Some remarks on spectral analysis of Markov chains: from Protein Dynamics to MCMC algorithms

Florian Maire, Olivier Roux [1] and Prof. Vio Buchete [2]

Statistical Learning Working Group, November 2016

[1] Centrale Marseille, France
[2] UCD, School of Physics

1 Introduction

2 Protein Dynamics and Approximations

3 Connection discrete time – continuous time Markov chain

4 Spectral Analysis of Markov operators

Outline

1 Introduction

2 Protein Dynamics and Approximations

3 Connection discrete time – continuous time Markov chain

4 Spectral Analysis of Markov operators

Context

- Protein is a chain of amino acids
- Proteins are assumed to be found in a (possibly large) number of states $X \in \mathcal{S}$
- A state encompasses information such as the distance/angle between amino acids, energy levels... (so $|\mathcal{S}| \gg 1$)

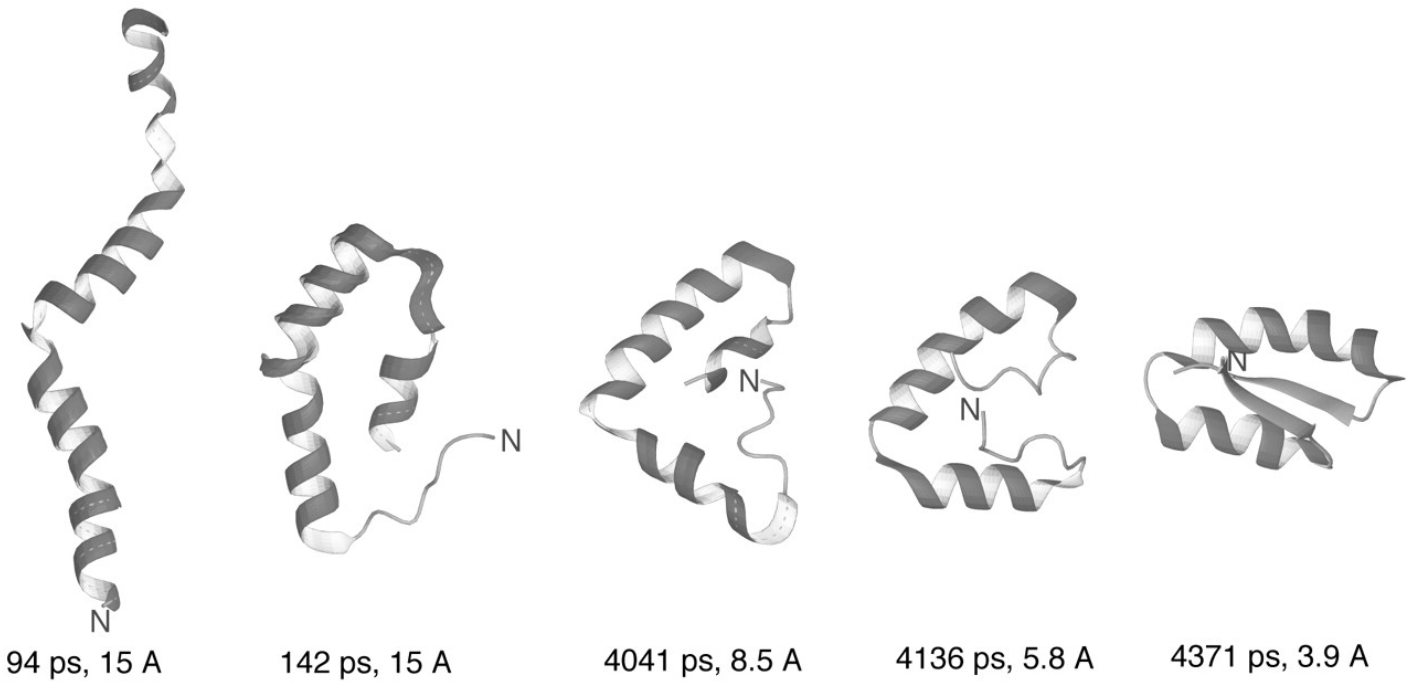


Figure: Protein folding pathway of 1E0G obtained in Langevin dynamics simulations (A. Liwo et al, PNAS, 2005)

Inference of protein dynamics

A model for protein dynamics (typically a Markovian process $\{X_t\}$ $(t > 0)$) can be fitted from experimental measurements

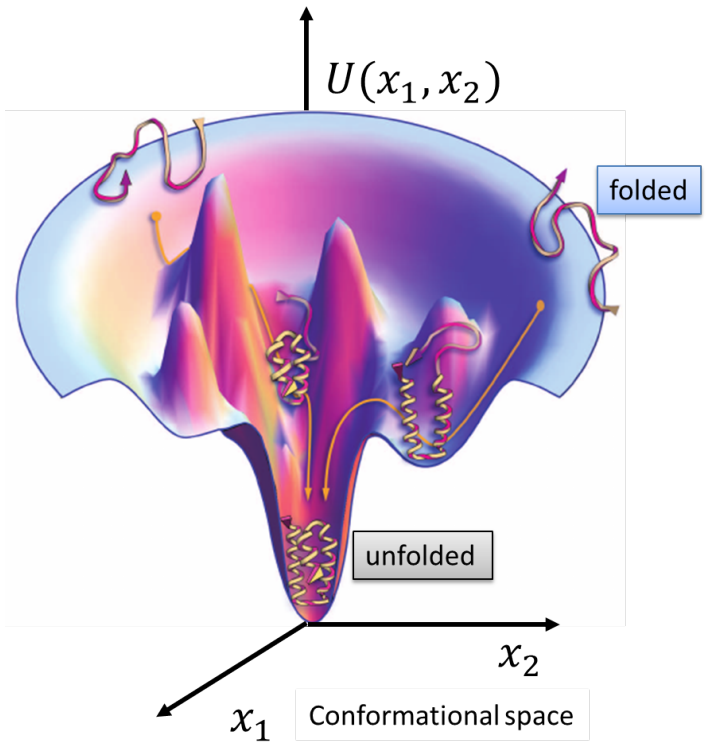


Figure: Model of a protein energy landscape (H. Ma et al, PNAS, 2006)

Equilibrium distribution of $\{X_t\}$ is

$$\Pr(X_t = x) \propto \exp\{-H(x)\}$$

Inference of protein dynamics

Interest lies in getting summaries from the model

- Is there a simpler model? *i.e.* $X \in \mathcal{S}'$ and $|\mathcal{S}'| \ll |\mathcal{S}|$
  $\Rightarrow$ Is there an equivalent two state model
  $\mathcal{S} = \{\text{Folded}\}, \{\text{Unfolded}\}$?
- If the protein is at an intermediate state what is the chance that it will first Fold before Unfold?

Spectral representation of the Markov operator ruling the dynamic of $\{X_t\}$ reveals some answers to those questions.

