Some ordering results for a class of nonreversible Markov (chain & process) Monte Carlo algorithms

Christophe Andrieu
joint with Sam Livingstone

September 4, 2018
Outline

Motivation

Reversible scenario & known results

Modified/Skew detailed balance and Yaglom reversibility

Some applications

Tangential (but interesting) results

The continuous time scenario
Outline

Motivation

Reversible scenario & known results

Modified/Skew detailed balance and Yaglom reversibility

Some applications

Tangential (but interesting) results

The continuous time scenario
Overview

- Assume one is interested in sampling from a probability distribution \( \pi \), defined on \((X, \mathcal{X})\)
Overview

- Assume one is interested in sampling from a probability distribution \( \pi \), defined on \((X, \mathcal{X})\).

- There are numerous reasons to want to do this, such as the numerical approximation of expectations of functions \( f : X \to \mathbb{R} \) with respect to \( \pi \).
Overview

- Assume one is interested in sampling from a probability distribution $\pi$, defined on $(X, \mathcal{X})$

- There are numerous reasons to want to do this, such as the numerical approximation of expectations of functions $f : X \to \mathbb{R}$ with respect to $\pi$.

- Markov chain Monte Carlo methods is a vast class of versatile methods (algorithms) designed to achieve this.
Overview

- Assume one is interested in sampling from a probability distribution $\pi$, defined on $(X, \mathcal{X})$.
- There are numerous reasons to want to do this, such as the numerical approximation of expectations of functions $f : X \to \mathbb{R}$ with respect to $\pi$.
- Markov chain Monte Carlo methods is a vast class of versatile methods (algorithms) designed to achieve this.
- The main idea is to simulate realisations $\{X_i, i \geq 0\}$ of Markov chains with the properties
  - $\mathbb{P}(X_i \in A) \to \pi(A)$ as $i \to \infty$
  - and/or, in some sense,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \mathbb{E}_\pi(f(X)).$$
Overview

- Assume one is interested in sampling from a probability distribution $\pi$, defined on $(X, \mathcal{X})$.

- There are numerous reasons to want to do this, such as the numerical approximation of expectations of functions $f : X \rightarrow \mathbb{R}$ with respect to $\pi$.

- Markov chain Monte Carlo methods is a vast class of versatile methods (algorithms) designed to achieve this.

- The main idea is to simulate realisations $\{X_i, i \geq 0\}$ of Markov chains with the properties

  - $\mathbb{P}(X_i \in A) \rightarrow \pi(A)$ as $i \rightarrow \infty$
  - and/or, in some sense,

$$
\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \mathbb{E}_\pi(f(X)).
$$

- There are many known ways to design such Markov chains and an issue is that of choosing the right Markov chain.
Overview

- Understanding the theoretical properties of these processes is important in order to guide the design of such algorithms (for example, for finite $N$, what is the variability of the estimator above?)
Overview

- Understanding the theoretical properties of these processes is important in order to guide the design of such algorithms (for example, for finite $N$, what is the variability of the estimator above?)
- Reversibility plays a central rôle among existing methods and
  - numerous theoretical results exist (I will review some of them)
Overview

▶ Understanding the theoretical properties of these processes is important in order to guide the design of such algorithms (for example, for finite $N$, what is the variability of the estimator above?)

▶ Reversibility plays a central rôle among existing methods and
  ▶ numerous theoretical results exist (I will review some of them)
  ▶ but reversibility is not necessarily a desirable property.
Overview

- Understanding the theoretical properties of these processes is important in order to guide the design of such algorithms (for example, for finite $N$, what is the variability of the estimator above?)

- Reversibility plays a central rôle among existing methods and
  - numerous theoretical results exist (I will review some of them)
  - but reversibility is not necessarily a desirable property.

- There has recently been renewed interest in designing algorithms relying on nonreversible Markov chains, and processes.
Overview

- Understanding the theoretical properties of these processes is important in order to guide the design of such algorithms (for example, for finite $N$, what is the variability of the estimator above?)

- Reversibility plays a central rôle among existing methods and
  - numerous theoretical results exist (I will review some of them)
  - but reversibility is not necessarily a desirable property.

- There has recently been renewed interest in designing algorithms relying on nonreversible Markov chains, and processes.

- The presentation is about extending known results to this scenario, but also to their continuous time counterparts which have received some interest recently,
Overview

▶ Understanding the theoretical properties of these processes is important in order to guide the design of such algorithms (for example, for finite $N$, what is the variability of the estimator above?)

▶ Reversibility plays a central rôle among existing methods and

▶ numerous theoretical results exist (I will review some of them)
▶ but reversibility is not necessarily a desirable property.

▶ There has recently been renewed interest in designing algorithms relying on nonreversible Markov chains, and processes.

▶ The presentation is about extending known results to this scenario, but also to their continuous time counterparts which have received some interest recently,

▶ Along the way we also show some results of independent interest which compare performance of nonreversible chains to that of their reversible counterparts.
Outline

Motivation

Reversible scenario & known results

Modified/Skew detailed balance and Yaglom reversibility

Some applications

Tangential (but interesting) results

The continuous time scenario
The Metropolis-Hastings algorithm

- $\pi$ a probability distribution defined on $(X, \mathcal{X})$
The Metropolis-Hastings algorithm

- $\pi$ a probability distribution defined on $(X, \mathcal{X})$
- Choose a family $\{q(x, \cdot), x \in X\}$ of probability distributions on $(X, \mathcal{X})$ (user defined),

The MH transition probability $P(x, dy)$

Given $x$, 

The Metropolis-Hastings algorithm

- \( \pi \) a probability distribution defined on \((X, \mathcal{X})\)
- Choose a family \(\{q(x, \cdot), x \in X\}\) of probability distributions on \((X, \mathcal{X})\) (user defined),

The MH transition probability \( P(x, dy) \)

Given \(x\),

- Sample a proposal \(z \sim q(x, \cdot)\)
The Metropolis-Hastings algorithm

- \( \pi \) a probability distribution defined on \((X, \mathcal{X})\)
- Choose a family \( \{q(x, \cdot), x \in X\} \) of probability distributions on \((X, \mathcal{X})\) (user defined),

The MH transition probability \( P(x, dy) \)

Given \( x \),

- Sample a proposal \( z \sim q(x, \cdot) \)
  - Compute

\[
\alpha(x, z) = \min \left\{ 1, \frac{\pi(dz)q(z, dx)}{\pi(dx)q(x, dz)} \right\} = \min \{1, r(x, z)\}.
\]
The Metropolis-Hastings algorithm

- $\pi$ a probability distribution defined on $(X, \mathcal{X})$
- Choose a family $\{q(x, \cdot), x \in X\}$ of probability distributions on $(X, \mathcal{X})$ (user defined),

The MH transition probability $P(x, dy)$

Given $x$,

- Sample a proposal $z \sim q(x, \cdot)$
  - Compute

$$\alpha(x, z) = \min \left\{ 1, \frac{\pi(dz)q(z, dx)}{\pi(dx)q(x, dz)} \right\} = \min \{1, r(x, z)\}.$$  

- Set $y = z$ with probability $\alpha(x, z)$, otherwise set $y = x$. 

Detailed balance condition

- Denote $P(x, dy)$ this Markov chain transition probability,
Detailed balance condition

- Denote $P(x, dy)$ this Markov chain transition probability,
- It is not hard to check that the MH transition probability satisfies “detailed balance”, that is for any $x, y \in X$

$$
\pi(dx)P(x, dy) = \pi(dy)P(y, dx),
$$
Detailed balance condition

- Denote \( P(x, dy) \) this Markov chain transition probability,

- It is not hard to check that the MH transition probability satisfies "detailed balance", that is for any \( x, y \in X \)

\[
\pi(dx)P(x, dy) = \pi(dy)P(y, dx),
\]

- This automatically ensures that \( P \) leaves \( \pi \) invariant ("global balance condition"),

\[
\int \pi(dx)P(x, dy) = \pi(dy).
\]
Detailed balance condition

- Denote $P(x, dy)$ this Markov chain transition probability,
- It is not hard to check that the MH transition probability satisfies “detailed balance”, that is for any $x, y \in X$

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx),$$

- This automatically ensures that $P$ leaves $\pi$ invariant (“global balance condition”),

$$\int \pi(dx)P(x, dy) = \pi(dy).$$

- It also implies that the process $\{X_i, i \geq 0\}$ with $X_0 \sim \pi$ is reversible.
Detailed balance condition

- Denote $P(x, dy)$ this Markov chain transition probability,
- It is not hard to check that the MH transition probability satisfies "detailed balance", that is for any $x, y \in X$

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx),$$

- This automatically ensures that $P$ leaves $\pi$ invariant ("global balance condition"),

$$\int \pi(dx)P(x, dy) = \pi(dy).$$

- It also implies that the process $\{X_i, i \geq 0\}$ with $X_0 \sim \pi$ is reversible.
- Detailed balance plays a central rôle as most building blocks rely on this property, but it is not clearly desirable if "performance" is what we are interested in.
In order to sample from $\pi$ it is often useful to sample from another convenient distribution $\mu$ defined on $(E, \mathcal{E})$, and such that samples from $\pi$ can be recovered from samples of $\mu$ (e.g. $\pi$ is a marginal of $\mu$),
Operator on a Hilbert space

- In order to sample from $\pi$ it is often useful to sample from another convenient distribution $\mu$ defined on $(E, \mathcal{E})$, and such that samples from $\pi$ can be recovered from samples of $\mu$ (e.g. $\pi$ is a marginal of $\mu$),

- It is useful to consider the operator (also denoted $P$) associated with the Markov transition $P$, which is assumed to leave $\mu$ invariant.
Operator on a Hilbert space

- In order to sample from $\pi$ it is often useful to sample from another convenient distribution $\mu$ defined on $(E, \mathcal{E})$, and such that samples from $\pi$ can be recovered from samples of $\mu$ (e.g. $\pi$ is a marginal of $\mu$),

- It is useful to consider the operator (also denoted $P$) associated with the Markov transition $P$, which is assumed to leave $\mu$ invariant.

- The operator $P$ acts on functions to the right as $Pf(x) = \int f(y)P(x, dy)$,
Operator on a Hilbert space

- In order to sample from $\pi$ it is often useful to sample from another convenient distribution $\mu$ defined on $(E, \mathcal{G})$, and such that samples from $\pi$ can be recovered from samples of $\mu$ (e.g. $\pi$ is a marginal of $\mu$).

- It is useful to consider the operator (also denoted $P$) associated with the Markov transition $P$, which is assumed to leave $\mu$ invariant.

- The operator $P$ acts on functions to the right as $Pf(x) = \int f(y)P(x, dy)$.

- Define the standard function spaces
  
  \[ L^2(\mu) := \{ f : E \rightarrow \mathbb{R} : \mu(f^2) < \infty \} \text{ and } L^2_0(\mu) := \{ f \in L^2(\mu) : \mu(f) = 0 \} \]
Operator on a Hilbert space

▶ In order to sample from $\pi$ it is often useful to sample from another convenient distribution $\mu$ defined on $(E, \mathcal{E})$, and such that samples from $\pi$ can be recovered from samples of $\mu$ (e.g. $\pi$ is a marginal of $\mu$).

▶ It is useful to consider the operator (also denoted $P$) associated with the Markov transition $P$, which is assumed to leave $\mu$ invariant.

