Comparison of Disjoint Decomposition Theorems for Metropolis Algorithms on \( \mathbb{R} \).

Neal Madras* and Wai Kong Yuen†

September 19, 2006

Abstract
We review and generalize a disjoint decomposition result of Martin and Randall [6] to include non-positive semi-definite reversible Markov chains on continuous state spaces and decompositions with countably many pieces. It is then compared with another disjoint decomposition result of Jerrum et al. [3].

Key words and phrases: Markov chain, Metropolis algorithm, spectral gap, decomposition.

2000 Mathematics Subject Classification: Primary 60J05; secondary 65C05.

1 Introduction

In this paper, we generalize a disjoint decomposition result of Martin and Randall [6] to include non-positive semi-definite reversible Markov chains on continuous state spaces and decompositions with countably many pieces. This is then compared with another result of Jerrum et al. [3] in the continuous state space. Examples are all underlying Markov chains of Metropolis algorithms [7], [2].

*Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3 Canada; madras@mathstat.yorku.ca. Research supported in part by NSERC Canada.

†Department of Mathematics, Brock University, St. Catharines, Ontario L2S 3A1 Canada; jyuen@brock.ca. Research supported in part by NSERC Canada.
1.1 Spectral Gaps.

Let \( \Omega \) be the state space of a Markov chain. To discuss probability densities on our state space, we need a reference measure \( \zeta \) on \( \Omega \). (For example, if \( \Omega \) is discrete, then \( \zeta \) can be counting measure, while if \( \Omega = \mathbb{R}^n \), then \( \zeta \) can be Lebesgue measure.) Let \( \pi \) be a probability density function (with respect to \( \zeta \)) on \( \Omega \). It is also convenient to denote the associated probability measure by \( \pi \) as well, so that for measurable \( B \subset \Omega \),

\[
\pi(B) := \int_B \pi(x) \zeta(dx).
\]

(If \( \zeta \) is counting measure, then this says \( \pi(B) := \sum_{x \in B} \pi(x) \).)

Let \( P(x, dy) \) be the transition probability kernel of a Markov chain on \( \Omega \) that is reversible with respect to a probability density \( \pi \) i.e.,

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

In this case, \( \pi \) is a stationary distribution for the Markov chain. It is well known that the convergence rate of the Markov chain to \( \pi \) (see e.g. Madras and Yuen for detail) is closely related to the spectral gap of the chain, defined as

\[
\text{Gap}(P) = \inf_{f \in L^2(\pi), f \neq 0} \frac{\int \int |f(x) - f(y)|^2 \pi(dx)P(x, dy)}{\int \int |f(x) - f(y)|^2 \pi(dx)\pi(dy)} \quad (1)
\]

Generally speaking, the larger the spectral gap, the faster the convergence to equilibrium. The disjoint decomposition concerns bounding the spectral gap from below.
2 Generalizations of Disjoint Decomposition Theorems

Let $P$ be a Markov chain defined on $\Omega$, reversible with respect to $\pi$ as in Section 1.1. To state the first decomposition result, we first describe the “pieces” of the chain $P$. Let $I$ be a countable index set and suppose that $\Omega$ is partitioned into disjoint (measurable) subsets $\{A_i\}_{i \in I}$. For each $i \in I$, we define a new Markov chain on $A_i$ by rejecting any transition of $P$ that would leave $A_i$. The transition kernel $P_{[A_i]}$ of the new chain is given by

$$P_{[A_i]}(x, B) = P(x, B) + 1_{\{x \in B\}}(x, A_i^c) \quad \text{for } x \in A_i, B \subset A_i. \quad (2)$$

It is easy to see that $P_{[A_i]}$ is reversible (on the state space $A_i$) with respect to the measure whose density is proportional to the restriction of $\pi$ to $A_i$. Define a new state space $\{a_i\}_{i \in I}$ in which each point $a_i$ represents the set $A_i$, and define a projection chain $\bar{P}$ on $\{a_i\}_{i \in I}$ as follows:

$$\bar{P}(a_i, a_j) = \frac{1}{\pi(A_i)} \int_{x \in A_i} \int_{y \in A_j} \pi(dx) P(x, dy). \quad (3)$$

The chain $\bar{P}$ is reversible with respect to $\bar{\pi}$, the measure defined on $\{a_i\}_{i \in I}$ by $\bar{\pi}(a_i) = \pi(A_i)$. The following is a slightly improved version of the Martin-Randall disjoint decomposition theorem.

