Positive matrices

We write $P \geq Q$ if $P_{ij} \geq Q_{ij}$ for all i, j.
We write $P > Q$ if $P_{ij} > Q_{ij}$ for all i, j.

(Also, $P \geq 0, P > 0$.)

If $A \geq 0, B \geq 0$, then $AB \geq 0$.

Strict positivity trick

If $A > 0, \mathbf{v} \geq \mathbf{0}$, then $A\mathbf{v} > \mathbf{0}$, unless $\mathbf{v} = \mathbf{0}$.

If $A > 0$, $\mathbf{u} \geq \mathbf{w}$, then $A\mathbf{u} > A\mathbf{w}$, unless $\mathbf{u} = \mathbf{w}$.
(Proof: set $\mathbf{v} = \mathbf{u} - \mathbf{w}$ above.)

Perron-Frobenius theorem

Theorem

Let $P > 0$ be a square matrix. Then,

1. P has an eigenvalue $\lambda > 0$ with eigenvector $\mathbf{v} > \mathbf{0}$, i.e., $P\mathbf{v} = \lambda\mathbf{v}$.
2. λ is the largest eigenvalue in absolute value.
3. λ has multiplicity 1, and it is the only eigenvalue with eigenvector $\mathbf{v} > 0$ (up to multiples).

If $P \geq 0$, 1., 2., 4. hold with $\geq$ instead of $>$, but in many cases also the original statement holds (unless P is associated to a 'disconnected' or 'periodic' graph — we'll see it later.)

$\lambda = \rho(P), \mathbf{v}$ are called 'the Perron eigenvalue/eigenvector' of P. Also,

4. (monotonicity) If $0 \leq P \leq Q$, then $0 \leq \rho(P) \leq \rho(Q)$.

Proof (just a sketch while you are still awake)

We say that P stretches a vector $\mathbf{v} \geq \mathbf{0}$ by a factor $k > 0$ if $P\mathbf{v} \geq k\mathbf{v}$ (and k is the largest possible).

Example

$$P = \begin{bmatrix} 0.2 & 0.7 & 0.1 \\ 0.3 & 0.4 & 0.5 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad P\mathbf{v} = \begin{bmatrix} 1.9 \\ 2.6 \\ 2.1 \end{bmatrix}$$

P stretches $\mathbf{v}$ by a factor $\min(\frac{1.9}{1}, \frac{2.6}{2}, \frac{2.1}{3}) = 0.7$.

Indeed, $0.7 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \leq \begin{bmatrix} 1.9 \\ 2.6 \\ 2.1 \end{bmatrix}$.

Proof (cont.)

By the strict positivity trick,

$$k\mathbf{v} \leq P\mathbf{v} \implies kP\mathbf{v} < P(P\mathbf{v}),$$

i.e., P stretches $P\mathbf{v}$ by strictly more than it stretches $\mathbf{v}$; unless $k\mathbf{v} = P\mathbf{v}$.

Hence if we take the vector $\mathbf{v}_{\text{max}}$ which has the maximum stretch factor k_{max}, then it must be the case that $k_{\text{max}}\mathbf{v}_{\text{max}} = P\mathbf{v}_{\text{max}}$. This proves 1.

2. can be proved by taking another eigenvector $\mathbf{w}$ and considering the stretch factor of $|\mathbf{w}|$.

Proof (cont.)

3. can be proved by contradiction: assume P has another eigenpair $P\mathbf{w} = \mu\mathbf{w}$ with $\mathbf{w} > \mathbf{0}$ in addition to the Perron one $P\mathbf{v} = \lambda\mathbf{v}$.

The transposed matrix P^T has the same eigenvalue λ: $P^T\mathbf{u} = \lambda\mathbf{u}$, i.e., $\mathbf{u}^T P = \lambda\mathbf{u}^T$. (This is called sometimes a left eigenvector of P). By Part 1 of the theorem, $\mathbf{u} > \mathbf{0}$.

Compute in two ways:

$$\lambda\mathbf{u}^T\mathbf{w} = (\mathbf{u}^T P)\mathbf{w} = \mathbf{u}^T P\mathbf{w} = \mathbf{u}^T(P\mathbf{w}) = \mu\mathbf{u}^T\mathbf{w}.$$

4. follows by the fact that Q stretches the Perron vector $\mathbf{v}$ of P by a factor at least λ.

