Spectral Graph Sparsification

Following Spielman, Chapters 32,33.

Gil Cohen

January 11, 2021

Overview

The Loewner (partial) order

Recall from the problem set.

Definition

Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ symmetric matrices. We write $\mathbf{A} \succcurlyeq \mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is PSD (which recall we write as $\mathbf{A} - \mathbf{B} \succcurlyeq 0$).

It is best to first consider the definition for diagonal matrices $\mathbf{A}, \mathbf{B}$. Note that $\mathbf{A} \succcurlyeq \mathbf{B}$ iff $\mathbf{A}_{i,i} \geq \mathbf{B}_{i,i}$ for all $i \in [n]$.

A useful property that will allow us to prove a more general statement is

$$\mathbf{A} \succcurlyeq \mathbf{B} \implies \mathbf{C}^T\mathbf{A}\mathbf{C} \succcurlyeq \mathbf{C}^T\mathbf{B}\mathbf{C} \text{ for all } C.$$

The Loewner order and eigenvalues

Assume $\mathbf{A}, \mathbf{B}$ have a common basis of eigenvectors, with corresponding eigenvalues $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$, respectively. Then,

$$\mathbf{A} \succcurlyeq \mathbf{B} \iff \forall i \in [n] \;\; \alpha_i \geq \beta_i.$$

Assume $\alpha_1 \geq \cdots \geq \alpha_n$ are the eigenvalues of $\mathbf{A}$, and $\beta_1 \geq \cdots \geq \beta_n$ the eigenvalues of $\mathbf{B}$. Then,

$$\mathbf{A} \succcurlyeq \mathbf{B} \implies \forall i \in [n] \;\; \alpha_i \geq \beta_i$$

even under no assumption on the eigenvectors. The other direction is (generally) false.

Loewner order and the spectral norm

Claim

For every symmetric matrix $\mathbf{A}$,

$$\|\mathbf{A}\| \leq c \iff -c \leq \alpha_n \leq \alpha_1 \leq c$$
$$\iff -c\mathcal{I} \preccurlyeq \mathbf{A} \preccurlyeq c\mathcal{I}.$$

Recall that if $\mathbf{W}$ is the random walk matrix of G, then G is a $(1-\omega)$-spectral expander iff $\|\mathbf{W}-\mathbf{J}\| \leq \omega$. This is equivalent to

$$-\omega\mathcal{I} \preccurlyeq \mathbf{W}-\mathbf{J} \preccurlyeq \omega\mathcal{I}.$$

An equivalent formulation, focusing on the normalized Laplacian is the following multiplicative-type statement.

$$(1-\omega)(\mathcal{I}-\mathbf{J}) \preccurlyeq \mathcal{I}-\mathbf{W} \preccurlyeq (1+\omega)(\mathcal{I}-\mathbf{J}).$$

Overview

Spectral approximation

Definition

Let G, H be graphs on n vertices.

- We write $G \succcurlyeq H$ if $\mathbf{L}_G \succcurlyeq \mathbf{L}_H$.
- We say that H an ε-spectral approximation of G if

$$(1-\varepsilon)\mathbf{L}_G \preccurlyeq \mathbf{L}_H \preccurlyeq (1+\varepsilon)\mathbf{L}_G.$$

Spectral approximation is a strong notion. It implies closeness of

- conductances,
- eigenvalues,
- effective resistances.

Expanders as spectral approximators of the complete graph

For every $\varepsilon > 0$ there exists $c = c(\varepsilon)$ such that for every n there exists a graph G with at most cn edges that ε-approximates the complete graph with self loops. In particular, Ramanujan graphs achieve $c = O(1/\varepsilon^2)$.

Can we spectral-approximate every graph with a sparse graph?

Theorem

For every $\varepsilon > 0$ and every (undirected) graph G on n vertices there exists a weighted graph H with $O(n/\varepsilon^2)$ edges that ε-spectral approximates G.

Overview

Resistor networks and leverage scores

In the recitations we viewed graphs as resistor networks. In particular, the effective resistance between two vertices a, b was defined to be the voltage difference $\mathbf{v}(a) - \mathbf{v}(b)$ when flowing one unit of current to a and out of b ($\mathbf{i} = \mathbf{e}(a) - \mathbf{e}(b)$).

$$\mathrm{R_{eff}}(a, b) = (\mathbf{e}(a) - \mathbf{e}(b))^T \mathbf{L}^+ (\mathbf{e}(a) - \mathbf{e}(b)).$$

Definition

The leverage score of an edge e is defined by $\ell_e = w_e \mathrm{R_{eff}}(e)$.