Markov chain Monte Carlo algorithms

Class of algorithms that simulate discrete time Markov chains $\{X_k\}$ ($k \in \mathbb{N}$) to perform numerical integration (Bayesian inference in particular)

Quantitatively, those algorithms are usually assessed/ranked according to:

- Speed of convergence of the law of the chain to the target distribution
- Asymptotic variance of Monte Carlo estimators

Here again, those quantities can be revealed by the spectral analysis of the Markov operator that simulates $\{X_k\}$

Purpose of the talk

- Gain some insight on how to diagnose when a two state approximation of the protein dynamics model is relevant
- Quantify this approximation
- Compare spectral analysis of continuous time Markov chain vs discrete time Markov chains in general state space
- How the approximation of the chain distribution established in CTMC can be interpreted in the context of MCMC algorithms

Outline

1 Introduction

2 Protein Dynamics and Approximations

3 Connection discrete time – continuous time Markov chain

4 Spectral Analysis of Markov operators

Specification of the model

- Even though protein dynamics are naturally parameterized by continuous parameters (angles, momenta, energy), it is assume to be discretized *i.e.*

$$X_t \in \{1, 2, \ldots, N\}$$

with $N \gg 1$

- Each state i corresponds to a given range of angle, energy...
- The process $\{X_t\}$ is assumed to be memoryless (Master equation)

$$\begin{aligned}\Delta\{\text{particles in state } i \text{ during t,t+d}\} = \\ \#\{\text{new particles in state i during in t,t+dt}\} \\ - \#\{\text{particles leaving state i during in t,t+dt}\}\end{aligned}$$

and

$$\begin{aligned}\#\{\text{new particles in state i during in t,t+dt}\} \\ = \sum_{k \neq i} \alpha_{k,i} \#\{\text{particles in state } k \text{ at time t}\}\end{aligned}$$

Continuous Time Markov chain (CTMC)

The previous assumptions are equivalent to CTMC model.
Let $\{X_t\}$ $(t \geq 0)$ be a stochastic process on the discrete state $\mathcal{S} = \{1, \ldots, N\}$ that evolves as follows:

$$\mathbb{P}(X_{t+dt} = j \,|\, X_{0:t}, X_t = i) = \mathbb{P}(X_{t+dt} = j \,|\, X_t = i) = Q_{i,j}dt + o(dt)\,, \qquad i \neq j \quad (1)$$

In this assumption $\{X_t\}$ is a Continuous Time Markov chain (CTMC) and is characterized by:

1. an initial distribution π_0 on $(\mathcal{S}, \mathfrak{S})$
2. an infinitesimal generator $Q \in \mathcal{M}_N(\mathbb{R})$ defined as
   - $i \neq j$
   $$Q_{i,j} = \frac{\mathrm{d}}{\mathrm{d}t} \lim_{t \downarrow 0} \mathbb{P}(X_t = j \,|\, X_0 = j)$$
   - $i = j$
   $$Q_{i,i} = -\sum_{i \neq j} Q_{i,j}$$

Distribution of the CTMC

We denote by $P(t) = \{\mathbb{P}(X_t = j \,|\, X_0 = i)\}_{i,j}$ the matrix of probability of transition of the CTMC. It is related to Q by:

$$P(t) = \exp\{Qt\}$$

(from Kolmogorov equations)
Denoting $\pi(t) = \mathbb{P}(X_t \in \cdot) = \sum_{i=1}^{N} \mathbb{P}(X_t \in \cdot, X_0 = i)$, we have:

$$\pi(t) = \pi(0) \exp\{Qt\}$$

Assumption 1

We will assume that $\{X_t\}$

- *has an unique stationary distribution* π
- *is time reversible:*

$$\pi_i P_{i,j}(t) = \pi_j P_{j,i}(t)\,, \qquad \forall\, t > 0\,.$$

Spectral decomposition of Q

Under Assumption 1, we have:

$$\pi P(t) = \pi, \qquad P(t)\mathbf{1} = \mathbf{1}, \qquad \mathrm{sp}(P(t)) \subseteq (-1, 1)\,.$$

Spectral properties of $P(t)$ propagates to that of Q.

Proposition 1

$sp(P(t))$ and $sp(Q)$ are connected:

$$\lambda \in sp(Q) \Leftrightarrow \exp\{\lambda t\} \in sp(P(t))$$

Indeed, let y be a right eigenvector of Q with eigenvalue λ, then

$$P(t)y = \sum_{k=0}^{N} \frac{t^k}{k!} Q^k y = \sum_{k=0}^{N} \frac{(\lambda t)^k}{k!} y = \exp\{\lambda t\} y\,.$$

- $\mathbf{1} \in \mathrm{sp}(P(t))$ implies $0 \in \mathrm{sp}(Q)$
- $\pi Q = 0$ so $y_1^{(L)} = \pi$
- $Q\mathbf{1} = 0$ so $y_1^{(R)} = \mathbf{1}$
- And in fact $\mathrm{sp}(Q) \subset (-\infty, 0)$

Spectral decomposition of Q

Under Assumption 1, Q is diagonalizable:

$$Q = UDU^{-1}$$

where:

- $U = [y_1^{(R)} \cdots y_N^{(R)}]$ and $y_1^{(R)}, \ldots, y_N^{(R)}$ are Q right eigenvectors, with $y_1^{(R)} = \mathbf{1}_n$
- $U^{-1} = [y_1^{(L)'} \cdots y_N^{(L)'}]'$ and $y_1^{(L)}, \ldots, y_N^{(L)}$ are Q left eigenvectors, with $y_1^{(L)} = \pi$
- $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ with $\lambda_1 = 0$ and $\lambda_i < 0$ for $i > 1$

so that:

$$\pi(t) = \sum_{\ell=1}^{n} \left\langle \pi_0, y_\ell^{(R)} \right\rangle \exp\{\lambda_\ell t\} y_\ell^{(L)}$$

$$= \pi + \sum_{\ell=2}^{n} \left\langle \pi_0, y_\ell^{(R)} \right\rangle \exp\{\lambda_\ell t\} y_\ell^{(L)}$$

Interpretation of the chain distribution

First, note that $\sum_{i=1}^{N} y_\ell^{(L)}(i) = 0$

$$\pi(t) = \pi + \sum_{\ell=2}^{n} \left\langle \pi_0, y_\ell^{(R)} \right\rangle \exp\{\lambda_\ell t\} y_\ell^{(L)}$$

- one stationary process (π)
- $(N-1)$ transient processes ρ_ℓ

$$\rho_\ell(t) = \left\langle \pi_0, y_\ell^{(R)} \right\rangle \exp\{\lambda_\ell t\} y_\ell^{(L)}$$

that act as probability mass transfer.

If $\pi_0 = \pi + \delta y_\ell^{(L)}$, it would take $\tau_\ell := -1/\lambda_\ell$ time to absorb the perturbation (relaxation time).

