Lecture 17
Perron-Frobenius Theory

- Positive and nonnegative matrices and vectors
- Perron-Frobenius theorems
- Markov chains
- Economic growth
- Population dynamics
- Max-min and min-max characterization
- Power control
- Linear Lyapunov functions
- Metzler matrices

Positive and nonnegative vectors and matrices

we say a matrix or vector is

- *positive* (or *elementwise positive*) if all its entries are positive
- *nonnegative* (or *elementwise nonnegative*) if all its entries are nonnegative

we use the notation $x > y$ ($x \geq y$) to mean $x - y$ is elementwise positive (nonnegative)

warning: if A and B are square and symmetric, $A \geq B$ can mean:

- $A - B$ is PSD ($i.e.$, $z^T A z \geq z^T B z$ for all z), or
- $A - B$ elementwise positive ($i.e.$, $A_{ij} \geq B_{ij}$ for all $i,\ j$)

in this lecture, $>$ and $\geq$ mean elementwise

Application areas

nonnegative matrices arise in many fields, $e.g.$,

- economics
- population models
- graph theory
- Markov chains
- power control in communications
- Lyapunov analysis of large scale systems

Basic facts

if $A \geq 0$ and $z \geq 0$, then we have $Az \geq 0$

conversely: if for all $z \geq 0$, we have $Az \geq 0$, then we can conclude $A \geq 0$

in other words, matrix multiplication preserves nonnegativity if and only if the matrix is nonnegative

if $A > 0$ and $z \geq 0$, $z \neq 0$, then $Az > 0$

conversely, if whenever $z \geq 0$, $z \neq 0$, we have $Az > 0$, then we can conclude $A > 0$

if $x \geq 0$ and $x \neq 0$, we refer to $d = (1/\mathbf{1}^T x)x$ as its *distribution* or normalized form

$d_i = x_i / (\sum_j x_j)$ gives the fraction of the total of x, given by x_i

Regular nonnegative matrices

suppose $A \in \mathbf{R}^{n \times n}$, with $A \geq 0$

A is called *regular* if for some $k \geq 1$, $A^k > 0$

meaning: form directed graph on nodes $1, \ldots, n$, with an arc from j to i whenever $A_{ij} > 0$

then $(A^k)_{ij} > 0$ if and only if there is a path of length k from j to i

A is regular if for some k there is a path of length k from every node to every other node examples:

- any positive matrix is regular

- $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ are not regular

- $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ is regular

Perron-Frobenius theorem for regular matrices

suppose $A \in \mathbf{R}^{n \times n}$ is nonnegative and regular, $i.e.$, $A^k > 0$ for some k

then

- there is an eigenvalue $\lambda_{\rm pf}$ of A that is real and positive, with positive left and right eigenvectors
- for any other eigenvalue λ, we have $|\lambda| < \lambda_{\rm pf}$
- the eigenvalue $\lambda_{\rm pf}$ is simple, $i.e.$, has multiplicity one, and corresponds to a 1×1 Jordan block

the eigenvalue $\lambda_{\rm pf}$ is called the *Perron-Frobenius* (PF) eigenvalue of A

the associated positive (left and right) eigenvectors are called the (left and right) PF eigenvectors (and are unique, up to positive scaling)

Perron-Frobenius theorem for nonnegative matrices

suppose $A \in \mathbf{R}^{n \times n}$ and $A \geq 0$

then

- there is an eigenvalue $\lambda_{\rm pf}$ of A that is real and nonnegative, with associated nonnegative left and right eigenvectors
- for any other eigenvalue λ of A, we have $|\lambda| \leq \lambda_{\rm pf}$

$\lambda_{\rm pf}$ is called the *Perron-Frobenius* (PF) eigenvalue of A

the associated nonnegative (left and right) eigenvectors are called (left and right) PF eigenvectors

in this case, they need not be unique, or positive

Markov chains

we consider stochastic process $X_0,\ X_1,\ldots$ with values in $\{1,\ldots,n\}$