We proved that ℓ_e is the probability edge e will be included in a random spanning tree (suitably sampled according to the weights).

Overview

A probabilistic proof of a weaker statement

Theorem

For every $\varepsilon > 0$ and every weighted graph G on n vertices there exists a weighted graph H with $O(n\log(n)/\varepsilon^2)$ edges that ε-spectral approximates G.

Algorithm. For a parameter c to be set later on, include each edge e of G to H independently with probability $p_e = c\ell_e$ and set its weight to be $w_H(e) = w_G(e)/p_e$.

There is a technical issue - an edge e might have $p_e > 1$ (as c will be chosen larger than 1). To solve this we split every edge e to several edges (which does not affect the Laplacian).

Note, the above algorithm can be thought of as taking the union of c uniformly sampled MST (though not quite).

A probabilistic proof of a weaker statement - analysis

We first show that the graph is sparse in expectation.

$$\mathbb{E}[|E_H|] = \sum_e p_e = c \sum_e \ell_e.$$

Recall that

$$\ell_e = \Pr_T[e \in T] = \mathbb{E}_T[\mathbf{1}_{e \in T}].$$

Thus,

$$\sum_e \ell_e = \sum_e \mathbb{E}_T[\mathbf{1}_{e \in T}] = \mathbb{E}_T\left[\sum_e \mathbf{1}_{e \in T}\right] = n - 1.$$

Thus, $\mathbb{E}[|E_H|] = c(n-1)$. By the Chernoff bound, except with probability, $\exp(-cn)$, we have $|E_H| \leq 2cn$.

A probabilistic proof of a weaker statement - analysis

As for the Laplacian,

$$\mathbb{E}[\mathbf{L}_H] = \mathbb{E}\left[\sum_e w_H(e)\mathbf{L}_e\right] = \sum_e \mathbb{E}[w_H(e)]\mathbf{L}_e.$$

Now,

$$\mathbb{E}[w_H(e)] = p_e \cdot \frac{w_G(e)}{p_e} = w_G(e),$$

and so

$$\mathbb{E}[\mathbf{L}_H] = \sum_e w_G(e)\mathbf{L}_e = \mathbf{L}_G.$$

A probabilistic proof of a weaker statement - analysis

For a symmetric matrix **A** let $\alpha_{\min}(\mathbf{A}), \alpha_{\max}(\mathbf{A})$ denote the smallest and largest eigenvalues of **A**, respectively.

Theorem (Matrix Chernoff Bound)

Let $\mathbf{X}_1, \ldots, \mathbf{X}_m$ *be independent random PSD matrices of bounded norm* $\|\mathbf{X}_i\| \leq r$ *for all* $i \in [m]$*. Let* $\mathbf{X} = \sum_{i=1}^{m} \mathbf{X}_i$ *and denote* $\bar{\alpha}_{min} = \alpha_{min}(\mathbb{E}[\mathbf{X}])$, $\bar{\alpha}_{max} = \alpha_{max}(\mathbb{E}[\mathbf{X}])$*. Then, for every* $\varepsilon > 0$,

$$\Pr[\alpha_{min}(\mathbf{X}) \leq (1-\varepsilon)\bar{\alpha}_{min}] \leq n \cdot \exp\left(-\varepsilon^2 \bar{\alpha}_{min}/3r\right),$$

$$\Pr[\alpha_{max}(\mathbf{X}) \geq (1+\varepsilon)\bar{\alpha}_{max}] \leq n \cdot \exp\left(-\varepsilon^2 \bar{\alpha}_{max}/3r\right).$$

A probabilistic proof of a weaker statement - analysis

Consider the matrix

$$\mathbf{X} = \mathbf{L}_G^{+/2} \mathbf{L}_H \mathbf{L}_G^{+/2}$$

We have that

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}[\mathbf{L}_G^{+/2} \mathbf{L}_H \mathbf{L}_G^{+/2}] = \mathbf{L}_G^{+/2} \mathbb{E}[\mathbf{L}_H] \mathbf{L}_G^{+/2} = \mathbf{L}_G^{+/2} \mathbf{L}_G \mathbf{L}_G^{+/2} = \mathbf{\Pi},$$

where, note, $\mathbf{\Pi}$ is the projection to the image of $\mathbf{L}_G$.

We wish to apply the Chernoff bound but $\mathbf{\Pi}$ has 0 as an eigenvalue, rendering the (lower) Chernoff bound ineffective. To get around this technicality, we work in the image of $\mathbf{\Pi}$.