Spectral gap and the two state approximation

In protein dynamics, a Physicist is typically interested in knowing if $\{X_t\}_t$ could be represented by a two state system

$$\tilde{X}_t \in \{\text{Unfolded}, \text{Folded}\}$$

and what the transition rate between those two states look like.

Rule 1 (Buchete and Hummer, 2008)

If there is a "large enough" gap in the spectrum of Q, that is

$$\gamma = |\lambda_2 - \lambda_3|/|\lambda_1 - \lambda_2| \approx 10$$

then $\{\tilde{X}_t\}$ is a "good" approximation of $\{X_t\}$.

Remark:

$$\gamma = \left|\frac{\lambda_2}{\lambda_3} - 1\right| = \left|\frac{\tau_2}{\tau_3} - 1\right| .$$

Interpretation of this rule

Considering the time $t' = t/\tau_2$

$$\pi(t') = \pi + \beta_2 \exp\{-t'\} y_2^{(L)} + \beta_3 \exp\{-t'(\gamma+1)\} y_3^{(L)} + \cdots$$

with $\beta_i = \left\langle \pi_0, y_\ell^{(R)} \right\rangle$.
If $\gamma \approx 10$ then the $N-2$ slowest processes might be neglected without much error

$$\pi(t') \approx \tilde{\pi}(t') = \pi + \beta_2 \exp\{-t'\} y_2^{(L)}$$

Fact 1

If $\gamma \approx 10$, then the only probability mass transfer is between the two basins of potential, namely $\{Unfolded\}$ and $\{Folded\}$.

Simulating random rate matrices

To gain more insight on the 2 state approximations, we resorted to simulating random rate matrix:

- simulate a stationary probability π
- simulate the lower triangular rate matrix that has the m-first diagonals non-zeros filled with

$$Q_{i,j} \sim \text{expo}(\lambda)$$

- fill the upper triangular to ensure reversibility $Q_{j,i} = Q_{i,j}\pi_i/\pi_j$
- fill the diagonal so that $Q_{i,i} = -\sum_{j \neq i} Q_{i,j}$

The parameter λ was tuned in order to match the slowest relaxation time usually observed in protein dynamics.

Two state approximation

We compare on one example where $\gamma = 7.8$ ($N = 52$), the true distribution π and the two state approximation $\tilde{\pi}$

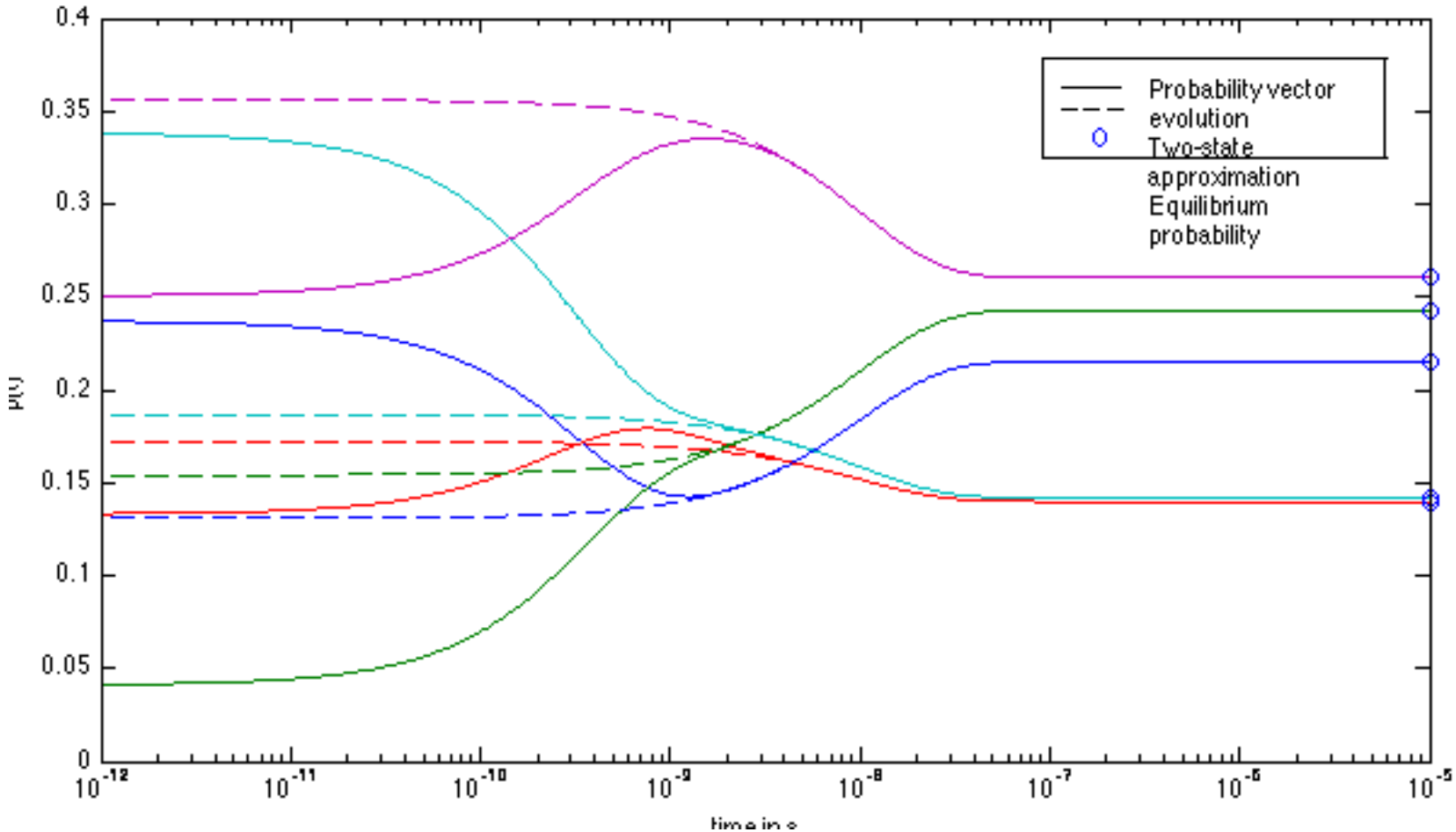


Figure: Representation of the probabilities $\pi(t)$ and $\tilde{\pi}(t)$ throughout time (in log scale) for four states.

Correlation between γ and error L_2

To quantify the approximation, we define

$$\text{Err} = \int \|\pi(t) - \tilde{\pi}(t)\|_2 \mathrm{d}t$$

and try to estimate the correlation with the gap.