▶ The operator $P$ acts on functions to the right as $Pf(x) = \int f(y) P(x, dy)$,

▶ Define the standard function spaces

$$L^2(\mu) := \{ f : E \to \mathbb{R} : \mu(f^2) < \infty \}$$
and
$$L^2_0(\mu) := \{ f \in L^2(\mu) : \mu(f) = 0 \}$$

▶ Introduce the inner product $\langle f, g \rangle_\mu := \int f(z)g(z)\mu(dz)$ and associated norm $\|f\|_\mu := \sqrt{\langle f, f \rangle_\mu}$,
In order to sample from $\pi$ it is often useful to sample from another convenient distribution $\mu$ defined on $(E, \mathcal{E})$, and such that samples from $\pi$ can be recovered from samples of $\mu$ (e.g. $\pi$ is a marginal of $\mu$).

It is useful to consider the operator (also denoted $P$) associated with the Markov transition $P$, which is assumed to leave $\mu$ invariant.

The operator $P$ acts on functions to the right as $Pf(x) = \int f(y)P(x, dy)$.

Define the standard function spaces

$$L^2(\mu) := \{ f : E \to \mathbb{R} : \mu(f^2) < \infty \} \text{ and } L^2_0(\mu) := \{ f \in L^2(\mu) : \mu(f) = 0 \}$$

Introduce the inner product $\langle f, g \rangle_\mu := \int f(z)g(z)\mu(dz)$ and associated norm $\| f \|_\mu := \sqrt{\langle f, f \rangle_\mu}$.

We let $P^*$ be the adjoint operator of $P$ i.e. $\langle f, Pg \rangle_\mu = \langle P^* f, g \rangle_\mu$ for all $f, g \in L^2(\mu)$. 
Self-adjointness and some consequences

- Detailed balance can be reframed in operator language in terms of self-adjointness, that is

\[ \langle f, Pg \rangle_\mu = \langle Pf, g \rangle_\mu = \langle P^* f, g \rangle_\mu \text{ for all } f, g \in L^2(\mu) \]
Self-adjointness and some consequences

- Detailed balance can be reframed in operator language in terms of self-adjointness, that is
  \[ \langle f, Pg \rangle_\mu = \langle Pf, g \rangle_\mu = \langle P^* f, g \rangle_\mu \text{ for all } f, g \in L^2(\mu) \]

- This is a strong property, which implies numerous interesting facts. For example,
Self-adjointness and some consequences

- Detailed balance can be reframed in operator language in terms of self-adjointness, that is

\[ \langle f, Pg \rangle_\mu = \langle Pf, g \rangle_\mu = \langle P^*f, g \rangle_\mu \text{ for all } f, g \in L^2(\mu) \]

- This is a strong property, which implies numerous interesting facts. For example,
  - There is a spectral theory extending the idea of diagonalization of symmetric matrices,
Self-adjointness and some consequences

- Detailed balance can be reframed in operator language in terms of self-adjointness, that is

\[ \langle f, Pg \rangle_\mu = \langle Pf, g \rangle_\mu = \langle P^* f, g \rangle_\mu \] for all \( f, g \in L^2(\mu) \)

- This is a strong property, which implies numerous interesting facts. For example,
  - There is a spectral theory extending the idea of diagonalization of symmetric matrices,
  - Implies the existence of a CLT under very mild and simple assumptions (Kipnis-Varadhan),

\[ \text{var}(f, P) := \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left[ \left( \sum_{k=1}^{T} f(X_k) - \pi(f) \right)^2 \right] \in [0, \infty) \]
Self-adjointness and some consequences

- Detailed balance can be reframed in operator language in terms of self-adjointness, that is

\[ \langle f, Pg \rangle_\mu = \langle Pf, g \rangle_\mu = \langle P^* f, g \rangle_\mu \text{ for all } f, g \in L^2(\mu) \]

- This is a strong property, which implies numerous interesting facts. For example,
  - There is a spectral theory extending the idea of diagonalization of symmetric matrices,
  - Implies the existence of a CLT under very mild and simple assumptions (Kipnis-Varadhan),
  - A nice fact is that there are nice ways of comparing performance of reversible Markov chains sharing a common invariant distribution.
Self-adjointness and some consequences

- Detailed balance can be reframed in operator language in terms of self-adjointness, that is
  \[ \langle f, Pg \rangle_\mu = \langle Pf, g \rangle_\mu = \langle P^* f, g \rangle_\mu \text{ for all } f, g \in L^2(\mu) \]

- This is a strong property, which implies numerous interesting facts. For example,
  - There is a spectral theory extending the idea of diagonalization of symmetric matrices,
  - Implies the existence of a CLT under very mild and simple assumptions (Kipnis-Varadhan),
  - A nice fact is that there are nice ways of comparing performance of reversible Markov chains sharing a common invariant distribution.

- With \( P \) a Markov transition kernel with invariant distribution \( \mu \), letting \( X_0 \sim \mu \) and \( X_n \sim P(X_{n-1}, \cdot) \) for \( n \geq 1 \),

\[
\text{var}(f, P) := \lim_{T \to \infty} T \mathbb{E} \left[ \left( \frac{1}{T} \sum_{k=1}^{T} f(X_k) - \pi(f) \right)^2 \right] \in [0, \infty].
\]
Metropolis-Hastings vs. Barker

- Choose a family \( \{ q(x, \cdot), x \in X \} \) of probability distributions on \( (X, \mathcal{X}) \) (user defined),

The MH transition probability

Given \( x \),

- Sample a proposal \( z \sim q(x, \cdot) \)
- Compute

\[
\alpha(x, z) = \min \left\{ 1, \frac{\pi(dz)q(z, dx)}{\pi(dx)q(x, dz)} \right\} = \min \{ 1, r(x, z) \}.
\]

- Set \( y = z \) with probability \( \alpha(x, z) \), otherwise set \( y = x \).

Had we used \( \alpha(x, z) = \frac{1}{1 + r(x, z)} \) this would still satisfy \( \pi \) - detailed balance.

Quizz: which acceptance probability should one choose?
Metropolis-Hastings vs. Barker

- Choose a family \( \{ q(x, \cdot), x \in X \} \) of probability distributions on \((X, \mathcal{X})\) (user defined),

The MH transition probability

Given \( x \),

- Sample a proposal \( z \sim q(x, \cdot) \)
- Compute

\[
\alpha(x, z) = \min \left\{ 1, \frac{\pi(dz)q(z, dx)}{\pi(dx)q(x, dz)} \right\} = \min \left\{ 1, r(x, z) \right\}.
\]

- Set \( y = z \) with probability \( \alpha(x, z) \), otherwise set \( y = x \).

- Had we used

\[
\alpha(x, z) = \frac{r(x, z)}{1 + r(x, z)}
\]

this would still satisfy \( \pi \)–detailed balance.
Metropolis-Hastings vs. Barker

- Choose a family \(\{q(x, \cdot), x \in X\}\) of probability distributions on \((X, \mathcal{X})\) (user defined),

The MH transition probability

Given \(x\),

- Sample a proposal \(z \sim q(x, \cdot)\)
- Compute

\[
\alpha(x, z) = \min \left\{ 1, \frac{\pi(dz)q(z, dx)}{\pi(dx)q(x, dz)} \right\} = \min \{1, r(x, z)\}.
\]

- Set \(y = z\) with probability \(\alpha(x, z)\), otherwise set \(y = x\).

- Had we used

\[
\alpha(x, z) = \frac{r(x, z)}{1 + r(x, z)}
\]

this would still satisfy \(\pi\)–detailed balance.

- Quizz: which acceptance probability should one choose?
Let $\Pi^{(1)}$ and $\Pi^{(2)}$ be two Markov kernel reversible with respect to some common invariant distribution $\mu$ on $(E, \mathcal{E})$. 

Theorem 1. Whenever for any $x \in E$ and $A \in \mathcal{E}$ such that $x \in A$,

\[ \Pi^{(1)}(x, A) \geq \Pi^{(2)}(x, A) \]

then for any $f : E \to \mathbb{R}$ such that $\text{var} \mu(f) < \infty$

\[ \text{var}(f, \Pi^{(1)}) \leq \text{var}(f, \Pi^{(2)}) \]

and

\[ \text{Gap}_R(\Pi^{(1)}) \geq \text{Gap}_R(\Pi^{(2)}) \].

This is a very interesting result as the criterion can be checked in numerous situations.

**Quizz (answer)**: one can check that $\min\{1, u\} \geq u/(1+u)$ for $u \geq 0$, so $\alpha(x, z) = \min\{1, r(x, z)\}$ is better for the criteria above. In fact it is optimum.
Let $\Pi^{(1)}$ and $\Pi^{(2)}$ be two Markov kernel reversible with respect to some common invariant distribution $\mu$ on $(E, \mathcal{E})$.

A well known result due originally to Peskun [Peskun, 1973] states that

**Theorem 1.** Whenever for any $x \in E$ and $A \in \mathcal{E}$ such that $x \notin A$, $\Pi^{(1)}(x, A) \geq \Pi^{(2)}(x, A)$ then for any $f : E \to \mathbb{R}$ such that $\text{var}_\mu(f) < \infty$

$$\text{var}(f, \Pi^{(1)}) \leq \text{var}(f, \Pi^{(2)}) \text{ and } \text{Gap}_R(\Pi^{(1)}) \geq \text{Gap}_R(\Pi^{(2)}).$$
Let $\Pi^{(1)}$ and $\Pi^{(2)}$ be two Markov kernel reversible with respect to some common invariant distribution $\mu$ on $(E, \mathcal{E})$.

A well known result due originally to Peskun [Peskun, 1973] states that

**Theorem 1.** Whenever for any $x \in E$ and $A \in \mathcal{E}$ such that $x \notin A$, $\Pi^{(1)}(x, A) \geq \Pi^{(2)}(x, A)$ then for any $f : E \rightarrow \mathbb{R}$ such that $\text{var}_\mu(f) < \infty$

$$\text{var}(f, \Pi^{(1)}) \leq \text{var}(f, \Pi^{(2)}) \text{ and } \text{Gap}_R(\Pi^{(1)}) \geq \text{Gap}_R(\Pi^{(2)}).$$

This is a very interesting result as the criterion can be checked in numerous situations.
Let $\Pi^{(1)}$ and $\Pi^{(2)}$ be two Markov kernel reversible with respect to some common invariant distribution $\mu$ on $(E, \mathcal{E})$.

A well known result due originally to Peskun [Peskun, 1973] states that

**Theorem 1.** Whenever for any $x \in E$ and $A \in \mathcal{E}$ such that $x \notin A$, $\Pi^{(1)}(x, A) \geq \Pi^{(2)}(x, A)$ then for any $f : E \to \mathbb{R}$ such that $\text{var}_\mu(f) < \infty$

$$\text{var}(f, \Pi^{(1)}) \leq \text{var}(f, \Pi^{(2)}) \quad \text{and} \quad \text{Gap}_R(\Pi^{(1)}) \geq \text{Gap}_R(\Pi^{(2)}).$$

This is a very interesting result as the criterion can be checked in numerous situations.

**Quizz (answer):** one can check that $\min\{1, u\} \geq u/(1 + u)$ for $u \geq 0$, so $\alpha(x, z) = \min\{1, r(x, z)\}$ is better for the criteria above. In fact it is optimum.
Peskun’s result is in fact a consequence of a more general result which establishes the following.

\[
\text{For } \Pi, f \in L^2(\mu), \text{ define the Dirichlet form } E(f, \Pi) := \langle f, (I - \Pi)f \rangle_{\mu}.
\]

Note that this is \( \|f\|_2^2 - \langle f, \Pi f \rangle_{\mu} \) where the last term is the first order autocovariance coefficient for \( f \in L^2_0(\mu) \).