A probabilistic proof of a weaker statement - analysis

Let $\psi_1, \ldots, \psi_n$ be an orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of $\mathbf{L}_G$ with corresponding eigenvalues $0 = \mu_1 < \mu_2 \leq \cdots \leq \mu_n$. So,

$$\mathbf{L}_G = \sum_{i=2}^{n} \mu_i \psi_i \psi_i^T .$$

Let $\mathbf{B}$ be the $n \times (n-1)$ matrix with column $i \in [n-1]$ equal ψ_{i+1}. Note that

$$\mathbf{\Pi} = \mathbf{L}_G^{+/2} \mathbf{L}_G \mathbf{L}_G^{+/2} = \sum_{i=2}^{n} \psi_i \psi_i^T$$

and that

$$\mathcal{I}_{n-1} = \mathbf{B}^T \Pi \mathbf{B}.$$

A probabilistic proof of a weaker statement - analysis

For applying Chernoff, it will be convenient to write

$$\mathbf{B}^T\mathbf{X}\mathbf{B} = \sum_e \mathbf{B}^T\mathbf{X}_e\mathbf{B},$$

where

$$\mathbf{X}_e = \begin{cases} \frac{w_e}{p_e}\mathbf{L}_G^{+/2}\mathbf{L}_e\mathbf{L}_G^{+/2} & \text{with probability } p_e \\ 0 & \text{with probability } 1-p_e, \end{cases}$$

and we recall

$$\mathbb{E}[\mathbf{B}^T\mathbf{X}\mathbf{B}] = \mathbf{B}^T\mathbb{E}[\mathbf{X}]\mathbf{B} = \mathbf{B}^T\mathbf{\Pi}\mathbf{B} = \mathcal{I}_{n-1}.$$

A probabilistic proof of a weaker statement - analysis

We turn to bound the norm of $\mathbf{B}^T\mathbf{X}_{a,b}\mathbf{B}$. For that it suffices to bound the norm of $\mathbf{X}_{a,b}$.

$$
\begin{aligned}
\left\|\mathbf{L}_G^{+/2}\mathbf{L}_{a,b}\mathbf{L}_G^{+/2}\right\| &= \left\|\mathbf{L}_G^{+/2}(\mathbf{e}(a)-\mathbf{e}(b))(\mathbf{e}(a)-\mathbf{e}(b))^T\mathbf{L}_G^{+/2}\right\| \\
&= \left((\mathbf{e}(a)-\mathbf{e}(b))^T\mathbf{L}_G^{+/2}\right)\left(\mathbf{L}_G^{+/2}(\mathbf{e}(a)-\mathbf{e}(b))\right) \\
&= (\mathbf{e}(a)-\mathbf{e}(b))^T\mathbf{L}_G^{+}(\mathbf{e}(a)-\mathbf{e}(b)) \\
&= \mathrm{R}_{\mathrm{eff}}(a,b).
\end{aligned}
$$

Recall that

$$
p_{a,b} = c\ell_{a,b} = cw_{a,b}\mathrm{R}_{\mathrm{eff}}(a,b),
$$

and so $\|\mathbf{B}^T\mathbf{X}_{a,b}\mathbf{B}\| \leq \|\mathbf{X}_{a,b}\| \leq \frac{1}{c}$.

A probabilistic proof of a weaker statement - analysis

By the matrix Chernoff bound, except with probability $2n \cdot \exp(-c\varepsilon^2/3)$, we have that

$$1 - \varepsilon \leq \alpha_{\text{min}}(\mathbf{B}^T\mathbf{X}\mathbf{B}) \leq \alpha_{\text{max}}(\mathbf{B}^T\mathbf{X}\mathbf{B}) \leq 1 + \varepsilon.$$

Hence,

$$(1 - \varepsilon)\mathcal{I}_{n-1} \preccurlyeq \mathbf{B}^T\mathbf{X}\mathbf{B} \preccurlyeq (1 + \varepsilon)\mathcal{I}_{n-1}.$$

Note $\mathbf{B}\mathbf{B}^T = \mathbf{\Pi}$. Hitting with $\mathbf{B}$ and $\mathbf{B}^T$ from the left and right, respectively, we conclude

$$(1 - \varepsilon)\mathbf{\Pi} \preccurlyeq \mathbf{\Pi}\mathbf{X}\mathbf{\Pi} \preccurlyeq (1 + \varepsilon)\mathbf{\Pi}.$$

