A Proof of Theorem 3

Given the setup in Theorem 3, we first restate (Fill, 1991, Theorem 2.1) (note that the norm in (Fill, 1991) is twice the total variation distance):

$$\left\|P^t(\sigma,\cdot)-\pi\right\|_{TV}^2 \le \frac{\left(1-\lambda(R(P))\right)^t}{\pi(\sigma)}. \tag{10}$$

Let $\lambda := \lambda(R(P))$ and $T := \log\left(\frac{4e^2}{\pi_{\min}}\right) T_{rel}(P) = \frac{1}{1-\sqrt{1-\lambda}} \log\left(\frac{4e^2}{\pi_{\min}}\right)$. Then it is easy to verify that

$$T \ge \frac{2}{\lambda}\log\left(\frac{2e}{\sqrt{\pi_{\min}}}\right)$$

and by (10), we have that

$$\begin{aligned}
\max_{\sigma\in\Omega}\left\|P^T(\sigma,\cdot)-\pi\right\|_{TV} &\le \frac{(1-\lambda)^{T/2}}{\sqrt{\pi_{\min}}} \\
&\le \frac{(1-\lambda)^{\lambda^{-1}\log\left(\frac{2e}{\sqrt{\pi_{\min}}}\right)}}{\sqrt{\pi_{\min}}} \\
&\le \frac{e^{-\log\left(\frac{2e}{\sqrt{\pi_{\min}}}\right)}}{\sqrt{\pi_{\min}}} \\
&= \frac{1}{2e}.
\end{aligned}$$

In other words,

$$T_{mix}(P) \le T = \log\left(\frac{4e^2}{\pi_{\min}}\right) T_{rel}(P).$$

B Operator Norms and the Spectral Gap

We also view the transition matrix P as an operator that mapping functions to functions. More precisely, let f be a function $f:\Omega\to\mathbb{R}$ and P acting on f is defined as

$$Pf(x) := \sum_{y\in\Omega} P(x,y)f(y).$$

This is also called the *Markov operator* corresponding to P. We will not distinguish the matrix P from the operator P as it will be clear from the context. Note that $Pf(x)$ is the expectation of f with respect to the distribution $P(x,\cdot)$. We can regard a function f as a column vector in $\mathbb{R}^\Omega$, in which case Pf is simply matrix multiplication. Recall (4) and P^* is also called the *adjoint operator* of P. Indeed, P^* is the (unique) operator that satisfies $\langle f, Pg\rangle_\pi = \langle P^*f, g\rangle_\pi$. It is easy to verify that if P satisfies the detailed balanced condition (1), then P is *self-adjoint*, namely $P = P^*$.

The Hilbert space $L_2(\pi)$ is given by endowing $\mathbb{R}^\Omega$ with the inner product

$$\langle f, g\rangle_\pi := \sum_{x\in\Omega} f(x)g(x)\pi(x),$$

where $f, g\in\mathbb{R}^\Omega$. In particular, the norm in $L_2(\pi)$ is given by

$$\|f\|_\pi := \langle f, f\rangle_\pi.$$

The spectral gap (2) can be rewritten in terms of the operator norm of P, which is defined by

$$\|P\|_\pi := \max_{\|f\|_\pi\neq 0}\frac{\|Pf\|_\pi}{\|f\|_\pi}.$$

Indeed, the operator norm equals the largest eigenvalue (which is just 1 for a transition matrix P), but we are interested in the second largest eigenvalue. Define the following operator

$$S_\pi(\sigma,\tau) := \pi(\tau). \tag{11}$$

It is easy to verify that $S_\pi f(\sigma) = \langle f, \mathbf{1}\rangle_\pi$ for any σ. Thus, the only eigenvalues of S_π are 0 and 1, and the eigenspace of eigenvalue 0 is $\{f \in L_2(\pi) : \langle f, \mathbf{1}\rangle_\pi = 0\}$. This is exactly the union of eigenspaces of P excluding the eigenvalue 1. Hence, the operator norm of $P - S_\pi$ equals the second largest eigenvalue of P, namely,

$$\lambda(P) = 1 - \|P - S_\pi\|_\pi . \tag{12}$$

The expression in (12) can be found in, for example, (Ullrich, 2014, Eq. (2.8)). In particular, using (12), we show that the definition (5) coincides with (3) when P is reversible.

