Optimization with nonlinear Perron eigenvectors

Francesco Tudisco
Gran Sasso Science Institute (Italy)

SIAM LA21 Minitutorial
Applied Nonlinear Perron-Frobenius Theory

May 18, 2021

Outline

I will present a Perron-Frobenius type result for nonlinear operators. This result is stated as a global optimization algorithm for a class of constrained opt problems.

1. Motivation: graph-based unsupervised learning aka graph partitioning
2. Perron–Frobenius theorem for (sub) multihomogeneous maps
3. Some example applications

Matrix singular vectors, once again

Consider the constrained optimization problem

$$\begin{cases} \text{optimize} & \mathbf{x}_1^\top M \mathbf{x}_2 \\ \text{subject to} & \|\mathbf{x}_1\| = \|\mathbf{x}_2\| = 1 \end{cases}$$

In general, $f(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^\top M \mathbf{x}_2$ is not convex. However, we know how to compute global max and global min:

singular vectors and singular values of M:

$$M_i \mathbf{x}_i = \lambda \mathbf{x}_i, \qquad i = 1, 2$$

with $M_1 = MM^\top$, $M_2 = M^\top M$.

Nonlinear singular vectors

For *sufficiently smooth* homogeneous functions f, g, the problem

$$\begin{cases} \text{optimize} & f(\boldsymbol{x}) \\ \text{subject to} & g(\boldsymbol{x}_1) = g(\boldsymbol{x}_2) = 1 \end{cases}$$

with $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, can be brought down to

nonlinear singular vector problem

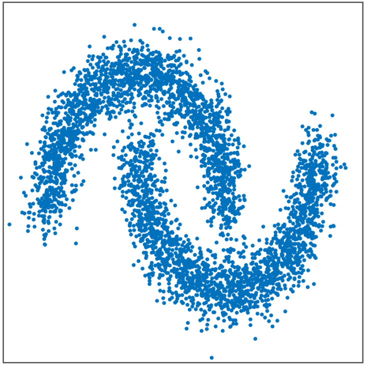
$$M_i(\boldsymbol{x})\boldsymbol{x}_i = \lambda \boldsymbol{x}_i, \quad i = 1, 2$$

where $M_i(\boldsymbol{x})$ are matrix-valued mappings, obtained differentiating f twice.

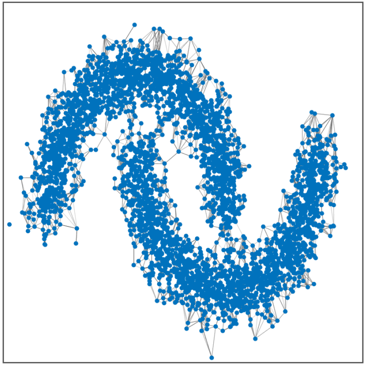
However, global max/min can be NP-hard...

Motivating example: graph clustering

Graph clustering

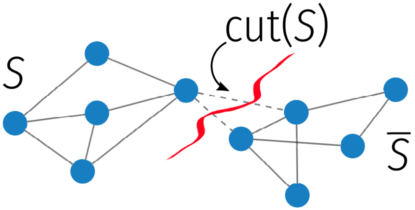


Graph clustering



$$\mathcal{G} = (V, E),\ V = \{1, \ldots, n\},\ E \subseteq V \times V$$

Balanced cut problem



$$\gamma(\mathcal{G}) = \min_{S \subseteq V} \frac{\mathrm{cut}(S)}{\min\{|S|, |\overline{S}|\}}$$

$$\mathrm{cut}(S) = \{ij \in E : i \in S, j \in \overline{S}\}, \qquad \overline{S} = V \setminus S$$

General ratio-of-set-functions problem

$\gamma(\mathcal{G}) = \min_S \Phi(S)/\Psi(S)$ is the global mimimum of the ratio of two set functions $\Phi(S) = \mathrm{cut}(S)$, $\Psi(S) = \min\{|S|, |\overline{S}|\}$ such that:

1. Φ, Ψ are nonnegative
2. $\Phi(V) = \Psi(V) = 0$

General ratio-of-set-functions problem

$\gamma(\mathcal{G}) = \min_S \Phi(S)/\Psi(S)$ is the global mimimum of the ratio of two set functions $\Phi(S) = \mathrm{cut}(S)$, $\Psi(S) = \min\{|S|, |\overline{S}|\}$ such that:

1. Φ, Ψ are nonnegative
2. $\Phi(V) = \Psi(V) = 0$

In general, consider the problem

$$\min_{S \subseteq V} \vartheta(S), \qquad \vartheta(S) = \frac{\Phi(S)}{\Psi(S)}$$

with $\Phi, \Psi : 2^V \to \mathbb{R}$ such that 1 and 2 hold

Homogeneous exact relaxation

Computing $\min \vartheta$ is in general NP-hard .... can we approximate it?

