Nonlinear Perron-Frobenius theory and applications

Francesco Tudisco

Gran Sasso Science Institute

Numerical Linear Algebra days – Due giorni di Algebra Linear Numerica
GSSI – May 10, 2023

May 10, 2022

May 10, 2022

May 10, 2022

May 10, 2022

Introduction:

Cone theoretic proof of the PF theorem for positive matrices

Perron eigenvector

Consider the problem:

Find x such that $Ax = \lambda x$

Perron eigenvector

Consider the problem:

Find x such that $Ax = \lambda x$

Perron–Frobenius theorem

If A is positive then there exists a unique solution $x^\star > 0$, $\|x^\star\| = 1$ and

$$x_{m+1} = Ax_m / \|Ax_m\| \xrightarrow{m\to\infty} x^\star$$

for any choice of $x_0 > 0$.

Perron eigenvector

Consider the problem:

Find x such that $Ax = \lambda x$

Perron–Frobenius theorem

If A is positive then there exists a unique solution $x^\star > 0$, $\|x^\star\| = 1$ and

$$x_{m+1} = Ax_m/\|Ax_m\| \xrightarrow{m\to\infty} x^\star$$

for any choice of $x_0 > 0$.

How do you prove that?

Cone-theoretic proof 1/4

Hilbert-Birkhoff projective metric

x, y positive vectors $\qquad d_H(x,y) = \log\left(\max_i \frac{x_i}{y_i}\ \max_i \frac{y_i}{x_i}\right)$

d_H is projective i.e. $d_H(x,y) = d_H(\alpha x, \beta y)$, $\forall \alpha, \beta > 0$. Thus

Cone-theoretic proof 2/4

Observation 1

Eigenvector	$\iff$	Fixed point
$Ax = \lambda x$	$\iff$	$d_H(Ax, x) = 0$

Observation 2

Take any norm $\|\cdot\|$ and let $S_{++} = \{x > 0 : \|x\| = 1\}$
Then (S_{++}, d_H) is a complete metric space.

Cone-theoretic proof 3/4

Birkhoff-Hopf theorem

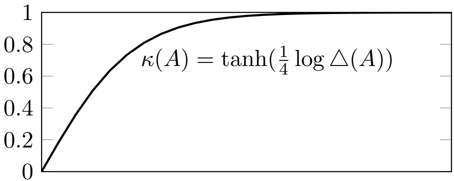
Let A be any positive matrix. Then

$$d_H(Ax, Ay) \leq \kappa(A)\, d_H(x, y) \qquad \forall x, y > 0$$

where $\kappa(A) = \tanh(\frac{1}{4}\mathrm{diam}(A)) = \tanh(\frac{1}{4} \log \triangle(A))$, and

$$\triangle(A) = \triangle(A^T) = \max_{ijhk} \frac{A_{ij} A_{hk}}{A_{ik} A_{hj}}$$

$$\triangle(A) \geq \frac{A_{12} A_{12}}{A_{12} A_{12}} = 1$$



Birkhoff contraction ratio - example

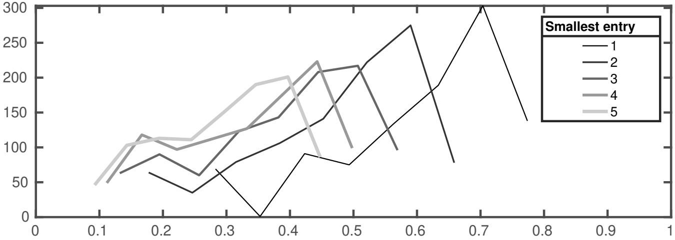


Figure: Each line shows the distribution of $\kappa(A)$ over 1000 random matrices A with entries between s and 10. Different curves correspond to different values of $s \in \{1, 2, \dots, 5\}$.

Cone-theoretic proof 4/4

If

$$G(x) = Ax/\|Ax\|$$

then

$$d_H(G(x), G(y)) = d_H(Ax, Ay) \leq \kappa(A) d_H(x, y)$$

Cone-theoretic proof 4/4

If

$$G(x) = Ax/\|Ax\|$$

then

$$d_H(G(x), G(y)) = d_H(Ax, Ay) \leq \kappa(A) d_H(x, y)$$

Thus: if A is a positive matrix, G is a contraction in the metric space (S_{++}, d_H), with contraction constant $\kappa(A)$.

Cone-theoretic proof 4/4

If

$$G(x) = Ax/\|Ax\|$$

then

$$d_H(G(x), G(y)) = d_H(Ax, Ay) \leq \kappa(A) d_H(x, y)$$

Thus: if A is a positive matrix, G is a contraction in the metric space (S_{++}, d_H), with contraction constant $\kappa(A)$.

Using the Banach fixed point theorem we conclude that:

$$Ax_m/\|Ax_m\| \xrightarrow{m\to\infty} x^\star \text{ as } \kappa(A)^m$$

and $x^\star$ is the unique positive eigenvector of A

Introduction:

Nonlinear matrix eigenvectors

Nonlinear Perron eigenvector

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be such that $x > 0 \Longrightarrow f(x) > 0$

Consider the problem

> Find x such that $Af(x) = \lambda x$

For the time being: $f(x) = x^\alpha =$ component-wise power, $\alpha \neq 0$

Nonlinear Perron eigenvector

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be such that $x > 0 \Longrightarrow f(x) > 0$

Consider the problem

Find x such that $Af(x) = \lambda x$

For the time being: $f(x) = x^{\alpha} =$ component-wise power, $\alpha \neq 0$

Can we do the same?

Dilatations

$$\alpha \in \mathbb{R} \quad \Longrightarrow \quad d_H(x^\alpha, y^\alpha) = |\alpha|\, d_H(x,y)$$

Therefore

$$d_H(Af(x), Af(y)) \leq \kappa(A) d_H(f(x), f(y)) = \underbrace{|\alpha|\kappa(A)}_{\text{contraction constant}} d_H(x,y)$$

Nonlinear Perron eigenvector

Consider the problem

Find x such that $Ax^\alpha = \lambda x$

Theorem

If $|\alpha|\kappa(A) < 1$ then there exists a unique solution $x^\star > 0$, $\|x^\star\| = 1$ and

$$x_{m+1} = Ax_m^\alpha / \|Ax_m^\alpha\| \xrightarrow{m\to\infty} x^\star \text{ as } \big(|\alpha|\kappa(A)\big)^m$$

What about other *spectral equations*?

