Computing a closest approximant to a stochastic matrix given linear constraints

R Iskandar
rowan_iskandar@brown.edu

TA Trikalinos
thomas_trikalinos@brown.edu

Abstract

To do

***K*eywords** discrete time point process · approximation

1 Introduction

To do

2 Initial statement of the problem

Consider a process with m states. Let $\boldsymbol{P}$ be the $m \times m$ transition probability matrix for a homogeneous point process in discrete time. Its elements P_{ij} are the probabilities of transitioning from the state i to state j at each time cycle; thus, $\boldsymbol{P}$ is row stochastic.[1]

We want to approximate $\boldsymbol{P}$ with a transition probability matrix $\boldsymbol{Q}$ that has some user imposed structure. Without loss of generality, we will consider the practical case of forcing structural zeros on $\boldsymbol{Q}$. Starting from the same state $\boldsymbol{x}_0$, and considering all cycles indexed by $n \in \{1, \ldots, N\} = [N]$, we want the process based on $\boldsymbol{P}$ and the process based on $\boldsymbol{Q}$ to be close in some well-defined sense.

A first description of our problem is as follows:

$$
\begin{aligned}
\underset{\boldsymbol{Q}}{\text{minimize}} \quad & \sum_{n=1}^{N} ||\boldsymbol{x}_0' \boldsymbol{P}^n - \boldsymbol{x}_0' \boldsymbol{Q}^n||_p && \text{(1a)}\\
\text{subject to} \quad & \boldsymbol{Q}\,\mathbf{1} = \mathbf{1}, && \text{(1b)}\\
& \boldsymbol{Q} \geq \mathbf{0}^{m\times m}, && \text{(1c)}\\
& \boldsymbol{Q} \leq \boldsymbol{M} \leq \mathbf{1}^{m\times m}, && \text{(1d)}\\
& \boldsymbol{x}_0 \geq 0, && \text{(1e)}\\
& \boldsymbol{x}_0\,\mathbf{1} = \mathbf{1}, && \text{(1f)}
\end{aligned}
$$

with $||\boldsymbol{x}||_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$ an L_p vector norm. From now on, write $\mathcal{Q}$ for the set of $\boldsymbol{Q}$ matrices that satisfy constraints (1b), (1c), and (1d).

Above,

- The objective (1a) minimizes the total of the sizes (L_p norms) of the differences between the states $\boldsymbol{x}_{\boldsymbol{P},n}$ with transition matrix $\boldsymbol{P}$ and the states $\boldsymbol{x}_{\boldsymbol{Q},n}$ with transition matrix $\boldsymbol{Q}$ over all N cycles; in each cycle n $(\boldsymbol{x}_{\boldsymbol{P},n} - \boldsymbol{x}_{\boldsymbol{Q},n})' = \boldsymbol{x}_0'(\boldsymbol{P}^n - \boldsymbol{Q}^n)$.

[1] Setting up $\boldsymbol{P}$ so that P_{ij} refers to transitions from state j to state i does not change the essence of the steps that follow; see Section 5 for a remark.

- Constraints (1b) and (1c) make $\boldsymbol{Q}$ row-stochastic.
- Matrix $\boldsymbol{M}$ in (1d) has elements 0 or 1. The constraint imposes structural zeros on $\boldsymbol{Q}$.
- Constraints (1e) and (1f) restrict the starting state $\boldsymbol{x}_0$ on an m-dimensional simplex.

3 Some closely related problems

Problem (1) motivates the use of an induced (subordinate) matrix norm to measure the size of each $(\boldsymbol{P}^n - \boldsymbol{Q}^n)$. From the definition of induced matrix norms,

$$||\boldsymbol{A}||_p = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{||\boldsymbol{A}\boldsymbol{x}||_p}{||\boldsymbol{x}||_p}$$

measures the size of the matrix $\boldsymbol{A}$ as the maximum change in size (in the sense of $||\cdot||_p$) that the linear operator of $\boldsymbol{A}$ confers to an arbitrary non-zero vector $\boldsymbol{x}$ [1, pg 327].