Homogeneous exact relaxation

Computing $\min \vartheta$ is in general NP-hard .... can we approximate it?

Theorem.

Given Φ, Ψ and $p \geq 1$, there exist homogeneous functions $f, g : \mathbb{R}^n \to \mathbb{R}$ of degree p such that, if λ is a solution to

$$\lambda = \begin{cases} \min_{\boldsymbol{x} \in \mathbb{R}^n} & f(\boldsymbol{x}) \\ \text{subject to} & g(\boldsymbol{x}) = 1, \end{cases}$$

then $\lambda \leq \min \vartheta \leq C^{p-1} \lambda^{1/p}$, in particular $\lambda \xrightarrow{p \to 1} \min \vartheta$.

Proof's sketch (for $p = 1$)

Consider the Lovasz extensions f, g of the functions Φ and Ψ.

1. $f(\mathbb{1}_S) = \Phi(S)$, $g(\mathbb{1}_S) = \Psi(S)$

$$\min_{S \subseteq V} \frac{\Phi(S)}{\Psi(S)} \geq \min_{\boldsymbol{x} \in \mathbb{R}^n} \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} = \min_{\boldsymbol{x} \in \mathbb{R}^n} f\Big(\frac{\boldsymbol{x}}{g(\boldsymbol{x})}\Big)$$

2. $f(\boldsymbol{x}) = \sum_{i=0}^{n-1} \Phi(S_{x_i})|x_{i+1} - x_i| = \int_{-\infty}^{+\infty} \Phi(S_t)dt$ where $S_t = \{k : x_k > t\}$.

$$\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} = \frac{\int_{-\infty}^{+\infty} \Phi(S_t)dt}{\int_{-\infty}^{+\infty} \Psi(S_t)dt} \geq \inf_t \frac{\Phi(S_t)}{\Psi(S_t)} \geq \min_{S \subseteq V} \frac{\Phi(S)}{\Psi(S)}$$

Based on

Hein, Setzer, NeurIPS 2012

Back to the clustering problem

$\Phi(S) = \mathrm{cut}(S)$, $\Psi(S) = \min\{|S|, |\overline{S}|\}$. The homogeneous functions f, g are:

$$f(\boldsymbol{x}) = \frac{1}{2}\sum_{ij=1}^{n} A_{ij}|x_i - x_j|^p \qquad g(\boldsymbol{x}) = \|\boldsymbol{x} - \mathrm{mean}(\boldsymbol{x})\mathbb{1}\|_p^p = 1$$

and we have that

$$\lambda = \begin{cases} \min_{\boldsymbol{x}\in\mathbb{R}^n} & f(\boldsymbol{x}) \\ \text{subject to} & g(\boldsymbol{x}) = 1 \end{cases} \iff \begin{array}{c} \lambda = \text{smallest nonzero sol of:} \\ M_p(\boldsymbol{x})\boldsymbol{x} = \lambda \boldsymbol{x} \end{array}$$

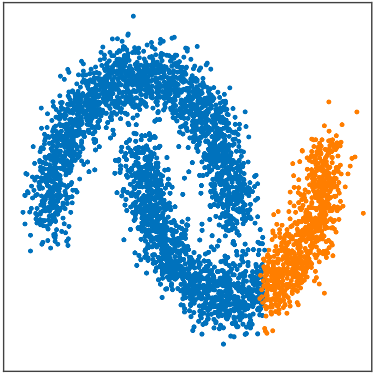
where M_p is a matrix-valued mapping, based on the graph p-Laplacian

Cheeger inequality

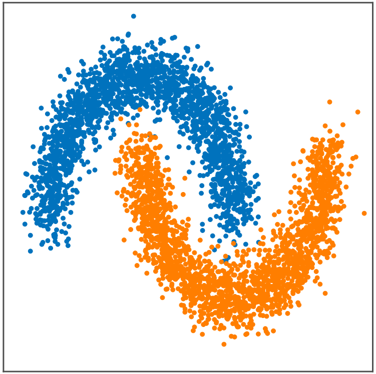
For $p = 2$ we obain the famous Cheeger inequality