Singular vectors:

PF theorem for singular vectors

Singular vectors

Consider the problem

Find (x, y) such that $\begin{cases} Ay = \lambda x \\ A^T x = \lambda y \end{cases}$

$$\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} & A \\ A^T & \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

Problem: Even if A is positive, it holds

$$\kappa(A) = \kappa(A^T) < 1 \quad \text{but} \quad \kappa\left(\begin{bmatrix} & A \\ A^T & \end{bmatrix}\right) = 1$$

Singular vectors

Consider the problem

Find (x, y) such that $\begin{cases} Ay = \lambda x \\ A^T x = \lambda y \end{cases}$

$$\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} & A \\ A^T & \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

Problem: Even if A is positive, it holds

$$\kappa(A) = \kappa(A^T) < 1 \quad \text{but} \quad \kappa\left(\begin{bmatrix} & A \\ A^T & \end{bmatrix}\right) = 1$$

Solution: “Higher-order” Hilbert metric

"Order-2" Birkhoff-Hopf theorem

$$\begin{bmatrix} d_H(Ay, Av) \\ d_H(A^T x, A^T u) \end{bmatrix} \leq \begin{bmatrix} & \kappa(A) \\ \kappa(A^T) & \end{bmatrix} \begin{bmatrix} d_H(x,u) \\ d_H(y,v) \end{bmatrix} = K(\mathcal{A}) \begin{bmatrix} d_H(x,u) \\ d_H(y,v) \end{bmatrix}$$

We call $K(\mathcal{A})$ "Lipschitz matrix" of $\mathcal{A} = \begin{bmatrix} & A \\ A^T & \end{bmatrix} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$

"Order-2" Birkhoff-Hopf theorem

$$\begin{bmatrix} d_H(Ay, Av) \\ d_H(A^T x, A^T u) \end{bmatrix} \leq \begin{bmatrix} & \kappa(A) \\ \kappa(A^T) & \end{bmatrix} \begin{bmatrix} d_H(x, u) \\ d_H(y, v) \end{bmatrix} = K(\mathcal{A}) \begin{bmatrix} d_H(x, u) \\ d_H(y, v) \end{bmatrix}$$

We call $K(\mathcal{A})$ "Lipschitz matrix" of $\mathcal{A} = \begin{bmatrix} & A \\ A^T & \end{bmatrix} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$

Theorem

For any positive matrix A there exists a metric δ_H on $\mathbb{R}^n \times \mathbb{R}^n$ st

$$\delta_H\Big(\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix}, \mathcal{A}\begin{bmatrix} u \\ v \end{bmatrix}\Big) \leq \rho(K(\mathcal{A}))\, \delta_H\Big(\begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} u \\ v \end{bmatrix}\Big)$$

for all $(x, y) > 0$, $(u, v) > 0$.

δ_H is defined in terms of any positive eigenvector of $\mathcal{A}$

… moreover

Observation 1

Take any norm $\|\cdot\|$ and let $S^2_{++} = \{(x, y) > 0 : \|x\| = \|y\| = 1\}$

Then (S^2_{++}, δ_H) is a complete metric space

… moreover

Observation 1

Take any norm $\|\cdot\|$ and let $S^2_{++} = \{(x,y) > 0 : \|x\| = \|y\| = 1\}$

Then (S^2_{++}, δ_H) is a complete metric space

Observation 2

If G is the "normalized version" of $\mathcal{A}$

$$G\Big(\begin{bmatrix} x \\ y \end{bmatrix}\Big) = \begin{bmatrix} Ay/\|Ay\| \\ A^T x/\|A^T x\| \end{bmatrix}$$

then

$$\delta_H\Big(G\Big(\begin{bmatrix} x \\ y \end{bmatrix}\Big), G\Big(\begin{bmatrix} u \\ v \end{bmatrix}\Big)\Big) = \delta_H\Big(\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix}, \mathcal{A}\begin{bmatrix} u \\ v \end{bmatrix}\Big)$$

… moreover

Observation 3

For this particular case we have

$$\rho(K(\mathcal{A})) = \rho\Big(\begin{bmatrix} & \kappa(A) \\ \kappa(A^T) & \end{bmatrix} \Big) = \kappa(A)$$

Perron-Frobenius theorem for singular values

Consider the problem

Find (x, y) such that $\begin{cases} Ay = \lambda x \\ A^T x = \lambda y \end{cases}$

Theorem

If A is positive then there exists a unique solution $x^\star, y^\star > 0$ such that $\|x^\star\| = \|y^\star\| = 1$ and

$$\begin{bmatrix} x_{m+1} \\ y_{m+1} \end{bmatrix} = \begin{bmatrix} Ay_m/\|Ay_m\| \\ A^T x_m/\|A^T x_m\| \end{bmatrix} \xrightarrow{m\to\infty} \begin{bmatrix} x^\star \\ y^\star \end{bmatrix} \text{ as } \kappa(A)^m$$

Singular vectors:

Nonlinear matrix singular vectors

More in general

$\circledast$ Find (x, y) such that $\begin{cases} Ay^{\beta} = \lambda x \\ Bx^{\alpha} = \mu y \end{cases}$

More in general

$\circledast$ Find (x,y) such that $\begin{cases} Ay^{\beta} = \lambda x \\ Bx^{\alpha} = \mu y \end{cases}$

Application examples:

- Matrix norms:

  Compute $\|A\|_{p,q} = \max_{x\neq 0} \|Ax\|_q / \|x\|_p$
  Boils down to $\circledast$ for $\alpha = 1/(p-1)$ and $\beta = q/(q-1)$

- Matrix rescaling (matrix Sibkhorn method) and entropy minimization:

  Find a diagonal D_1, D_2 such that D_1AD_2 is doubly stochastic
  Boils down to $\circledast$ for $\alpha = \beta = -1$

Birkhoff-Hopf theorem

$$\begin{bmatrix} d_H(Ay^\beta, Av^\beta) \\ d_H(Bx^\alpha, Bu^\alpha) \end{bmatrix} \leq \underbrace{\begin{bmatrix} & \kappa(A) \\ \kappa(B) & \end{bmatrix} \begin{bmatrix} |\alpha| & \\ & |\beta| \end{bmatrix}}_{K} \begin{bmatrix} d_H(x,u) \\ d_H(y,v) \end{bmatrix}$$

Theorem

There exists a metric δ_H such that

$$\delta_H\Big(\begin{bmatrix} Ay^\beta \\ Bx^\alpha \end{bmatrix}, \begin{bmatrix} Av^\beta \\ Bu^\alpha \end{bmatrix} \Big) \leq \rho(K)\, \delta_H\Big(\begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} u \\ v \end{bmatrix} \Big)$$

and $\rho(K) = \sqrt{|\alpha\beta|\kappa(A)\kappa(B)}$.