M-matrices

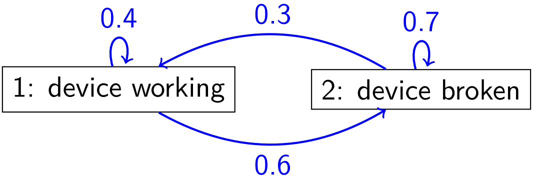
Similarly, matrices with sign pattern

$$\begin{bmatrix} + & - & - & - \\ - & + & - & - \\ - & - & + & - \\ - & - & - & + \end{bmatrix}$$

(possibly with zeros) can be seen as $sI - P$, for some scalar $s \geq 0$ and matrix $P \geq 0$, and some results on their eigenvalues can be derived from this (just a heads-up).

Markov chains

TL;DR: finite state automaton + transition probabilities.



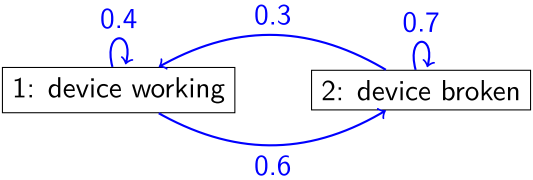
At every (discrete) 'time step', we follow an arrow exiting from the current state.

Markov property: transition probabilities do not depend on 'what happened earlier / where I am coming from'.

$$\mathbb{P}[s_k = j \mid s_1 = i_1, s_2 = i_2, \dots, s_{k-1} = i_{k-1}] = \mathbb{P}[s_k = j \mid s_{k-1} = i_{k-1}].$$

Homogeneity: transition probabilities do not change with step k.

Markov chains



Transition probability matrix: $P_{ij} = \mathbb{P}[\text{transition } i \to j]$, e.g.,

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix}.$$

P is row-stochastic, i.e., $\sum_j P_{ij} = 1$ for each row i. Or, in other words, $P\mathbf{1} = \mathbf{1}$ for the vector $\mathbf{1}$ of all ones (Perron vector with eigenvalue 1!)

Markov chain and linear algebra

Key idea: computing transition probabilities = matrix multiplication.

If $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_n]$ (often with $\boldsymbol{\pi} \cdot \mathbf{1} = 1$) contains probabilities of being in each state at a certain time t, then $\boldsymbol{\pi} P$ contains the probabilities at time $t + 1$.

Proof:

$$\mathbb{P}[s_{t+1} = j] = \sum_i \mathbb{P}[s_t = i] P_{ij} = \boldsymbol{\pi} P.$$

Markov chain and matrix products

Starting from a certain initial probability $\boldsymbol{\pi}$, the probability of observing a transition with probability P_1 first, then one with probability P_2, ..., then one with probability P_k, is

$$\boldsymbol{\pi} P_1 P_2 \dots P_k.$$

This is true not only when $\boldsymbol{\pi}, P_1, \dots, P_k$ are scalar, but also when they are vectors / matrices.

Hitting probabilities

Example: hitting probabilities.
Markov chain with a set of n_B 'bad' states, and n_G 'good' states.
Start out in a good state (with probabilities resp. $\boldsymbol{\pi} = [\pi_1, p_2, \dots, p_{n_G}]$); what is the probability of reaching ('hitting') a bad state? Which one of them is reached first?

$$P = \begin{bmatrix} P_{GG} & P_{GB} \\ P_{BG} & P_{BB} \end{bmatrix}$$

To reach a bad state, we can either:

- transition directly to a bad state, with probability $\boldsymbol{\pi} P_{GB}$;
- transition to a good state once, then to a bad state, $\boldsymbol{\pi} P_{GG} P_{GB}$;
- transition through 2 good states, then to a bad state, $\boldsymbol{\pi} P_{GG}^2 P_{GB}$; etc.

Total probability:

$$\boldsymbol{\pi}(I + P_{GG} + P_{GG}^2 + P_{GG}^3 + \dots)P_{GB} = \boldsymbol{\pi}(I - P_{GG})^{-1} P_{GB}.$$

Hitting probabilities

$$F_{GB} = (I + P_{GG} + P_{GG}^2 + P_{GG}^3 + \dots)P_{GB} = (I - P_{GG})^{-1} P_{GB}.$$

$(F_{GB})_{ij}$ gives the probabilities that the Markov chain first enters the set of bad states B in its jth state, starting from the ith good state.

Remarks:

- The formula $(I + M + M^2 + M^3 + \dots) = (I - M)^{-1}$ holds for each square matrix M with $\rho(M) < 1$.
- P_{GG} satisfies $P_{GG}\mathbf{1} \leq \mathbf{1}$, and the $\leq$ is not an equal (unless $P_{GB} = 0$). So by monotonicity $\rho(P_{GG}) < 1$.

Also, one can prove that the 'mean hitting time' is $(I - P_{GG})^{-1}\mathbf{1}$.