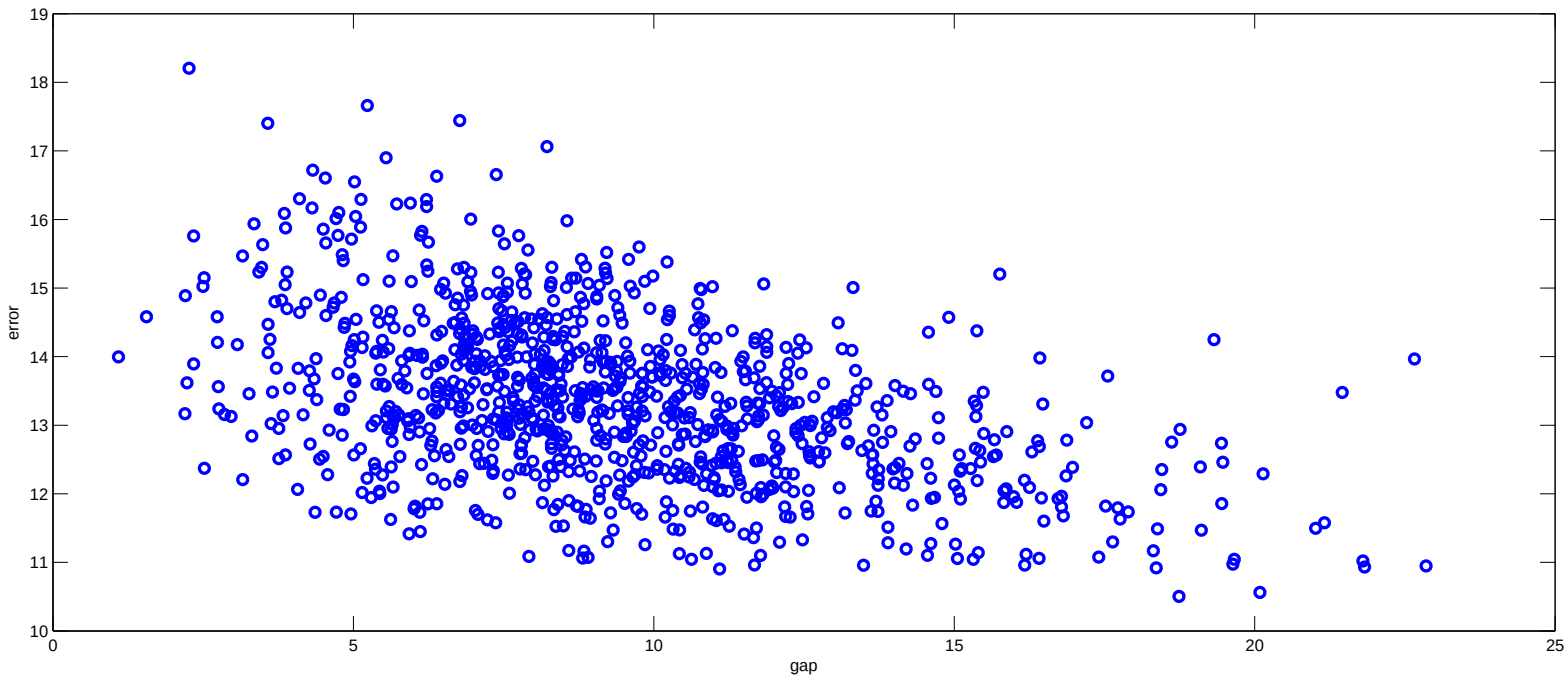
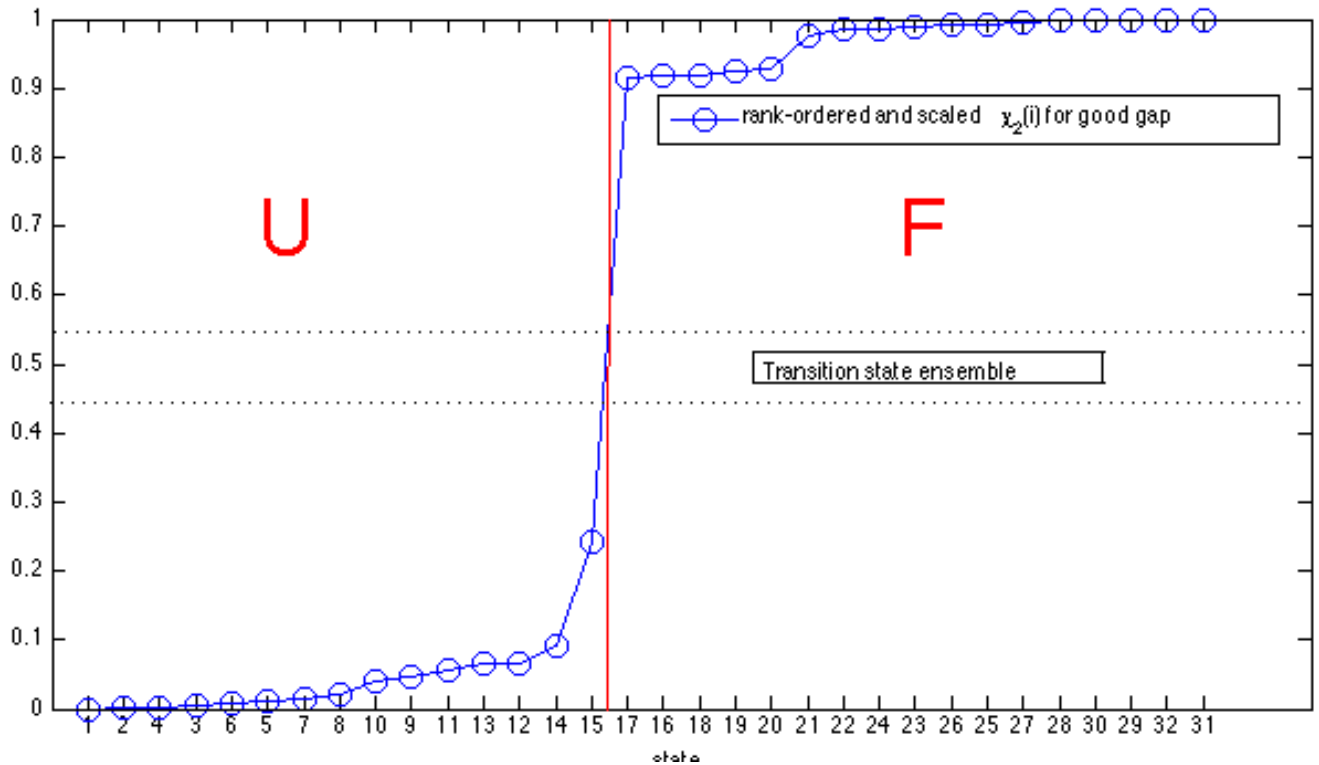