$\mathbf{Prob}(X_{t+1}=i|X_t=j)=P_{ij}$

P is called the *transition matrix*; clearly $P_{ij}\geq 0$

let $p_t\in\mathbf{R}^n$ be the distribution of X_t, $i.e.$, $(p_t)_i=\mathbf{Prob}(X_t=i)$

then we have $p_{t+1}=Pp_t$

note: standard notation uses transpose of P, and row vectors for probability distributions

P is a *stochastic matrix*, $i.e.$, $P\geq 0$ and $\mathbf{1}^TP=\mathbf{1}^T$

so $\mathbf{1}$ is a left eigenvector with eigenvalue 1, which is in fact the PF eigenvalue of P

Equilibrium distribution

let π denote a PF (right) eigenvector of P, with $\pi \geq 0$ and $\mathbf{1}^T\pi = 1$

since $P\pi = \pi$, π corresponds to an *invariant distribution* or *equilibrium distribution* of the Markov chain

now suppose P is regular, which means for some k, $P^k > 0$

since $(P^k)_{ij}$ is $\mathbf{Prob}(X_{t+k} = i | X_t = j)$, this means there is positive probability of transitioning from any state to any other in k steps

since P is regular, there is a unique invariant distribution π, which satisfies $\pi > 0$

the eigenvalue 1 is simple and dominant, so we have $p_t \to \pi$, no matter what the initial distribution p_0

in other words: the distribution of a regular Markov chain always converges to the unique invariant distribution

Rate of convergence to equilibrium distribution

rate of convergence to equilibrium distribution depends on second largest eigenvalue magnitude, $i.e.$,

$$\mu = \max\{|\lambda_2|, \ldots, |\lambda_n|\}$$

where λ_i are the eigenvalues of P, and $\lambda_1 = \lambda_{\rm pf} = 1$

(μ is sometimes called the SLEM of the Markov chain)

the *mixing time* of the Markov chain is given by

$$T = \frac{1}{\log(1/\mu)}$$

(roughly, number of steps over which deviation from equilibrium distribution decreases by factor e)

Dynamic interpretation

consider $x_{t+1} = Ax_t$, with $A \geq 0$ and regular

then by PF theorem, $\lambda_{\rm pf}$ is the unique dominant eigenvalue

let $v,\ w > 0$ be the right and left PF eigenvectors of A, with $\mathbf{1}^T v = 1$, $w^T v = 1$

then as $t \to \infty$, $(\lambda_{\rm pf}^{-1} A)^t \to vw^T$

for any $x_0 \geq 0$, $x_0 \neq 0$, we have

$$\frac{1}{\mathbf{1}^T x_t} x_t \to v$$

as $t \to \infty$, *i.e.*, the distribution of x_t converges to v

we also have $(x_{t+1})_i/(x_t)_i \to \lambda_{\rm pf}$, *i.e.*, the one-period growth factor in each component always converges to $\lambda_{\rm pf}$

Economic growth

we consider an economy, with activity level $x_i \geq 0$ in sector i, $i = 1, \ldots, n$

given activity level x in period t, in period $t+1$ we have $x_{t+1} = Ax_t$, with $A \geq 0$

$A_{ij} \geq 0$ means activity in sector j does not decrease activity in sector i, $i.e.$, the activities are mutually noninhibitory

we'll assume that A is regular, with PF eigenvalue λ_{pf}, and left and right PF eigenvectors $w,\ v$, with $\mathbf{1}^T v = 1$, $w^T v = 1$

PF theorem tells us:

- $(x_{t+1})_i/(x_t)_i$, the growth factor in sector i over the period from t to $t+1$, each converge to λ_{pf} as $t \to \infty$
- the distribution of economic activity ($i.e.$, x normalized) converges to v

- asymptotically the economy exhibits (almost) balanced growth, by the factor λ_{pf}, in each sector

these hold independent of the original economic activity, provided it is nonnegative and nonzero

what does left PF eigenvector w mean?

for large t we have

$$x_t \sim \lambda_{\mathrm{pf}}^t w^T x_0 v$$

where $\sim$ means we have dropped terms small compared to dominant term

so asymptotic economic activity is scaled by $w^T x_0$

in particular, w_i gives the relative *value* of activity i in terms of long term economic activity

Population model

$(x_t)_i$ denotes number of individuals in group i at period t

groups could be by age, location, health, marital status, etc.