Nonlinear Perron singular vector

Find (x, y) such that $\begin{cases} Ay^{\beta} = \lambda x \\ Bx^{\alpha} = \mu y \end{cases}$

Theorem

If $|\alpha\beta|\kappa(A)\kappa(B) < 1$ then there exists a unique solution $x^\star, y^\star > 0$ such that $\|x^\star\| = \|y^\star\| = 1$ and

$$\begin{bmatrix} x_{m+1} \\ y_{m+1} \end{bmatrix} = \begin{bmatrix} Ay_m^{\beta}/\|Ay_m^{\beta}\| \\ Bx_m^{\alpha}/\|Bx_m^{\alpha}\| \end{bmatrix} \xrightarrow{m\to\infty} \begin{bmatrix} x^\star \\ y^\star \end{bmatrix} \text{ as } \big(|\alpha\beta|\kappa(A)\kappa(B)\big)^{m/2}$$

Moreover: If $A = B^T$ then $\lambda = \mu$

Singular vectors:

Example of application: matrix operator norm

Mixed matrix norms

Compute $\|A\| \quad = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|}$

Mixed matrix norms

Compute $\|A\|_{p,q} = \max_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_q}, \quad 1 < p, q < +\infty$

NP-hard to approximate if $p = q \neq 1, 2, \infty$ [Hendrickx, Olshevsky, 2009]

Mixed matrix norms

Compute $\|A\|_{p,q} = \max_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_q}, \quad 1 < p, q < +\infty$

NP-hard to approximate if $p = q \neq 1, 2, \infty$ [Hendrickx, Olshevsky, 2009]

Observations

1. If $A > 0$ then $\max_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_q} = \max_{x>0} \frac{\|Ax\|_p}{\|x\|_q}$

Mixed matrix norms

Compute $\|A\|_{p,q} = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_q}, \quad 1 < p, q < +\infty$

NP-hard to approximate if $p = q \neq 1, 2, \infty$ [Hendrickx, Olshevsky, 2009]

Observations

1. If $A > 0$ then $\max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_q} = \max_{x > 0} \frac{\|Ax\|_p}{\|x\|_q}$
2. $\nabla_x \left(\frac{\|Ax\|_p}{\|x\|_q} \right) = 0 \iff A^T (Ax)^{p-1} = \sigma \, x^{q-1}$

$$A^T(Ax)^{p-1} = \sigma\, x^{q-1} \iff \begin{cases} (Ax)^{p-1} = \lambda\, y \\ A^T y = \mu\, x^{q-1} \end{cases}$$

$$A^T(Ax)^{p-1} = \sigma\, x^{q-1} \iff \begin{cases} (Ax)^{p-1} = \lambda\, y \\ A^T y = \mu\, x^{q-1} \end{cases}$$

$$\iff \begin{cases} Ax = \widetilde{\lambda}\, y^{\frac{1}{p-1}} \\ A^T y = \mu\, x^{q-1} \end{cases}$$

$$A^T(Ax)^{p-1} = \sigma\, x^{q-1} \iff \begin{cases} (Ax)^{p-1} = \lambda\, y \\ A^T y = \mu\, x^{q-1} \end{cases}$$

$$\iff \begin{cases} Ax = \widetilde{\lambda}\, y^{\frac{1}{p-1}} \\ A^T y = \mu\, x^{q-1} \end{cases}$$

$$\left(\begin{array}{c} u = x^{q-1} \\ v = y^{\frac{1}{p-1}} \end{array} \right) \iff \begin{cases} Au^{\frac{1}{q-1}} = \widetilde{\widetilde{\lambda}}\, v \\ A^T v^{p-1} = \widetilde{\mu}\, u \end{cases}$$

$$A^T(Ax)^{p-1} = \sigma\, x^{q-1} \iff \begin{cases} (Ax)^{p-1} = \lambda\, y \\ A^T y = \mu\, x^{q-1} \end{cases}$$

$$\iff \begin{cases} Ax = \widetilde{\lambda}\, y^{\frac{1}{p-1}} \\ A^T y = \mu\, x^{q-1} \end{cases}$$

$$\left(\begin{array}{c} u = x^{q-1} \\ v = y^{\frac{1}{p-1}} \end{array} \right) \iff \begin{cases} Au^{\frac{1}{q-1}} = \widetilde{\widetilde{\lambda}}\, v \\ A^T v^{p-1} = \widetilde{\mu}\, u \end{cases}$$

Computing $\|A\|_{p,q}$ is equivalent to solving $\begin{cases} Au^{\alpha} = \lambda\, v \\ A^T v^{\beta} = \mu\, u \end{cases}$

Mixed matrix norms: Theorem

Compute $\|A\|_{p,q} = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_q}, \quad 1 < p, q < +\infty$

Theorem

We can compute $\|A\|_{p,q}$ if $q > \kappa(A)^2(p-1)+1$.

Mixed matrix norms: Theorem

$$\text{Compute } \|A\|_{p,q} = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_q}, \qquad 1 < p, q < +\infty$$

Theorem

We can compute $\|A\|_{p,q}$ if $q > \kappa(A)^2(p-1)+1$.

The condition is **necessary and sufficient** for $A_\varepsilon = \begin{bmatrix} \varepsilon & 1 \\ 1 & \varepsilon \end{bmatrix}$, $\varepsilon > 0$

Mixed matrix norms: Theorem

Compute $\|A\|_{p,q} = \max_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_q}, \quad 1 < p, q < +\infty$

Theorem

We can compute $\|A\|_{p,q}$ if $q > \kappa(A)^2(p-1)+1$.