Proposition 7. *Let P be the transition matrix of a reversible matrix with the stationary distribution π. Then*

$$\frac{1}{\lambda(P)} = \frac{1}{1 - \sqrt{1 - \lambda(R(P))}}.$$

Proof. Since P is reversible, P is self-adjoint, namely, $P^* = P$. Hence $(P - S_\pi)^* = P^* - S_\pi$ and

$$\begin{aligned}(P - S_\pi)(P - S_\pi)^* &= (P - S_\pi)(P^* - S_\pi)\\ &= PP^* - PS_\pi - S_\pi P^* + S_\pi S_\pi\\ &= PP^* - S_\pi,\end{aligned}$$

where we use the fact that $PS_\pi = S_\pi P^* = S_\pi S_\pi = S_\pi$. It implies that

$$\begin{aligned}1 - \lambda(R(P)) &= \|R(P) - S_\pi\|_\pi && \text{(by (12))}\\ &= \|PP^* - S_\pi\|_\pi\\ &= \left\|(P - S_\pi)(P - S_\pi)^*\right\|_\pi\\ &= \|P - S_\pi\|_\pi^2\\ &= (1 - \lambda(P))^2 .\end{aligned}$$

Rearranging the terms yields the claim. □

C Proof of Theorem 1

The transition matrix of updating a particular variable x is the following

$$T_x(\sigma,\tau) = \begin{cases} \frac{\pi(\sigma^{x,s})}{\sum_{s\in S}\pi(\sigma^{x,s})} & \text{if } \tau = \sigma^{x,s} \text{ for some } s \in S;\\ 0 & \text{otherwise.}\end{cases} \tag{13}$$

Moreover, let I be the identity matrix that $I(\sigma,\tau) = \mathbb{1}(\sigma,\tau)$.

Lemma 8. *Let π be a bipartite distribution, and P_{RU}, P_{AS}, T_x be defined as above. Then we have that*

1. $P_{RU} = \dfrac{I}{2} + \dfrac{1}{2n}\sum_{x\in V} T_x$.
2. $P_{AS} = \prod_{i=1}^{n_1} T_{x_i} \prod_{j=1}^{n_2} T_{y_j}$.

Proof. Note that T_x is the transition matrix of resampling $\sigma(x)$. For P_{RU}, the term $\frac{I}{2}$ comes from the fact that the chain is "lazy". With the other 1/2 probability, we resample $\sigma(x)$ for a uniformly chosen $x \in V$. This explains the term $\frac{1}{2n}\sum_{x\in V} T_x$.

For P_{AS}, we sequentially resample all variables in V_1 and then all variables in V_2, which yields the expression. □

Lemma 9. *Let π be a bipartite distribution and T_x be defined as above. Then we have that*

1. *For any $x \in V$, T_x is a self-adjoint operator and idempotent. Namely, $T_x = T_x^*$ and $T_x T_x = T_x$.*
2. *For any $x \in V$, $\|T_x\|_\pi = 1$.*
3. *For any $x, x' \in V_i$ where $i = 1$ or 2, T_x and $T_{x'}$ commute. In other words $T_{x'} T_x = T_x T_{x'}$ if $x, x' \in V_i$ for $i = 1$ or 2.*

Proof. For Item 1, the fact that T_x is self-adjoint follows from the detailed balance condition (1). Idempotence is because updating the same vertex twice is the same as a single update.

Item 2 follows from Item 1. This is because

$$\|T_x\|_\pi = \|T_x T_x\|_\pi = \|T_x T_x^*\|_\pi = \|T_x\|_\pi^2 .$$

For Item 3, suppose $i = 1$. Since π is bipartite, resampling x or x' only depends on σ_2. Therefore the ordering of updating x or x' does not matter as they are in the same partition. $\square$

Define

$$P_{GS1} := \frac{I}{2} + \frac{1}{2n_1} \sum_{i=1}^{n_1} T_{x_i}, \qquad \text{and} \qquad P_{GS2} := \frac{I}{2} + \frac{1}{2n_2} \sum_{j=1}^{n_2} T_{y_j}.$$

Then, since $n_1 + n_2 = n$,

$$P_{RU} = \frac{1}{n} \left(n_1 P_{GS1} + n_2 P_{GS2} \right). \tag{14}$$

Similarly, define

$$P_{AS1} := \prod_{i=1}^{n_1} T_{x_i}, \qquad \text{and} \qquad P_{AS2} := \prod_{j=1}^{n_2} T_{y_j}.$$

Then

$$P_{AS} = P_{AS1} P_{AS2}. \tag{15}$$

With this notation, Lemma 9 also implies the following.