population dynamics is given by $x_{t+1} = Ax_t$, with $A \geq 0$

A_{ij} gives the fraction of members of group j that move to group i, or the number of members in group i created by members of group j ($e.g.$, in births)

$A_{ij} \geq 0$ means the more we have in group j in a period, the more we have in group i in the next period

- if $\sum_i A_{ij} = 1$, population is preserved in transitions out of group j
- we can have $\sum_i A_{ij} > 1$, if there are births (say) from members of group j
- we can have $\sum_i A_{ij} < 1$, if there are deaths or attrition in group j

now suppose A is regular

- PF eigenvector v gives asymptotic population distribution
- PF eigenvalue λ_{pf} gives asymptotic growth rate (if > 1) or decay rate (if < 1)
- $w^T x_0$ scales asymptotic population, so w_i gives relative value of initial group i to long term population

Path count in directed graph

we have directed graph on n nodes, with adjacency matrix $A \in \mathbf{R}^{n\times n}$

$$A_{ij} = \begin{cases} 1 & \text{there is an edge from node } j \text{ to node } i \\ 0 & \text{otherwise} \end{cases}$$

$\left(A^k\right)_{ij}$ is number of paths from j to i of length k

now suppose A is regular

then for large k,

$$A^k \sim \lambda_{\mathrm{pf}}^k v w^T = \lambda_{\mathrm{pf}}^k (\mathbf{1}^T w) v (w/\mathbf{1}^T w)^T$$

($\sim$ means: keep only dominant term)

$v,\ w$ are right, left PF eigenvectors, normalized as $\mathbf{1}^T v = 1$, $w^T v = 1$

total number of paths of length k: $\mathbf{1}^T A^k \mathbf{1} \approx \lambda_{\rm pf}^k (\mathbf{1}^T w)$

for k large, we have (approximately)

- $\lambda_{\rm pf}$ is factor of increase in number of paths when length increases by one
- v_i: fraction of length k paths that end at i
- $w_j/\mathbf{1}^T w$: fraction of length k paths that start at j
- $v_i w_j/\mathbf{1}^T w$: fraction of length k paths that start at j, end at i

- v_i measures importance/connectedness of node i as a *sink*
- $w_j/\mathbf{1}^T w$ measures importance/connectedness of node j as a *source*

(Part of) proof of PF theorem for positive matrices

suppose $A > 0$, and consider the optimization problem

$$\begin{array}{ll} \text{maximize} & \delta \\ \text{subject to} & Ax \geq \delta x \text{ for some } x \geq 0, \quad x \neq 0 \end{array}$$

note that we can assume $\mathbf{1}^T x = 1$

interpretation: with $y_i = (Ax)_i$, we can interpret y_i/x_i as the 'growth factor' for component i

problem above is to find the input distribution that maximizes the minimum growth factor

let λ_0 be the optimal value of this problem, and let v be an optimal point, $i.e.$, $v \geq 0$, $v \neq 0$, and $Av \geq \lambda_0 v$

we will show that λ_0 is the PF eigenvalue of A, and v is a PF eigenvector

first let's show $Av = \lambda_0 v$, *i.e.*, v is an eigenvector associated with λ_0

if not, suppose that $(Av)_k > \lambda_0 v_k$

now let's look at $\tilde{v} = v + \epsilon e_k$

we'll show that for small $\epsilon > 0$, we have $A\tilde{v} > \lambda_0 \tilde{v}$, which means that $A\tilde{v} \geq \delta \tilde{v}$ for some $\delta > \lambda_0$, a contradiction

for $i \neq k$ we have

$$(A\tilde{v})_i = (Av)_i + A_{ik}\epsilon > (Av)_i \geq \lambda_0 v_i = \lambda_0 \tilde{v}_i$$

so for any $\epsilon > 0$ we have $(A\tilde{v})_i > \lambda_0 \tilde{v}_i$

$$\begin{aligned} (A\tilde{v})_k - \lambda_0 \tilde{v}_k &= (Av)_k + A_{kk}\epsilon - \lambda_0 v_k - \lambda_0 \epsilon \\ &= (Av)_k - \lambda_0 v_k - \epsilon(\lambda_0 - A_{kk}) \end{aligned}$$