A probabilistic proof of a weaker statement - analysis

As $\mathbf{X} = \mathbf{L}_G^{+/2}\mathbf{L}_H\mathbf{L}_G^{+/2}$ and since $\mathbf{\Pi}\mathbf{L}_G^{+/2} = \mathbf{L}_G^{+/2} = \mathbf{L}_G^{+/2}\mathbf{\Pi}$,

$$\mathbf{\Pi}\mathbf{X}\mathbf{\Pi} = \mathbf{\Pi}\mathbf{L}_G^{+/2}\mathbf{L}_H\mathbf{L}_G^{+/2}\mathbf{\Pi} = \mathbf{L}_G^{+/2}\mathbf{L}_H\mathbf{L}_G^{+/2}.$$

Thus,

$$(1-\varepsilon)\mathbf{\Pi} \preccurlyeq \mathbf{L}_G^{+/2}\mathbf{L}_H\mathbf{L}_G^{+/2} \preccurlyeq (1+\varepsilon)\mathbf{\Pi}.$$

Hitting by $\mathbf{L}_G^{1/2}$ on the left and right,

$$(1-\varepsilon)\mathbf{L}_G \preccurlyeq \mathbf{\Pi}\mathbf{L}_H\mathbf{\Pi} \preccurlyeq (1+\varepsilon)\mathbf{L}_G,$$

from which it follows that

$$(1-\varepsilon)\mathbf{L}_G \preccurlyeq \mathbf{L}_H \preccurlyeq (1+\varepsilon)\mathbf{L}_G.$$

A probabilistic proof of a weaker statement - analysis

To conclude the proof, we need to choose c such that, say, $2n \cdot \exp(-c\varepsilon^2/3) < 1/2$. We thus take $c = O(\log(n)/\varepsilon^2)$. Hence,

$$|E_H| \leq O(n\log(n)/\varepsilon^2).$$

Overview

Linear Sized Sparsifiers

As in the previous section, we consider

$$\begin{aligned}
\mathbf{\Pi} &= \mathbf{L}_G^{+/2} \mathbf{L}_G \mathbf{L}_G^{+/2} \\
&= \mathbf{L}_G^{+/2} \left(\sum_{ab \in E} w_{ab} \mathbf{L}_{ab} \right) \mathbf{L}_G^{+/2} \\
&= \sum_{ab \in E} w_{ab} \mathbf{L}_G^{+/2} (\mathbf{e}(a) - \mathbf{e}(b))(\mathbf{e}(a) - \mathbf{e}(b))^T \mathbf{L}_G^{+/2} \\
&= \sum_{ab \in E} \psi_{ab}^T \psi_{ab},
\end{aligned}$$

where

$$\psi_{ab} = \sqrt{w_{ab}} \mathbf{L}_G^{+/2} (\mathbf{e}(a) - \mathbf{e}(b))$$

Linear Sized Sparsifiers

Thus, it suffices to solve the following problem. Given $\varepsilon > 0$ and vectors $\psi_1, \dots, \psi_m \in \mathbb{R}^n$ in isotropic position

$$\sum_{i=1}^{m} \psi_i \psi_i^T = \mathcal{I},$$

find a subset $S \subseteq [m]$, of size $|S| = s = O(n/\varepsilon^2)$, and weights $(c_i)_{i \in S}$ such that

$$(1-\varepsilon)\mathcal{I} \preccurlyeq \sum_{i \in S} c_i \psi_i \psi_i^T \preccurlyeq (1+\varepsilon)\mathcal{I}.$$

Proof strategy

The algorithm for computing S and the weights is iterative. At iteration $j = 1, \ldots, s$, an element ψ_{i_j} will be added to S ($i_j \in [m]$) with a suitable weight c_j. An element may be chosen more than once.

We will maintain the invariant that for every j, the matrix

$$\mathbf{A}_j = \sum_{k=1}^{j} c_k \psi_{i_k} \psi_{i_k}^T$$

satisfies

$$-n + \lambda j \leq \alpha_{\min}(\mathbf{A}_j) \leq \alpha_{\max}(\mathbf{A}_j) \leq n + \upsilon j.$$

for some parameters $\lambda, \upsilon > 0$.

The barrier functions

Tracking only the smallest and largest eigenvalues does not seem to carry sufficient amount of information. Instead, the key idea is to record a suitably chosen potential function of all eigenvalues.

