# ACM 216, Problem Set 3

Ethan Wilk

February 23, 2023

## 1 Problem 1

**Theorem 1.** *An irreducible Markov chain with a finite state space and transition matrix $P$ is reversible if and only if $P = DS$ for some symmetric matrix $S$ and diagonal matrix with strictly positive entries $D$.*

*Proof.* Let $X = (X_n)_{n\in\mathbb{N}_0}$ be our Markov chain with finite state space $E$ and invariant distribution $\pi = \pi P = (\pi_1, ..., \pi_n)$. We know that $\pi$ exists and is unique with strictly positive entries since the chain is irreducible and has a finite state space. We will now prove both directions of the claim.

( $\implies$ ) Suppose $P$ is reversible. Then for any states $i, j \in E$, we have that

$$\pi_i P(i,j) = \pi_j P(j,i). \tag{1}$$

Now let $D = \text{diag}(\frac{1}{\pi_1}, ..., \frac{1}{\pi_n}), S = D^{-1}P$. Since $D$ is a diagonal matrix with strictly positive entries, we have easily that $D^{-1} = \text{diag}(\pi_1, ..., \pi_n)$ so $DD^{-1} = I$. Then we have that

$$S = \begin{pmatrix} P(1,1)\pi_1 & P(1,2)\pi_1 & \dots & P(1,n)\pi_1 \\ P(2,1)\pi_2 & P(2,2)\pi_2 & \dots & P(2,n)\pi_2 \\ \vdots & \vdots & \ddots & \vdots \\ P(n,1)\pi_n & P(n,2)\pi_n & \dots & P(n,n)\pi_n \end{pmatrix}.$$

But due to (1), we may replace $S_{i,j} = P(i,j)\pi_i$ by $P(j,i)\pi_j$. Do this for all $i < j$ in the matrix, and $S$ will in fact be symmetric. Then we have easily that

$$DS = DD^{-1}P = IP = P.$$

( $\impliedby$ ) Suppose $\exists$ $D$ a diagonal matrix with strictly positive diagonal entries and $S$ a symmetric matrix such that $P = DS$. Denote $D(i,i) = d_i, S(i,j) = s_{i,j}$. Then $P(i,j) = s_{i,j}d_i = s_{j,i}d_i$. Let $\pi^* = (\frac{1}{d_1}, ..., \frac{1}{d_n})$. Observe that

$$\pi^* P = \pi^* DS = \begin{pmatrix} s_{1,1} + s_{2,1} + \dots + s_{n,1} \\ s_{1,2} + s_{2,2} + \dots + s_{2,n} \\ \vdots \\ s_{1,n} + s_{2,n} + \dots + s_{n,n} \end{pmatrix}^\top.$$

Since the rows of $P$ must each sum to 1, we have that

$$\sum_{j=1}^{n} s_{i,j} d_i = 1 \implies \sum_{j=1}^{n} s_{i,j} = \frac{1}{d_i},$$

so since $S$ is symmetric,

$$s_{i,1} + s_{i,2} + ... + s_{i,n} = s_{1,i} + s_{2,i} + ... + s_{n,i} = \frac{1}{d_i}$$

for every $1 \leq i \leq n$. Then $\pi^* P$ may be written

$$\pi^* P = \pi^* DS = \begin{pmatrix} \frac{1}{d_1} & \frac{1}{d_2} & \cdots & \frac{1}{d_n} \end{pmatrix} = \pi^*,$$

so we have that, for the normalization constant $c = \sum_i \pi_i^*$, $\frac{\pi^*}{c} = \pi$, the invariant distribution of $P$. Then, since we saw $P(i,j) = s_{i,j} d_i$, we simply check that

$$\pi_i P(i,j) = \pi_j P(j,i) \iff \frac{P(i,j)}{d_i} = \frac{P(j,i)}{d_j}$$

$$\iff P(i,j) d_j = P(j,i) d_i \iff d_i s_{i,j} d_j = d_j s_{j,i} d_i,$$

which of course holds due to the symmetry of $S$.

□

**Theorem 2.** *If an irreducible Markov chain $X$ with a finite state space is reversible, its transition matrix $P$ has real eigenvalues.*

*Proof.* Let $X$ be an irreducible, reversible Markov chain with a finite state space $E$ and transition matrix $P$. We will show that $P$ is similar to some symmetric matrix $Q$ so that $Q$ (and thus $P$) has real eigenvalues.