since $(Av)_k - \lambda_0 v_k > 0$, we conclude that for small $\epsilon > 0$, $(A\tilde{v})_k - \lambda_0 \tilde{v}_k > 0$

to show that $v > 0$, suppose that $v_k = 0$

from $Av = \lambda_0 v$, we conclude $(Av)_k = 0$, which contradicts $Av > 0$ (which follows from $A > 0$, $v \geq 0$, $v \neq 0$)

now suppose $\lambda \neq \lambda_0$ is another eigenvalue of A, $i.e.$, $Az = \lambda z$, where $z \neq 0$

let $|z|$ denote the vector with $|z|_i = |z_i|$

since $A \geq 0$ we have $A|z| \geq |Az| = |\lambda||z|$

from the definition of λ_0 we conclude $|\lambda| \leq \lambda_0$

(to show strict inequality is harder)

Max-min ratio characterization

proof shows that PF eigenvalue is optimal value of optimization problem

$$\begin{array}{ll} \text{maximize} & \min_i \frac{(Ax)_i}{x_i} \\ \text{subject to} & x > 0 \end{array}$$

and that PF eigenvector v is optimal point:

- PF eigenvector v maximizes the minimum growth factor over components
- with optimal v, growth factors in all components are equal (to λ_{pf})

in other words: by maximizing minimum growth factor, we actually achieve balanced growth

Min-max ratio characterization

a related problem is

$$\begin{array}{ll} \text{minimize} & \max_i \frac{(Ax)_i}{x_i} \\ \text{subject to} & x > 0 \end{array}$$

here we seek to minimize the maximum growth factor in the coordinates

the solution is surprising: the optimal value is λ_{pf} and the optimal x is the PF eigenvector v

- if A is nonnegative and regular, and $x > 0$, the n growth factors $(Ax)_i/x_i$ 'straddle' λ_{pf}: at least one is $\geq \lambda_{\mathrm{pf}}$, and at least one is $\leq \lambda_{\mathrm{pf}}$
- when we take x to be the PF eigenvector v, all the growth factors are equal, and solve both max-min and min-max problems

Power control

we consider n transmitters with powers $P_1, \ldots, P_n > 0$, transmitting to n receivers

path gain from transmitter j to receiver i is $G_{ij} > 0$

signal power at receiver i is $S_i = G_{ii} P_i$

interference power at receiver i is $I_i = \sum_{k \neq i} G_{ik} P_k$

signal to interference ratio (SIR) is

$$S_i / I_i = \frac{G_{ii} P_i}{\sum_{k \neq i} G_{ik} P_k}$$

how do we set transmitter powers to maximize the minimum SIR?

we can just as well minimize the maximum interference to signal ratio, *i.e.*, solve the problem

$$\begin{array}{ll} \text{minimize} & \max_i \frac{(\tilde{G}P)_i}{P_i} \\ \text{subject to} & P > 0 \end{array}$$

where

$$\tilde{G}_{ij} = \begin{cases} G_{ij}/G_{ii} & i \neq j \\ 0 & i = j \end{cases}$$

since $\tilde{G}^2 > 0$, $\tilde{G}$ is regular, so solution is given by PF eigenvector of $\tilde{G}$

PF eigenvalue $\lambda_{\rm pf}$ of $\tilde{G}$ is the optimal interference to signal ratio, *i.e.*, maximum possible minimum SIR is $1/\lambda_{\rm pf}$

with optimal power allocation, all SIRs are equal

note: $\tilde{G}$ is the matrix of ratios of interference to signal path gains

Nonnegativity of resolvent

suppose A is nonnegative, with PF eigenvalue $\lambda_{\rm pf}$, and $\lambda \in \mathbf{R}$

then $(\lambda I - A)^{-1}$ exists and is nonnegative, if and only if $\lambda > \lambda_{\rm pf}$

for any square matrix A the power series expansion

$$(\lambda I - A)^{-1} = \frac{1}{\lambda} I + \frac{1}{\lambda^2} A + \frac{1}{\lambda^3} A^2 + \cdots$$

converges provided $|\lambda|$ is larger than all eigenvalues of A

if $\lambda > \lambda_{\rm pf}$, this shows that $(\lambda I - A)^{-1}$ is nonnegative