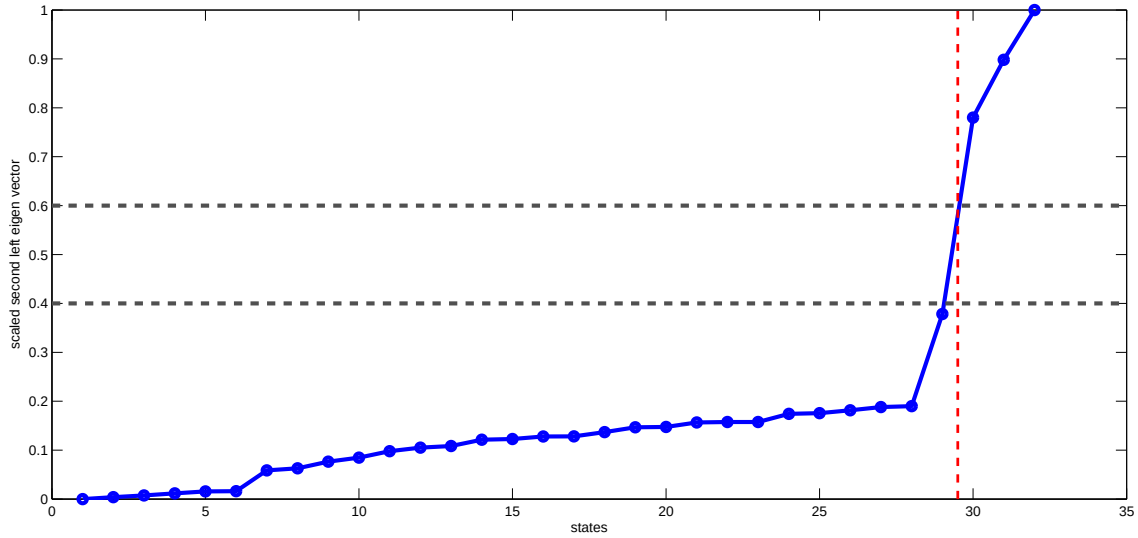
Figure: Realizations of gap plotted against Err.

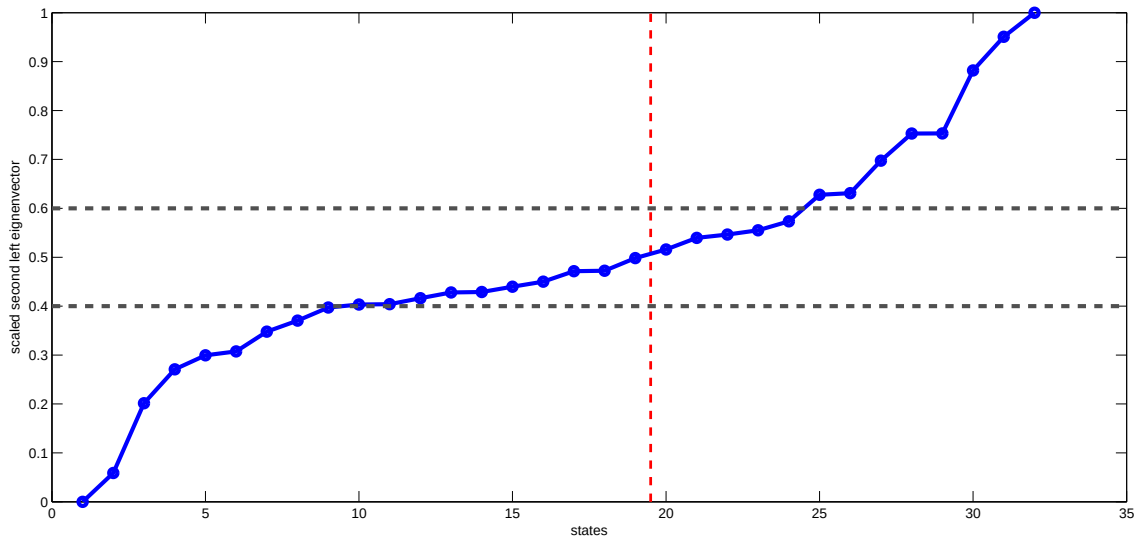
Illustration of the two states probability mass transfer channel

Representation of the scaled second left eigenvector of Q: $y_2^{(L)}$ (in this example $\gamma = 25.2$).



Representation of the scaled second left eigenvector of Q: $y_2^{(L)}$($\gamma = 8.2$ at the top and $\gamma = 0.65$ at the bottom)





Gap distribution ($N = 32$)