3.1 Simpler but related problems

Instead of (1), consider the related problems (one per norm)[2] that minimizes the size of differences of powers of matrices irrespective of the starting condition $\boldsymbol{x}_0$.

$$\underset{\boldsymbol{Q} \in \mathcal{Q}}{\text{minimize}} \quad \sum_{n=1}^{N} ||\boldsymbol{P}^n - \boldsymbol{Q}^n||_p \tag{2}$$

There are three induced matrix norms of practical interest for p values of $1, 2$ and ∞. It can be shown that for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{r \times c}$

$$\begin{aligned}
||\boldsymbol{A}||_1 &= \max_j \sum_i |A_{ij}|, && \text{"max column sum"} \\
||\boldsymbol{A}||_2 &= \sigma_{\max}(\boldsymbol{A}), && \text{spectral norm,} \\
||\boldsymbol{A}||_\infty &= \max_i \sum_j |A_{ij}| = ||\boldsymbol{A}'||_1, && \text{"max row sum"} \\
||\boldsymbol{A}\boldsymbol{B}||_p &\leq ||\boldsymbol{A}||_p \, ||\boldsymbol{B}||_p, && \text{consistency (all three norms)}
\end{aligned}$$

where $\sigma_{\max}(\boldsymbol{A})$ is the maximum singular value of $\boldsymbol{A}$ [1, p327].

3.2 Even simpler related problems

Each term in the objective of (2) is bounded above by $nb||\boldsymbol{P} - \boldsymbol{Q}||_p$, where, for each norm, b is a function of m as per Lemma 1 in the Appendix.

$$\sum_{n=1}^{N} ||\boldsymbol{P}^n - \boldsymbol{Q}^n||_p \quad \leq \quad \sum_{n=1}^{N} bn||\boldsymbol{P} - \boldsymbol{Q}||_p \; = \; \frac{bN(N+1)}{2}||\boldsymbol{P} - \boldsymbol{Q}||_p, \text{ with } b = \begin{cases} m & \text{if } p = 1, \\ \sqrt{m} & \text{if } p = 2, \text{ and} \\ 1 & \text{if } p = \infty. \end{cases} \tag{3}$$

Changing the objective of (2) to minimize the right hand side of (3) yields the problem

$$\underset{\boldsymbol{Q} \in \mathcal{Q}}{\text{minimize}} \quad ||\boldsymbol{P} - \boldsymbol{Q}||_p. \tag{4}$$

(The constants $bn > 0$ were dropped because they do not affect the optimal $\boldsymbol{Q}$.)

[2] Why *related* rather than *reformulated*? We use closed form formulas for induced matrix norms. These hold over all non-zero $\boldsymbol{x}$, or, equivalently, over all normalized $\boldsymbol{x}/||\boldsymbol{x}||_p$. However, problem (1) restricts $\boldsymbol{x}_0$ on a simplex. If we were to show that the formulas are the same for this restriction, we would have a *reformulation*.

4 Reformulation of the problems in (4)

4.1 The case $p = 1$

(See Section 4.3/ Figure 1, comment in last block stating $||\boldsymbol{A}||_1 = ||\boldsymbol{A}'||_\infty$).

4.2 The case $p = 2$

(To do need to remember SVD – this must be solved a bazillion times)

4.3 The case $p = \infty$

See Figure 1, block #3.