Let $D = \text{diag}(\sqrt{\pi_1}, ..., \sqrt{\pi_n}), Q = DPD^{-1}$. Then we have that

$$Q(i,j) = \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} P(i,j) = \frac{\pi_i}{\sqrt{\pi_i \pi_j}} P(i,j),$$

$$Q(j,i) = \frac{\sqrt{\pi_j}}{\sqrt{\pi_i}} P(j,i) = \frac{\pi_j}{\sqrt{\pi_j \pi_i}} P(j,i),$$

so since $X$ is reversible, $P(i,j)\pi_i = P(j,i)\pi_j$, and thus

$$Q(i,j) = \frac{\pi_j}{\sqrt{\pi_i \pi_j}} P(j,i) = Q(j,i),$$

so $Q$ is symmetric. Then $P$ is indeed similar to a symmetric matrix $Q$, and so since $Q$ has real eigenvalues, $P$ does as well. □

# 2 Problem 2

The Jupyter notebook for this problem is linked here: `https://github.com/ewilk0/Caltech-ACM-216/blob/master/ACM%20216%20PS%203/ACM%20216%20PS%203%20(Ising%20Model).ipynb`. (The images didn't load when I tried to paste it here.) It includes test cases and information about the model. Let me know if you need anything else.

# 3 Problem 3

## 3.1 Part 3a

Images for some instances of the trajectories under each of the transition matrices are attached below (next page on Gradescope).

## 3.2 Part 3b

(1) $\mathbb{Q}^{(1)}$ is both irreducible and reversible (with respect to the uniform distribution on $E$). Any permutation of the elements of a vector is the product of transpositions. Thus any permutation can be reached from any other by applying transpositions in the correct ordering, and so the chain is irreducible.

We also have that, for any neighbor permutations $x, y, \mathbb{Q}^{(1)}(x, y) = \binom{N}{2}^{-1}$, since there are $\binom{N}{2}$ unique transpositions that can be made at any jump. Thus

$$\frac{\mathbb{Q}^{(1)}(x, y)}{\mathbb{Q}^{(1)}(y, x)} = 1 = \frac{\pi_y^{(1)}}{\pi_x^{(1)}},$$

for $\pi^{(1)} = \{\pi_x^{(1)} = \frac{1}{|E|} : x \in E\}$. Furthermore, if $z$ is some state that is not a neighbor of $x$, then $\mathbb{Q}^{(1)}(x, z) = \mathbb{Q}^{(1)}(z, x) = 0$, so we have easily that the detailed balance equations hold for these states, since $0(\pi_x^{(1)}) = 0(\pi_z^{(1)})$ will hold for any $\pi_x^{(1)}, \pi_z^{(1)}$. We therefore have that the chain is reversible with respect to the uniform distribution on $E$.

(2) $\mathbb{Q}^{(2)}$ is both irreducible and reversible. It is irreducible since any transposition between entries $x_i$ and $x_j$ can be expressed as a chain of adjacent transpositions, and therefore $\mathbb{Q}^{(2)}$ will be irreducible for the same reasons $\mathbb{Q}^{(1)}$ is. It is reversible with respect to the uniform distribution $\pi^{(2)} = \{\pi_x^{(2)} = \frac{1}{|E|} : x \in E\}$.

This is because the chain will choose a random integer $i \in [1, N-1]$ that will solely determine the jump made, so if swapping $x_i, x_{i+1}, i \in [1, N-1]$ in $x$ yields $y$, then the probability of transition from $x$ to $y$ is in fact $\frac{1}{N-1}$; the converse is also true. Thus

$$\frac{\mathbb{Q}^{(2)}(x, y)}{\mathbb{Q}^{(2)}(y, x)} = 1 = \frac{|E|}{|E|}.$$

Furthermore, for any state $z$ that is not a neighbor of $x$, we have that

$$\frac{\mathbb{Q}^{(2)}(x,z)}{|E|} = 0 = \frac{\mathbb{Q}^{(2)}(z,x)}{|E|},$$

so the chain is reversible with respect to the uniform distribution on $E$. Note, however, that since $\mathbb{Q}^{(2)}$ is not irreducible, this does not mean that the chain will converge in distribution to $\pi^{(2)}$.

(3) $\mathbb{Q}^{(3)}$ is irreducible and reversible. Suppose we start from some state $x = (x_0, ..., x_{N+1})$. Then we can achieve any single transposition between indices $i, j \in [1, N], i < j$ by first flipping all elements $x_i, ..., x_j$ so we transition to some state $y$ and then flipping all elements $y_{i+1}, ..., y_{j-1}$. This results in the state $z$ which is a single transposition of indices $i, j$ away from $x$. Thus for the same reasons $\mathbb{Q}^{(1)}$ was irreducible, $\mathbb{Q}^{(3)}$ is irreducible.

