(A)$$

# Markov Chain Monte Carlo

If $\mathcal{L}$ is a rate matrix, then the empirical measures of sample paths of chains with generator $\mathcal{L}$ approach the solution to

$$\mathcal{L}^* \phi = 0$$

These eigenvectors are exactly stationary distributions that satisfy.

$$\mathbb{P}_\pi(X_t \in A) = \pi(A)$$

In MCMC, we want to draw samples from a stationary distribution, typically one with many states, or compute expectations over them. To do this, we construct dynamics that we know have the stationary distribution and sample *(Metropolis-Hastings)*.

In MCMC, we want to draw samples from a stationary distribution, typically one with many states, or compute expectations over them. To do this, we construct dynamics that we know have the stationary distribution and sample *(Metropolis-Hastings)*.

**Gibbs distribution**:

$$\phi \propto e^{-\frac{\Phi}{\beta}}$$

$\beta$ is a temperature parameter

# Generalization

If $\mathcal{L}$ is just any matrix with positive off-diagonal elements, we can turn it into a rate matrix by taking from diagonal elements, so without loss of generality we can solve

$$\mathcal{L}^* \phi(x) + c(x)\phi(x) = \alpha\phi(x) \quad \forall x \in S$$

Where $\mathcal{L}$ is a rate matrix.

# Quasi-Stationary Distributions

What if $(X_t)$ takes values over $S = D \cup \partial D$, and once particles enter $\partial D$, they never leave? The chain is no longer irreducible, but we can look at **Quasi-Stationary Distributions (QSD)**:

$$\mathbb{P}_\nu(X_t \in A | \tau > t) = \nu(A)$$

$$\tau = \inf\{t > 0 : X_t \in \partial D\}$$

# Quasi-Stationary Distributions

What if $(X_t)$ takes values over $S = D \cup \partial D$, and once particles enter $\partial D$, they never leave? The chain is no longer irreducible, but we can look at **Quasi-Stationary Distributions (QSD)**:

$$\mathbb{P}_\nu(X_t \in A | \tau > t) = \nu(A)$$

$$\tau = \inf\{t > 0 : X_t \in \partial D\}$$

If $c(x) = q(x, \partial D)$ and $\mathcal{L}$ is the generator for the process in $D$ without transitions into $\partial D$, our QSD will solve

$$\mathcal{L}^* \phi(x) + c(x)\phi(x) = \alpha\phi(x) \quad x \in D$$

for a maximal eigenvalue $\alpha$ and $\|\phi\|_1 = 1$.

# Quasi-Stationary Distributions

What if $(X_t)$ takes values over $S = D \cup \partial D$, and once particles enter $\partial D$, they never leave? The chain is no longer irreducible, but we can look at **Quasi-Stationary Distributions (QSD)**:

$$\mathbb{P}_\nu(X_t \in A | \tau > t) = \nu(A)$$

$$\tau = \inf\{t > 0 : X_t \in \partial D\}$$

If $c(x) = q(x, \partial D)$ and $\mathcal{L}$ is the generator for the process in $D$ without transitions into $\partial D$, our QSD will solve

$$\mathcal{L}^* \phi(x) + c(x)\phi(x) = \alpha\phi(x) \quad x \in D$$

for a maximal eigenvalue $\alpha$ and $\|\phi\|_1 = 1$.

*Perron-Frobenius theorem: Existence & uniqueness.*

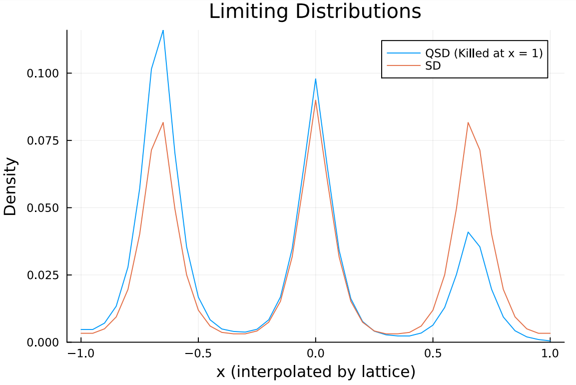
# Fleming-Viot Particle Systems

Several ways of estimating QSD. One is (based on) a **Fleming-Viot** system, an **Interacting Particle System**:

- Take $N$ particles distributed in $D$
- Evolve independently according to $\mathcal{L}$
- When a particle is killed according to the killing rate $c$, resample uniformly over the other $N-1$ particles.