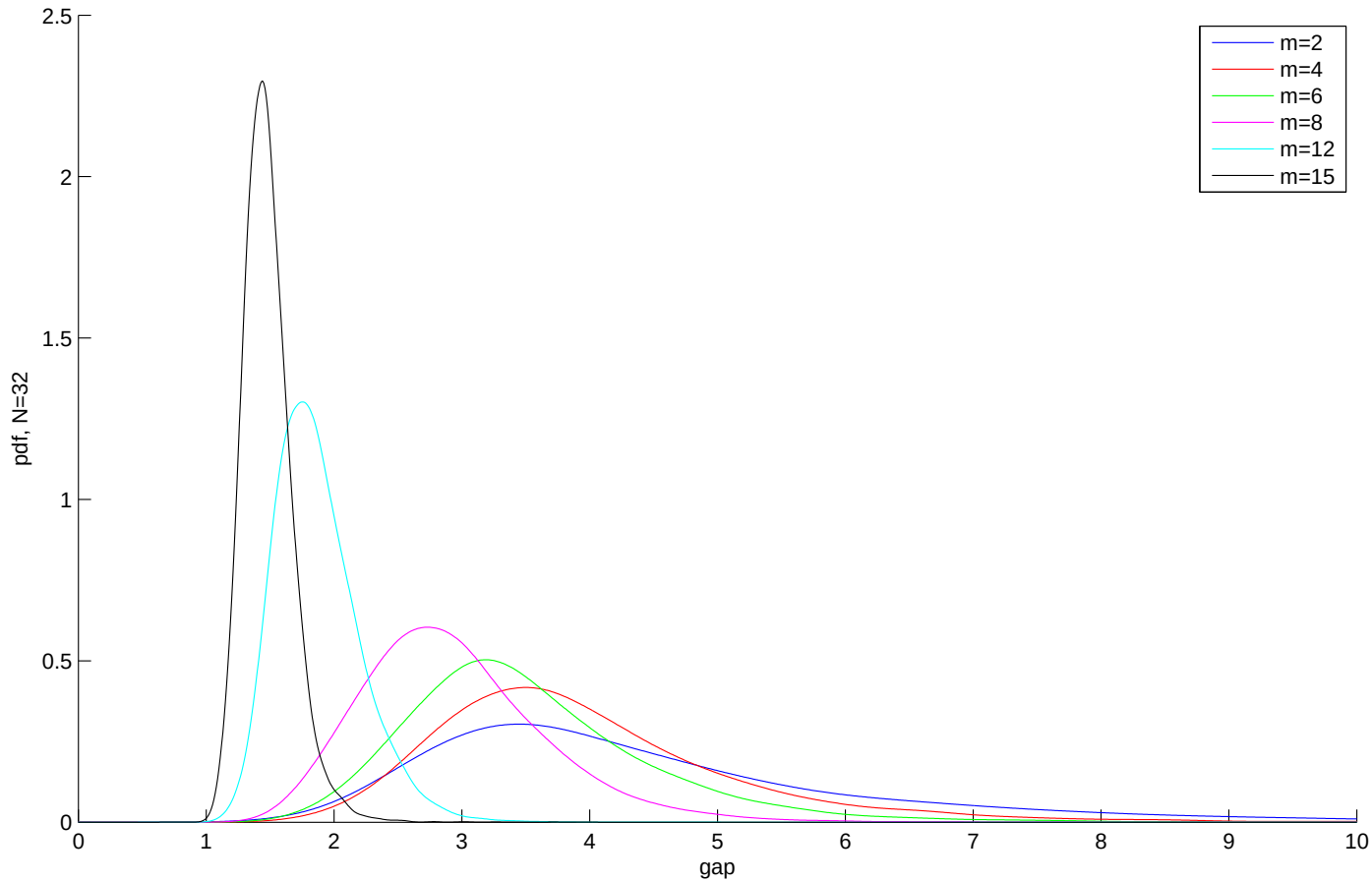


Figure: Gap probability density function for band diagonal random rate matrix of different structures.

Gap distribution ($N = 64$)

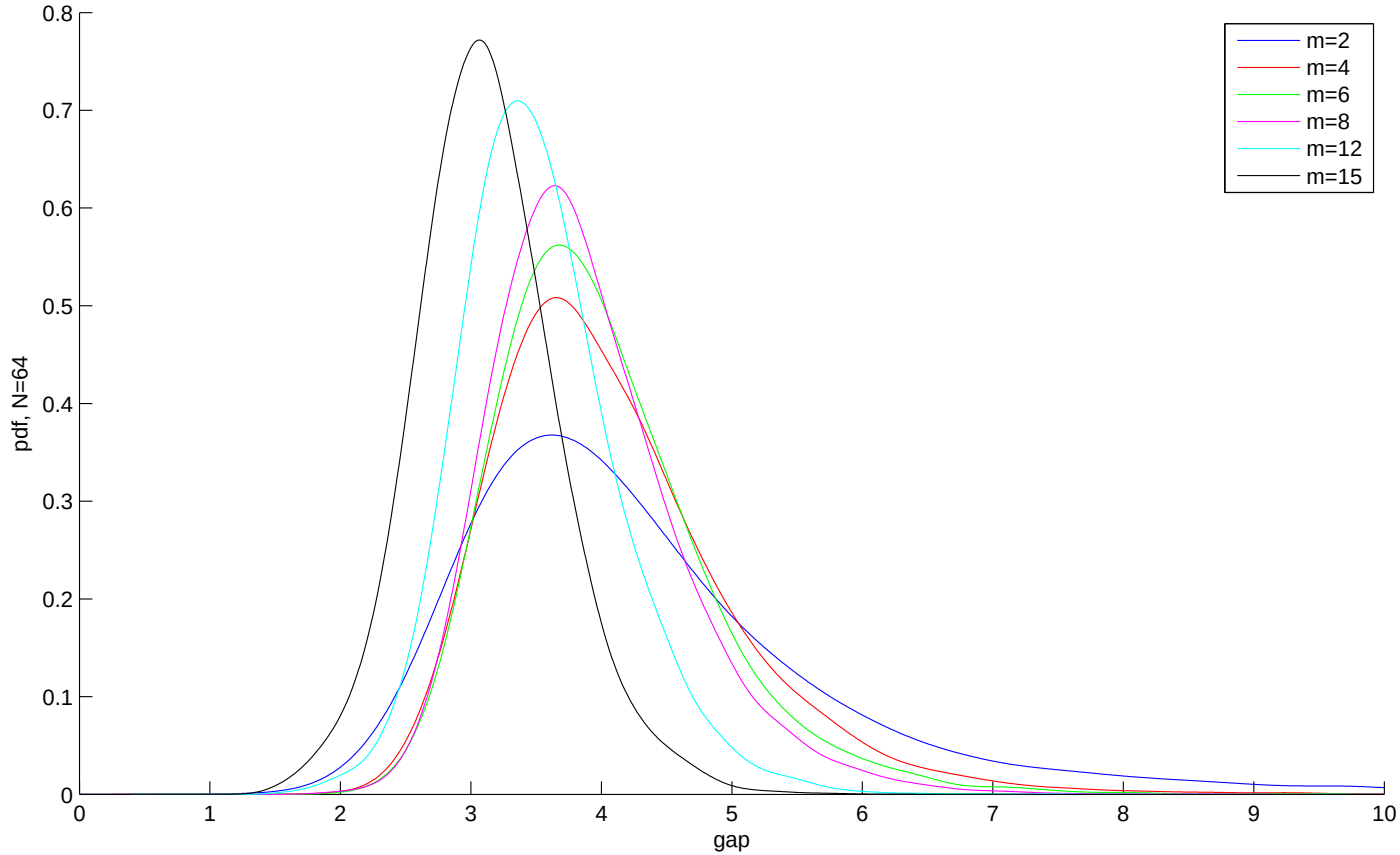


Figure: Gap probability density function for band diagonal random rate matrix of different structures.

Outline

1 Introduction

2 Protein Dynamics and Approximations

3 Connection discrete time – continuous time Markov chain

4 Spectral Analysis of Markov operators

Discrete Time Markov chain (DTMC)

Let $\{X_k\}$ $(k \in \mathbb{N})$ be a stochastic process defined on the discrete state space $\mathcal{S} = \{1, \ldots, N\}$ that evolves as follows:

$$\text{for all } (i,j) \in \mathcal{S}^2, \quad \mathbb{P}(X_{k+1} = j \,|\, X_{0:k-1}, X_k = i) = \mathbb{P}(X_{k+1} = j \,|\, X_k = i) = P_{i,j}\,,$$

Under this assumption $\{X_k\}$ is a Discrete Time Markov chain (DTMC) and is characterized by:

1. an initial distribution π_0 on $(\mathcal{S}, \mathfrak{S})$
2. a probability transition matrix $P \in \mathcal{M}_N(\mathbb{R})$ satisfying:
   - for all $(i,j) \in \mathcal{S}^2$
   $$P_{i,j} \in (0,1)$$
   - for all $i \in \mathcal{S}$
   $$\sum_{j \in \mathcal{S}} P_{i,j} = 1$$

A first observation

$$\mathbb{P}(X_{t+dt} = j \,|\, X_t = i) = Q_{i,j}dt + o(dt)\,, \qquad i \neq j$$

$$\Updownarrow$$

$$\mathbb{P}(X_{t+dt} \neq i \,|\, X_t = i) = \underbrace{\sum_{j \neq i} Q_{i,j}}_{\lambda_i} dt + o(dt) \qquad (2)$$