Furthermore, $\mathbb{Q}^{(3)}$ is reversible with respect to the distribution $\pi^{(3)} = \{\pi_x^{(3)} = \binom{N}{2}^{-1} : x \in E\}$. From any state $x$, we will choose two integers $i, j \in [1, N]$ from $N$ to determine our next state. Thus each state $x$ has $\binom{N}{2}$ neighbors, so for $y$ a neighbor of $x$, $\mathbb{Q}^{(3)}(x,y) = \mathbb{Q}^{(3)}(y,x) = \binom{n}{2}^{-1}$, since there is only one unique choice of $i, j$ that will transition from $x$ to $y$, and so only one choice of $j, i$ to revert back to $x$. Then

$$\frac{\mathbb{Q}^{(3)}(x,z)}{\mathbb{Q}^{(3)}(z,x)} = 1 = \frac{\pi_z^{(3)}}{\pi_x^{(3)}}.$$

Now, if $z$ is some state not a neighbor of $x$, then $\mathbb{Q}^{(3)}(x,z) = \mathbb{Q}^{(3)}(z,x) = 0$. Thus the detailed balance equations for $\pi^{(3)}$ will again hold, and so the chain is indeed reversible with respect to $\pi^{(3)}$.

(4) $\mathbb{Q}^{(4)}$ is irreducible and reversible. Note that if we choose $i \in [1, N], j \in [0, N]$ such that $j = i + 1$, then we are effectively choosing a single $i \in [1, N]$ and flipping $x_i, x_{i+1}$, so $\mathbb{Q}^{(4)}$ is irreducible for the same reasons $\mathbb{Q}^{(2)}$ is.

Moreover, $\mathbb{Q}^{(4)}$ is also reversible with respect to the uniform distribution on $E : \pi^{(4)} = \{\pi_x^{(4)} = \frac{1}{|E|}, x \in E\}$. Since there is a uniform chance of transitioning to any neighbor $y$ of $x$, and we can observe that one can transition from $y = (x_0, ..., x_{i-1}, x_{i+1}, ..., x_j, x_i, x_{j+1}, ..., x_N, x_{N+1})$ back to $x = (x_0, ..., x_{i-1}, x_i, x_{i+1}, ..., x_j, x_{j+1}, ..., x_N, x_{N+1})$ by simply choosing to place the $x_i$ entry in front of the $x_{i-1}$ entry in $y$. Then $\mathbb{Q}^{(4)}(x,y) = \mathbb{Q}^{(4)}(y,x)$, so

$$\frac{\mathbb{Q}^{(4)}(x,y)}{|E|} = \frac{\mathbb{Q}^{(4)}(y,x)}{|E|},$$

and thus the chain is reversible.

## 3.3 Part 3c

We will provide the outline for an implementation of the Metropolis algorithm to complete this task.

**Algorithm 1** Metropolis for TSP

**Require** number of iterations $n$
**Let** $\mathbb{Q}$ be the uniform transition matrix on $E$
**Pick** $X_0 \in E$ uniformly
**for** i in $\{1, ..., n\}$ **do**
  **Pick** $y$ a neighbor of $X_i = x$ according to $\mathbb{Q}$
  **Compute**

$$\frac{\pi_T(y)}{\pi_T(x)} = \exp\left(\frac{1}{T}(f(x) - f(y))\right) = r(x, y)$$

  **if** $r(x, y) > 1$ **do**
    $r(x, y) = 1$
  **end if**
  **Generate** $u$ from $U[0, 1]$
  **if** $r(x, y) > u$ **do**
    $X_{i+1} = y$
  **else do**
    $X_{i+1} = x$
  **end if**
**end for**

To avoid computation of the $Z$ term in the formula given for $\pi_T$, we use the fact that, for any states $x, y \in E$,

$$\frac{\pi_T(y)}{\pi_T(x)} = \exp\left(\frac{1}{T}(f(x) - f(y))\right).$$

Then by defining $\frac{\pi_T(y)}{\pi_T(x)}$ to be our acceptance ratio for this implementation of the Metropolis algorithm, we will generate a Markov chain that will begin to sample from $\pi_T$ after a large number of iterations.