to show converse, suppose $(\lambda I - A)^{-1}$ exists and is nonnegative, and let $v \neq 0$, $v \geq 0$ be a PF eigenvector of A

then we have

$$(\lambda I - A)^{-1} v = \frac{1}{\lambda - \lambda_{\rm pf}} v \geq 0$$

and it follows that $\lambda > \lambda_{\rm pf}$

Equilibrium points

consider $x_{t+1} = Ax_t + b$, where A and b are nonnegative

equilibrium point is given by $x_{\mathrm{eq}} = (I - A)^{-1}b$

by resolvent result, if A is stable, then $(I - A)^{-1}$ is nonnegative, so equilibrium point x_{eq} is nonnegative for any nonnegative b

moreover, equilibrium point is monotonic function of b: for $\tilde{b} \geq b$, we have $\tilde{x}_{\mathrm{eq}} \geq x_{\mathrm{eq}}$

conversely, if system has a nonnegative equilibrium point, for every nonnegative choice of b, then we can conclude A is stable

Iterative power allocation algorithm

we consider again the power control problem

suppose γ is the desired or target SIR

simple iterative algorithm: at each step t,

1. first choose $\tilde{P}_i$ so that

$$\frac{G_{ii}\tilde{P}_i}{\sum_{k\neq i} G_{ik}(P_t)_k} = \gamma$$

   $\tilde{P}_i$ is the transmit power that would make the SIR of receiver i equal to γ, *assuming none of the other powers change*

2. set $(P_{t+1})_i = \tilde{P}_i + \sigma_i$, where $\sigma_i > 0$ is a parameter
   $i.e.$, add a little extra power to each transmitter)

each receiver only needs to know its current SIR to adjust its power: if current SIR is α dB below (above) γ, then increase (decrease) transmitter power by α dB, then add the extra power σ

$i.e.$, this is a *distributed algorithm*

question: does it work? (we assume that $P_0 > 0$)

answer: yes, if and only if γ is less than the maximum achievable SIR, $i.e.$, $\gamma < 1/\lambda_{\mathrm{pf}}(\tilde{G})$

to see this, algorithm can be expressed as follows:

- in the first step, we have $\tilde{P} = \gamma \tilde{G} P_t$
- in the second step we have $P_{t+1} = \tilde{P} + \sigma$

and so we have

$$P_{t+1} = \gamma \tilde{G} P_t + \sigma$$

a linear system with constant input

PF eigenvalue of $\gamma\tilde{G}$ is $\gamma\lambda_{\rm pf}$, so linear system is stable if and only if $\gamma\lambda_{\rm pf} < 1$

power converges to equilibrium value

$$P_{\rm eq} = (I - \gamma\tilde{G})^{-1}\sigma$$

(which is positive, by resolvent result)

now let's show this equilibrium power allocation achieves SIR at least γ for each receiver

we need to verify $\gamma\tilde{G}P_{\rm eq} \leq P_{\rm eq}$, *i.e.*,

$$\gamma\tilde{G}(I - \gamma\tilde{G})^{-1}\sigma \leq (I - \gamma\tilde{G})^{-1}\sigma$$

or, equivalently,

$$(I - \gamma\tilde{G})^{-1}\sigma - \gamma\tilde{G}(I - \gamma\tilde{G})^{-1}\sigma \geq 0$$

which holds, since the lefthand side is just σ

Linear Lyapunov functions

suppose $A \geq 0$

then $\mathbf{R}_+^n$ is invariant under system $x_{t+1} = Ax_t$

suppose $c > 0$, and consider the linear Lyapunov function $V(z) = c^T z$

if $V(Az) \leq \delta V(z)$ for some $\delta < 1$ and all $z \geq 0$, then V proves (nonnegative) trajectories converge to zero

fact: a nonnegative regular system is stable if and only if there is a linear Lyapunov function that proves it

to show the 'only if' part, suppose A is stable, *i.e.*, $\lambda_{\rm pf} < 1$

take $c = w$, the (positive) left PF eigenvector of A

then we have $V(Az) = w^T Az = \lambda_{\rm pf} w^T z$, *i.e.*, V proves all nonnegative trajectories converge to zero