- $M_2 = L = \mathrm{diag}(A\mathbb{1}) - A$ = Graph Laplacian Matrix
- λ = Fiedler eigenvalue (or algebraic connectivity)
- $\min \vartheta = \gamma(\mathcal{G})$ = graph Cheeger constant
- $\lambda \leq \gamma(\mathcal{G}) \leq C\sqrt{\lambda}$

Linear (vs) nonlinear spectral clustering



$p = 2$



$p = 1$

Drawback

f and g are nonlinear and nonconvex in general, solving

$$\lambda = \begin{cases} \min_{\boldsymbol{x} \in \mathbb{R}^n} & f(\boldsymbol{x}) \\ \text{subject to} & g(\boldsymbol{x}) = 1, \end{cases}$$

can be very challenging.

E.g. think at the result $\lambda \xrightarrow{p \to 1}$ Cheeger constant

Drawback

f and g are nonlinear and nonconvex in general, solving

$$\lambda = \begin{cases} \min_{\boldsymbol{x}\in\mathbb{R}^n} & f(\boldsymbol{x}) \\ \text{subject to} & g(\boldsymbol{x}) = 1, \end{cases}$$

can be very challenging.

E.g. think at the result $\lambda \xrightarrow{p\to 1}$ Cheeger constant

However, we can compute λ to an arbitrary accuracy when f and g are **nonnegative** and **sub-multihomogeneous**.

Perron–Frobenius theorem for sub-multihomogeneous mappings

Setting

We are going to consider the

optimization of $f : \mathbb{R}^n \to \mathbb{R}$ such that
the gradient of f is **sub-multihomogeneous**

To this end, we first introduce this concept.

Multihomogeneous mappings (two variables)

Suppose $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^n$ with $\boldsymbol{x}_1 \in \mathbb{R}^{n_1}$, $\boldsymbol{x}_2 \in \mathbb{R}^{n_2}$.

Partition the gradient of $f : \mathbb{R}^n \to \mathbb{R}$ as:

$$\partial f = \left[\begin{array}{c} \partial_1 f \\ \hline \partial_2 f \end{array}\right], \quad \partial_i f = \partial_{\boldsymbol{x}_i} f = \text{partial derivative w.r.t. variables in } \boldsymbol{x}_i$$

If there exists a 2×2 matrix Θ such that

$$\begin{cases} \partial_1 f(\lambda \boldsymbol{x}_1, \boldsymbol{x}_2) = \Theta_{11} \partial_1 f(\boldsymbol{x}) & \partial_1 f(\boldsymbol{x}_1, \lambda \boldsymbol{x}_2) = \Theta_{12} \partial_1 f(\boldsymbol{x}) \\ \partial_2 f(\lambda \boldsymbol{x}_1, \boldsymbol{x}_2) = \Theta_{21} \partial_2 f(\boldsymbol{x}) & \partial_2 f(\boldsymbol{x}_1, \lambda \boldsymbol{x}_2) = \Theta_{22} \partial_2 f(\boldsymbol{x}) \end{cases}$$

then ∂f is multihomogeneous. We write this compactly as $f \in \mathrm{hom}'(\Theta)$.

Multihomogeneous functions have multihomogeneous gradient

An example of $f \in \mathrm{hom}'(\Theta)$ are multihomogeneous functions.
In fact, if f is such that

$$f(\boldsymbol{x}_1, \dots, \lambda \boldsymbol{x}_j, \dots, \boldsymbol{x}_m) = \lambda^{\delta_j} f(\boldsymbol{x})$$

then it is easy to veryfy that

$$\partial_j f(\boldsymbol{x}_1, \dots, \lambda \boldsymbol{x}_j, \dots, \boldsymbol{x}_m) = \lambda^{\Theta_{ij}} \partial_j f(\boldsymbol{x})$$

with

$$\Theta = \begin{bmatrix} \delta_1 - 1 & \delta_2 & \dots & \delta_s \\ \delta_1 & \delta_2 - 1 & \dots & \delta_s \\ \vdots & & \ddots & \vdots \\ \delta_1 & \dots & \delta_{s-1} & \delta_s - 1 \end{bmatrix} = \mathbb{1}\delta^\top - I\,.$$

Euler's characterization (two-variables)

Suppose f is twice differentiable.