Let's now observe the behavior of our model at extreme values of $T$. We have that

$$\exp\left(\frac{1}{T}(f(x) - f(y))\right) \xrightarrow{T \to +\infty} 1,$$

and

$$f(x) - f(y) > 0 \implies \exp\left(\frac{1}{T}(f(x) - f(y))\right) \xrightarrow{T \to 0} +\infty,$$

$$f(x) - f(y) < 0 \implies \exp\left(\frac{1}{T}(f(x) - f(y))\right) \xrightarrow{T \to 0} -\infty.$$

We see that, at large values of $T$, our acceptance probability approaches 1, so we accept very frequently regardless of whether tour $y$ costs less than $x$. At low values of $T$, our acceptance probability will either blow up to $+\infty$ if the cost of candidate tour $y$ is smaller than that of $x$ or to $-\infty$ if the cost of $y$ is larger, so we will rarely ever jump to higher cost candidates.

### 3.4 Part 3d

We have that, for any state $x \in E$,

$$\pi_T(x) = \frac{1}{Z} \exp\left(-\frac{f(x)}{T}\right),$$

so the larger the cost $f(x)$ of tour $x$, the higher $\pi_T(x)$ will be. Thus finding the best tour is simply a matter of finding the mode of the invariant distribution $\pi_T$.

We can do this by running the algorithm from part (3c) for a sufficiently large number $n$ of iterations, and then using the states of the chain from $n$ onward to be approximate samples from the invariant distribution. (We can assert this since we know that the chain generated by the Metropolis algorithm will converge to the given $\pi_T$ used in the computation of the acceptance probabilities.)

Then the most frequently observed state for large realizations of the chain will converge toward that with the largest invariant probability of occurring: That is, the tour with the lowest cost.

### 3.5 Part 3e

Let $t_1, ..., t_P$ be the minimum times to visit all cities in countries $1, ..., P$, respectively. Then, no matter our ordering of the visits to the countries, we will always have that our total time to visit all countries is given by

$$T = \sum_{i=1}^{P} t_i + T^*(x),$$

where $T^*$ is the time it takes for us to complete the traversal of the path $x = (x_0, x_1, ..., x_P, x_{P+1})$ between the countries.

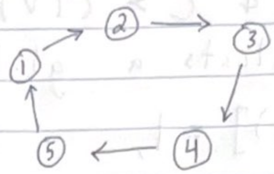
In essence, then, our problem is very similar to that from above: We solve the TSP with countries instead of cities, and add back the minimum times $t_1, ..., t_P$ to visit all cities in each country, yielding the minimum total time to visit all cities and each country in the map.

To determine the minimum time to traverse between all the countries in the world map, we will again employ a simulated annealing algorithm that randomly elects to permute between either the order in which countries are traversed or the order in which cities are traversed. Paths with lower costs (shorter distances) will more often be accepted than those with higher costs, in line with the general approach to Gibbs-based Metropolis algorithms.

### 3.6 Part 3f

The Jupyter notebook for this part is linked here: `https://github.com/ewilk0/Caltech-ACM-216/blob/master/ACM%20216%20PS%203/ACM%20216%20PS%203%20(TSP).ipynb`. It includes test cases with known solutions.

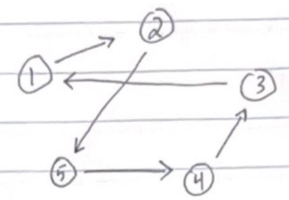
3. a) Assume starting path is



$x = (1, 2, 3, 4, 5, 1)$

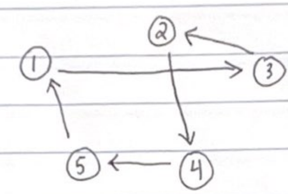
Then some sample transformations are

$\underline{Q^{(1)}}$:



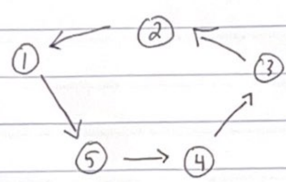
$x' = (1, 2, 5, 4, 3, 1)$

$\underline{Q^{(2)}}$:



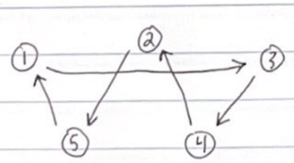
$x' = (1, 3, 2, 4, 5, 1)$

$\underline{Q^{(3)}}$:



$x' = (1, 5, 4, 3, 2, 1)$

$\underline{Q^{(4)}}$:



$x' = (1, 3, 4, 2, 5, 1)$