Theorem 2. If for any \( g \in L^2(\mu) \),
\[
E(g, \Pi(1)) \geq E(g, \Pi(2)),
\]
then for any \( f \in L^2(\mu) \),
\[
\text{var}(f, \Pi(1)) \leq \text{var}(f, \Pi(2)), \quad \text{Gap}_R(\Pi(1)) \geq \text{Gap}_R(\Pi(2)).
\]

This is not as transparent as Peskun’s criterion, but there are situations where the latter cannot be used, while the former can.
Standard ordering of reversible chains—Tierney

- Peskun’s result is in fact a consequence of a more general result which establishes the following.

- For $\Pi$ and $f \in L^2(\mu)$, define the Dirichlet form

$$\mathcal{E}(f, \Pi) := \langle f, (I - \Pi)f \rangle_\mu.$$  

Note that this is $\|f\|_\mu^2 - \langle f, \Pi f \rangle_\mu$ where the last term is the first order autocovariance coefficient for $f \in L^2_0(\mu)$. 
Peskun’s result is in fact a consequence of a more general result which establishes the following.

For $\Pi$ and $f \in L^2(\mu)$, define the Dirichlet form

$$\mathcal{E}(f, \Pi) := \langle f, (I - \Pi)f \rangle_\mu.$$

Note that this is $\|f\|_{\mu}^2 - \langle f, \Pi f \rangle_\mu$ where the last term is the first order autocovariance coefficient for $f \in L^2_0(\mu)$.

**Theorem 2.** If for any $g \in L^2(\mu)$, $\mathcal{E}(g, \Pi^{(1)}) \geq \mathcal{E}(g, \Pi^{(2)})$, then for any $f \in L^2(\mu)$

$$\text{var}(f, \Pi^{(1)}) \leq \text{var}(f, \Pi^{(2)})$$

and

$$\text{Gap}_R(\Pi^{(1)}) \geq \text{Gap}_R(\Pi^{(2)})$$
Standard ordering of reversible chains—Tierney

- Peskun’s result is in fact a consequence of a more general result which establishes the following.

- For $\Pi$ and $f \in L^2(\mu)$, define the Dirichlet form

$$E(f, \Pi) := \langle f, (I - \Pi)f \rangle_\mu.$$  

Note that this is $\|f\|_\mu^2 - \langle f, \Pi f \rangle_\mu$ where the last term is the first order autocovariance coefficient for $f \in L^2_0(\mu)$.

**Theorem 2.** If for any $g \in L^2(\mu)$, $E(g, \Pi^{(1)}) \geq E(g, \Pi^{(2)})$, then for any $f \in L^2(\mu)$

$$\text{var}(f, \Pi^{(1)}) \leq \text{var}(f, \Pi^{(2)}) \text{ and Gap}_R(\Pi^{(1)}) \geq \text{Gap}_R(\Pi^{(2)})$$

This is not as transparent as Peskun’s criterion, but there are situations where the latter cannot be used, while the former can.
Importance of these results

- I have shown you an example of application of this comparison result, \(\min\{1, r\} \text{ vs } r/(1 + r)\) in the MH update?
Importance of these results

- I have shown you an example of application of this comparison result, \( \min\{1, r\} \) vs \( r/(1 + r) \) in the MH update?

- Applications are numerous (and more exciting!).
Importance of these results

- I have shown you an example of application of this comparison result, \( \min\{1, r\} \) vs \( r/(1 + r) \) in the MH update?

- Applications are numerous (and more exciting!).

- As pointed out earlier there has been recent interest in developing nonreversible Monte Carlo Markov chain & processes.
Importance of these results

▶ I have shown you an example of application of this comparison result, $\min\{1, r\}$ vs $r/(1 + r)$ in the MH update?

▶ Applications are numerous (and more exciting!).

▶ As pointed out earlier there has been recent interest in developing nonreversible Monte Carlo Markov chain & processes.

▶ In the rest of the presentation I will show that they share a property similar to detailed balance / self-adjointness.
Importance of these results

- I have shown you an example of application of this comparison result, $\min\{1, r\}$ vs $r/(1 + r)$ in the MH update?

- Applications are numerous (and more exciting!).

- As pointed out earlier there has been recent interest in developing nonreversible Monte Carlo Markov chain & processes.

- In the rest of the presentation I will show that they share a property similar to detailed balance / self-adjointness.

- Simple comparison theorems can be established.
Hilbert space techniques summary

▶ One can introduce, for \( \lambda \in [0, 1] \),

\[
\text{var}(f, \lambda \Pi) := 2 \left\langle f, (I - \lambda \Pi)^{-1}f \right\rangle_{\mu} - \|f\|_{\mu}^2.
\]

▶ And the “Dirichlet forms” \( E(f, \lambda \Pi) := \left\langle f, (I - \lambda \Pi)f \right\rangle_{\mu} \).
Hilbert space techniques summary

- One can introduce, for $\lambda \in [0, 1]$,
  \[
  \text{var}(f, \lambda \Pi) := 2 \langle f, (I - \lambda \Pi)^{-1} f \rangle_{\mu} - \| f \|_{\mu}^2 .
  \]

- And the “Dirichlet forms” $E(f, \lambda \Pi) := \langle f, (I - \lambda \Pi) f \rangle_{\mu}$,

- Now for $\Pi_1$ and $\Pi_2$ reversible w.r.t $\mu$ the property underpinning the Peskun-Caracciolo-Pelissetto-Sokal-Tierney result is essentially
  \[
  \forall f \in L^2(\mu) \quad \langle f, (I - \lambda \Pi_2)^{-1} f \rangle_{\mu} \geq \langle f, (I - \lambda \Pi_1)^{-1} f \rangle_{\mu} \]
  \[\iff\]
  \[
  \forall g \in L^2(\mu) \quad \langle g, (I - \lambda \Pi_2) g \rangle_{\mu} \leq \langle g, (I - \lambda \Pi_1) g \rangle_{\mu} ,
  \]

Hilbert space techniques summary

- One can introduce, for $\lambda \in [0, 1]$,
  \[
  \text{var}(f, \lambda \Pi) := 2 \langle f, (I - \lambda \Pi)^{-1} f \rangle_{\mu} - \|f\|_{\mu}^2.
  \]

- And the “Dirichlet forms” $\mathcal{E}(f, \lambda \Pi) := \langle f, (I - \lambda \Pi) f \rangle_{\mu}$,

- Now for $\Pi_1$ and $\Pi_2$ reversible w.r.t $\mu$ the property underpinning the Peskun-Caracciolo-Pelissetto-Sokal-Tierney result is essentially
  \[
  \left[ \forall f \in L^2(\mu) \quad \langle f, (I - \lambda \Pi_2)^{-1} f \rangle_{\mu} \geq \langle f, (I - \lambda \Pi_1)^{-1} f \rangle_{\mu} \right]
  \iff \left[ \forall g \in L^2(\mu) \quad \langle g, (I - \lambda \Pi_2) g \rangle_{\mu} \leq \langle g, (I - \lambda \Pi_1) g \rangle_{\mu} \right],
  \]

- Or more explicitly
  \[
  \left[ \forall f \in L^2(\mu) \quad \text{var}(f, \lambda \Pi_2) \geq \text{var}(f, \lambda \Pi_1) \right]
  \iff \left[ \forall g \in L^2(\mu) \quad \mathcal{E}(f, \lambda \Pi_2) \leq \mathcal{E}(f, \lambda \Pi_1) \right].
  \]
Outline

Motivation

Reversible scenario & known results

Modified/Skew detailed balance and Yaglom reversibility

Some applications

Tangential (but interesting) results

The continuous time scenario
The standard random walk Metropolis

- For simplicity assume that $X = \mathbb{Z}$ and we are interested in sampling from $\pi$.

RWM

Given $x$,

- Sample an increment $z \sim \mathcal{U}(-1, 1)$
- Compute
  \[ \alpha(x, x + z) = \min \left\{ 1, \frac{\pi(x + z)}{\pi(x)} \right\} \]
- Set $y = x + z$ with probability $\alpha(x, x + z)$, otherwise set $y = x$. 

The standard random walk Metropolis

For simplicity assume that $X = \mathbb{Z}$ and we are interested in sampling from $\pi$.

**RWM**

Given $x$,

- Sample an increment $z \sim U(-1, 1)$
- Compute
  \[
  \alpha(x, x + z) = \min\left\{1, \frac{\pi(x + z)}{\pi(x)}\right\}
  \]
- Set $y = x + z$ with probability $\alpha(x, x + z)$, otherwise set $y = x$.

The chain can “backtrack”, which we may want to avoid (remember that $\pi(x)P(x, y) = \pi(y)P(y, x)$).
Motivation—the guided random walk

Motivation—the guided random walk

- The guided walk Metropolis works for $\pi$ defined on $X = \mathbb{R}^d$, but $X = \mathbb{Z}$ captures the important features required in what follows.
Motivation—the guided random walk

- The guided walk Metropolis works for \( \pi \) defined on \( X = \mathbb{R}^d \), but \( X = \mathbb{Z} \) captures the important features required in what follows,
- We introduce the auxiliary variable \( v \in \{-1, 1\} = V \), \( E = X \times V \), the new target distribution \( \mu(x, v) := \pi(x) \frac{1}{2} \), and the Markov transition \( P(x, v; y, w) \) given by:

Guided random walk

Given \((x, v)\),

- Calculate the acceptance ratio \( r(x, v) := \frac{\mu(x+v, v)}{\mu(x, v)} = \frac{\pi(x+v)}{\pi(x)} \).
- Set \((w, y) = (x+v, v)\) with probability \( \min\{1, r(x, v)\} \), otherwise set \((w, y) = (x, -v)\).

The process will travel in the same direction until a rejection occurs.
Motivation–the guided random walk

- The guided walk Metropolis works for $\pi$ defined on $X = \mathbb{R}^d$, but $X = \mathbb{Z}$ captures the important features required in what follows.
- We introduce the auxiliary variable $v \in \{-1, 1\} = V$, $E = X \times V$, the new target distribution $\mu(x, v) := \pi(x)^{1/2}$, and the Markov transition $P(x, v; y, w)$ given by:

Guided random walk
Given $(x, v)$,
Motivation—the guided random walk

- The guided walk Metropolis works for $\pi$ defined on $X = \mathbb{R}^d$, but $X = \mathbb{Z}$ captures the important features required in what follows,
- We introduce the auxiliary variable $v \in \{-1, 1\} = V$, $E = X \times V$, the new target distribution $\mu(x, v) := \pi(x) \frac{1}{2}$, and the Markov transition $P(x, v; y, w)$ given by:

Guided random walk
Given $(x, v)$,

- Calculate the acceptance ratio

$$r(x, v) := \frac{\mu(x + v, v)}{\mu(x, v)} = \frac{\pi(x + v)}{\pi(x)}.$$
Motivation—the guided random walk

- The guided walk Metropolis works for $\pi$ defined on $X = \mathbb{R}^d$, but $X = \mathbb{Z}$ captures the important features required in what follows,
- We introduce the auxiliary variable $v \in \{-1, 1\} = V$, $E = X \times V$, the new target distribution $\mu(x, v) := \pi(x)\frac{1}{2}$, and the Markov transition $P(x, v; y, w)$ given by:

**Guided random walk**

**Given** $(x, v)$,

- Calculate the acceptance ratio

$$r(x, v) := \frac{\mu(x + v, v)}{\mu(x, v)} = \frac{\pi(x + v)}{\pi(x)}.$$

- Set $(w, y) = (x + v, v)$ with probability $\min \{1, r(x, v)\}$, otherwise set $(w, y) = (x, -v)$. 
Motivation—the guided random walk

- The guided walk Metropolis works for $\pi$ defined on $X = \mathbb{R}^d$, but $X = \mathbb{Z}$ captures the important features required in what follows,
- We introduce the auxiliary variable $v \in \{-1, 1\} = V$, $E = X \times V$, the new target distribution $\mu(x, v) := \pi(x)^{\frac{1}{2}}$, and the Markov transition $P(x, v; y, w)$ given by:

Guided random walk

Given $(x, v)$,

- Calculate the acceptance ratio

$$r(x, v) := \frac{\mu(x + v, v)}{\mu(x, v)} = \frac{\pi(x + v)}{\pi(x)}.$$

- Set $(w, y) = (x + v, v)$ with probability $\min \{1, r(x, v)\}$, otherwise set $(w, y) = (x, -v)$.