Then we can partition the Hessian of f accordingly

$$\partial^2 f = \left[\begin{array}{c|c} \partial_1\partial_1 f & \partial_2\partial_1 f \\ \hline \partial_1\partial_2 f & \partial_2\partial_2 f \end{array}\right]$$

Euler's theorem applied block-wise gives us

$$f \in \hom'(\Theta) \iff \begin{cases} \partial_1\partial_1 f(\boldsymbol{x})x_1 = \Theta_{11}\partial_1 f(\boldsymbol{x}) & \partial_2\partial_1 f(\boldsymbol{x})x_2 = \Theta_{12}\partial_1 f(\boldsymbol{x}) \\ \partial_1\partial_2 f(\boldsymbol{x})x_1 = \Theta_{21}\partial_2 f(\boldsymbol{x}) & \partial_2\partial_2 f(\boldsymbol{x})x_2 = \Theta_{22}\partial_2 f(\boldsymbol{x}) \end{cases}$$

for all $\boldsymbol{x} \succ 0$

Euler's characterization of multihomogeneous mappings

Definition.

The gradient of $f : \mathbb{R}^n \to \mathbb{R}$ is multihomogeneous if for some m there exists a partition of the variable $\boldsymbol{x} \in \mathbb{R}^n$

$$\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_m), \qquad \boldsymbol{x}_i \in \mathbb{R}^{n_i}, \textstyle\sum_i n_i = m$$

and a matrix $\Theta \in \mathbb{R}^{m \times m}$ such that

$$\partial_i \partial_j f(\boldsymbol{x}) \boldsymbol{x}_i = \Theta_{ij} \boldsymbol{x}_i$$

for all $i, j = 1, \dots, m$ and all positive vectors $\boldsymbol{x} \succ 0$

Sub-multihomogeneity

$f : \mathbb{R}^n \to \mathbb{R}$ twice differentiable is sub-multihomogeneous if there exists a partition $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ and a matrix Θ such that

$$|\Theta_{ij}| = \min\left\{\lambda \geq 0 \;:\; |\partial_i \partial_j f(\boldsymbol{x})|\boldsymbol{x}_i \leq \lambda \left|\partial_i f(\boldsymbol{x})\right|\right\}$$

$$\text{for all } i, j = 1, \ldots, m \text{ and } \boldsymbol{x} \succ 0$$

where $|\cdot|$ denotes absolute value taken component-wise.

We write this compactly $f \in \text{subhom}'(\Theta)$.

Global optimization with nonlinear Perron eigenvectors

Consider the problem

$$(*) \qquad \begin{cases} \max_{\boldsymbol{x}\in\mathbb{R}^n} & f(\boldsymbol{x}_1,\dots,\boldsymbol{x}_m) \\ \text{subject to} & g_1(\boldsymbol{x}_1)=\cdots=g_m(\boldsymbol{x}_m)=1, \end{cases}$$

where:

- $f \in \mathrm{subhom}'(\Theta)$
- $g_i \in \mathrm{hom}(1+\alpha_i)$, $\alpha_i \neq 0$
- ∂g_i is invertible on $\mathbb{R}^{n_i}_{++}$
- Both $(\partial g_i)^{-1} : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$ and $\partial_i f : \mathbb{R}^n \to \mathbb{R}^{n_i}$ are positive mappings, i.e. map positive vectors into positive vector

Perron–Frobenius theorem

Let $B = \mathrm{Diag}(\alpha_1, \dots, \alpha_m)^{-1}\Theta$.
If $\rho(|B|) < 1/2$ or $\rho(|B|) < 1$ and $\partial^2 f$ is a positive map, then

- There exists a unique solution $\boldsymbol{x}^* \in \mathbb{R}^n$ to $(*)$ and $\boldsymbol{x}^* \succ 0$
- $\boldsymbol{x}_i^*$ are nonlinear singular vectors, solution to

  $$M(\boldsymbol{x}^*)\boldsymbol{x}_i^* = \lambda_i \boldsymbol{x}_i^*$$

  corresponding to the largest nonlinear sinuglar values $(\lambda_1, \dots, \lambda_m)$
- The nonlinear power method

  $$\begin{cases} \boldsymbol{y} = M(\boldsymbol{x}^{(k)})\boldsymbol{x}_i^{(k)} = (\partial g_i)^{-1} \circ \partial_i f(\boldsymbol{x}^{(k)}) \\ \boldsymbol{x}^{(k+1)} = \begin{bmatrix} \frac{\boldsymbol{y}_1}{g_1(\boldsymbol{y}_1)} & \dots & \frac{\boldsymbol{y}_m}{g_m(\boldsymbol{y}_m)} \end{bmatrix} \end{cases} \qquad k = 0, 1, 2, \dots$$

  converges to $\boldsymbol{x}^*$ as $O(\rho(|B|)^k)$, for any $\boldsymbol{x}_0 \succ 0$.