Weighted ℓ_1-norm Lyapunov function

to make the analysis apply to *all* trajectories, we can consider the weighted sum absolute value (or weighted ℓ_1-norm) Lyapunov function

$$V(z) = \sum_{i=1}^{n} w_i |z_i| = w^T |z|$$

then we have

$$V(Az) = \sum_{i=1}^{n} w_i |(Az)_i| \leq \sum_{i=1}^{n} w_i (A|z|)_i = w^T A |z| = \lambda_{\mathrm{pf}} w^T |z|$$

which shows that V decreases at least by the factor λ_{pf}

conclusion: a nonnegative regular system is stable if and only if there is a weighted sum absolute value Lyapunov function that proves it

SVD analysis

suppose $A \in \mathbf{R}^{m \times n}$, $A \geq 0$

then $A^T A \geq 0$ and $AA^T \geq 0$ are nonnegative

hence, there are nonnegative left & right singular vectors v_1, w_1 associated with σ_1

in particular, there is an optimal rank-1 approximation of A that is nonnegative

if $A^T A$, AA^T are regular, then we conclude

- $\sigma_1 > \sigma_2$, $i.e.$, maximum singular value is isolated
- associated singular vectors are positive: $v_1 > 0$, $w_1 > 0$

Continuous time results

we have already seen that $\mathbf{R}_+^n$ is invariant under $\dot{x} = Ax$ if and only if $A_{ij} \geq 0$ for $i \neq j$

such matrices are called *Metzler matrices*

for a Metzler matrix, we have

- there is an eigenvalue $\lambda_{\rm metzler}$ of A that is real, with associated nonnegative left and right eigenvectors
- for any other eigenvalue λ of A, we have $\Re\lambda \leq \lambda_{\rm metzler}$

  $i.e.$, the eigenvalue $\lambda_{\rm metzler}$ is dominant for system $\dot{x} = Ax$
- if $\lambda > \lambda_{\rm metzler}$, then $(\lambda I - A)^{-1} \geq 0$

the analog of the stronger Perron-Frobenius results:

if $(\tau I + A)^k > 0$, for some τ and some k, then

- the left and right eigenvectors associated with eigenvalue $\lambda_{\mathrm{metzler}}$ of A are positive
- for any other eigenvalue λ of A, we have $\Re\lambda < \lambda_{\mathrm{metzler}}$

  i.e., the eigenvalue $\lambda_{\mathrm{metzler}}$ is strictly dominant for system $\dot{x} = Ax$

Derivation from Perron-Frobenius Theory

suppose A is Metzler, and choose τ s.t. $\tau I + A \geq 0$
($e.g.$, $\tau = 1 - \min_i A_{ii}$)

by PF theory, $\tau I + A$ has PF eigenvalue $\lambda_{\rm pf}$, with associated right and left eigenvectors $v \geq 0$, $w \geq 0$

from $(\tau I + A)v = \lambda_{\rm pf} v$ we get $Av = (\lambda_{\rm pf} - \tau)v = \lambda_0 v$, and similarly for w

we'll show that $\Re\lambda \leq \lambda_0$ for any eigenvalue λ of A

suppose λ is an eigenvalue of A

suppose $\tau + \lambda$ is an eigenvalue of $\tau I + A$

by PF theory, we have $|\tau + \lambda| \leq \lambda_{\rm pf} = \tau + \lambda_0$

this means λ lies inside a circle, centered at $-\tau$, that passes through λ_0

which implies $\Re\lambda \leq \lambda_0$

Linear Lyapunov function

suppose $\dot{x} = Ax$ is stable, and A is Metzler, with $(\tau I + A)^k > 0$ for some τ and some k

we can show that all nonnegative trajectories converge to zero using a linear Lyapunov function

let $w > 0$ be left eigenvector associated with dominant eigenvalue $\lambda_{\rm metzler}$

then with $V(z) = w^T z$ we have

$$\dot{V}(z) = w^T A z = \lambda_{\rm metzler} w^T z = \lambda_{\rm metzler} V(z)$$

since $\lambda_{\rm metzler} < 0$, this proves $w^T z \to 0$