The process will travel in the same direction until a rejection occurs.
The modified/skew detailed balance

- One can show that this update satisfies the following modified, or skew, detailed balance condition

\[ \mu(x, v) P(x, v; y, w) = \mu(y, w) P(y, -w; x, -v), \]

This property implies the global balance condition (i.e. \( P \) leaves \( \mu \) invariant),

Surprisingly, despite the familiar structure, relatively little seems to be known about these processes.

It will be easier to think in terms of operators—for example consider \( Q \) such that

\[ Qf(x, v) = f(x, -v) \]

(flip the velocity).

The modified detailed balance now reads

\[ \mu(x, v) P(x, v; y, w) = \mu(y, w) P(y, -w; x, -v), \]

One can in fact be more general, but again it is useful to use the operator formalism, and note that the above suggests that \( QPQ \) is the adjoint of \( P \).
The modified/skew detailed balance

- One can show that this update satisfies the following modified, or skew, detailed balance condition

\[
\mu(x, v) P(x, v; y, w) = \mu(y, w) P(y, -w; x, -v),
\]

- This property implies the global balance condition (i.e. \(P\) leaves \(\mu\) invariant),
The modified/skew detailed balance

▶ One can show that this update satisfies the following modified, or skew, detailed balance condition

$$\mu(x, v) P(x, v; y, w) = \mu(y, w) P(y, -w; x, -v),$$

▶ This property implies the global balance condition (i.e. $P$ leaves $\mu$ invariant),

▶ Surprisingly, despite the familiar structure, relatively little seems to be known about these processes.
The modified/skew detailed balance

- One can show that this update satisfies the following modified, or skew, detailed balance condition

\[ \mu(x, v) P(x, v; y, w) = \mu(y, w) P(y, -w; x, -v), \]

- This property implies the global balance condition (i.e. \( P \) leaves \( \mu \) invariant),

- Surprisingly, despite the familiar structure, relatively little seems to be known about these processes.

- It will be easier to think in terms of operators—for example consider \( Q \) such that \( Qf(x, v) = f(x, -v) \) (flip the velocity).
The modified/skew detailed balance

- One can show that this update satisfies the following modified, or skew, detailed balance condition

\[ \mu(x, v) P(x, v; y, w) = \mu(y, w) P(y, -w; x, -v), \]

- This property implies the global balance condition (i.e. \( P \) leaves \( \mu \) invariant),

- Surprisingly, despite the familiar structure, relatively little seems to be known about these processes.

- It will be easier to think in terms of operators—for example consider \( Q \) such that \( Qf(x, v) = f(x, -v) \) (flip the velocity).

- The modified detailed balance now reads

\[ \mu(x, v) P(x, v; y, w) = \mu(y, w) QPQ(y, w; x, v). \]
The modified/skew detailed balance

- One can show that this update satisfies the following modified, or skew, detailed balance condition

\[ \mu(x, v) P(x, v; y, w) = \mu(y, w) P(y, -w; x, -v), \]

- This property implies the global balance condition (i.e. \( P \) leaves \( \mu \) invariant),

- Surprisingly, despite the familiar structure, relatively little seems to be known about these processes.

- It will be easier to think in terms of operators—for example consider \( Q \) such that \( Qf(x, v) = f(x, -v) \) (flip the velocity).

- The modified detailed balance now reads

\[ \mu(x, v) P(x, v; y, w) = \mu(y, w) QPQ(y, w; x, v). \]

- One can in fact be more general, but again it is useful to use the operator formalism, and note that the above suggests that \( QPQ \) is the adjoint of \( P \).
The notion of $(\mu, Q)$—self adjointness

**Definition 3.** We call an operator $Q : L^2(\mu) \to L^2(\mu)$ an *isometric involution* if

1. $\langle f, g \rangle_\mu = \langle Qf, Qg \rangle_\mu$ for all $f, g \in L^2(\mu)$
2. $Q^2 = I$

Note that from the first property $\langle f, Qg \rangle_\mu = \langle Qf, Q^2g \rangle_\mu$, and from the second property this is equivalent to $\langle Qf, g \rangle_\mu$, meaning that $Q$ is $\mu$—self-adjoint (and therefore $\mu Q = \mu$).
The notion of \((\mu, Q)\)--self adjointness

**Definition 3.** We call an operator \(Q : L^2(\mu) \to L^2(\mu)\) an *isometric involution* if

1. \(\langle f, g \rangle_\mu = \langle Qf, Qg \rangle_\mu\) for all \(f, g \in L^2(\mu)\)
2. \(Q^2 = I\)

Note that from the first property \(\langle f, Qg \rangle_\mu = \langle Qf, Q^2g \rangle_\mu\), and from the second property this is equivalent to \(\langle Qf, g \rangle_\mu\), meaning that \(Q\) is \(\mu\)--self-adjoint (and therefore \(\mu Q = \mu\)).

**Definition 4.** We say that an operator \(P\) is \((\mu, Q)\)-self-adjoint if there is an involutive isometry \(Q\) such that for all \(f, g \in L^2(\mu)\)

\[
\langle Pf, g \rangle_\mu = \langle f, QPQg \rangle_\mu.
\]
A $Q$–self adjoint operator is always a composition

**Fact 5.** $QP$ and $PQ$ are $\mu$–self-adjoint, and $P = Q(QP) = (PQ)Q$. 
A $Q$–self adjoint operator is always a composition

**Fact 5.** $QP$ and $PQ$ are $\mu$–self-adjoint, and $P = Q(QP) = (PQ)Q$.

**Proof.** Using that $Q$ is self-adjoint and involutive and that $P^* = QPQ$ gives $(QP)^* = P^*Q^* = QP$, as required.

We compare $P$ and $QP$ in more detail later on.
Tensorized and cyclic chains

\[ P(x_1, x_2; d(y_1, y_2)) := P_1(x_1, dy_2)P_2(x_2, dy_1) \]

\( P_i \) assumed \( \pi \)-reversible.

- So in this case \( P^* = QPQ \) with \( Q \) the “swap components” operator.
- This allows the study a type of nonhomogeneous MC (cycle of reversible kernels) with time-homogeneous MC tools.
- This is the structure behind the results of [Maire et al., 2014] as noted in [Andrieu, 2016].
- There are other scenarios where the \( Q \)-operator is a “swap” [CA & Livingstone 2018].
Ordering of asymptotic variances

It can be shown that in some circumstances

\[
\text{var} (f, P) := \lim_{T \to \infty} T \mathbb{E} \left[ \left( \frac{1}{T} \sum_{k=1}^{T} f(Z_k) - \pi(f) \right)^2 \right],
\]

can be written in terms of \((I - P)^{-1} := \sum_{k=0}^{\infty} P^k\)

\[
\text{var} (f, P) = \|f\|_\mu^2 + 2 \sum_{k=1}^{\infty} \langle f, P^k f \rangle_\mu = 2 \langle f, (I - P)^{-1} f \rangle_\mu - \|f\|_\mu^2.
\]
Ordering of asymptotic variances

- It can be shown that in some circumstances
  \[
  \text{var}(f, P) := \lim_{T \to \infty} T \mathbb{E} \left[ \left( \frac{1}{T} \sum_{k=1}^{T} f(Z_k) - \pi(f) \right)^2 \right],
  \]
  can be written in terms of \((I - P)^{-1} := \sum_{k=0}^{\infty} P^k\)
  \[
  \text{var}(f, P) = \|f\|_\mu^2 + 2 \sum_{k=1}^{\infty} \langle f, P^k f \rangle_\mu = 2 \langle f, (I - P)^{-1} f \rangle_\mu - \|f\|_\mu^2.
  \]

- We will consider quantities, guaranteed to exist, defined for \(\lambda \in [0, 1)\),
  \[
  \text{var}_\lambda(f, P) := 2 \langle f, (I - \lambda P)^{-1} f \rangle_\mu - \|f\|_\mu^2.
  \]

**Theorem 6.** Let \(P_1\) and \(P_2\) be \((\mu, Q)\)-self-adjoint. If

\[
\forall g \in L^2(\mu), \langle g, (I - QP_2)g \rangle_\mu \leq \langle g, (I - QP_1)g \rangle_\mu
\]

\[
\iff \forall g \in L^2(\mu), \langle g, (I - P_2 Q)g \rangle_\mu \leq \langle g, (I - P_1 Q)g \rangle_\mu
\]

then for any \(f \in L^2_0(\mu)\) such that \(Qf = f\), and for any \(\lambda \in [0, 1)\) it holds that

\[
\text{var}_\lambda(f, P_1) \leq \text{var}_\lambda(f, P_2).
\]
How the proof works

▶ There are at least two ways one can establish this result. The first one mirrors that for the Tierney-Carraciolo-Pelisseto-Sokal result.
How the proof works

- There are at least two ways one can establish this result. The first one mirrors that for the Tierney-Carraciolo-Pelisseto-Sokal result.

- Consider the \((\mu, Q)\)-self-adjoint kernels \(P(\beta) := \beta P_1 + (1 - \beta)P_2\) for \(\beta \in [0, 1]\) and study the sign of \(\partial_\beta \text{var}_\lambda(P(\beta), f)\).
How the proof works

▶ There are at least two ways one can establish this result. The first one mirrors that for the Tierney-Carraciolo-Pelisseto-Sokal result.

▶ Consider the \((\mu, Q)\)-self-adjoint kernels \(P(\beta) := \beta P_1 + (1 - \beta) P_2\) for \(\beta \in [0, 1]\) and study the sign of \(\partial_\beta \text{var}_\lambda(P(\beta), f)\).

▶ Using the representation \(\text{var}_\lambda(P(\beta), f) = 2 \langle f, [I - \lambda P(\beta)]^{-1} f \rangle_{\mu} - \|f\|_{\mu}^2\) we focus on

\[
\partial_\beta \langle f, [I - \lambda P(\beta)]^{-1} f \rangle_{\mu} = \lambda \langle f, [I - \lambda P(\beta)]^{-1} \partial_\beta P(\beta) [I - \lambda P(\beta)]^{-1} f \rangle_{\mu}.
\]
How the proof works

▶ There are at least two ways one can establish this result. The first one mirrors that for the Tierney-Carraciolo-Pelisseto-Sokal result.

▶ Consider the \((\mu, Q)\)–self-adjoint kernels \(P(\beta) := \beta P_1 + (1 - \beta) P_2\) for \(\beta \in [0, 1]\) and study the sign of \(\partial_\beta \text{var}_\lambda(P(\beta), f)\).