# Metastability

**Metastability**: When there exist areas of the state space that do not communicate well.



# Metastability

One way to overcome metastability when estimating Gibbs measures is Parallel Tempering also called Replica Exchange MCMC.

- Swap a particle between different temperatures at appropriate rates (maintaining stationary distributions).

# Metastability

One way to overcome metastability when estimating Gibbs measures is **Parallel Tempering** also called **Replica Exchange MCMC**.

- Swap a particle between different temperatures at appropriate rates (maintaining stationary distributions).

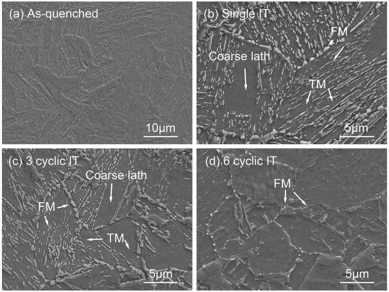



**Figure 1:** Tempering in metallurgy [2]

# Dual Problems

If $\mathcal{L}$ is a generator for the process, $c$ is the killing rate, $h$ is the normalizing term that turns $\mathcal{L}^*$ to a rate matrix, and $d = c + h$, then there are two eigenvalues associated with the QSD problem:

$$\begin{cases} -\mathcal{L}\psi(x) + c(x)\psi(x) & = \lambda\psi(x) \\ -\mathcal{L}^*\phi(x) + d(x)\phi(x) & = \lambda\phi(x) \end{cases} \tag{1}$$

$\phi$ is the QSD, that is computed with the dynamics of $\mathcal{L}$ and killing $c$, and $\psi$ is a vector that represents the state-dependent exit/decay rate, computed via the $\mathcal{L}^*$ dynamics and killing by $d$.

# Swapping

The dynamics associated with $\mathcal{L}^*$ are reversals of $\mathcal{L}$, so we expect the dynamics to spend more time in low energy areas of the state space, exactly where the original chain does not explore.

Question: can we develop a swapping scheme that respects the distribution of the independent forward/backward systems $(\phi \times \psi)$?

# Swapping

Yes! Pair up particles, and swap them according to the rates

$$r_{x,y} = e^{-(\Psi(y)+\Phi(x)-\Psi(x)-\Phi(y))^+}$$

where $\Phi$ and $\Psi$ are the energy potentials from (1), and $r_{x,y}$ is the rate that a pair at positions $(x, y)$ is swapped to positions $(y, x)$.

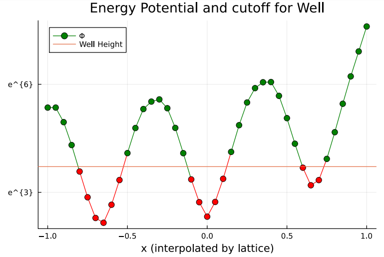
---

**Theorem (Particle Swapping Rates)**

$$e^{-(\Psi(y)+\Phi(x)-\Psi(x)-\Phi(y))^+} = \left[\prod_1^n \frac{q(z_{i+1}, z_i)}{q(z_i, z_{i+1})}\right] \vee 1 = \frac{\pi(y)}{\pi(x)} \vee 1$$

*Where $\pi$ is the stationary distribution in D prior to killing, and $\{z_i\}_1^n$ is a path with positive probability from $x$ to $y$, or alternatively a sequence with $q(z_i, z_{i+1}) > 0\ \forall i \in [n-1]$ where $x = z_1$ and $z_n = y$.*

# Improvements in Metastability



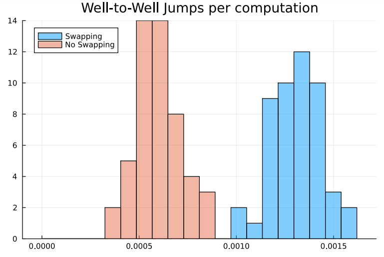




Figure 2: The well test for comparing metastability of chains

# Consistency

From the literature [1], we have

**Corollary (Uniform Consistency Bound)**
*We can find $K_0, \gamma > 0$ for which*

$$\sup_{t\geq 0} \sup_{\|\varphi\|_\infty \leq 1} \mathbb{E}_\eta \left[ |m(\eta_t)(\varphi) - m(\eta) T_t \varphi| \right] \leq \frac{K_0}{N^\gamma}$$