Let $\mathbf{A}$ be an $n \times n$ symmetric matrix with eigenvalues $\alpha_1 \leq \cdots \leq \alpha_n$. We define the upper and lower barrier functions

$$\Phi^u(\mathbf{A}) = \sum_{i=1}^{n} \frac{1}{u - \alpha_i} = \operatorname{Tr}\left((u\mathcal{I} - \mathbf{A})^{-1}\right),$$

$$\Phi_\ell(\mathbf{A}) = \sum_{i=1}^{n} \frac{1}{\alpha_i - \ell} = \operatorname{Tr}\left((\mathbf{A} - \ell\mathcal{I})^{-1}\right).$$

The barrier functions

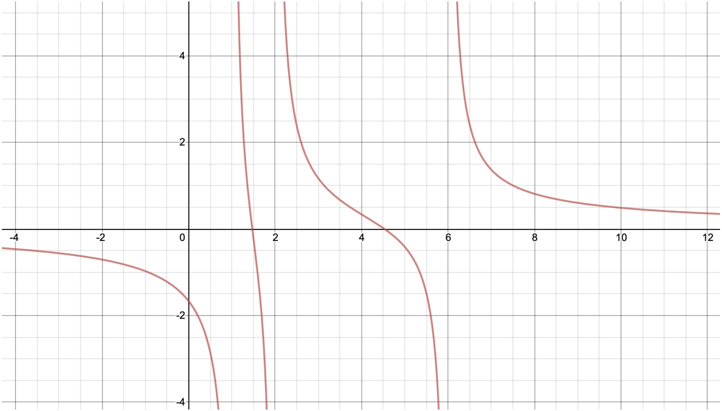


Figure: The upper barrier function with $(\alpha_1, \alpha_2, \alpha_3) = (1, 2, 6)$.

The barrier functions

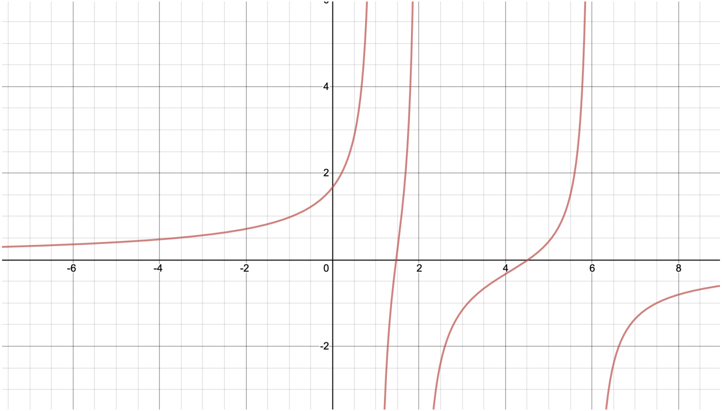


Figure: The lower barrier function with $(\alpha_1, \alpha_2, \alpha_3) = (1, 2, 6)$.

The barrier functions

Note that for every $u > \alpha_n$ and $\ell < \alpha_1$,

$$\alpha_n \leq u - \frac{1}{\Phi^u(\mathbf{A})} \qquad \alpha_1 \geq \ell + \frac{1}{\Phi_\ell(\mathbf{A})}.$$

Instead of only considering $\alpha_{\min}, \alpha_{\max}$, we maintain an invariant on the barrier functions. For $j = 0, 1, \ldots, s$, we define

$$u_j = n + \upsilon j$$
$$\ell_j = -n + \lambda j,$$

and maintain the invariant

$$\Phi^{u_j}(\mathbf{A}_j) \leq 1,$$
$$\Phi_{\ell_j}(\mathbf{A}_j) \leq 1.$$

Initialization

Initially, we set $S = \emptyset$, and so $\mathbf{A}_0 = 0$. Hence,

$$\Phi^{u_0}(\mathbf{A}_0) = \sum_{i=1}^{n} \frac{1}{u_0} = \frac{n}{u_0} = 1,$$

$$\Phi_{\ell_0}(\mathbf{A}_0) = \sum_{i=1}^{n} \frac{1}{-\ell_0} = -\frac{n}{\ell_0} = 1.$$

The upper barrier functions

How does the upper barrier function changes under a rank one update? This you resolved in the problem set. In particular, you proved

Lemma (Sherman-Morrison)

Let $\mathbf{B}$ *be a nonsingular symmetric matrix. Let* $\psi \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$*. Then,*

$$(\mathbf{B} - c\psi\psi^T)^{-1} = \mathbf{B}^{-1} + \frac{c}{1 - c\psi^T\mathbf{B}^{-1}\psi} \cdot \mathbf{B}^{-1}\psi\psi^T\mathbf{B}^{-1}$$