▶ Using the representation \(\text{var}_\lambda(P(\beta), f) = 2\langle f, [I - \lambda P(\beta)]^{-1} f \rangle_\mu - \|f\|_\mu^2\)

we focus on

\[
\partial_\beta \langle f, [I - \lambda P(\beta)]^{-1} f \rangle_\mu = \lambda \langle f, [I - \lambda P(\beta)]^{-1} \partial_\beta P(\beta) [I - \lambda P(\beta)]^{-1} f \rangle_\mu.
\]

▶ One can show that \([I - \lambda P(\beta)]^{-1}\) is \((\mu, Q)\)–self-adjoint and \(\partial_\beta P(\beta) = P_1 - P_2\)

\[
\partial_\beta \langle f, [I - \lambda P(\beta)]^{-1} f \rangle_\mu = \langle Q[I - \lambda P(\beta)]^{-1} Qf, (P_1 - P_2)[I - \lambda P(\beta)]^{-1} f \rangle_\mu.
\]
How the proof works

- There are at least two ways one can establish this result. The first one mirrors that for the Tierney-Carraciolo-Pelisseto-Sokal result.

- Consider the \((\mu, Q)\)-self-adjoint kernels \(P(\beta) := \beta P_1 + (1 - \beta)P_2\) for \(\beta \in [0, 1]\) and study the sign of \(\partial_\beta \text{var}_\lambda(P(\beta), f)\).

- Using the representation \(\text{var}_\lambda(P(\beta), f) = 2\langle f, [I - \lambda P(\beta)]^{-1}f\rangle_\mu - \|f\|_\mu^2\) we focus on
  \[\partial_\beta \langle f, [I - \lambda P(\beta)]^{-1}f\rangle_\mu = \lambda \langle f, [I - \lambda P(\beta)]^{-1} \partial_\beta P(\beta) [I - \lambda P(\beta)]^{-1} f\rangle_\mu.\]

- One can show that \([I - \lambda P(\beta)]^{-1}\) is \((\mu, Q)\)-self-adjoint and \(\partial_\beta P(\beta) = P_1 - P_2\)
  \[\partial_\beta \langle f, [I - \lambda P(\beta)]^{-1}f\rangle_\mu = \langle Q[I - \lambda P(\beta)]^{-1} Qf, (P_1 - P_2)[I - \lambda P(\beta)]^{-1} f\rangle_\mu.\]

- Recalling that \(Qf = f\), with \(g := [I - \lambda P(\beta)]^{-1}f\) we deduce that
  \[\partial_\beta \langle f, [I - \lambda P(\beta)]^{-1}f\rangle_\mu = \langle Qg, (P_1 - P_2)g\rangle_\mu,\]
  \[= \langle g, (I - QP_2)g\rangle_\mu - \langle g, (I - QP_1)g\rangle_\mu.\]
In [Maire et al., 2014] the authors are interested in the study of inhomogeneous Markov chains where one cycles through two Markov transitions $P_1$ and $P_2$, each reversible w.r.t. $\mu$. The remarkable result is that Peskun type ideas extend to this scenario, i.e.

$$E(g, P_1, i) \geq E(g, P_2, i)$$

for all $g \in L^2(\mu)$ and $i \in \{1, 2\}$ implies

$$\text{var} \lambda(f, \{P_1, 1, P_1, 2\}) \leq \text{var} \lambda(f, \{P_2, 1, P_2, 2\}).$$

As a by-product they find that

Lemma 7. (Lemma 18 of [Maire et al., 2014]). For $i = 1, 2$, let $P_1, i$ and $P_2, i$ be $\mu$-reversible Markov kernels on $(E, E)$ such that $E(g, P_1, i) \geq E(g, P_2, i) \geq 0$ for all $g \in L^2(\mu)$. Consider the kernels $P_1 := P_1, 1 P_1, 2$ and $P_2 := P_2, 1 P_2, 2$. If for any $x, v \in E$ we have

$$P_1, 2(x, v; \{x\} \times V) = P_2, 2(x, v; \{x\} \times V) = 1,$$

then for any $f \in L^2(\mu)$ such that for any $x \in X$, $v \mapsto f(x, v)$ is constant and both

$$\sum_{k=1}^\infty |\langle f, P_k^1 f \rangle_\mu| < \infty$$

and

$$\sum_{k=1}^\infty |\langle f, P_k^2 f \rangle_\mu| < \infty,$$

$$\text{var} \lambda(P_1, f) \leq \text{var} \lambda(P_2, f).$$

Our result can be obtained by writing $P = (PQ)$ and using that $Q$ and $PQ$ are reversible. More on this later!
In [Maire et al., 2014] the authors are interested in the study of inhomogeneous Markov chains where one cycles through two Markov transitions $P_1$ and $P_2$, each reversible w.r.t. $\mu$.

The remarkable result is that Peskun type ideas extend to this scenario i.e. $\mathcal{E}(g, P_{1,i}) \geq \mathcal{E}(g, P_{2,i})$ for all $g \in L^2(\mu)$ and $i \in \{1, 2\}$ implies $\text{var}_\lambda(f, \{P_{1,1}, P_{1,2}\}) \leq \text{var}_\lambda(f, \{P_{2,1}, P_{2,2}\})$. 

Our result can be obtained by writing $P = (PQ)Q$ and using that $Q$ and $PQ$ are reversible. More on this later!
Ordering for “cyclic MCMC”

▶ In [Maire et al., 2014] the authors are interested in the study of inhomogeneous Markov chains where one cycles through two Markov transitions $P_1$ and $P_2$, each reversible w.r.t. $\mu$.

▶ The remarkable result is that Peskun type ideas extend to this scenario i.e. $\mathcal{E}(g, P_{1,i}) \geq \mathcal{E}(g, P_{2,i})$ for all $g \in L^2(\mu)$ and $i \in \{1, 2\}$ implies $\text{var}_\lambda(f, \{P_{1,1}, P_{1,2}\}) \leq \text{var}_\lambda(f, \{P_{2,1}, P_{2,2}\})$.

▶ As a by-product they find that

**Lemma 7.** (Lemma 18 of [Maire et al., 2014]). For $i = 1, 2$, let $P_{1,i}$ and $P_{2,i}$ be $\mu$—reversible Markov kernels on $(E, \mathcal{E})$ such that $\mathcal{E}(g, P_{1,i}) \geq \mathcal{E}(g, P_{2,i}) \geq 0$ for all $g \in L^2(\mu)$. Consider the kernels $P_1 := P_{1,1}P_{1,2}$ and $P_2 := P_{2,1}P_{2,2}$. If for any $x, v \in E$ we have $P_{1,2}(x, v; \{x\} \times V) = P_{2,2}(x, v; \{x\} \times V) = 1$, then for any $f \in L^2(\mu)$ such that for any $x \in X$, $v \mapsto f(x, v)$ is constant and both $\sum_{k=1}^{\infty} |\langle f, P_1^k f \rangle_\mu| < \infty$ and $\sum_{k=1}^{\infty} |\langle f, P_2^k f \rangle_\mu| < \infty$,

$$\text{var}(P_1, f) \leq \text{var}(P_2, f).$$
Ordering for “cyclic MCMC”

In [Maire et al., 2014] the authors are interested in the study of inhomogeneous Markov chains where one cycles through two Markov transitions $P_1$ and $P_2$, each reversible w.r.t. $\mu$.

The remarkable result is that Peskun type ideas extend to this scenario i.e. $\mathcal{E}(g, P_{1,i}) \geq \mathcal{E}(g, P_{2,i})$ for all $g \in L^2(\mu)$ and $i \in \{1, 2\}$ implies $\text{var}_\lambda(f, \{P_{1,1}, P_{1,2}\}) \leq \text{var}_\lambda(f, \{P_{2,1}, P_{2,2}\})$.

As a by-product they find that

**Lemma 7.** (Lemma 18 of [Maire et al., 2014]). For $i = 1, 2$, let $P_{1,i}$ and $P_{2,i}$ be $\mu-$reversible Markov kernels on $(E, \mathcal{E})$ such that $\mathcal{E}(g, P_{1,i}) \geq \mathcal{E}(g, P_{2,i}) \geq 0$ for all $g \in L^2(\mu)$. Consider the kernels $P_1 := P_{1,1}P_{1,2}$ and $P_2 := P_{2,1}P_{2,2}$. If for any $x, v \in E$ we have $P_{1,2}(x, v; \{x\} \times V) = P_{2,2}(x, v; \{x\} \times V) = 1$, then for any $f \in L^2(\mu)$ such that for any $x \in X$, $v \mapsto f(x, v)$ is constant and both $\sum_{k=1}^{\infty} |\langle f, P_1^k f \rangle_\mu| < \infty$ and $\sum_{k=1}^{\infty} |\langle f, P_2^k f \rangle_\mu| < \infty$,

$$\text{var}(P_1, f) \leq \text{var}(P_2, f).$$

Our result can be obtained by writing $P = (PQ)Q$ and using that $Q$ and $PQ$ are reversible. More on this later!
Outline

Motivation

Reversible scenario & known results

Modified/Skew detailed balance and Yaglom reversibility

Some applications

Tangential (but interesting) results

The continuous time scenario
Acceptance probability in the guided walk...

- For the guided walk Metropolis on $\mathbb{R}$ one multiplies $v \in \{-1, 1\}$ with $z \sim q$ such that $z \geq 0$ and the acceptance probability is

$$\alpha(x, v, z) = \min \left\{ 1, \frac{\pi(x + vz)}{\pi(x)} \right\}$$

- One could as well use

$$\alpha(x, v, z) = \frac{\pi(x + vz)}{1 + \frac{\pi(x + vz)}{\pi(x)}}$$
Acceptance probability in the guided walk...

- For the guided walk Metropolis on $\mathbb{R}$ one multiplies $\nu \in \{-1, 1\}$ with $z \sim q$ such that $z \geq 0$ and the acceptance probability is

$$\alpha(x, \nu, z) = \min\left\{1, \frac{\pi(x + \nu z)}{\pi(x)}\right\}$$

- One could as well use

$$\alpha(x, \nu, z) = \frac{\frac{\pi(x + \nu z)}{\pi(x)}}{1 + \frac{\pi(x + \nu z)}{\pi(x)}}$$

- The conclusion is similar to that in the standard reversible scenario.
Acceptance probability in the guided walk...

- For the guided walk Metropolis on $\mathbb{R}$ one multiplies $v \in \{-1, 1\}$ with $z \sim q$ such that $z \geq 0$ and the acceptance probability is

$$
\alpha(x, v, z) = \min \left\{ 1, \frac{\pi(x + vz)}{\pi(x)} \right\}
$$

- One could as well use

$$
\alpha(x, v, z) = \frac{\pi(x + vz)}{1 + \frac{\pi(x + vz)}{\pi(x)}}
$$

- The conclusion is similar to that in the standard reversible scenario.