Sketch of the proof

Let $H : \mathbb{R}^n \to \mathbb{R}^n$ be the iterator of the nonlinear power method

1. $\boldsymbol{x}^*$ is solution of $(*)$ if and only if $H(\boldsymbol{x}^*) = \boldsymbol{x}^*$
2. $H(\mathcal{K}) \subseteq \mathcal{K}$ where $\mathcal{K} = \{\boldsymbol{x} \succ 0 : g_i(\boldsymbol{x}_i) = 1, \forall i\}$
3. $\mathcal{K}$ is a complete metric space with respect to the Thompson metric δ_T
4. $\delta_T(H(\boldsymbol{x}), H(\boldsymbol{y})) \leq C\, \delta_T(\boldsymbol{x}, \boldsymbol{y})$ with

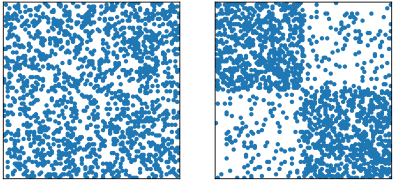
$$C = \sup_{\boldsymbol{x} \in \mathcal{K}} \left\| \mathrm{diag}(H(\boldsymbol{x}))^{-1} |M(\boldsymbol{x})| \boldsymbol{x} \right\|_\infty$$

5. $C \leq \rho(|B|)$

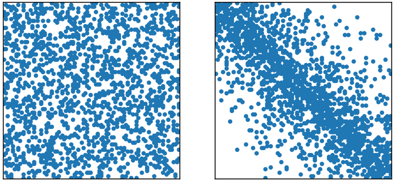
Core–Periphery detection in networks

Matrix reordering problems

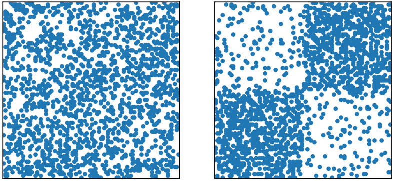
Clusters (communities)



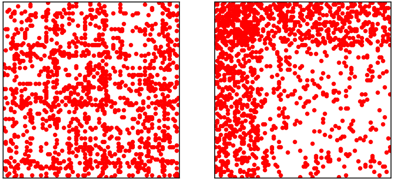
Lattice (small–world)



Bipartite (anti–communities)



Core–periphery



Core–periphery

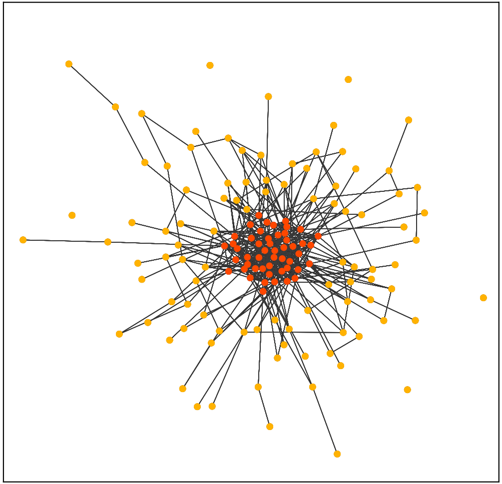
Borgatti, Everett, *Social Networks*, 1999

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

Csermely, London, Wu, Uzzi, *J. of Complex Networks*, 2013

Rombach, Porter, Fowler, Mucha, *SIAM Review*, 2018

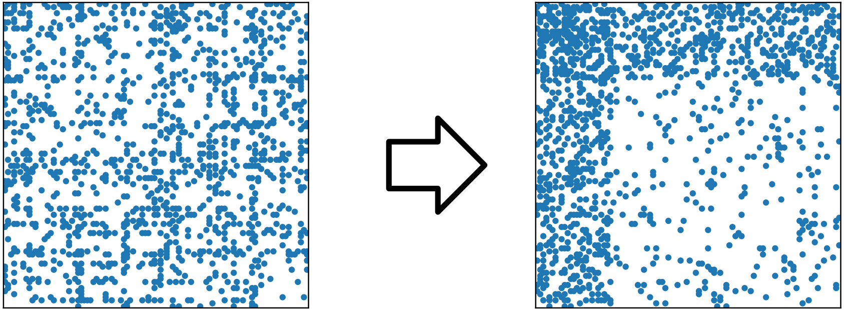
Core–periphery visualization



Core–periphery detection problem

Tasks:

1. **Reorder** nodes to reveal core–periphery structure
2. assign **coreness score** to nodes



Core–periphery kernel optimization

Core–score vector $\boldsymbol{u}$ is such that :

if $u_i > u_j \quad \Longrightarrow \quad i$ is closer to the core than j

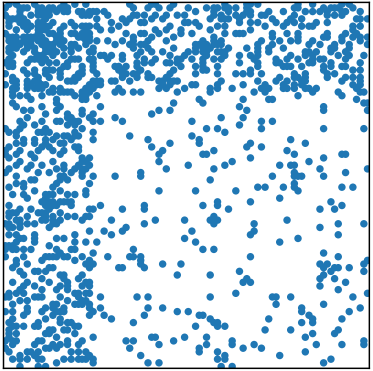
F T, D J Higham, SIAM Math of Data Science, 2019

Core–score vector as solution of the following constrained optimization

$$\text{(cp)} \qquad \begin{cases} \text{maximize} & f_\alpha(\boldsymbol{u}) = \sum_{i,j=1}^{n} A_{ij}\kappa_\alpha(u_i, u_j) \\ \text{subject to} & \|\boldsymbol{u}\|_p = 1, \boldsymbol{u} \succeq 0 \end{cases}$$

with A = adjacency matrix and $\kappa_\alpha(x, y) = \left(\frac{x^\alpha + y^\alpha}{2}\right)^{1/\alpha}$

Core-periphery kernel

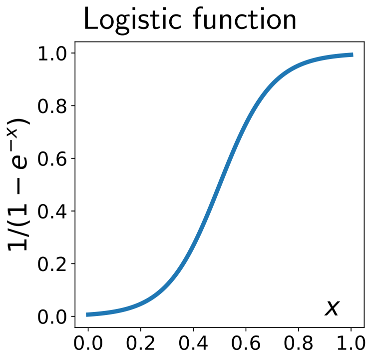


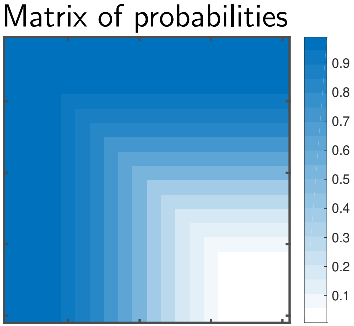
α large $\Rightarrow \kappa_\alpha(x, y) \approx \max\{x, y\}$

$f_\alpha(\boldsymbol{u}) = \sum_{ij} A_{ij}\kappa_\alpha(u_i, u_j)$ is large when edges $A_{ij} = 1$ involve at least one node with large core-score

Logistic core–periphery (LCP) random model

$$\text{Random graph: } \Pr(i \sim j) = \frac{1}{1 + e^{-\kappa_\alpha(u_i, u_j)}} =: p_{ij}(\boldsymbol{u})$$





Connection to CP kernel

Suppose we have a sample from the LCP random graph model, with nodes in arbitrary order.

Find $\boldsymbol{u}$ that maximizes the **likelihood** $\lambda(\boldsymbol{u}) = \prod_{i \sim j} p_{ij}(\boldsymbol{u}) \prod_{i \not\sim j} (1 - p_{ij}(\boldsymbol{u}))$

Connection to CP kernel

Suppose we have a sample from the LCP random graph model, with nodes in arbitrary order.

Find $\boldsymbol{u}$ that maximizes the **likelihood** $\lambda(\boldsymbol{u}) = \prod_{i \sim j} p_{ij}(\boldsymbol{u}) \prod_{i \not\sim j} (1 - p_{ij}(\boldsymbol{u}))$

Theorem:

If $\boldsymbol{u}$ is a node labeling (permutation) then

$$\boldsymbol{u} \text{ solves (cp)} \iff \boldsymbol{u} \text{ maximizes the likelihood } \lambda$$

(Useful for testing core–periphery detection algorithms)

Connection with node degrees

If $p = 2$ and $\alpha = 1$ then $\kappa_1 =$ arithmetic mean

$$\text{(cp)} \iff \max_{\boldsymbol{u} \geq 0} \frac{||A\boldsymbol{u}||_1}{||\boldsymbol{u}||_2} = ||A||_{2\to 1}$$

and the maximizer is

$$\boldsymbol{u} = \text{degree vector}$$

Connection with eigenvector centrality

If $p = 1$ and $\alpha = 0$ then $\kappa_0 =$ geometric mean

$$(\text{cp}) \iff \max_{\boldsymbol{u} \geq 0} \frac{\boldsymbol{u}^T A \boldsymbol{u}}{\boldsymbol{u}^T \boldsymbol{u}} = \rho(A)$$

and the maximizer is

$$\boldsymbol{u} = \text{Perron eigenvector of } A$$

What about the general case ......