The upper barrier functions

Substituting $\mathbf{B} = u\mathcal{I} - \mathbf{A}$, we get

$$\begin{aligned}\Phi^u(\mathbf{A} + c\psi\psi^T) &= \mathrm{Tr}\left((u\mathcal{I} - \mathbf{A} - c\psi\psi^T)^{-1}\right)\\ &= \mathrm{Tr}\left((u\mathcal{I} - \mathbf{A})^{-1}\right) + \Delta(u)\\ &= \Phi^u(\mathbf{A}) + \Delta(u),\end{aligned}$$

where

$$\begin{aligned}\Delta(u) &= \frac{c}{1 - c\psi^T(u\mathcal{I} - \mathbf{A})^{-1}\psi} \cdot \mathrm{Tr}\left((u\mathcal{I} - \mathbf{A})^{-1}\psi\psi^T(u\mathcal{I} - \mathbf{A})^{-1}\right)\\ &= \frac{c\psi^T(u\mathcal{I} - \mathbf{A})^{-2}\psi}{1 - c\psi^T(u\mathcal{I} - \mathbf{A})^{-1}\psi} = \frac{\psi^T(u\mathcal{I} - \mathbf{A})^{-2}\psi}{1/c - \psi^T(u\mathcal{I} - \mathbf{A})^{-1}\psi}.\end{aligned}$$

The upper barrier functions

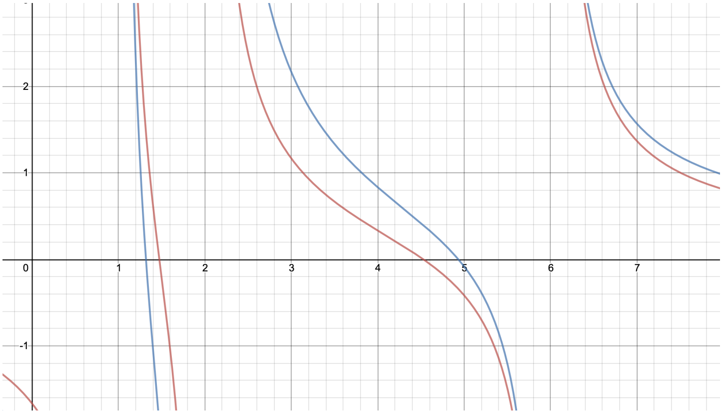
The upper potential function thus increases under a rank one update by $\Delta(u)$. We want to counteract this increase by increasing u. Namely, we want to be able to choose ψ_{i_j}, c_j such that

$$\Phi^{u_j}(\mathbf{A}_j) \geq \Phi^{u_j+\upsilon}(\mathbf{A}_j + c_j \psi_{i_j} \psi_{i_j}^T).$$

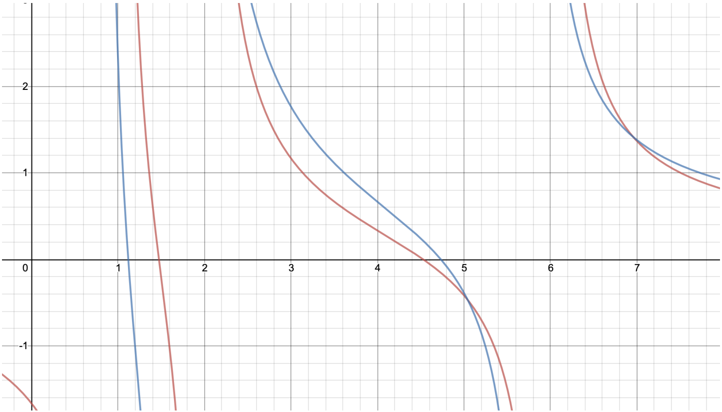
We compute

$$\begin{aligned} \Phi^{u+\upsilon}(\mathbf{A} + c\psi\psi^T) - \Phi^u(\mathbf{A}) &= \Phi^{u+\upsilon}(\mathbf{A} + c\psi\psi^T) - \Phi^{u+\upsilon}(\mathbf{A}) \\ &\quad + \Phi^{u+\upsilon}(\mathbf{A}) - \Phi^u(\mathbf{A}) \\ &= \Delta(u+\upsilon) + \Phi^{u+\delta}(\mathbf{A}) - \Phi^u(\mathbf{A}). \end{aligned}$$