- That’s because

$$
\langle g, (I - PQ)g \rangle_{\mu} =
$$

$$
= \frac{1}{2} \int \mu(d(x, v)) PQ((x, v); d(y, w)) [g(x, v) - g(y, w)]^2
$$

$$
= \frac{1}{2} \int \mu(d(x, v)) P((x, v); d(y, w)) [g(x, v) - Qg(y, w)]^2
$$

$$
= \frac{1}{2} \int \mu(d(x, v)) q(dz) \alpha(x, v, z) [g(x, v) - g(x + vz, -v, )]^2.
$$
Ordering of cyclic MCMC

- We have established earlier that a result of [Maire et al., 2014] implies our main result. In fact one can show that these results are equivalent.
Ordering of cyclic MCMC

- We have established earlier that a result of [Maire et al., 2014] implies our main result. In fact one can show that these results are equivalent.
- The main idea consists of embedding the inhomogeneous Markov chain into a bivariate homogeneous Markov chain as in [Andrieu, 2016],

Define \( \mu(A \times B) := \pi(A) \pi(B) \) for any \( A, B \in X \) and the \( \mu \)-reversible transitions \( P_i,\diamondsuit P_i \),

\[
P_i,\diamondsuit P_i(x_1, x_2; d(y_1, y_2)) = P_i,\diamondsuit P_i(x_1, d y_2) \pi(x_2) P_i,\diamondsuit P_i(x_2, d y_1) \pi(x_1)
\]

and we can apply Theorem 6 to functions \( g \in L^2(\mu) \) such that \( Qg = g \).
Ordering of cyclic MCMC

- We have established earlier that a result of [Maire et al., 2014] implies our main result. In fact one can show that these results are equivalent.
- The main idea consists of embedding the inhomogeneous Markov chain into a bivariate homogeneous Markov chain as in [Andrieu, 2016],
- Define \( \mu(A \times B) := \pi(A) \pi(B) \) for any \( A, B \in \mathcal{X} \) and the \( \mu \)-reversible transitions \( P_{i,1} \diamond P_{i,2}(x_1, x_2; d(y_1, y_2)) := P_{i,1}(x_1, dy_2) P_{i,2}(x_2, dy_1) \) for \( i \in \{1, 2\} \) and for all \( (x_1, x_2), (y_1, y_2) \in \mathcal{X}^2 \),
Ordering of cyclic MCMC

- We have established earlier that a result of [Maire et al., 2014] implies our main result. In fact one can show that these results are equivalent.
- The main idea consists of embedding the inhomogeneous Markov chain into a bivariate homogeneous Markov chain as in [Andrieu, 2016],
- Define \( \mu(A \times B) := \pi(A)\pi(B) \) for any \( A, B \in \mathcal{X} \) and the \( \mu \)-reversible transitions \( P_{i,1} \diamond P_{i,2}(x_1, x_2; d(y_1, y_2)) := P_{i,1}(x_1, d y_2)P_{i,2}(x_2, d y_1) \) for \( i \in \{1, 2\} \) and for all \( (x_1, x_2), (y_1, y_2) \in X^2 \),
- Consider the isometric involution \( Q \) such that \( Qg(x_1, x_2) := g(x_2, x_1) \) for all \( g \in L^2(\mu) \).
Ordering of cyclic MCMC

- We have established earlier that a result of [Maire et al., 2014] implies our main result. In fact one can show that these results are equivalent.
- The main idea consists of embedding the inhomogeneous Markov chain into a bivariate homogeneous Markov chain as in [Andrieu, 2016],
- Define \( \mu(A \times B) := \pi(A)\pi(B) \) for any \( A, B \in \mathcal{X} \) and the \( \mu \)-reversible transitions \( P_{i,1} \diamond P_{i,2}(x_1, x_2; d(y_1, y_2)) := P_{i,1}(x_1, dy_2)P_{i,2}(x_2, dy_1) \) for \( i \in \{1, 2\} \) and for all \((x_1, x_2), (y_1, y_2) \in X^2\),
- Consider the isometric involution \( Q \) such that \( Qg(x_1, x_2) := g(x_2, x_1) \) for all \( g \in L^2(\mu) \).
- For \( i \in \{1, 2\} \) \( P_{i,1} \diamond P_{i,2} \), are \((\mu, Q)\)-self-adjoint

\[
\mu(d(x_1, x_2))P_{i,1} \diamond P_{i,2}(x_1, x_2; d(y_1, y_2)) = \\
= \pi(dx_1)\pi(dx_2)P_{i,1}(x_1, dy_2)P_{i,2}(x_2, dy_1) \\
= \pi(dx_1)P_{i,1}(x_1, dy_2)\pi(dx_2)P_{i,2}(x_2, dy_1) \\
= \pi(dy_2)P_{i,1}(y_2, dx_1)\pi(dy_1)P_{i,2}(y_1, dx_2) \\
= \pi(dy_1)\pi(dy_2)P_{i,1} \diamond P_{i,2}(y_2, y_1; d(x_2, x_1)) \\
= \mu(dy_1, y_2)QP_{i,1} \diamond P_{i,2}Q(y_1, y_2; d(x_1, x_2))
\]

and we can apply Theorem 6 to functions \( g \in L^2(\mu) \) such that \( Qg = g \).
Theorem 6 says that if for any $h \in L^2(\mu)$, 
\[ \mathcal{E}(QP_{1,1} \diamond P_{1,2}, h) \geq \mathcal{E}(QP_{2,1} \diamond P_{2,2}, h) \]
then for any functions $g \in L^2(\mu)$ such that $Qg = g$

\[ \text{var}_\lambda(g, P_{1,1} \diamond P_{1,2}) \leq \text{var}_\lambda(g, P_{2,1} \diamond P_{2,2}). \]
Ordering of cyclic MCMC

- Theorem 6 says that if for any $h \in L^2(\mu)$,
  $\mathcal{E}(QP_{1,1} \diamond P_{1,2}, h) \geq \mathcal{E}(QP_{2,1} \diamond P_{2,2}, h)$ then for any functions $g \in L^2(\mu)$ such that $Qg = g$

  $$\text{var}_\lambda(g, P_{1,1} \diamond P_{1,2}) \leq \text{var}_\lambda(g, P_{2,1} \diamond P_{2,2}).$$

- In [Andrieu, 2016, Corollary 1] it is established that with
  $g(x_1, x_2) = f(x_1) + f(x_2)$,

  $$\text{var}_\lambda(g, P_{i,1} \diamond P_{i,2}) = \text{var}(f) + 2\text{var}_\lambda(f, \{P_{i,1}, P_{i,2}\})$$

  and one can conclude with a variance decomposition identity

  $$\mathcal{E}(Q(P_{i,1} \diamond P_{i,2}), h) = \mathcal{E}(P_{i,1}, \mathbb{E}_\pi[h(X, \cdot)]) + \mathbb{E}_\pi[\mathcal{E}(P_{i,2}, h(X, \cdot))].$$
Ordering of cyclic MCMC

- Theorem 6 says that if for any $h \in L^2(\mu)$,
  \[ \mathcal{E}(QP_{1,1} \diamond P_{1,2}, h) \geq \mathcal{E}(QP_{2,1} \diamond P_{2,2}, h) \] then for any functions $g \in L^2(\mu)$ such that $Qg = g$
  \[ \text{var}_\lambda(g, P_{1,1} \diamond P_{1,2}) \leq \text{var}_\lambda(g, P_{2,1} \diamond P_{2,2}). \]

- In [Andrieu, 2016, Corollary 1] it is established that with $g(x_1, x_2) = f(x_1) + f(x_2)$,
  \[ \text{var}_\lambda(g, P_{i,1} \diamond P_{i,2}) = \text{var}(f) + 2\text{var}_\lambda(f, \{P_{i,1}, P_{i,2}\}) \]
  and one can conclude with a variance decomposition identity
  \[ \mathcal{E}(Q(P_{i,1} \diamond P_{i,2}), h) = \mathcal{E}(P_{i,1}, \mathbb{E}_\pi[h(X, \cdot)]) + \mathbb{E}_\pi[\mathcal{E}(P_{i,2}, h(X, \cdot))]. \]

- An analog of this proof was given in [Andrieu, 2016, Theorem 2 and Lemma 3] but the central rôle played by $(\mu, Q)$—self-adjointness was not realized.
Ordering of cyclic MCMC

- Theorem 6 says that if for any \( h \in L^2(\mu) \),
  \[ \mathcal{E}(QP_{1,1} \diamond P_{1,2}, h) \geq \mathcal{E}(QP_{2,1} \diamond P_{2,2}, h) \]
  then for any functions \( g \in L^2(\mu) \) such that \( Qg = g \)
  \[ \text{var}_\lambda(g, P_{1,1} \diamond P_{1,2}) \leq \text{var}_\lambda(g, P_{2,1} \diamond P_{2,2}). \]

- In [Andrieu, 2016, Corollary 1] it is established that with
  \( g(x_1, x_2) = f(x_1) + f(x_2) \),
  \[ \text{var}_\lambda(g, P_{i,1} \diamond P_{i,2}) = \text{var}(f) + 2\text{var}_\lambda(f, \{P_{i,1}, P_{i,2}\}) \]
  and one can conclude with a variance decomposition identity
  \[ \mathcal{E}(Q(P_{i,1} \diamond P_{i,2}), h) = \mathcal{E}(P_{i,1}, \mathbb{E}_\pi[h(X, \cdot)]) + \mathbb{E}_\pi[\mathcal{E}(P_{i,2}, h(X, \cdot))]. \]

- An analog of this proof was given in [Andrieu, 2016, Theorem 2 and Lemma 3] but the central rôle played by \((\mu, Q)\)—self-adjointness was not realized.

- In fact we show a more general results, where \( P_{1,1}, P_{1,2}, P_{2,1}, P_{2,2} \) are assumed \((\mu, Q)\)—self-adjoint. Can be applied to Horowitz’s generalized HMC algorithm [Horowitz, 1991].
1. One can show equivalence with the result of [Maire et al., 2014] concerned with a Peskun type result for nonhomogeneous Markov chains where one cycles through two $\mu-$reversible transitions $P_1$ and $P_2$, and in fact generalize to $(\mu, Q)-$reversible kernels,

2. Applicable to general Metropolis-Hastings updates based on the flow $t \mapsto \psi_t$ of a dynamical system $dz_t/dt = F(z_t)$ with time reversal symmetry, i.e. such that $\psi_{-t} = Q \circ \psi_t \circ Q$ (in particular Horowitz’s generalized HMC [Horowitz, 1991]),

3. Delayed rejection versions of the above,

4. The lifted-algorithm [Turitsyn et al., 2011],

5. Neal’s no-backtracking algorithm [Neal, 2004],

6. The extra-chance algorithm [Campos and Sanz-Serna, 2015],

Our result also holds when $P^* = Q^{-1}PQ$ (i.e. we allow for $Q^{-1} \neq Q$), but we are not aware of applications?
Outline

Motivation

Reversible scenario & known results

Modified/Skew detailed balance and Yaglom reversibility

Some applications

Tangential (but interesting) results

The continuous time scenario
Additional results

**Theorem 8.** For $f \in L^2(\mu)$ with $Qf = f$, it holds that

$$\text{var}(P_{\text{guided}}, f) \leq \text{var}\left(\frac{1}{2}(P_{\text{guided}} + P^*_{\text{guided}}), f\right) \leq \text{var}(QP, f).$$

It turns out that $\frac{1}{2}(P_{\text{guided}} + P^*_{\text{guided}})$ is closely related to the random walk Metropolis ("the $x$ component has the same law").
Outline

Motivation

Reversible scenario & known results

Modified/Skew detailed balance and Yaglom reversibility

Some applications

Tangential (but interesting) results

The continuous time scenario
Motivation: pdmp based MC methods

- Earlier we have heard about the Zig-Zag process:
Motivation: pdmp based MC methods

▶ Earlier we have heard about the Zig-Zag process:

▶ This process can be shown to be a scaling limit of a version of Gustafson’s algorithm, so it is natural to expect $(\mu, Q)$—self-adjointness.
Informally known PDMP-MC are continuous time processes
\{ (X_t, V_t), t \in \mathbb{R}_+ \} “targeting” \( \mu(x, v) = \pi(x)\nu(v) \) on \( E = X \times V \) with the following characteristics:

- \((X_t, V_t)\) follows a deterministic trajectories for random time intervals \([T_{k-1}, T_k)\) with \( k \in \mathbb{N} \) and \( T_0 = 0 \),
- at event times \( T_k \), arising from a nonhomogeneous Poisson process, \( X_{T_k} := \lim_{t \uparrow T_k} X_t \) and the velocity \( V_t \) is updated.