The condition is **necessary and sufficient** for $A_\varepsilon = \begin{bmatrix} \varepsilon & 1 \\ 1 & \varepsilon \end{bmatrix}$, $\varepsilon > 0$

Moreover, if for example $\varepsilon = 3/4$, we have

Classical condition ([1–4]): $(q-1) \geq (p-1)$

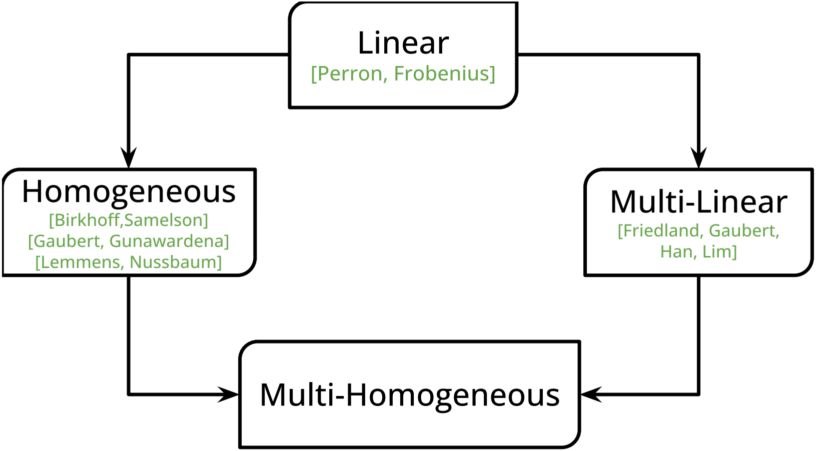
New condition: $(q-1) > 0.0016 \cdot (p-1)$

[1] Boyd ('73), [2] N.Higham ('92), [3] Bhaskara, Vijayaraghavan ('11), [4] Gautier, Hein ('16)

Multihomogeneous PF theorem:

The contractive case

Perron–Frobenius theory



Multihomogeneous operators

The spectral equation

$$\text{Find } (x, y) \text{ such that } \begin{cases} Ag(y) = Ay^{\beta} = \lambda x \\ Bf(x) = Bx^{\alpha} = \mu y \end{cases}$$

is an example of a multihomogeneous spectral problem, just like

$$\text{Find } x \text{ such that } Af(x) = Ax^{\alpha} = \lambda x$$

is an example of a homogeneous spectral problem.

Multihomogeneous operators

Homogeneity

$F : \mathbb{R}^n \to \mathbb{R}^n$ is θ-homogeneous, in symbols $F \in \hom(\theta)$, if

$$F(\lambda x) = \lambda^\theta F(x) \text{ for all } \lambda > 0 \text{ and all } x$$

$\theta \in \mathbb{R}$ is called **homogeneity degree**

Multihomogeneity $X_d := \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$

$F : X_d \to X_d$ is Θ-homogeneous, in symbols $F \in \hom(\Theta)$, if

$$F_i(x_1, \ldots, \lambda x_j, \ldots, x_d) = \lambda^{\Theta_{ij}} F_i(x) \text{ for all } \lambda > 0 \text{ and all } x \in X_d$$

$\Theta \in \mathbb{R}^{d \times d}$ is called **homogeneity matrix**

Examples: Linear and multilinear operators

- $F(x) = Ax \in \hom(1)$

Examples: Linear and multilinear operators

- $F(x) = Ax \in \hom(1)$
- $F(x) = Ax^\alpha \in \hom(\alpha)$

Examples: Linear and multilinear operators

- $F(x) = Ax \in \hom(1)$
- $F(x) = Ax^{\alpha} \in \hom(\alpha)$
- $F(x) = Af(x) \in \hom(\theta)$ provided $f \in \hom(\theta)$

Examples: Linear and multilinear operators

- $F(x) = Ax \in \hom(1)$
- $F(x) = Ax^{\alpha} \in \hom(\alpha)$
- $F(x) = Af(x) \in \hom(\theta)$ provided $f \in \hom(\theta)$
- $F(x,y) = \begin{bmatrix} Ay \\ Bx \end{bmatrix} \in \hom\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right)$

Examples: Linear and multilinear operators

- $F(x) = Ax \in \hom(1)$
- $F(x) = Ax^{\alpha} \in \hom(\alpha)$
- $F(x) = Af(x) \in \hom(\theta)$ provided $f \in \hom(\theta)$
- $F(x, y) = \left[\begin{smallmatrix} Ay \\ Bx \end{smallmatrix}\right] \in \hom\left(\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]\right)$
- $F(x, y) = \left[\begin{smallmatrix} Ay^{\beta} \\ Bx^{\alpha} \end{smallmatrix}\right] \in \hom\left(\left[\begin{smallmatrix} 0 & \beta \\ \alpha & 0 \end{smallmatrix}\right]\right)$

Examples: Linear and multilinear operators

- $F(x) = Ax \in \hom(1)$
- $F(x) = Ax^{\alpha} \in \hom(\alpha)$
- $F(x) = Af(x) \in \hom(\theta)$ provided $f \in \hom(\theta)$
- $F(x,y) = \begin{bmatrix} Ay \\ Bx \end{bmatrix} \in \hom\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right)$
- $F(x,y) = \begin{bmatrix} Ay^{\beta} \\ Bx^{\alpha} \end{bmatrix} \in \hom\left(\begin{bmatrix} 0 & \beta \\ \alpha & 0 \end{bmatrix}\right)$
- $F(x,y) = \begin{bmatrix} Af(x,y) \\ Bg(x,y) \end{bmatrix} \in \hom\left(\begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{bmatrix}\right)$ provided $f \in \hom([\alpha_1, \alpha_2]), g \in \hom([\beta_1, \beta_2])$

Examples: Linear and multilinear operators

If T is a 3rd-order tensor, define $z = Txy$ as $z_i = \sum_{jk} T_{jk} x_j y_k$

- $F(x, y, z) = \begin{bmatrix} T_1 yz \\ T_2 xz \\ T_3 xy \end{bmatrix} \in \hom\left(\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\right)$

Examples: Linear and multilinear operators

If T is a 3rd-order tensor, define $z = Txy$ as $z_i = \sum_{jk} T_{jk} x_j y_k$

- $F(x,y,z) = \begin{bmatrix} T_1 yz \\ T_2 xz \\ T_3 xy \end{bmatrix} \in \hom\left(\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\right)$
- $F(x,y,z) = \begin{bmatrix} T_1 f(x,y,z) \\ T_2 g(x,y,z) \\ T_3 h(x,y,z) \end{bmatrix} \in \hom\left(\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\right)$ provided $f \in \hom([a_1, a_2, a_3]), g \in \hom([b_1, b_2, b_3]), h \in \hom([c_1, c_2, c_3])$

Examples: Linear and multilinear operators

If T is a 3rd-order tensor, define $z = Txy$ as $z_i = \sum_{jk} T_{jk} x_j y_k$

- $F(x, y, z) = \begin{bmatrix} T_1 yz \\ T_2 xz \\ T_3 xy \end{bmatrix} \in \hom\left(\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\right)$
- $F(x, y, z) = \begin{bmatrix} T_1 f(x,y,z) \\ T_2 g(x,y,z) \\ T_3 h(x,y,z) \end{bmatrix} \in \hom\left(\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}\right)$ provided $f \in \hom([a_1, a_2, a_3]), g \in \hom([b_1, b_2, b_3]), h \in \hom([c_1, c_2, c_3])$

General tensor mapping

For a collection of tensors $\boldsymbol{T} = [T_1, \ldots, T_d]$, $x = [x_1, \ldots, x_d] \in X_d$,

$$F(x) = \boldsymbol{T} f(x) \in \hom(\Theta) \text{ provided } f \in \hom(\Theta)$$

Multihomogeneous eigenproblem

We want to consider systems of spectral equations...