The barrier functions



The barrier functions



The upper barrier functions

Recall that

$$\Delta(u+\upsilon) = \frac{\psi^T((u+\upsilon)\mathcal{I} - \mathbf{A})^{-2}\psi}{1/c - \psi^T((u+\upsilon)\mathcal{I} - \mathbf{A})^{-1}\psi}.$$

Thus, it suffices to have

$$\begin{aligned}\frac{1}{c} &\geq \psi^T((u+\upsilon)\mathcal{I} - \mathbf{A})^{-1}\psi + \frac{\psi^T((u+\upsilon)\mathcal{I} - \mathbf{A})^{-2}\psi}{\Phi^u(\mathbf{A}) - \Phi^{u+\upsilon}(\mathbf{A})} \\ &= \psi^T \mathbf{U_A} \psi,\end{aligned}$$

where

$$\mathbf{U_A} = ((u+\upsilon)\mathcal{I} - \mathbf{A})^{-1} + \frac{((u+\upsilon)\mathcal{I} - \mathbf{A})^{-2}}{\Phi^u(\mathbf{A}) - \Phi^{u+\upsilon}(\mathbf{A})}.$$

The upper barrier functions

Thus, we have found a clean condition that imply we can add ψ to S with weight c by increasing u by υ and without increasing the upper barrier function. We summarize this in the following claim.

Claim

$$\frac{1}{c} \geq \psi^T \mathbf{U_A} \psi \quad \Longrightarrow \quad \Phi^{u+\upsilon}(\mathbf{A} + c\psi\psi^T) \leq \Phi^u(\mathbf{A}).$$

The lower barrier functions

Define

$$\mathbf{L}_A = \frac{(\mathbf{A} - (\ell + \lambda)\mathcal{I})^{-2}}{\Phi_{\ell+\lambda}(\mathbf{A}) - \Phi_{\ell}(\mathbf{A})} - (\mathbf{A} - (\ell + \lambda)\mathcal{I})^{-1}.$$

Similar to the previous claim, one can show

Claim

$$\frac{1}{c} \leq \psi^T \mathbf{L}_{\mathbf{A}} \psi \quad \Longrightarrow \quad \Phi_{\ell+\lambda}(\mathbf{A} + c\psi\psi^T) \leq \Phi_{\ell}(\mathbf{A}).$$

The inductive argument

It remains to show that there exist ψ_i and a weight c such that

$$\Phi^{u+\upsilon}(\mathbf{A} + c\psi_i\psi_i^T) \leq \Phi^u(\mathbf{A}),$$
$$\Phi_{\ell+\lambda}(\mathbf{A} + c\psi_i\psi_i^T) \leq \Phi_\ell(\mathbf{A}).$$

By the two claims, it suffices to prove that there exists $i \in [m]$ such that

$$\psi_i^T \mathbf{U}_\mathbf{A} \psi_i \leq \psi_i^T \mathbf{L}_\mathbf{A} \psi_i.$$

We can then take any weight c in between. By an averaging argument, it suffices to prove that

$$\sum_{i=1}^m \psi_i^T \mathbf{U}_\mathbf{A} \psi_i \leq \sum_{i=1}^m \psi_i^T \mathbf{L}_\mathbf{A} \psi_i.$$

The inductive argument

We first prove the following claim.

Claim

For every matrix $\mathbf{B}$,

$$\sum_{i=1}^{m} \psi_i^T \mathbf{B} \psi_i = \mathrm{Tr}(\mathbf{B}).$$

As $\psi^T \mathbf{B} \psi = \mathrm{Tr}(\psi^T \mathbf{B} \psi) = \mathrm{Tr}(\psi \psi^T \mathbf{B})$, we have

$$\sum_{i=1}^{m} \psi_i^T \mathbf{B} \psi_i = \sum_{i=1}^{m} \mathrm{Tr}(\psi_i \psi_i^T \mathbf{B}) = \mathrm{Tr}\left(\left(\sum_{i=1}^{m} \psi_i \psi_i^T\right) \mathbf{B}\right) = \mathrm{Tr}(\mathbf{B}).$$

The inductive argument

Recall,

$$\mathbf{U_A} = ((u+\upsilon)\mathcal{I} - \mathbf{A})^{-1} + \frac{((u+\upsilon)\mathcal{I} - \mathbf{A})^{-2}}{\Phi^{u}(\mathbf{A}) - \Phi^{u+\upsilon}(\mathbf{A})}.$$

Claim

$$\sum_{i=1}^{m} \psi_i^T \mathbf{U_A} \psi_i \leq \frac{1}{\upsilon} + \Phi^{u}(\mathbf{A}).$$