There are numerous degrees of freedom involved in the choice of the deterministic trajectories, the intensities and updates of the velocity.

Here is a high level review which covers most known scenarios and to which our results apply.

For convenience we introduce \( U: X \to \mathbb{R} \) such that \( \pi(x) \propto \exp (-U(x)) \) and \( K: V \to \mathbb{R} \) such that \( \nu(v) \propto \exp (-\kappa(v)) \).
Known PDMPs: deterministic trajectories

- The deterministic trajectories are defined as solutions of an ODE,
- In fact for known PDMPs they are the solutions of Hamilton’s equations for a potential $H_0(x, v)$ e.g. $H_0(x, v) = U_0(x) + \kappa_0(v)$,
  - for $U_0(x) = U(x)$ and $\kappa_0(v) = \kappa(v)$ follows the isocontours of $\mu$,
  - for $U_0(x) = 0$ and arbitrary $\kappa(v)$, $x$ follows straight lines and $v$ remains constant,
  - any compatible combination will work ($U_0$ nonconstant and $\kappa_0$ constant may not be possible).

- The inter-event times are defined through a nonhomogeneous Poisson process of (properly chosen) intensity $\lambda(x, v)$, that is with $\Lambda(t, x, v) := \int_0^t \lambda(x + uv, v)du$ and for $\tau \geq 0$

$$\mathbb{P}(T_i \geq \tau \mid T_{i-1}, X_{T_{i-1}}, V_{T_{i-1}}) = \exp \left(-\Lambda(\tau - T_{i-1}, X_{T_{i-1}}, V_{T_{i-1}})\right).$$
Known PDMPs: velocity updates

- There are numerous possibilities to update the velocity. We focus here on mixtures of
  
  1. deterministic updates of the following type, for unit vectors  
     \[ \{ n_k(x) \in V, k \in [1, K] \} \]
     
     \[ v_{\text{new}} \leftarrow v_{\text{old}} - 2[v_{\text{old}} n_k(x)]n_k(x) \]
     
     in other words reflections through the hyperplane orthogonal to the unit vector \( n_k(x) \).

  2. sampling afresh from \( \nu \), we note the corresponding operator \( \Pi \) (we also treat the Ornstein-Uhlenbeck scenario...).

- The update is chosen randomly according to
  
  \[ \mathbb{P}(M = k \mid X_{T_k}, V_{T_{k-1}}) = \lambda_k(X_{T_k}, V_{T_{k-1}}) / \lambda(X_{T_k}, V_{T_{k-1}}) \]
  
  for \( \lambda_k : E \to \mathbb{R}_+ \) such that \( \sum_{k=0}^K \lambda_k = \lambda \).

- More on the actual choice of \( \{ n_k(x) \in V, k \in [1, K] \} \) and associated intensities later on.
Summary of a PDMP-MC \((U_0 = 0)\)

- Initialization \(z(0) = (x(0), v(0))\) and \(T_0 = 0\)
- Repeat
  1. Draw \(T_i\) such that
     \[
     \mathbb{P}(T_i \geq \tau \mid T_{i-1}, X_{T_{i-1}}, V_{T_{i-1}}) = \exp (-\Lambda(\tau - T_{i-1}, X_{T_{i-1}}, V_{T_{i-1}}))
     \]
  2. \((X_t, V_t) = (X_{T_{i-1}} + (t - T_{i-1})V_{T_{i-1}}, V_{T_{i-1}})\) for \(t \in [T_{i-1}, T_i)\),
  3. \(X_{T_i} = \lim_{t \uparrow T_i} X_t\) and with
     \[M \sim \mathcal{P}(\lambda_0(X_{T_i}, V_{T_{i-1}}), \lambda_1(X_{T_i}, V_{T_{i-1}}), \ldots, \lambda_K(X_{T_i}, V_{T_{i-1}}))\] set
     \(V_{T_i} \sim \nu\) if \(M = 0\), otherwise
     \[
     V_{T_i} = V_{T_{i-1}} - 2[V_{T_{i-1}}^\top n_M(x)]n_M(x).
     \]
  4. \(k \leftarrow k + 1\),
Zig-zag is a particular continuous time Markov process designed to sample from $\pi$. The name was coined in [Bierkens and Roberts, 2015] and further extended in [Bierkens et al., 2016] to the multivariate scenario, and can be understood as being a particular case of the process studied [Faggionato et al., 2009].
Zig-zag is a particular continuous time Markov process designed to sample from $\pi$. The name was coined in [Bierkens and Roberts, 2015] and further extended in [Bierkens et al., 2016] to the multivariate scenario, and can be understood as being a particular case of the process studied [Faggionato et al., 2009].

The bouncy particle [Bouchard-Côté et al., 2015] algorithm belongs to the same family of piecewise deterministic processes, which target distributions of the type $\mu(d(x, v)) = \pi(dx)\nu(dv)$ where $\nu(\cdot)$ is “isotropic”.
Zig-zag is a particular continuous time Markov process designed to sample from $\pi$. The name was coined in [Bierkens and Roberts, 2015] and further extended in [Bierkens et al., 2016] to the multivariate scenario, and can be understood as being a particular case of the process studied [Faggionato et al., 2009].

The bouncy particle [Bouchard-Côté et al., 2015] algorithm belongs to the same family of piecewise deterministic processes, which target distributions of the type $\mu(d(x, v)) = \pi(dx)\nu(dv)$ where $\nu(\cdot)$ is “isotropic”.

All known algorithms have got generators of the form, for some $k \in \mathbb{N}$ and $f \in D(L)$,

$$Lf = Df + \sum_{i=1}^{k} \lambda_i \cdot [R_i f - f]$$

where for any $(x, v) \in E$,

$$Df(x, v) := \lim_{t \downarrow 0} \frac{f(x + tv, v) - f(x, v)}{t}.$$
One can extend the notion of \((\mu, Q)\)-self-adjointness to continuous time processes and the practical condition is of the form

\[
\langle Lf, g \rangle_\mu = \langle f, QLQg \rangle_\mu
\]

for \(f, g \in \mathcal{A}\), where \(\mathcal{A} \subset L^2(\mu)\) consists of “sufficiently regular and rich” functions...
$(\mu, Q)$—self-adjointness characterization

- One can extend the notion of $(\mu, Q)$—self-adjointness to continuous time processes and the practical condition is of the form

$$\langle Lf, g \rangle_\mu = \langle f, QLQg \rangle_\mu$$

for $f, g \in \mathcal{A}$, where $\mathcal{A} \subset L^2(\mu)$ consists of “sufficiently regular and rich” functions...

- This property implies in particular that $\mu$ is invariant by the corresponding semi-group.
$(\mu, Q)$—self-adjointness characterization

- One can extend the notion of $(\mu, Q)$—self-adjointness to continuous time processes and the practical condition is of the form

$$\langle Lf, g \rangle_\mu = \langle f, QLQg \rangle_\mu$$

for $f, g \in \mathcal{A}$, where $\mathcal{A} \subset L^2(\mu)$ consists of “sufficiently regular and rich” functions...

- This property implies in particular that $\mu$ is invariant by the corresponding semi-group.

- If the continuous time algorithms are limits of Markov chains with $(\mu, Q)$—self-adjoint transition, one would expect them to be $(\mu, Q)$—self-adjoint.
One can extend the notion of \((\mu, Q)\)-self-adjointness to continuous time processes and the practical condition is of the form

\[ \langle Lf, g \rangle_\mu = \langle f, QLQg \rangle_\mu \]

for \(f, g \in \mathcal{A}\), where \(\mathcal{A} \subset L^2(\mu)\) consists of “sufficiently regular and rich” functions...

This property implies in particular that \(\mu\) is invariant by the corresponding semi-group.

If the continuous time algorithms are limits of Markov chains with \((\mu, Q)\)-self-adjoint transition, one would expect them to be \((\mu, Q)\)-self-adjoint.

This is satisfied for the earlier PDMPs if the three following conditions (with \(U(x) := -\log \pi(x)\)),

1. \(D^*f = QDQf + f \cdot DU\) (property inherited from any dynamical system with time-reversal symmetry)
2. \(\sum_{i=1}^{k} \lambda_i - Q\lambda_i = DU\)
3. for \(i \in [1, k]\), \(R_{x,i}\) is \((\lambda_i(x, \cdot)\nu(\cdot), Q)\)-self-adjoint.
One can extend the notion of \((\mu, Q)\)–self-adjointness to continuous time processes and the practical condition is of the form

\[
\langle Lf, g \rangle_{\mu} = \langle f, QLQg \rangle_{\mu}
\]

for \(f, g \in \mathcal{A}\), where \(\mathcal{A} \subset L^2(\mu)\) consists of “sufficiently regular and rich” functions...

This property implies in particular that \(\mu\) is invariant by the corresponding semi-group.

If the continuous time algorithms are limits of Markov chains with \((\mu, Q)\)–self-adjoint transition, one would expect them to be \((\mu, Q)\)–self-adjoint.

This is satisfied for the earlier PDMPs if the three following conditions (with \(U(x) := -\log \pi(x)\)),

1. \(D^*f = QDQf + f \cdot DU\) (property inherited from any dynamical system with time-reversal symmetry)
2. \(\sum_{i=1}^{k} \lambda_i - Q\lambda_i = DU\)
3. for \(i \in [1, k]\), \(R_{x,i}\) is \((\lambda_i(x, \cdot)\nu(\cdot), Q)\)–self-adjoint.

These conditions are satisfied for all the algorithms we are aware of.
Again the three basic conditions are

1. $D^* f = QDQf + f \cdot DU$
2. $\sum_{i=1}^{k} \lambda_i - Q\lambda_i = DU$
3. for $i \in [1, k]$, $R_{x,i}$ is $(\lambda_i(x, \cdot)\nu(\cdot), Q)$—self-adjoint.
$(\mu, Q)$—self-adjointness characterization

- Again the three basic conditions are
  1. $D^* f = QDQf + f \cdot DU$
  2. $\sum_{i=1}^{k} \lambda_i - Q\lambda_i = DU$
  3. for $i \in [1, k]$, $R_{x,i}$ is $(\lambda_i(x, \cdot)\nu(\cdot), Q)$—self-adjoint.

- The second condition raises algorithmical design questions.
(µ, Q)—self-adjointness characterization

- Again the three basic conditions are
  1. \( D^* f = Q D Q f + f \cdot D U \)
  2. \( \sum_{i=1}^{k} \lambda_i - Q \lambda_i = D U \)
  3. for \( i \in [1, k] \), \( R_{x,i} \) is \((\lambda_i(x, \cdot)\nu(\cdot), Q)\)—self-adjoint.

- The second condition raises algorithmical design questions.

- For simplicity assume \( k = 1 \), then notice that
$(\mu, Q)$—self-adjointness characterization

- Again the three basic conditions are
  1. $D^* f = QDQf + f \cdot DU$
  2. $\sum_{i=1}^{k} \lambda_i - Q\lambda_i = DU$
  3. for $i \in [1, k]$, $R_{x,i}$ is $(\lambda_i(x, \cdot)\nu(\cdot), Q)$—self-adjoint.

- The second condition raises algorithmical design questions.