Multihomogeneous eigenproblem

We want to consider systems of spectral equations...

$x = (x_1, \dots, x_d) \in X_d$ and $\boldsymbol{\lambda} \in \mathbb{R}^d$ are an eigenpair for the multhom operator $F : X_d \to X_d$ if

$$\begin{cases} F_1(x) = \lambda_1 x_1 \\ F_2(x) = \lambda_2 x_2 \\ \vdots \\ F_d(x) = \lambda_d x_d \end{cases}$$

In symbols

$$F(x) = \boldsymbol{\lambda} \otimes x$$

Multihomogeneous PF theorem (1)

Let $F : X_d \to X_d \in \hom(\Theta)$ be such that

- Continuous
- Positive: $F(x) > 0$ if $x > 0$
- Order-preserving: $F(x) \geq F(y)$ if $x \geq y$
- $\rho(|\Theta|) < 1$, $|\cdot|$ = component wise
- There exists $v > 0$ such that $v^T|\Theta| = \lambda v^T$, $\|v\| = 1$

Then

- there exists a unique $x^* \in S_d$ st $x^* > 0$, $F(x^*) = \boldsymbol{\lambda} \otimes x^*$
- $x^{(k+1)} = G(x^{(k)}) \to x^*$ as $\rho(|\Theta|)^k$

where $G_i(x) = F_i(x)/\|F_i(x)\|_i$, and $S_d = \{x \in X_d : \|x_i\|_i = 1\}$

Multihomogeneous BH theorem

Multihomogeneous BH theorem

A. It applies to maps of the form $F(x) = \boldsymbol{T} f(x) \in \hom(\Theta)$.

Multihomogeneous BH theorem

A. It applies to maps of the form $F(x) = \boldsymbol{T} f(x) \in \hom(\Theta)$.

B. It is based on the mode-j tensor contraction ratio

$$\kappa_j(T) = \tanh\left(\frac{1}{4}\log \triangle_j(T)\right)$$

with $$\triangle_j(T) = \max_{\substack{i_1,\ldots,i_m\\ k_1,\ldots,k_m}} \frac{T_{i_1,\ldots,i_{j-1},i_j,i_{j+1},\ldots,i_m}}{T_{i_1,\ldots,i_{j-1},k_j,i_{j+1},\ldots,i_m}} \frac{T_{k_1,\ldots,k_{j-1},k_j,k_{j+1},\ldots,k_m}}{T_{k_1,\ldots,k_{j-1},i_j,k_{j+1},\ldots,k_m}}$$

Multihomogeneous BH theorem

A. It applies to maps of the form $F(x) = \boldsymbol{T} f(x) \in \hom(\Theta)$.

B. It is based on the mode-j tensor contraction ratio

$$\kappa_j(T) = \tanh\left(\frac{1}{4}\log \triangle_j(T)\right)$$

with $$\triangle_j(T) = \max_{\substack{i_1,\ldots,i_m \\ k_1,\ldots,k_m}} \frac{T_{i_1,\ldots,i_{j-1},i_j,i_{j+1},\ldots,i_m}}{T_{i_1,\ldots,i_{j-1},k_j,i_{j+1},\ldots,i_m}} \frac{T_{k_1,\ldots,k_{j-1},k_j,k_{j+1},\ldots,k_m}}{T_{k_1,\ldots,k_{j-1},i_j,k_{j+1},\ldots,k_m}}$$

C. It states that

There exists $K(\boldsymbol{T},\Theta) \in \mathbb{R}^{d\times d}$ defined in terms of $\kappa_j(T_i)$ and Θ, such that Θ in the PF theorem can be replaced by $K(\boldsymbol{T},\Theta)$:

- if there exists $T_i > 0$, then $\rho(K(\boldsymbol{T},\Theta)) < \rho(\Theta)$
- otherwise, $\kappa_j(T_i) = 1$ for all i, j and $K(\boldsymbol{T},\Theta) = \Theta$

Example: tensor singular vectors (1)

Find (x, y, z) such that $\begin{cases} T_1 y^\beta z^\gamma = \lambda x \\ T_2 x^\alpha z^\gamma = \mu y \\ T_3 x^\alpha y^\beta = \sigma z \end{cases}$

Tensor norm: $\|T\|_{p,q,r} = \max_{x,y\neq 0} \frac{\|Txy\|_p}{\|x\|_q \|y\|_r}$ [Friedland, Gaubert, Han, 2016]

$$K(\boldsymbol{T}, \Theta) = \begin{bmatrix} 0 & \kappa_2(T_1) & \kappa_3(T_1) \\ \kappa_2(T_2) & 0 & \kappa_3(T_2) \\ \kappa_2(T_3) & \kappa_3(T_3) & 0 \end{bmatrix} \begin{bmatrix} |\alpha| & & \\ & |\beta| & \\ & & |\gamma| \end{bmatrix}$$

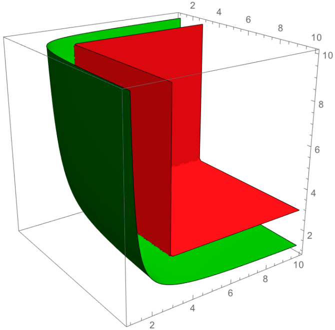
Comparison with Gaubert and Friedland

Theorem [Friedland, Gaubert, Han], [Lim]

If T is "weakly irreducible" and $p, q, r \geq 3$ then there exist unique $(x, y, z) > 0$ that realize that maximum

$\rho(\Theta) < 1$

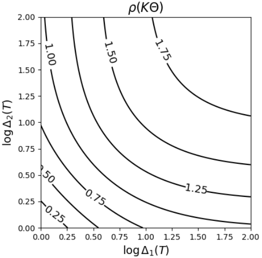
$p, q, r \geq 3$



Example: tensor singular vectors (2)

Find (x, y) such that $\begin{cases} Tx^{\alpha}y^{\beta} = \lambda x \\ Sx^{\alpha}x^{\alpha} = \mu y \end{cases}$

$$K(\boldsymbol{T}, \Theta) = \begin{bmatrix} \kappa_2(T_1) & \kappa_3(T_1) \\ \kappa_2(T_2) + \kappa_3(T_2) & 0 \end{bmatrix} \begin{bmatrix} |\alpha| & \\ & |\beta| \end{bmatrix}$$



Multihomogeneous PF theorem:

Non-expansive case

Matrix Perron–Frobenius theorem

Let $A \in \mathbb{R}^{n\times n}$ be such that $A \geq 0$. Then:

- There exists $x \geq 0$ such that $Ax = \lambda x$
- $\lambda \geq 0$ and $\lambda = \rho(A)$

if additionally, A is irreducible, then

- $\rho(A) > 0$ and x is unique

if additionally, A is primitive (aperiodic), then

- $x^{(k+1)} = G(x^{(k)}) = Ax^{(k)}/\|Ax^{(k)}\| \longrightarrow x$, as $k \to \infty$

Linear operators $F(x) = Ax$ are continuous and $\mathrm{hom}(1)$.
Moreover, F is positive and order-preserving **iff** $A \geq 0$.