By the previous claim,

$$\sum_{i=1}^{m} \psi_i^T \mathbf{U_A} \psi_i = \mathrm{Tr}(\mathbf{U_A}) = \Phi^{u+\upsilon}(\mathbf{A}) + \frac{\mathrm{Tr}\left(((u+\upsilon)\mathcal{I} - \mathbf{A})^{-2}\right)}{\Phi^{u}(\mathbf{A}) - \Phi^{u+\upsilon}(\mathbf{A})}.$$

The inductive argument

Now, as we consider $u \geq \alpha_{\max}(\mathbf{A})$, we have $\Phi^{u+v}(\mathbf{A}) \leq \Phi^{u}(\mathbf{A})$. As for the second term,

$$\frac{\partial}{\partial u}\Phi^{u}(\mathbf{A}) = -\sum_{i=1}^{m} \frac{1}{(u-\alpha_i)^2} = -\mathrm{Tr}\left((u\mathcal{I} - \mathbf{A})^{-2}\right).$$

By convexity,

$$\frac{\Phi^{u+v}(\mathbf{A}) - \Phi^{u}(\mathbf{A})}{v} \leq \frac{\partial}{\partial u}\Phi^{u+v}(\mathbf{A}).$$

Hence,

$$\frac{\mathrm{Tr}\left(((u+v)\mathcal{I} - \mathbf{A})^{-2}\right)}{\Phi^{u}(\mathbf{A}) - \Phi^{u+v}(\mathbf{A})} \leq \frac{1}{v}.$$

The inductive argument

Similarly, one can prove that

$$\sum_{i=1}^{m} \psi_i^T \mathbf{L_A} \psi_i \geq \frac{1}{\lambda} - \frac{1}{\frac{1}{\Phi_\ell(\mathbf{A})} - \lambda}.$$

Try to prove that by yourself. This time, the analog of the statement $\Phi^{u+v}(\mathbf{A}) > \Phi^u(\mathbf{A})$ for all $u > \alpha_{\max}(\mathbf{A})$ is a bit trickier, and is given by the following claim.

Claim

For every $\ell < \alpha_{min}(\mathbf{A})$ and $\lambda < 1/\Phi_\ell(\mathbf{A})$, it holds that

$$\Phi_{\ell+\lambda}(\mathbf{A}) \leq \frac{1}{\frac{1}{\Phi_\ell(\mathbf{A})} - \lambda}.$$

Setting the parameters

By the above, we can take any υ, λ such that

$$\frac{1}{\upsilon} + \Phi^{u_j}(\mathbf{A}_j) \leq \frac{1}{\lambda} - \frac{1}{\frac{1}{\Phi_{\ell_j}(\mathbf{A}_j)} - \lambda}$$

for all $j = 0, 1, \ldots, s = cn$. Recall,

$$\alpha_{\max}(\mathbf{A}_s) \leq u_s - \frac{1}{\Phi^{u_s}(\mathbf{A}_s)} = n + \upsilon cn - \frac{1}{\Phi^{u_s}(\mathbf{A}_s)} \leq (\upsilon c + 1)n - 1,$$

$$\alpha_{\min}(\mathbf{A}_s) \geq \ell_s + \frac{1}{\Phi_{\ell_s}(\mathbf{A}_s)} = -n + \lambda cn + \frac{1}{\Phi_{\ell_s}(\mathbf{A}_s)} \geq (\lambda c - 1)n + 1.$$

Setting the parameters

By our invariant, we can take any υ, λ for which

$$\frac{1}{\upsilon} + 1 \leq \frac{1}{\lambda} - \frac{1}{1-\lambda}.$$

For every such choice, we have

$$\frac{\alpha_{\mathrm{max}}(\mathbf{A}_s)}{\alpha_{\mathrm{min}}(\mathbf{A}_s)} \leq \frac{n + \nu cn - 1}{-n + \lambda cn + 1} \leq \frac{\upsilon c + 1}{\lambda c - 1}.$$

Setting the parameters

The example presented in Spielman considers $\lambda = \frac{1}{3}$ which leads us to take $\nu = 2$. Setting, say, $c = 13$ yields a ratio of 13. By dividing all weights by $\sqrt{13}$ we get

$$\frac{1}{\sqrt{13}}\mathbf{L}_G \preccurlyeq \mathbf{L}_H \preccurlyeq \sqrt{13}\mathbf{L}_G$$

You are encouraged to play with the numbers to improve the ratio.