Using Perron–Frobenius

$$
\text{(cp)} \qquad \begin{cases} \text{maximize} & f_\alpha(\boldsymbol{u}) = \sum_{i,j=1}^n A_{ij}\kappa_\alpha(u_i, u_j) \\ \text{subject to} & \|\boldsymbol{u}\|_p = 1,\ \boldsymbol{u} \succeq 0 \end{cases}
$$

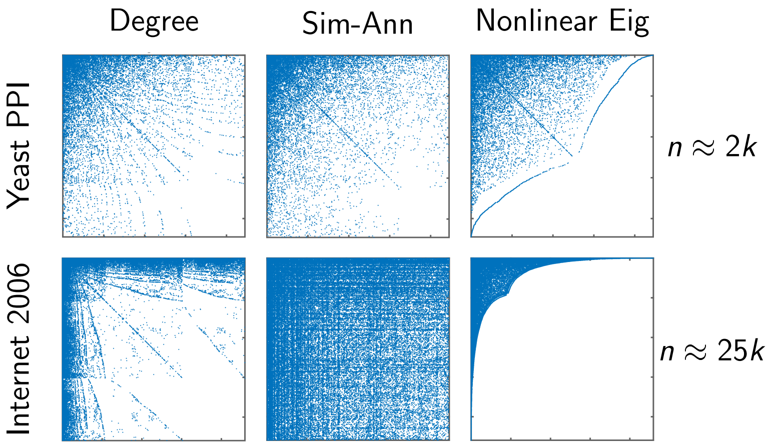
A direct computation reveals that

- $f_\alpha \in \text{subhom}'(\alpha - 1)$ and $\partial^2 f$ is positive
- $g(\boldsymbol{u}) = \|\boldsymbol{u}\|_p^p \in \text{hom}(p)$ and ∂g is invertible on $\mathbb{R}^n_{++}$

Then, if $\rho(|B|) = |\alpha - 1|/|p - 1| < 1$ we have that (cp) has a **unique positive solution**, which we can **compute to an arbitrary precision** and which coincides with the **nonlinear Perron eigenvector**

$$
M(\boldsymbol{u})\boldsymbol{u} = \lambda \boldsymbol{u} \quad \text{with} \quad M(\boldsymbol{u})_{ij} = \frac{1}{u_i^{1-\alpha}} \frac{A_{ij}\, u_j^{\alpha-2}}{(u_i^\alpha / u_j^\alpha + 1)^{\frac{1-\alpha}{\alpha}}}
$$

Qualitative results



Degree coincides with (cp) for $\alpha = 1$ and $p = 2$

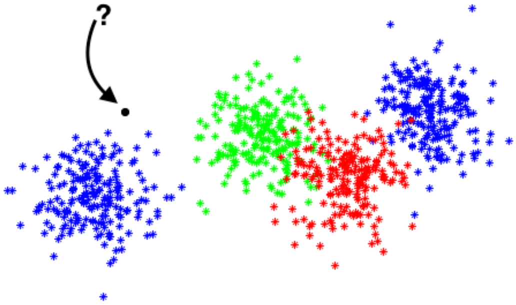
Convergence in a **few seconds** vs **several minutes** with Sim-Ann.

Beyond matrix-like forms

$f(x) = \sum_{ij} A_{ij} \kappa_\alpha(x_i, x_j)$ still looks like a nonlinear variation of a matrix form....