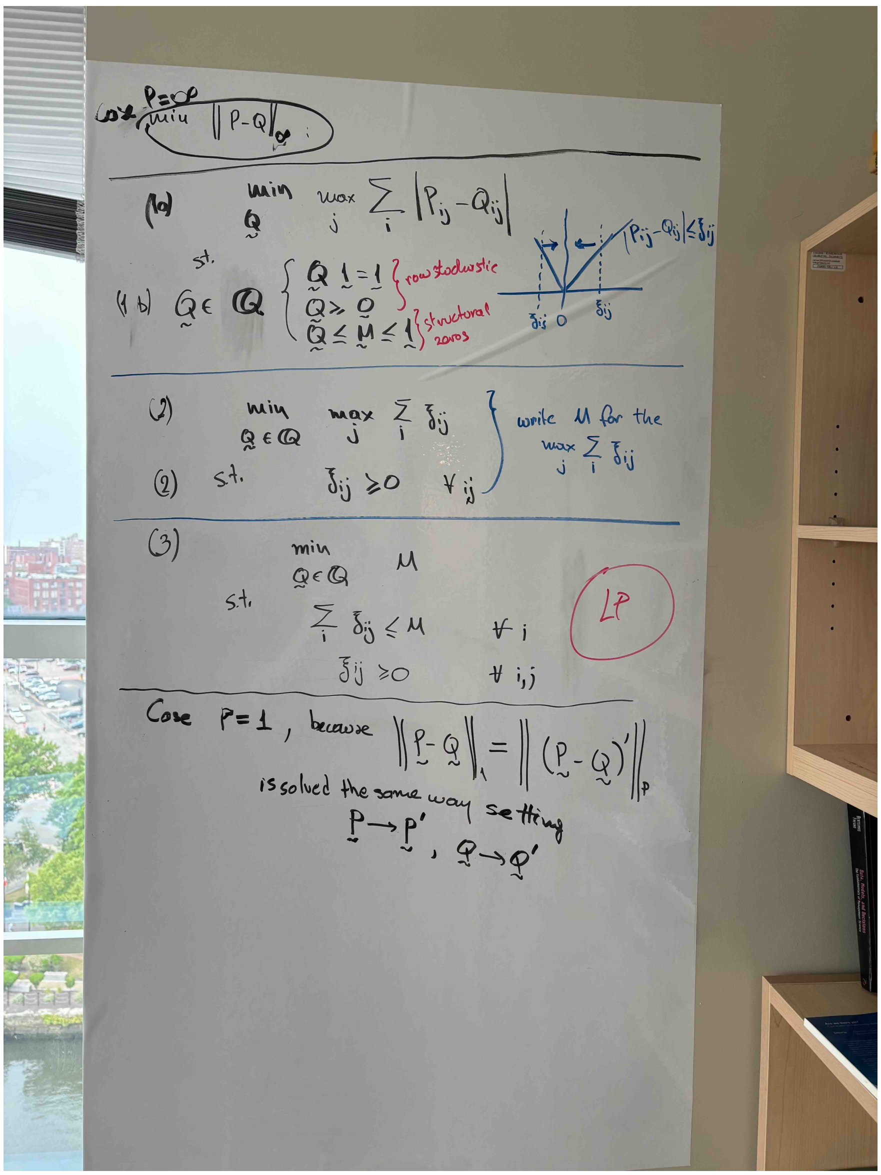

Figure 1: Reformulation of the case $p = \infty$.

5 Remarks

Appendix

Lemma 1. *If $\boldsymbol{P}, \boldsymbol{Q}$ are row stochastic $m \times m$ matrices,*

- $||\boldsymbol{P}^n - \boldsymbol{Q}^n||_1 \leq nm||\boldsymbol{P} - \boldsymbol{Q}||_1$
- $||\boldsymbol{P}^n - \boldsymbol{Q}^n||_2 \leq n\sqrt{m}||\boldsymbol{P} - \boldsymbol{Q}||_2$
- $||\boldsymbol{P}^n - \boldsymbol{Q}^n||_\infty \leq n||\boldsymbol{P} - \boldsymbol{Q}||_\infty$