All the above "irreducibility assumptions" are not needed if
$F \in \mathrm{hom}(\Theta)$ and $\rho(\Theta) < 1$ (or $\rho(\boldsymbol{T}, \Theta) < 1$)

Spectral radius

Observation

Let $F(x) \in \hom(\theta)$. Then $F(x) = \lambda x$ iff $F(\alpha x) = (\alpha^{\theta-1}\lambda)\alpha x$.
$\Longrightarrow$ Each eigenvector has infinitely many eigenvalues.

The notion of spectral radius makes sense only for $\theta = 1$

For multihomogeneous operators we will need $\Theta \geq 0$ and $\rho(\Theta) = 1$

Spectral radius via Gelfand formula

Definition *Homogeneous spectral radius*

Let $F \in \hom(1)$ be continuous and order-preserving. Define

$$\rho(F) = \lim_{k\to\infty} \|F^k\|_+^{1/k} := \lim_{k\to\infty} \sup_{x\geq 0, x\neq 0} \left(\frac{\|F^k(x)\|}{\|x\|}\right)^{1/k}$$

Irreducible and primitive matrices

There are different ways to define irreducibility / primitivity of a nonnegative matrix $A \in \mathbb{R}^{n\times n}$.

Irreducible and primitive matrices

There are different ways to define irreducibility / primitivity of a nonnegative matrix $A \in \mathbb{R}^{n \times n}$.

Irreducible iff

$\sum_{k=0}^{n} A^k x > 0$ for all $x \geq 0$, $x \neq 0$

Primitive iff

There exists k st $A^k x > 0$ for all $x \geq 0$, $x \neq 0$

Remark: primitive $\Longrightarrow$ irreducible

Homogeneous PF theorem

$F \in \mathrm{hom}(1)$ continuous, order-preserving is **irreducible** iff $\sum_{k=0}^{n} F^k(x) > 0$ for all $x \geq 0$, $x \neq 0$.

Homogeneous PF theorem

$F \in \hom(1)$ continuous, order-preserving is **irreducible** iff $\sum_{k=0}^{n} F^k(x) > 0$ for all $x \geq 0$, $x \neq 0$.

Theorem [Nussbaum, Eveson, Lemmens, Gaubert, ...]

Let $F \in \hom(1)$ continuous, order-preserving, **irreducible**. Then,

- there exists $x^* > 0$ st $F(x^*) = \lambda x^*$
- $\lambda = \rho(F) > 0$ and $\rho(F) = \max\{|\mu| : F(x) = \mu x\}$

if moreover the Jacobian $F'(x^*)$ is irreducible, then

- x^* is the unique positive eigenvector of F

if moreover the Jacobian $F'(x^*)$ is primitive, then

- $x^{(k+1)} = G(x^{(k)}) = F(x^{(k)})/\|F(x^{(k)})\| \longrightarrow x$, as $k \to \infty$

Note: $F \in \hom(1)$ is order-preserving iff $F'(x) \geq 0$ for all $x \geq 0$

Example: Matrix-inheritance

$F(x) = g(Af(x)) \in \mathrm{hom}(1)$ with $f \in \mathrm{hom}(\theta), g \in \mathrm{hom}(\theta^{-1})$.

For example, $F(x) = (Ax^{\alpha})^{1/\alpha}$.

Then

If A is irreducible, then

- there exists $x^* > 0$ st $F(x^*) = \lambda x^*$
- $\lambda = \rho(F) > 0$ and $\rho(F) = \max\{|\mu| : F(x) = \mu x\}$
- x^* is the unique positive eigenvector of F

If moreover A is primitive, then

- $x^{(k+1)} = G(x^{(k)}) = F(x^{(k)})/\|F(x^{(k)})\| \longrightarrow x^*$, as $k \to \infty$

Multihomogeneous spectral radius

Definition *Multihomogeneous spectral radius*

Let $F : X_d \to X_d \in \hom(\Theta)$ be continuous and order-preserving. Assume that

- $\Theta \geq 0$ and $\rho(\Theta) = 1$
- there exists $v > 0$ such that $v^T\Theta = \lambda v^T$, $\|v\| = 1$

Define

$$\rho_v(F) = \lim_{k\to\infty} \|F^k\|_{+,d}^{1/k}$$
$$= \lim_{k\to\infty} \sup_{\substack{x=(x_1,\ldots,x_d)\geq 0\\ x\neq 0}} \left(\frac{\|F_1^k(x)\|^{v_1}\cdots\|F_d^k(x)\|^{v_d}}{\|x_1\|^{v_1}\cdots\|x_d\|^{v_d}}\right)^{1/k}$$

Multihomogeneous PF theorem

Theorem

Let $F \in \hom(\Theta)$ continuous, order-preserving, **irreducible**.
Assume $\Theta \geq 0$, $\rho(\Theta) = 1$ and $v^T \Theta = \lambda v^T$ with $v > 0$, $\|v\| = 1$.
For $\boldsymbol{\lambda} \in \mathbb{R}^d$, define

$$|||\boldsymbol{\lambda}|||_v = |\lambda_1|^{v_1} \cdots |\lambda_d|^{v_d} = \exp \| \log(|\boldsymbol{\lambda}|)\|_{1,v}$$

Multihomogeneous PF theorem

Theorem

Let $F \in \hom(\Theta)$ continuous, order-preserving, **irreducible**.
Assume $\Theta \geq 0$, $\rho(\Theta) = 1$ and $v^T \Theta = \lambda v^T$ with $v > 0$, $\|v\| = 1$.
For $\boldsymbol{\lambda} \in \mathbb{R}^d$, define

$$|||\boldsymbol{\lambda}|||_v = |\lambda_1|^{v_1} \cdots |\lambda_d|^{v_d} = \exp \| \log(|\boldsymbol{\lambda}|)\|_{1,v}$$