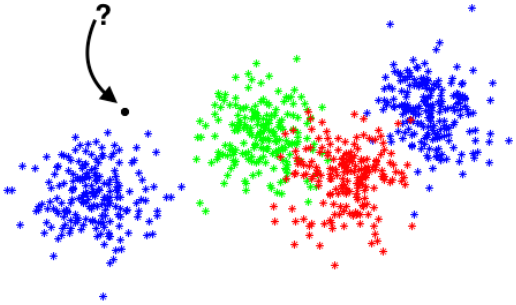
Polynomial neural networks

Supervised learning: training a nnet



Training points: $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(d)} \in \mathbb{R}^n$, $c_i \in \{1, 2, 3\}$ = class of $\boldsymbol{a}^{(i)}$
Activation matrices: (our variable) $X = (X_1, \ldots, X_m)$

Supervised learning: training a nnet



Training points: $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(d)} \in \mathbb{R}^n$, $c_i \in \{1, 2, 3\}$ = class of $\boldsymbol{a}^{(i)}$

Activation matrices: (our variable) $X = (X_1, \ldots, X_m)$

$$f(X) = \frac{1}{d} \sum_{i=1}^{d} \left[L\left(c_i, \varphi(X)(\boldsymbol{a}^{(i)}) \right) + \mathbb{1}^\top \varphi(X)(\boldsymbol{a}^{(i)}) \right]$$

$$L(j, \boldsymbol{z}) = z_j - \log\left(e^{z_1} + e^{z_2} + e^{z_3} \right) \text{ (cross–entropy loss)}$$

Polynomial activation function

A Gautier, Q Nguyen, M Hein, NeurIPS16

Train a classifier via

$$\begin{cases} \min_{X_1,\dots,X_m} & f(X_1,\dots,X_m) \\ \text{subject to} & \|X_1\|_{p_1} = \cdots = \|X_m\|_{p_m} = 1, \end{cases}$$

with **activation functions** $\varphi_j(\boldsymbol{u}) = \boldsymbol{u}^{\boldsymbol{b}_j} = (u_1^{(\boldsymbol{b}_j)_1}, \dots, u_{n_j}^{(\boldsymbol{b}_j)_{n_j}})$

Polynomial activation function

A Gautier, Q Nguyen, M Hein, NeurIPS16

Train a classifier via

$$\begin{cases} \min_{X_1,\dots,X_m} & f(X_1,\dots,X_m) \\ \text{subject to} & \|X_1\|_{p_1} = \cdots = \|X_m\|_{p_m} = 1, \end{cases}$$

with **activation functions** $\varphi_j(\boldsymbol{u}) = \boldsymbol{u}^{\boldsymbol{b}_j} = (u_1^{(\boldsymbol{b}_j)_1}, \dots, u_{n_j}^{(\boldsymbol{b}_j)_{n_j}})$

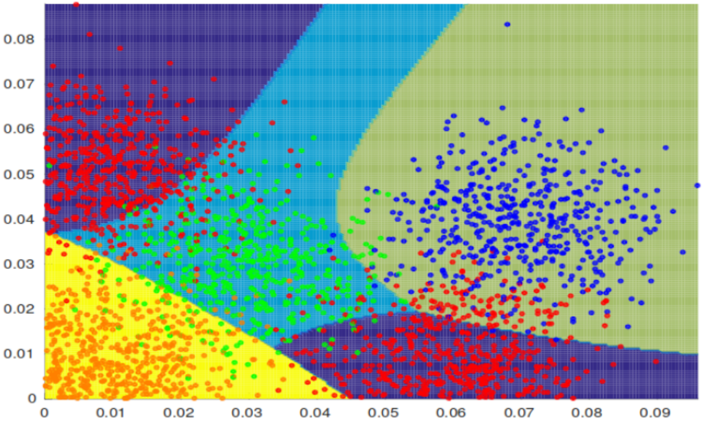
Theorem

There exists $\Theta = \Theta(\boldsymbol{b})$ such that

$$f(X) = \frac{1}{d}\sum_i \Big[L\Big(c_i, \varphi(X)(\boldsymbol{a}^{(i)})\Big) + \mathbb{1}^\top \varphi(X)(\boldsymbol{a}^{(i)})\Big]$$

is $\text{subhom}'(\Theta)$ and $\partial^2 f$ is positive. Conditions on $\boldsymbol{b}$ to get $\rho(\Theta) < 1$.

Example 2D decision boundary



Conclusions

Conclusions

If you have a "nonnegative problem" – check out the nonlinear PF theory

- Website – `ftudisco.github.io/siam-nonlinear-pf-tutorial` (feedback is very welcome)
- Book – soon (next year) to come

Some important questions concern:

- Convergence rate of nonlinear power method (e.g. for $Af(x) = \lambda x$)
- More advanced (faster) numerical eigensolvers

Thank you very much for your attention!