*Furthermore, if $\eta$ is distributed according to the stationary distribution of the system $\eta_N$, then there exist $K_0 > 0$ and $\gamma > 0$ such that*

$$\mathbb{E} \left[ |m(\eta)(\varphi) - \nu(\varphi)| \right] \leq \frac{K_0}{N^\gamma}$$

**Conjecture**

*If $(\eta, \mathfrak{r})_{A,N}$ is distributed according to the invariant distribution of a N-particle swapped Fleming-Viot system with swapping rate A with first marginal $\eta_{A,N}$, then there exists a sequence $(A_N) > 0$, such that $A_N \to \infty$ and for any sequence $B_N \leq A_N$*

$$\lim_{N\to\infty} \mathbb{E}\left[|m(\eta_{B_N,N})(\varphi) - \nu(\varphi)|\right] = 0$$

*for any $\|\varphi\|_\infty \leq 1$.*

The theory of two-time scale Markov processes asks what happens when

$$Q = \frac{1}{\varepsilon}\widetilde{Q} + \widehat{Q}$$

and $\varepsilon \to 0$.

# Infinite Swapping

The theory of two-time scale Markov processes asks what happens when

$$Q = \frac{1}{\varepsilon}\widetilde{Q} + \widehat{Q}$$

and $\varepsilon \to 0$.

To decompose the rates into dynamics/swapping, we need to define a new variable, $(S_t^{(n)}) \in \{-1, 1\}$ representing the order of the pair $n$ at time $t$.

# Infinite Swapping

The theory of two-time scale Markov processes asks what happens when

$$Q = \frac{1}{\varepsilon}\widetilde{Q} + \widehat{Q}$$

and $\varepsilon \to 0$.

To decompose the rates into dynamics/swapping, we need to define a new variable, $(S_t^{(n)}) \in \{-1, 1\}$ representing the order of the pair $n$ at time $t$.

*Why infinite swapping?* Computational efficiency.

# INS Algorithm

$$\nu_{x,y} = \frac{r_{x,y}}{r_{x,y} + r_{y,x}}$$

- If $(X, Y)^{(n)} = (x, y)$ then $X$ moves through its dynamics to $z$ with rate
$$\nu_{x,y} Q_{x,z} + \nu_{y,x} Q_{z,x}$$
.
- If $(X, Y)^{(n)} = (x, y)$ then $Y$ moves through its dynamics to $z$ with rate
$$\nu_{x,y} Q_{z,y} + \nu_{y,x} Q_{y,z}$$
.

# INS Algorithm

- If $(X, Y)^{(n)} = (x, y)$, then $X$ is killed with rate

$$\nu_{x,y} c(x) + \nu_{y,x} d(y)$$

  Once killed, it chooses another pair uniformly, and then, if the pair is at positions $(q, p)$, it moves to $q$ with probability $\nu_{q,p}$ and $p$ with probability $\nu_{p,q}$
- Similar for $Y$

An asymptotically consistent estimator of the QSD $\phi$:

$$\phi(dx) = \frac{1}{TN} \int_{t=0}^{T} \sum_{n=1}^{N} \nu_{X_t^{(n)}, Y_t^{(n)}} \, \delta_{X_t^{(n)}}(dx) + \nu_{Y_t^{(n)}, X_t^{(n)}} \delta_{Y_t^{(n)}}(dx) \; dt$$

# References i


B. Cloez and M.-N. Thai.
**Quantitative results for the fleming-viot particle system and quasi-stationary distributions in discrete space.**
*Stochastic Processes and their Applications*, 2014.

H. W. J. G. Wei Hou, Qingdong Liu.
**Effect of cyclic intercritical tempering on the microstructure and mechanical properties of a low-carbon cu-bearing 7ni steel.**
*Metallurgical and Materials Transactions A*, 2020.

Questions?