Example: a game with coins

We toss a coin repeatedly. I win if three consecutive tosses give THH, you win if you get HHT first. Who is at an advantage?

Set up transition matrix over all 8 possible sequences of 3 tosses, compute hitting probabilities for the set of bad states THH, HHT.

From the initial set $\boldsymbol{\pi} = [1/8, 1/8, \dots, 1.8]$, we get

$$\boldsymbol{\pi}_B + \boldsymbol{\pi}_G F_{GB} = \dots$$

Censoring

Censoring: rewrite transition history 'pretending the set G does not exist'.

$$b_1, g_1, b_2, b_2, g_2, g_3, g_1, g_2, b_1, b_3, g_1, g_2, b_2, \dots$$

becomes

$$b_1, b_2, b_2, b_1, b_3, b_2, \dots$$

Transition matrix of the censored chain $n_B \times n_B$:

$$P_{BB} + P_{BG}(I - P_{GG})^{-1} P_{GB}$$

Interpretation: we either transition from B to B directly, or we transition to G, stay inside it for $0, 1, 2, \dots$ time steps, and then get back out to B.

Interesting interpretation: Gaussian elimination on $I - P \iff$ censoring states $1, 2, 3, \dots$ in sequence.

Stationary probabilities

Suppose $P > 0$ for now (we'll see what changes if there are zero entries).

Theorem

For any initial probabilities $\boldsymbol{\pi}$, the probabilities $\boldsymbol{\pi} P, \boldsymbol{\pi} P^2, \dots, \boldsymbol{\pi} P^k, \dots$ of being in each state after k steps converge to a fixed vector $\boldsymbol{\mu}$ when $k \to \infty$.
This vector is the left eigenvector with eigenvalue $\rho(P) = 1$, i.e., the Perron vector of P^T.

Proof 1: this is simply the power method on P^T.

Proof 2: if $P = V\Lambda V^{-1}$ is diagonalizable, recall that all other eigenvalues apart from $1 = \lambda_1 = \rho(P)$ have modulus $|\lambda_i| < 1$, hence (cont.)

$$\boldsymbol{\pi}(V\Lambda V^{-1})^k = \boldsymbol{\pi} V \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix} V^{-1}$$

which converges when $k \to \infty$ to

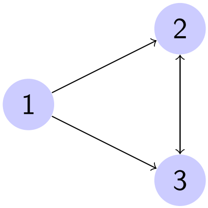
$$\boldsymbol{\pi}(V\Lambda V^{-1})^k = \boldsymbol{\pi} V \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} V^{-1}$$

which is a multiple of the first row of V^{-1}.

One can prove directly that the first row of V^{-1} is a left eigenvector, exactly like one proves that the columns of V are the right eigenvectors. Otherwise, take limits in both sides of $(\boldsymbol{\pi} P^k)P = \boldsymbol{\pi} P^{k+1}$.

What happens if there are zeros

Some states (called transient states) are visited only a few times and then ‘abandoned forever’, e.g.,



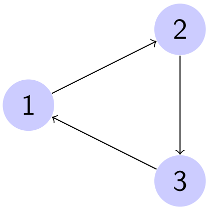
Key to understand it: doubly connected components of the graph with adjacency matrix P.

If a DCC has outgoing edges, eventually they will be taken with probability 1 $\implies$ the DCC will be abandoned with probability 1.

Each DCC without outgoing nodes is a possible final class.

Periodic chains

Another edge case: periodic chains. Suppose P is irreducible, but all closed paths have lengths that are multiple of a certain integer $d > 1$; e.g.,



In general, their transition matrices can be written as

$$\begin{bmatrix} & P_{12} & & & \\ & & P_{23} & & \\ & & & \ddots & \\ & & & & P_{d-1,d} \\ P_{d1} & & & & \end{bmatrix}.$$

(These may be blocks.)

General Perron–Frobenius

Periodic chains do not have a limit distribution (easy to see also in the example; starting from state 1 one 'loops indefinitely').

Another characterization: a chain is aperiodic if $P^k > 0$ for some k.

Periodic chains have d eigenvalues with modulus 1 (at the dth roots of 1). In particular, 'Perron-Frobenius with strict inequalities' does not hold for them.

The missing hypothesis

The Perron–Frobenius theorem 'with strict inequalities' holds for matrices $P \geq 0$ that have a doubly-connected graph (irreducible chains/matrices) and are aperiodic.

References

Meyer, *Matrix Analysis and Applied Linear Algebra*, chapter 8. — gentler introduction

Berman, Plemmons, *Nonnegative matrices in the mathematical sciences* — more technical; includes a whopping "50 equivalent conditions for a matrix to be a nonsingular M-matrix" (!)