**Theorem 2.1 (Martin-Randall [MnR] Disjoint Decomposition).** Let $P_{[A_i]}$ and $\bar{P}$ be defined as above. Then

$$\text{Gap}(P) \geq \frac{1}{3} \text{Gap}(\bar{P}) \inf_{i \in I} \text{Gap}(P_{[A_i]}). \quad (4)$$

In addition, if $P$ is positive semidefinite, then

$$\text{Gap}(P) \geq \frac{1}{2} \text{Gap}(\bar{P}) \inf_{i \in I} \text{Gap}(P_{[A_i]}). \quad (5)$$
Proof. The paper of Martin and Randall [6] assumes that $P$ is positive semidefinite. Thus we obtain (5) directly from Theorem 4.2 of [6] after we replace their min by inf. For (4), we define $V = (P + I)/2$. Then $V$ is a positive semidefinite transition probability operator, and $\text{Gap}(V) = \text{Gap}(P)/2$. Define the operator $\tilde{V}$ exactly as in equation (3) but with $P$ replaced by $V$. Then $\tilde{V} = (\hat{P} + I)/2$.

Now, the theorem of Caracciolo, Pelissetto and Sokal [1, 4, 6] tells us that

$$\text{Gap}(V^{1/2}P V^{1/2}) \geq \text{Gap}(\tilde{V}) \inf_{i \in I} \text{Gap}(P_{[A_i]}).$$

(6)

Now,

$$V^{1/2}PV^{1/2} = \frac{1}{2} ((I + P)^{1/2}[(I + P) - I](I + P)^{1/2}$$

$$= \frac{1}{2} [((I + P)^2 - (I + P)]$$

$$= \frac{1}{2} (P + P^2).$$

(7)

Let $\lambda$ be a number in the spectrum of $P_{[1]}$. Then $(\lambda + \lambda^2)/2$ is in the spectrum of $\frac{1}{2}(P + P^2)_{[1]}$, so

$$\text{Gap} \left[ \frac{1}{2}(P + P^2) \right] \leq 1 - \frac{1}{2} \lambda - \frac{1}{2} \lambda^2$$

$$= \left( 1 + \frac{1}{2} \lambda \right) (1 - \lambda)$$

$$\leq \frac{3}{2} (1 - \lambda).$$

(8)

The infimum of the right hand side of (8) over all $\lambda$ in spec$(P_{[1]})$ is $\frac{3}{2} \text{Gap}(P)$.

Combining this with (6) and (7) proves Equation (4). \qed

Remarks. (5) is the original result by Martin and Randall. In their paper, they assume implicitly that $P$ is positive semi-definite, which is not true in
many practical applications. For non-positive semi-definite \( P \), they consider lazy Markov chains \( \tilde{P} = \frac{1}{2}(P + I) \) so that \( \tilde{P} \) is positive semi-definite. In our examples, we do not make such an assumption and always apply (4).

The proof of the above result was based on a result of Caracciolo, Pelissetto and Sokal [1, 4]. Madras and Randall [4] used the same result to prove a state decomposition theorem, in with the pieces are not disjoint. Madras and Yuen [5] improved their result slightly and compared it to another disjoint decomposition theorem by Jerrum et al. [3] in a continuous setting. In this paper, we focus on comparing Theorem 2.1 and the disjoint decomposition theorem by Jerrum et al. [3]. The following is a general state space version of that result stated in [5]:

**Theorem 2.2 (Jerrum et al. [JSTV] Disjoint Decomposition).** Define \( P_{[A_i]} \) and \( \bar{P} \) as in Theorem 2.1, and \( \gamma \) as in (10). Then

\[
\text{Gap}(P) \geq \min \left\{ \frac{\text{Gap}(\bar{P})}{3}, \frac{\text{Gap}(\bar{P}) \inf_{i \in I} \text{Gap}(P_{[A_i]})}{3\gamma + \text{Gap}(\bar{P})}\right\},
\]

(9)

where

\[
\gamma := \max_{i \in I} \text{ess-sup}_{x \in A_i} P(x, A_i^c)
\]

(10)

and ess-sup is the essential supremum taken with respect to the measure \( \pi \).

Although both theorems can be applied to very general reversible Markov chain, the decomposition techniques arise naturally in underlying Markov chains for some Metropolis algorithms [7], [2]. We shall compare the applications of the two bounds in the next section.
3 Examples

3.1 A Symmetric Random Walk Metropolis Algorithm

We redo the first example of [5]. Let $\Omega = \mathbb{R}_+$ and $\pi_1(x) = e^{-x}$ for $x \geq 0$, the exponential distribution. Consider the symmetric random walk Metropolis algorithm with increment proposal distribution $q$ being the uniform distribution on $[-\rho, \rho]$, $\rho > 0$. Let $P_1$ denote the underlying Markov chain.