- same type of assumption than in a Poisson Process
  $\Rightarrow$ we know that the time until a change of state (*holding time*) is $\tau \sim \text{expo}(\lambda_i)$ and

$$X_{t:t+\tau^-} = i$$

- next state satisfies

$$\mathbb{P}(X_{t+\tau} = j \,|\, X_{t+\tau^-} = i, X_{t+\tau} \neq i)$$

$$= \begin{cases} 0 & \text{if } i = j \\ \frac{\mathbb{P}(X_{t+\tau}=j \,|\, X_{t+\tau^-}=i)}{\mathbb{P}(X_{t+\tau}\neq i \,|\, X_{t+\tau^-}=i)} = Q_{i,j}/\lambda_i & \text{otherwise} \end{cases}$$

Algorithms to simulate a CTMC

From the previous observation, we deduce that Algorithm 1 simulates the CTMC of interest:

Algorithm 1

(1) *draw an initial state* $Y_0 \sim \pi_0$ *(say* $Y_0 \rightsquigarrow i$*)*

(2) *draw a holding time (given* $Y_0 = i$*)* $\tau_0 \sim expo(\lambda_i)$

(3) *set* $X_{0:\tau_0} = i$ *and draw a new state* $Y_1 \sim \bar{P}_{i,\cdot}$ *where*

$$\bar{P}_{i,j} = \begin{cases} 0 & \text{if } i = j \\ Q_{i,j}/\lambda_i & \text{otherwise} \end{cases}$$

And then iterate (2)–(3)

Algorithms to simulate a CTMC

Let $\lambda^* = \max_i \lambda_i$ and consider the following modification of Algorithm 1:

Algorithm 2

(1) *draw an initial state* $Y_0 \sim \pi_0$ *(say* $Y_0 \rightsquigarrow i$*)*

(2) *draw a holding time* $\tau_0 \sim expo(\lambda^*)$

(3) *set* $X_{0:\tau_0} = i$ *and draw a new state* $Y_1 \sim \bar{R}_{i,\cdot}$ *where*

$$\bar{R}_{i,j} = \begin{cases} 1 - \lambda_i/\lambda^* & \text{if } i = j \\ Q_{i,j}/\lambda^* & \text{otherwise} \end{cases}$$

And then iterate (2)–(3)

CTMC–DTMC

Algorithm 2 can be decoupled:

- simulate a Poisson process $\{N_t\}$ with parameters λ^*
- simulate a DTMC $\{Y_k\}$ with transition matrix $\bar{R}$

Proposition 2

Simulate $\{Y_k\}$ and $\{N_t\}$ as above. Define $\{X_t\}$ as

$$\forall\, t \geq 0\,, \qquad X_t = Y_{N_t}\,.$$

Then $\{X_t\}$ is the desired CTMC.

	a counting process: N_t	a stochastic matrix: $\bar{R}_{i,j}$ $(i \neq j)$
DTMC	$\delta_{\mathbb{N}}$	$P_{i,j}$
CTMC	$\mathrm{PP}(\lambda^*)$	$Q_{i,j}/\lambda^*$

CTMC–DTMC: chain distribution

For the DTMC, it is straightforward to show that

$$\pi_n = \mathbb{P}(X_n \in \cdot) = \sum_{i=1}^{N} \mathbb{P}(X_n \in \cdot, X_0 = i) = \sum_{i=1}^{N} \pi_0(i) P^n(i, \cdot) = \pi_0 P^n$$

and for the CTMC, inspired by Algorithm 2, we write:

$$\pi(t) = \sum_{i=1}^{N} \sum_{n=0}^{\infty} \mathbb{P}(X_n \in \cdot, X_0 = i, N_t = n) = \sum_{n=0}^{\infty} \pi_0 \bar{R}^n \frac{(\lambda^* t)^n}{n!} \exp\{-\lambda^*\}\,.$$

Outline

1 Introduction

2 Protein Dynamics and Approximations

3 Connection discrete time – continuous time Markov chain

4 Spectral Analysis of Markov operators

(Discrete Time) Markov chain on general state space

Let $\{X_k\}$ ($k \in \mathbb{N}$) be a stochastic process defined on the general state space $(\mathbb{R}, \mathcal{B})$ that evolves as follows:

$$\text{for all } x_k \in \mathbb{R},\ A \in \mathcal{B}\,, \qquad \mathbb{P}(X_{k+1} \in A \,|\, X_{0:k-1}, X_k = x_k) = \mathbb{P}(X_{k+1} \in A \,|\, X_k = x_k) = P(x_k, A)\,.$$

Under this assumption $\{X_k\}$ is a (Discrete Time) Markov chain on general state space and is characterized by:

1. an initial distribution π_0 on $(\mathbb{R}, \mathcal{B})$
2. a conditional probability distribution $P(x, \cdot)$ on $(\mathbb{R}, \mathcal{B})$, determined by a function $p : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$, satisfying:
   - for all $x \in \mathbb{R}$
   $$P(x, \mathbb{R}) = \int_{\mathbb{R}} p(x, \mathrm{d}y) = 1$$
   - for all $x \in \mathbb{R}$, $A \in \mathcal{B}$
   $$P(x, A) = \int_{A} p(x, \mathrm{d}y) \in (0, 1)$$

Transition matrix as operator

For DTMC the operator is the transition matrix P and for CTMC the operator is $P(t)$

- left operator: let μ is a probability measure on $\mathcal{S}$,

$$P : \mu \mapsto \mu P$$

  is also a measure on $\mathcal{S}$, it is interpreted as

$$\mu \mapsto \mathbb{P}(X_1 \in \cdot \,|\, X_0 \sim \mu)$$

- right operation: let ω be a vector of $\mathbb{R}^n$,

$$P : \omega \mapsto P\omega$$

  is interpreted as for all $i \in \mathcal{S}$:

$$\omega_i \mapsto \mathbb{E}\{\omega_{X_1} \,|\, X_0 = i\}$$

Markov kernel as operator

For Markov chains on general state space, the operator is $P(x, \cdot)$

- left operation: let μ is a probability measure on $(\mathbb{R}, \mathcal{B})$,

$$P : \mu \mapsto \mu P = \int_{\mathbb{R}} \mu(\mathrm{d}x) P(x, \cdot)$$

  is also a measure on $(\mathbb{R}, \mathcal{B})$, it is interpreted as

$$\mu \mapsto \mathbb{P}(X_1 \in \cdot \,|\, X_0 \sim \mu)$$

- right operation: let $f : \mathbb{R} \to \mathbb{R}$ be a function,

$$P : f \mapsto Pf = \int_{\mathbb{R}} P(\cdot\,, \mathrm{d}x) f(x)$$

  is interpreted as:

$$f \mapsto \mathbb{E}\{f(X_1) \,|\, X_0 = \cdot\}$$

Markov operator on general state space

Consider the (infinite-dimensional) space of functions $\mathcal{L}^2(\pi)$ defined as:

$$\mathcal{L}^2(\pi) = \left\{ f : \mathbb{R} \to \mathbb{R}, \int \pi(\mathrm{d}x) f(x)^2 < \infty \right\}$$

equipped with the scalar product

$$\langle f, g \rangle = \int \pi(\mathrm{d}x) f(x) g(x) \,.$$

Assumption 2

We assume that $\{X_k\}$ is π-reversible:

$$\int_A \pi(\mathrm{d}x) P(x, B) = \int_B \pi(\mathrm{d}x) P(x, A)$$

(ie detailed balance condition holds).