Then, for $S_d = \{x \in X_d : \|x_1\|_1 = \cdots = \|x_d\|_d = 1\}$, it holds

- there exist $x^* \in S_d$, $x^* > 0$ and $\boldsymbol{\lambda} \in \mathbb{R}^d$ st $F(x^*) = \boldsymbol{\lambda} \otimes x^*$
- $\boldsymbol{\lambda} > 0$ and $|||\boldsymbol{\lambda}|||_v = \rho_v(F) = \max\{|||\boldsymbol{\mu}|||_v : F(x) = \boldsymbol{\mu} \otimes x\}$

if moreover the Jacobian $F'(x^*)$ is irreducible, then

- $x^* \in S_d$ is the unique positive eigenvector of F in X_d

if moreover the Jacobian $F'(x^*)$ is primitive, then

- $x^{(k+1)} = G(x^{(k)}) = \left(\frac{F_1(x^{(k)})}{\|F_1(x^{(k)})\|_1}, \ldots, \frac{F_d(x^{(k)})}{\|F_d(x^{(k)})\|_d} \right) \longrightarrow x^* \in S_d$

Example applications:

Constrained homogeneous optimization

Matrix singular vectors, again

Consider the constrained optimization problem

$$\begin{cases} \text{optimize} & x_1^T A x_2 \\ \text{subject to} & \|x_1\|_2 = \|x_2\|_2 = 1 \end{cases}$$

In general, the function $f : X_2 \to \mathbb{R}$, $f(x) = x_1^T A x_2$ is not convex.
However, we know how to compute global max and global min:

Singular values and singular vectors of A

Homogeneous singular vectors

For sufficiently smooth homogeneous functions $f : X_d \to \mathbb{R}$, $g_i \in \mathbb{R}^{n_i} \to \mathbb{R}$

$$\begin{cases} \text{optimize} & f(x) \\ \text{subject to} & g_1(x_1) = \cdots = g_d(x_d) = 1 \end{cases}$$

can be transformed into $F(x) = \boldsymbol{\lambda} \otimes x$ for a multihomogeneous F

However, global max/min can be NP-hard

Example: graph clustering

- $A =$ adjacency matrix of a graph with n nodes
- $f(x) = \frac{1}{2}\sum_{i=1}^{n} A_{ij}|x_i - x_j|$ graph total variation
- $g(x) = \|x - \mathrm{mean}(x)1\|_1$

Then

- Graph clustering $\leftrightarrow \min f(x)$ st $g(x) = 1$
- Modularity maximization $\leftrightarrow \max f(x)$ st $g(x) = 1$

[T., Zhang, *Nonlinear Spectral Duality*]
[T., Mercado, Hein, *Nonlinear modularity eigenvectors*]

Positive news

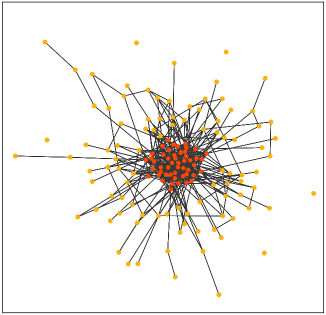
When f has nonegativity / order-preserving properties, we can solve these optimization problems globally

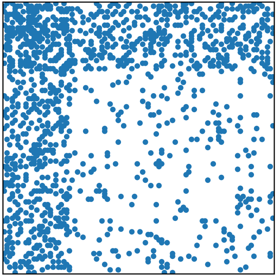
Two examples:

- Core-periphery detection
- Semi-supervised classification

Core–periphery classification

Core: nodes strongly connected across the whole network
Periphery: nodes strongly connected only to the core





Core-periphery score

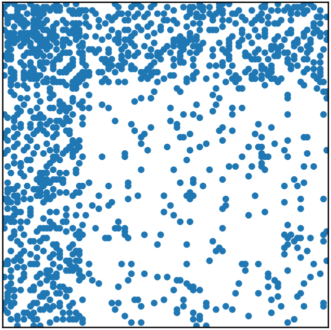
$$\begin{cases} \max f(x) := \sum_{ij} A_{ij}(|x_i|^\alpha + |x_j|^\alpha)^{1/\alpha} \\ \text{subject to } \|x\|_p = 1 \text{ and } x \geq 0 \end{cases} \qquad (\alpha \text{ large})$$

Coreness score: "$x_i > x_j$ if i is more in the core than j"

T., D. Higham, *SIMODS*, 2019

C. Higham, D. Higham, T., *KDD*, 2022

Core-periphery kernel



α large $\Rightarrow (x^\alpha + y^\alpha)^{1/\alpha} \approx \max\{x, y\}$

For a positive $x > 0$,
$f(x) = \sum_{ij} A_{ij}(x_i^\alpha + x_j^\alpha)^{1/\alpha}$ is large when edges $A_{ij} = 1$ involve at least one node with large core-score

Connection with node degrees

If $p = 2$ and $\alpha = 1$ then *(arithmetic mean)*

$$\max f(x) \text{ st } \|x\|_p = 1 \iff \max_{x \geq 0} \frac{\|Ax\|_1}{\|x\|_2} = \|A\|_{2\to 1}$$

and the maximizer is

$$x = \text{degree vector}$$

Connection with eigenvector centrality

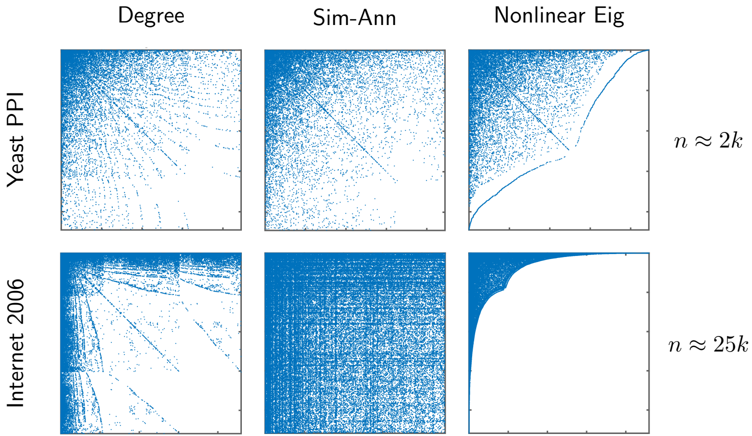
If $p = 1$ and $\alpha = 0$ then *(geometric mean)*

$$\max f(x) \text{ st } \|x\|_p = 1 \iff \max_{x \geq 0} \frac{x^T A x}{x^T x} = \rho(A)$$

and the maximizer is

$$x = \text{Perron eigenvector of } A$$

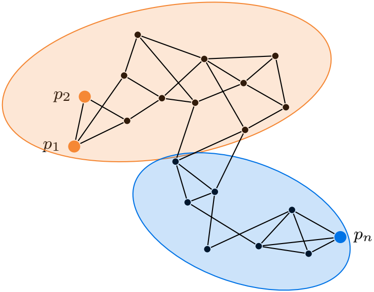
Qualitative results



Degree coincides with NEig for $\alpha = 1$ and $p = 2$

Convergence in a **few seconds** vs **several minutes** with Sim-Ann.