Corollary 10. *The following holds:*

1. *$\|P_{AS1}\|_\pi \le 1$ and $\|P_{AS2}\|_\pi \le 1$.*
2. *$P_{AS1} P_{GS1} = P_{AS1}$ and $P_{GS2} P_{AS2} = P_{AS2}$.*

Proof. For Item 1, by the submultiplicity of operator norms:

$$\begin{aligned} \|P_{AS1}\|_\pi = \left\| \prod_{i=1}^{n_1} T_{x_i} \right\|_\pi &\le \prod_{i=1}^{n_1} \|T_{x_i}\|_\pi \\ &= 1. \qquad \text{(By Item 2 of Lemma 9)} \end{aligned}$$

The claim $\|P_{AS2}\|_\pi \le 1$ follows similarly.

Item 2 follows from Item 1 and 3 of Lemma 9. We verify the first case as follows.

$$\begin{aligned}
P_{AS1}P_{GS1} &= \prod_{i=1}^{n_1} T_{x_i} \left(\frac{I}{2} + \frac{1}{2n_1} \sum_{j=1}^{n_1} T_{x_j} \right) \\
&= \frac{1}{2} \cdot \prod_{i=1}^{n_1} T_{x_i} + \frac{1}{2n_1} \cdot \prod_{i=1}^{n_1} T_{x_i} \sum_{j=1}^{n_1} T_{x_j} \\
&= \frac{1}{2} \cdot \prod_{i=1}^{n_1} T_{x_i} + \frac{1}{2n_1} \cdot \sum_{j=1}^{n_1} T_{x_j} \prod_{i=1}^{n_1} T_{x_i} \\
&= \frac{1}{2} \cdot \prod_{i=1}^{n_1} T_{x_i} + \frac{1}{2n_1} \cdot \sum_{j=1}^{n_1} T_{x_1} T_{x_2} \cdots T_{x_j} T_{x_j} \cdots T_{x_{n_1}} && \text{(By Item 3 of Lemma 9)} \\
&= \frac{1}{2} \cdot \prod_{i=1}^{n_1} T_{x_i} + \frac{1}{2n_1} \cdot \sum_{j=1}^{n_1} \prod_{i=1}^{n_1} T_{x_i} && \text{(By Item 1 of Lemma 9)} \\
&= \frac{1}{2} \cdot \prod_{i=1}^{n_1} T_{x_i} + \frac{1}{2} \cdot \prod_{i=1}^{n_1} T_{x_i} \\
&= P_{AS1}.
\end{aligned}$$

The other case is similar. □

Item 2 of Corollary 10 captures the following intuition: if we sequentially update all variables in V_i for $i = 1, 2$, then an extra individual update either before or after does not change the distribution. Recall Eq. (5).

Lemma 11. *Let π be a bipartite distribution and P_{RU} and P_{AS} be defined as above. Then we have that*

$$\| R(P_{AS}) - S_\pi \|_\pi \leq \| P_{RU} - S_\pi \|_\pi^2 .$$

Proof. Recall (11), the definition of S_π, using which it is easy to see that

$$P_{AS1} S_\pi = S_\pi P_{AS2} = S_\pi S_\pi = S_\pi. \tag{16}$$

Thus,

$$\begin{aligned}
P_{AS1}(P_{RU} - S_\pi) P_{AS2} &= P_{AS1} \left(\frac{n_1}{n} P_{GS1} + \frac{n_2}{n} P_{GS2} - S_\pi \right) P_{AS2} && \text{(By (14))} \\
&= \frac{n_1}{n} P_{AS1} P_{GS1} P_{AS2} + \frac{n_2}{n} P_{AS1} P_{GS2} P_{AS2} - P_{AS1} S_\pi P_{AS2} \\
&= \frac{n_1}{n} P_{AS1} P_{AS2} + \frac{n_2}{n} P_{AS1} P_{AS2} - S_\pi && \text{(By Item 2 of Cor 10)} \\
&= P_{AS1} P_{AS2} - S_\pi \\
&= P_{AS} - S_\pi, && (17)
\end{aligned}$$