PROOF: The proof is partly based on Lemma B.4 in [1]. The identity

$$\boldsymbol{P}^n - \boldsymbol{Q}^n = \sum_{i=0}^{n-1} \boldsymbol{P}^i (\boldsymbol{P} - \boldsymbol{Q}) \boldsymbol{Q}^{n-i-1}$$

can be shown by induction. For the base case of $n = 1$ and defining $\boldsymbol{Q}^0$ to be the identity matrix, the equation holds trivially. Assuming that it holds for n, it holds for $n + 1$:

$$\begin{aligned}
-\boldsymbol{Q}^{n+1} + \boldsymbol{P}^{n+1} &= \boldsymbol{P}^n \boldsymbol{Q} - \boldsymbol{Q}^{n+1} - \boldsymbol{P}^n \boldsymbol{Q} + \boldsymbol{P}^{n+1} \\
&= (\boldsymbol{P}^n - \boldsymbol{Q}^n)\boldsymbol{Q} + \boldsymbol{P}^n(\boldsymbol{P} - \boldsymbol{Q}) \\
&= (\sum_{i=0}^{n-1} \boldsymbol{P}^i (\boldsymbol{P} - \boldsymbol{Q}) \boldsymbol{Q}^{n-i-1})\boldsymbol{Q} + \boldsymbol{P}^n(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{Q}^0 \\
&= \sum_{i=0}^{n-1} \boldsymbol{P}^i (\boldsymbol{P} - \boldsymbol{Q}) \boldsymbol{Q}^{n-i} + \boldsymbol{P}^n(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{Q}^0 \\
&= \sum_{i=0}^{n} \boldsymbol{P}^i (\boldsymbol{P} - \boldsymbol{Q}) \boldsymbol{Q}^{n-i}.
\end{aligned}$$

Taking norms,

$$\begin{aligned}
||\boldsymbol{P}^n - \boldsymbol{Q}^n||_p &= \left|\left| \sum_{i=0}^{n-1} \boldsymbol{P}^i (\boldsymbol{P} - \boldsymbol{Q}) \boldsymbol{Q}^{n-i-1} \right|\right|_p \\
&\leq \sum_{i=0}^{n-1} ||\boldsymbol{P}^i (\boldsymbol{P} - \boldsymbol{Q}) \boldsymbol{Q}^{n-i-1}||_p && \text{(subadditivity)} \\
&\leq \sum_{i=0}^{n-1} ||\boldsymbol{P}^i||_p \; ||\boldsymbol{P} - \boldsymbol{Q}||_p \; ||\boldsymbol{Q}^{n-i-1}||_p && \text{(consistency)}
\end{aligned}$$

yielding

$$||\boldsymbol{P}^n - \boldsymbol{Q}^n||_p \leq ||\boldsymbol{P} - \boldsymbol{Q}||_p \quad \sum_{i=0}^{n-1} ||\boldsymbol{P}^i||_p \;\; ||\boldsymbol{Q}^{n-i-1}||_p \tag{5}$$

All three norms are within a constant factor of each other that depends only on the dimensions of the matrix [1, p327]. Specifically, for an $m \times m$ matrix $\boldsymbol{A}$,

$$||\boldsymbol{A}||_1 \leq m\; ||\boldsymbol{A}||_\infty, \text{ and} \tag{6}$$

$$||\boldsymbol{A}||_2 \leq \sqrt{m}\; ||\boldsymbol{A}||_\infty. \tag{7}$$

Because $\boldsymbol{P}, \boldsymbol{Q}$ are row stochastic, their powers $\boldsymbol{P}^i, \boldsymbol{Q}^i$ for $i \in \mathbb{N}$ are also row stochastic.

- For $p = \infty$, $||\boldsymbol{P}^i||_\infty = ||\boldsymbol{P}^i||_\infty = 1$, because the maximum row sum is always 1. Thus, each of the n terms in the sum on the right hand side of (5) is 1, which yields the third bound.
- For $p = 1$, (6) yields the first bound.
- For $p = 2$, (7) yields the second bound.

References

[1] Nicholas J Higham. *Functions of matrices: theory and computation*. SIAM, 2008.