Semi-supervised graph clustering

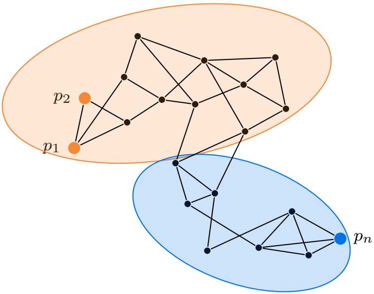


We are given points/nodes $\{p_i\}$; we know that they belong to K classes; and we know the class of **some of them**

Goal: assign classes to remaining points

Setting (cont.)

$$Y_{ij} = \begin{cases} 1 & \text{if } i \in \text{class } j \\ 0 & \text{otherwise} \end{cases}$$



$$Y = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & \vdots \\ \vdots & 0 \\ 0 & 1 \end{bmatrix}$$

Variance minimization

$$\begin{cases} \min f(X) := \left\| \frac{Y}{g(Y)} - X \right\|_F + \mu \sum_{e \in E} \left\{ A_e \sum_{i \in e} \left\| x_i - \left(\frac{\sum_{j \in e} x_j^p}{|e|} \right)^{1/p} \right\|_q \right\} \\ \text{subject to } g(X) = 1 \end{cases}$$

Example: $g(X) = \|X\|_F$

- [Flores, Calder, Lerman, Applied Comp Harmonic Analysis, 2022]
- [Slepcev, Thorpe, SIAP 2019]
- [Prokopchik, Benson, T., *ICML*, 2022]

Example applications:

Entropy minimization and optimal transport

Wasserstein distance

- $P(n) = \{x \in \mathbb{R}^n : x \geq 0, \mathrm{sum}(x) = 1\}$
- $a \in P(n_1)$, $b \in P(n_2)$, let $U(a,b) = \{M \geq 0 : M1 = 1, M^T 1 = b\}$
- $C \in \mathbb{R}^{n_1 \times n_2}$ weight matrix

$$W(a,b) = \min_{M \in U(a,b)} \sum_{ij} M_{ij} C_{ij}$$

Sand pile problem:

- n_1 piles of sand, with a_i units of sand each.
- n_2 target locations, each should receive b_j units
- $M \in U(a,b)$ represents a possible solution: move M_{ij} units from i to j, so that all initial sand $M1 = a$ is moved to the targets $M^T 1 = b$
- C_{ij} = cost of moving one unit of sand from i to j

Multimarginal optimal transport

- $U(a,b,c) = \{T \geq 0 \,:\, \sum_{jk} T_{ijk} = a_i, \sum_{ik} T_{ijk} = b_j, \sum_{ij} T_{ijk} = c_k\}$
- $C \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ weight tensor

$$W(a,b,c) = \min_{T \in U(a,b,c)} \sum_{ijk} T_{ijk} C_{ijk}$$

Sand pile problem:

- n_3 transportation means, with capacity c_k
- C_{ijk} = cost of moving one unit of sand from i to j by means of k

Computing $W(a,b,c)$ can be very expensive...

Entropy regularization

$$\Phi_\varepsilon(T) = \langle T, C\rangle - \varepsilon E(T) = \varepsilon KL(T|\exp(-C/\varepsilon))$$

- $E(T) = -\sum_{ijk} T_{ijk} \log T_{ijk}$ entropy
- $\mathrm{KL}(T|S) = \sum_{ijk} T_{ijk} \log(T_{ijk}/S_{ijk})$ Kullback–Leibler divergence

It can be shown that

$$\min_{T\in U(a,b,c)} \Phi_\varepsilon(T) \longrightarrow W(a,b,c) \text{ as } \varepsilon \to 0$$

From divergence to eigenvectors

Let $L(T, \alpha, \beta, \gamma)$ be the Lagrangian of Φ_ε. Imposing the first-order conditions $\partial L / \partial T_{ijk} = 0$ yields

$$\log T_{ijk} = \frac{1}{\varepsilon}\Big(\alpha_i + \beta_j + \gamma_k - C_{ijk}\Big)$$

Thus, taking the exponent, we can rewrite the multimarginal opt transport problem as finding $(x, y, z) = (\exp(\alpha/\varepsilon), \exp(\beta/\varepsilon), \exp(\gamma/\varepsilon))$ such that

$$T_{ijk} = \exp(-C_{ijk}/\varepsilon) x_i y_j z_k \in U(a, b, c)$$

which in turn is equivalent to

$$\begin{cases} \sum_{jk} a_i^{-1} \exp(-C_{ijk}/\varepsilon) y_j z_k = x_i^{-1} \\ \sum_{ik} b_j^{-1} \exp(-C_{ijk}/\varepsilon) x_i z_k = y_j^{-1} \\ \sum_{ij} c_k^{-1} \exp(-C_{ijk}/\varepsilon) x_i y_j = z_k^{-1} \end{cases}$$

Comments and open questions

- Everything can be generalized to “any” cone
- Better convergence bounds?
    - Use eigenvalues of A somewhere
    - Use different distances (e.g. Ergodicity coefficients)
- Speed up power method?

Some references

[1] A. Gautier, F. Tudisco, M. Hein, *Nonlinear Perron–Frobenius theorems for nonnegative tensors*, SIAM Review, 2023

[2] F. Tudisco, D. Zhang, *Nonlinear spectral duality*, arxiv:2209.06241, 2022

[3] F. Tudisco, A. R. Benson, K. Prokopchik, *Nonlinear higher-order label spreading*, WWW, 2021

[4] A. Gautier, F. Tudisco, M. Hein, *The Perron–Frobenius theorem for multhomogeneous mappings*, SIAM J. Matrix Analysis and Applications, 2019

[5] F. Tudisco, D. J. Higham, *A nonlinear spectral method for core-periphery detection in networks*, SIAM J. Mathematics of Data Science, 2019

[6] D. Fasino, F. Tudisco, *Ergodicity coefficients for higher-order stochastic processes*, SIAM J. Mathematics of Data Science, 2020