Proposition 3

Under Assumption 1, the operator P is a self adjoint operator on $\mathcal{L}^2(\pi)$.

Spectral analysis on $\mathcal{L}^2(\pi)$

Theorem 1 (spectral theorem)

Let P be a compact and self-adjoint operator on an Hilbert space $\mathcal{H}$, then

- *there exists an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of P*
- *the non-zero eigenvalues of P form a **finite** or **countably infinite** set $\{\lambda_k\}$ such that*

$$P = \sum_{\ell \geq 1} \lambda_\ell \Pi_\ell$$

where Π_ℓ is the projection onto the eigenspace with eigenvalue λ_k

So for any $f \in \mathcal{L}^2(\pi)$,

$$Pf = \left(\sum_{\ell \geq 1} \lambda_\ell \Pi_\ell \right) f = \sum_{\ell > 1} \lambda_\ell \Pi_\ell f$$

Case where the Markov kernel is compact

Let $\{X_k\}$ be generated by the Metropolis-Hastings algorithm:

1. propose $\tilde{X} \sim Q(X_k, \cdot)$
2. set $X_{k+1} = \tilde{X}$ w.p.

$$1 \wedge \frac{\pi(\tilde{X})Q(\tilde{X}, X_k)}{\pi(X_k)Q(X_k, \tilde{X})}$$

and $X_{k+1} = X_k$ otherwise.

Proposition 4 (Atchadé and Perron, 2002)

If

$$\iint Q(x, \mathrm{d}y)Q(y, \mathrm{d}x) < \infty$$

then P (the operator induced by M–H) is compact.

In particular this is always true for Independent M-H ($Q(x, \mathrm{d}y) = Q(\mathrm{d}y)$)

Chain distribution

To apply the Spectral Theorem to simplify

$$\mathbb{P}(X_n \in \cdot) = \pi_0 P^n(\cdot)$$

we need to map the left operator P to its right.

Proposition 5

Let $X_0 \sim \pi_0$ be the initial distribution of $\{X_k\}$. Assume $\{X_k\}$ is π-reversible, (P is self adjoint), then

$$\pi_0 P^n(A) = \int_A \pi(\mathrm{d}x) P^n f_0(x)\,,$$

where $f_0 = \mathrm{d}\pi_0/\mathrm{d}\pi$.

Corrolary 1

Let $\{X_k\}$ be a π-reversible Markov chain and P compact, then for any $A \in \mathcal{B}(\mathbb{R})$

$$\pi_n(A) = \int_A \pi(\mathrm{d}x) \sum_{\ell \geq 1} \lambda_\ell^n f_0^{(\ell)}(x)\,, \qquad f_0^{(\ell)} = \Pi_\ell f_0$$

Chain distribution

Denote by $\{\varepsilon_\ell\}_{\ell\geq 1}$ the eigenvectors of P.
Since $\lambda_1 = 1 \in \text{sp}(P)$ and its eigenvector is the constant function, $\varepsilon_1 = 1$, we have:

$$\pi_n(A) = \int_A \pi(\mathrm{d}x)\,\langle f_0, 1\rangle\, 1(x) + \varepsilon_\ell(x) + \sum_{\ell\geq 2} \lambda_\ell^n \int_A \pi(\mathrm{d}x)\,\langle f_0, \varepsilon_\ell\rangle\,\varepsilon_\ell(x)\,,$$

but $\langle f_0, 1\rangle = \int \pi(\mathrm{d}x) f_0(x) 1 = \int \pi_0(\mathrm{d}x) = 1$ so that the first term is simply $\pi_n(A)$.
Comparing discrete and continuous contexts we have for $i \in \{1, \ldots, n\}$ and $A \in \mathcal{B}(\mathbb{R})$:

$$\pi_t(i) = \pi(i) + \sum_{\ell=2}^{n} \exp\{\lambda_\ell t\} \left\langle \pi_0, y_\ell^{(R)} \right\rangle y_\ell^{(L)}(i) \tag{3}$$

$$\pi_n(A) = \pi(A) + \sum_{\ell\geq 2} \lambda_\ell^n \langle f_0, \varepsilon_\ell\rangle \int_A \pi(\mathrm{d}x)\varepsilon_\ell(x) \tag{4}$$

$\Rightarrow$ Under the self adjoint and compact assumption, a same interpretation of the probability mass transfer can be given to Markov chain on general state space.

Case where the Markov kernel is not compact

The spectral analysis is much more complicated.

Theorem 2 (Von Neumann's Spectral Theorem)

If P is self-adjoint and $(f, g) \in \mathcal{L}^2(\pi)$, then

$$\langle \phi(P) f, g \rangle = \int_{sp(P)} \phi(\lambda) \mathrm{d}\mu_{f,g,P}(\lambda)$$

where $\mathrm{d}\mu_{f,g,P}$ is a measure defined as $\mathrm{d}\mu_{f,g,P}(\lambda) = \langle E_P(\lambda) f, g \rangle$ where E_P is the spectral measure (a projection valued measure) of P.

Corrolary 2

If P is self-adjoint, then:

$$\pi_n(A) = \pi_0 P^n(A) = \int_{sp(P)} \lambda^n \mathrm{d} \langle E(\lambda) f_0, \mathbb{1}_A \rangle$$

$\Rightarrow$ since the spectrum is continuous it is not clear how to exhibit a similar decomposition than in the compact case.

Discussion

- spectral analysis of CTMC allows to derive reliable statistics giving insight on the process of interest:
  - a spectral gap between the second and third eigenvalue (a ratio of 10) seems to support a two state approximation (simulation)
  - but it is not true that if two models have two spectral gaps γ_1, γ_2 such that $\gamma_1 < \gamma_2$ then model 2 will yield a better two state approximation than model 1
- Interestingly, this analysis is very similar for Markov operator used to simulate Markov chains in MCMC algorithms
  - looks promising if the operator is compact
  - more work if we remove this assumption
- the decomposition of the MCMC distribution allows an interpretation of the chain dynamics in terms of flux of probability mass between high density regions