We choose a partition so that the spectral gap of the projection chain $\tilde{P}_1$ on $\mathbb{N}$ is computable. Notice that transition probabilities of $\tilde{P}_1$ are not easy to compute for a general decomposition, therefore we consider a natural partition $\{A_i\}_{i \in \mathbb{N}}$ where $A_i = [(i-1)\rho, i\rho)$ by matching the length of the interval $A_i$ with $\rho$. The computation of the bound using the JSTV decomposition (Theorem 2.2) was worked out in [5] and is given by

$$\delta_{\text{JSTV},1} \equiv \frac{\tilde{\delta}_1(1-e^{-\rho})}{\rho(3+2\tilde{\delta}_1)},$$

where

$$\tilde{\delta}_1 = (1-\alpha)(1-\sqrt{4q_0(1-q_0)}),$$

and

$$\alpha = \frac{\rho e^\rho - 3\rho + e^\rho - e^{-\rho}}{2\rho(e^\rho - 1)}$$

and 

$$q_0 = \frac{\rho - 1 + e^{-\rho}}{(\rho - e^\rho + \rho e^\rho + e^{-\rho})}, \quad (11)$$

The calculation of the bound using the MnR decomposition (Theorem 2.1) is similar and so we state the result without proof:

Proposition 3.1.

$$\text{Gap}(P_1) \geq \frac{1}{3}(1-\alpha)(1-\sqrt{4q_0(1-q_0)})\frac{1-e^{-\rho}}{2\rho} \equiv \delta_{\text{MnR},1}. \quad (12)$$

with $\alpha$ and $q_0$ defined in (11).
As can be seen from Figure 1, $\delta_{JSTV,1}$ is (at least 1.5 times) better than $\delta_{MnR,1}$ for all values of $\rho$.

**Remark.** Similarly, it is easy to compute the bounds if the target is a two-sided exponential distribution and the JSTV bounds are also better than the MnR bounds. In fact, many natural decompositions for a one dimensional symmetric random walk Metropolis algorithm have the property that $\gamma \leq 1/2$, as the probability of proposing an “escape” is only 1/2. Therefore, it is not surprising to find that the JSTV bounds usually perform better than the MnR bounds (e.g. the examples in [5]).
3.2 A Symmetric Metropolis Algorithm with Strong “Periodicity”

We consider a Metropolis algorithm on $\Omega = [0, 2)$ with target distribution

$$\pi_2(x) = \begin{cases} 
\beta, & \text{when } x \in [0, 1) \\
1 - \beta, & \text{when } x \in [1, 2) \\
0 & \text{otherwise}
\end{cases}$$

Given a current state $X$, we propose a move to $Y$ as follows: If $X \in [0, 1)$, we pick $Y \sim \text{Unif}[1, 2)$ with probability $p$ and $Y \sim \text{Unif}[0, 1)$ with probability $1 - p$. On the other hand, if $X \in [1, 2)$, we pick $Y \sim \text{Unif}[0, 1)$ with probability $p$ and $Y \sim \text{Unif}[1, 2)$ with probability $1 - p$. The proposal distribution is clearly symmetric, but not of a random walk type. Without loss of generality, we assume that $\beta \leq 1/2$.

Let $P_2$ denote the underlying Markov chain. To bound $\text{Gap}(P_2)$, we consider a natural decomposition $\{[0, 1), [1, 2)\}$ of $\Omega = [0, 2)$. Let $\delta_{\text{MnR},2}$ and $\delta_{\text{JSTV},2}$ denote the resulting bounds using Theorem 2.1 (bound (4)) and Theorem 2.2 (bound (9)) respectively under this decomposition. We have the following proposition:

**Proposition 3.2.**

$$p \geq \frac{3(1 - \beta)}{4 - 3\beta},$$

if and only if

$$\delta_{\text{JSTV},2} \leq \delta_{\text{MnR},2}.$$ 

In particular, the second inequality holds if $p \geq 3/4$ for any $0 < \beta < 1/2$.

**Proof.** We only prove the necessity. The sufficiency follows trivially. First note
that the projection chain $\hat{P}_2$ is a two state Markov chain with transition matrix
\[
\begin{pmatrix}
1 - p & p \\
\frac{p\beta}{1 - \beta} & 1 - \frac{p\beta}{1 - \beta}
\end{pmatrix}.
\]
With simple linear algebra, $\text{Gap}(\hat{P}_2) = p/(1 - \beta)$. Moreover, $\gamma = p$. Therefore, under the assumption, we have
\[
3\gamma + \text{Gap}(\hat{P}_2) = 3p + \frac{p}{1 - \beta} \geq 3.
\]
Hence,
\[
\delta_{JSTV,2} \leq \min \left\{ \frac{\text{Gap}(\hat{P})}{3}, \frac{\text{Gap}(\hat{P}) \min \{\text{Gap}(P_{[0,1]}), \text{Gap}(P_{[1,2]})\}}{3} \right\} = \delta_{MnR,2}
\]
since $\min \{\text{Gap}(P_{[0,1]}), \text{Gap}(P_{[1,2]})\} \leq 1$.

\textbf{Remark.} This is the first example in which the MnR bounds perform better than the JSTV bounds. It is easy to see that a similar result holds for a more general $\Omega$ and $\pi_2$, as long as there is a strong "periodicity" of the chain (large $p$ in our example).

\textbf{References}