- For simplicity assume $k = 1$, then notice that
  - the condition is satisfied with $\lambda_1(x, v) = \max\{0, DU(x, v)\}$ since
    \[
    QDU(x, v) = DU(x, -v) = \langle \nabla_x U(x), -v \rangle = -DU(x, v),
    \]
    and therefore
    \[
    \lambda_1(x, v) - Q\lambda_1(x, v) = \max\{0, DU(x, v)\} - \max\{0, -DU(x, v)\}.
    \]
\((\mu, Q)\)—self-adjointness characterization

- Again the three basic conditions are
  
  1. \(D^*f = QDQf + f \cdot DU\)
  
  2. \(\sum_{i=1}^{k} \lambda_i - Q\lambda_i = DU\)

  3. for \(i \in [1, k]\), \(R_{x,i}\) is \((\lambda_i(x, \cdot)\nu(\cdot), Q)\)–self-adjoint.

- The second condition raises algorithmical design questions.

- For simplicity assume \(k = 1\), then notice that
  
  - the condition is satisfied with \(\lambda_1(x, v) = \max\{0, DU(x, v)\}\) since
    
    \[QDU(x, v) = DU(x, -v) = \langle \nabla_x U(x), -v \rangle = -DU(x, v),\]

    and therefore

    \[\lambda_1(x, v) - Q\lambda_1(x, v) = \max\{0, DU(x, v)\} - \max\{0, -DU(x, v)\}.\]

  - but because the condition involves a difference
    
    \[\lambda_1(x, v) = \max\{0, DU(x, v)\} + \gamma(x)\] for \(\gamma(\cdot) : X \to \mathbb{R}_+\) would also work.
\((\mu, Q)\)–self-adjointness characterization

- Again the three basic conditions are
  
  1. \(D^* f = QDQf + f \cdot DU\)
  2. \(\sum_{i=1}^{k} \lambda_i - Q\lambda_i = DU\)
  3. for \(i \in [1, k]\), \(R_{x,i}\) is \((\lambda_i(x, \cdot)\nu(\cdot), Q)\)–self-adjoint.

- The second condition raises algorithmical design questions.

- For simplicity assume \(k = 1\), then notice that
  
  - the condition is satisfied with \(\lambda_1(x, v) = \max\{0, DU(x, v)\}\) since
    \[
    QDU(x, v) = DU(x, -v) = \langle \nabla_x U(x), -v \rangle = -DU(x, v),
    \]
  and therefore
    \[
    \lambda_1(x, v) - Q\lambda_1(x, v) = \max\{0, DU(x, v)\} - \max\{0, -DU(x, v)\}.
    \]

  - but because the condition involves a difference
    \(\lambda_1(x, v) = \max\{0, DU(x, v)\} + \gamma(x)\) for \(\gamma(\cdot) : X \to \mathbb{R}_+\) would also work.

- Quizz: how should we choose \(\lambda_1\)?.

Ordering variances

We are interested in comparing these quantities for two semi-groups \( \{P_{t,1}, t \geq 0\} \) and \( \{P_{t,2}, t \geq 0\} \) leaving \( \mu \) invariant and of generators \( L_1 \) and \( L_2 \) with domains \( D^2(L_1, \mu) \) and \( D^2(L_2, \mu) \).
Ordering variances

We are interested in comparing these quantities for two semi-groups \( \{P_{t,1}, t \geq 0\} \) and \( \{P_{t,2}, t \geq 0\} \) leaving \( \mu \) invariant and of generators \( L_1 \) and \( L_2 \) with domains \( D^2(L_1, \mu) \) and \( D^2(L_2, \mu) \).

As in the discrete time set-up the comparison relies on Dirichlet forms, which take the following form for a generator \( L \),

\[
\mathcal{E}(f, L) := \langle f, -Lf \rangle_\mu.
\]
Ordering variances

We are interested in comparing these quantities for two semi-groups \( \{P_{t,1}, t \geq 0\} \) and \( \{P_{t,2}, t \geq 0\} \) leaving \( \mu \) invariant and of generators \( L_1 \) and \( L_2 \) with domains \( D^2(L_1, \mu) \) and \( D^2(L_2, \mu) \).

As in the discrete time set-up the comparison relies on Dirichlet forms, which take the following form for a generator \( L \),

\[
\mathcal{E}(f, L) := \langle f, -Lf \rangle_\mu.
\]

Our result establishes that if for any \( g \in A \), \( \mathcal{E}(g, QL_1) \geq \mathcal{E}(g, QL_2) \) then for \( f \in L^2(\mu) \) such that \( Qf = f \) then

\[
\text{var}_\lambda(f, L_1) \leq \text{var}_\lambda(f, L_2).
\]
Ordering variances

- We are interested in comparing these quantities for two semi-groups \( \{P_{t,1}, t \geq 0\} \) and \( \{P_{t,2}, t \geq 0\} \) leaving \( \mu \) invariant and of generators \( L_1 \) and \( L_2 \) with domains \( \mathcal{D}^2(L_1, \mu) \) and \( \mathcal{D}^2(L_2, \mu) \).

- As in the discrete time set-up the comparison relies on Dirichlet forms, which take the following form for a generator \( L \),

\[
\mathcal{E}(f, L) := \langle f, -Lf \rangle_\mu.
\]

- Our result establishes that if for any \( g \in A, \mathcal{E}(g, QL_1) \geq \mathcal{E}(g, QL_2) \) then for \( f \in L^2(\mu) \) such that \( Qf = f \) then

\[
\text{var}_\lambda(f, L_1) \leq \text{var}_\lambda(f, L_2).
\]

- For Zig-zag one can show that with \( Q_i g(x, v) = g(x, v_{1:i-1}, -v_i, v_{i+1:d}) \),

\[
\mathcal{E}(g, QL_1) - \mathcal{E}(g, QL_2) = \frac{1}{2} \sum_{i=1}^{d} \int [\lambda_{1,i} - \lambda_{2,i}] \cdot (g - Q_i g)^2 d\mu,
\]

which tells us that the intensities should be as small as possible...
The end

Thank you!


For the bouncy particle the difference in Dirichlet forms is

\[
\mathcal{E}(f, QL_1) - \mathcal{E}(f, QL_2) = -\frac{1}{2} \int [\lambda_1 - \lambda_2] \cdot \left[ (f - Qf)^2 - (f - QBf)^2 \right] \, d\mu,
\]

The Dirichlet forms are not ordered for all functions, but this does not mean that this would not be the case for the “right” function.
Zig-zag (1D and mD)

- We are in a scenario where $E = X \times V$ and the distribution $\mu$ of interest has density (also denoted $\mu$)

$$\mu(x, v) \propto \exp(-U(x)) \nu(v)$$

with respect to some measure denoted $d(x, v)$, where $U : X = \mathbb{R}^d \to \mathbb{R}$ is an energy function and $\nu : V \subset \mathbb{R}^d \to \mathbb{R}_+$ are such that $\mu$ induces a probability distribution.
Zig-zag (1D and mD)

- We are in a scenario where $E = X \times V$ and the distribution $\mu$ of interest has density (also denoted $\mu$)

$$\mu(x, v) \propto \exp \left( - U(x) \right) \nu(v)$$

with respect to some measure denoted $d(x, v)$, where $U : X = \mathbb{R}^d \rightarrow \mathbb{R}$ is an energy function and $\nu : V \subset \mathbb{R}^d \rightarrow \mathbb{R}_+$ are such that $\mu$ induces a probability distribution.

- Zig-zag is a particular continuous time Markov process designed to sample from $\mu$. The name was coined in [Bierkens and Roberts, 2015] and further extended in [Bierkens et al., 2016], and can be understood as being a particular case of the process studied [Faggionato et al., 2009].
Zig-zag (1D and mD)

▶ We are in a scenario where $E = X \times V$ and the distribution $\mu$ of interest has density (also denoted $\mu$)

$$\mu(x, v) \propto \exp \left( - U(x) \right) \nu(v)$$

with respect to some measure denoted $d(x, v)$, where $U : X = \mathbb{R}^d \to \mathbb{R}$ is an energy function and $\nu : V \subset \mathbb{R}^d \to \mathbb{R}_+$ are such that $\mu$ induces a probability distribution.

▶ Zig-zag is a particular continuous time Markov process designed to sample from $\mu$. The name was coined in [Bierkens and Roberts, 2015] and further extended in [Bierkens et al., 2016], and can be understood as being a particular case of the process studied [Faggionato et al., 2009].

▶ Here $V := \{-1, 1\}^d$. For $i = 1, \ldots, d$ we introduce the sign flip operators $Q_i : \mathbb{R}^E \to \mathbb{R}^E$ $Q_i f(x, v) = f(x, v_{1:i-1}, -v_i, v_{i+1:d})$ and $Q := Q_1 Q_2 \cdots Q_d$ (note that the order of the composition does not matter).

▶ We will assume that $\nu$ is left invariant by $Q_i$, that is $\nu Q_i f = \nu(f)$. We let $\lambda_i$ be such that $\lambda_i(x, v) - Q_i \lambda_i(x, v) = \partial_i U(x) \cdot v_i$. Such $\lambda_i$s exist, for example

$$\lambda_i(x, v) = \left( \partial_i U(x) \cdot v_i \right)_+,$$

where $v_i$ is the $i$—th coordinate of $v$ and $\partial_i U(x)$ is the partial derivative of $U$ with respect to the $i$—th coordinate of $x$. 
Zig-zag (1D and mD)

- For $i \in \{1, \ldots, d\}$ we introduce
  \[
  \Lambda_i(t, x, v) := \int_0^t \lambda_i(x + uv, v)du,
  \]
  and $\Lambda(t, x, v) := \sum_{i=1}^d \Lambda_i(t, x, v)$. Below we let $X_{T_k} := \lim_{t \uparrow T_k} X_t$.
  Letting $Q_i(x, z) = Q_i f(x, z)$ for $i \in \{1, \ldots, d\}$ when $f$ is the identity.

- The Zig-zag process is defined as
  
  1. Draw $T_k$ such that
     \[
     P(T_k \geq \tau \mid T_{k-1}) = \exp(-\Lambda(\tau - T_{k-1}, X_{T_{k-1}}, V_{T_{k-1}})),
     \]
  2. $(X_t, V_t) = (X_{T_{k-1}} + (t - T_{k-1})V_{T_{k-1}}, V_{T_{k-1}})$ for $t \in [T_{k-1}, T_k)$,
  3. With probability
     \[
     M \sim P(\lambda_1(X_{T_{i-1}})/\lambda(X_{T_{i-1}}), \ldots, \lambda_d(X_{T_{i-1}})/\lambda(X_{T_{i-1}})) \text{ set } (X_{T_k}, V_{T_k}) = Q_M(X_{T_{k-1}}, V_{T_{k-1}}),
     \]
  4. $k \leftarrow k + 1$,
Theorem 9. The adjoint of the generator $L$ of Zig-zag is, for $f, g \in C_1(E) \cap L^2(\mu)$ such that $\nabla f, \nabla g \in L^2(\mu)$

\[ L^* f = -Df + \sum_{i=1}^{d} Q_i \lambda_i \cdot [Q_i f - f] \]

and is $Q$–self-adjoint, that is

\[ L^* f = QLQf. \]
**Theorem 10.** Consider two Zig-zag processes with intensities \( \lambda_i, \tilde{\lambda}_i : \to \mathbb{R}_+ \) such that for any \((x, v) \in E\) \(\tilde{\lambda}_i(x, v) \geq \lambda_i(x, v)\) and they both lead to semi-groups leaving \(\mu\) invariant. Then for any \(f \in D^2(L, \mu) \cap D^2(\tilde{L}, \mu)\) sufficiently regular,

\[
\mathcal{E}(f, QL) \geq \mathcal{E}(f, Q\tilde{L}) \quad \text{and} \quad \mathcal{E}(f, LQ) \geq \mathcal{E}(f, \tilde{L}Q).
\]

and

\[
\mathcal{E}(f, L) \leq \mathcal{E}(f, \tilde{L}).
\]