where in the last step we use (15). Moreover, we have that

$$\begin{aligned}
P_{AS}^* &= \left(\prod_{i=1}^{n_1} T_{x_i} \prod_{j=1}^{n_2} T_{y_j} \right)^* \\
&= \prod_{j=1}^{n_2} T^*_{y_{n_2+1-j}} \prod_{i=1}^{n_1} T^*_{x_{n_1+1-i}} \\
&= \prod_{j=1}^{n_2} T_{y_{n_2+1-j}} \prod_{i=1}^{n_1} T_{x_{n_1+1-i}} && \text{(By Item 1 of Lemma 9)} \\
&= \prod_{j=1}^{n_2} T_{y_j} \prod_{i=1}^{n_1} T_{x_i} && \text{(By Item 3 of Lemma 9)} \\
&= P_{AS2} P_{AS1}.
\end{aligned}$$

Hence, similarly to (17), we have that

$$\begin{aligned} P_{AS2}(P_{RU} - S_\pi)P_{AS1} &= P_{AS2}P_{AS1} - S_\pi \\ &= P^*_{AS} - S_\pi. \end{aligned} \tag{18}$$

Using (16), we further verify that

$$\begin{aligned} (P_{AS} - S_\pi)\,(P^*_{AS} - S_\pi) &= P_{AS}P^*_{AS} - P_{AS}S_\pi - S_\pi P^*_{AS} + S_\pi S_\pi \\ &= P_{AS}P^*_{AS} - S_\pi \end{aligned} \tag{19}$$

Combining (17), (18), and (19), we see that

$$\begin{aligned} \|R(P_{AS}) - S_\pi\|_\pi &= \|P_{AS}P^*_{AS} - S_\pi\|_\pi \\ &= \|(P_{AS} - S_\pi)\,(P^*_{AS} - S_\pi)\|_\pi \\ &= \|P_{AS1}\,(P_{RU} - S_\pi)\,P_{AS2}P_{AS2}\,(P_{RU} - S_\pi)\,P_{AS1}\|_\pi \\ &\le \|P_{AS1}\|_\pi\,\|P_{RU} - S_\pi\|_\pi\,\|P_{AS2}\|_\pi\,\|P_{AS2}\|_\pi\,\|P_{RU} - S_\pi\|_\pi\,\|P_{AS1}\|_\pi \\ &\le \|P_{RU} - S_\pi\|^2_\pi\,, \end{aligned}$$

where the first inequality is due to the submultiplicity of operator norms, and we use Item 1 of Corollary 10 in the last line. □

Remark. *The last inequality in the proof of Lemma 11 crucially uses the fact that the distribution is bipartite. If there are, say, k partitions, then the corresponding operators $P_{AS1}, \dots, P_{ASk}$ do not commute and the proof does not generalize.*

Proof of Theorem 1. For the first part, notice that the alternating-scan sampler is aperiodic. Any possible state σ of the chain must be in the state space Ω. Therefore $\pi(\sigma) > 0$ and the probability of staying at σ is strictly positive. Moreover, any single variable update can be simulated in the scan sampler, with small but strictly positive probability. Hence if the random-update sampler is irreducible, then so is the scan sampler.

To show that $T_{rel}(P_{AS}) \le T_{rel}(P_{RU})$, we have the following

$$\begin{aligned} T_{rel}(P_{AS}) &= \frac{1}{1 - \sqrt{1 - \lambda(R(P_{AS}))}} && \text{(By (5))} \\ &= \frac{1}{1 - \sqrt{\|R(P_{AS}) - S_\pi\|_\pi}} && \text{(By (12))} \\ &\le \frac{1}{1 - \|P_{RU} - S_\pi\|_\pi} && \text{(By Lemma 11)} \\ &= \frac{1}{\lambda(P_{RU})} && \text{(By (12))} \\ &= T_{rel}(P_{RU}). && \text{(By (3))} \end{aligned}$$

This completes the proof. □