Locally PSD Matrices and Hyperbolic Polynomials

Grigoriy Blekherman, Santanu Dey, Kevin Shu*, Shengding Sun

March 17, 2022

Georgia Institute of Technology

Part 1: Conic Optimization and Semidefinite Programming

Conic Optimization

Overview

- Powerful toolbox for solving problems in engineering and computer science.
- Related to convex optimization and linear programming.
- Semidefinite programming is a special kind of conic optimization that is very useful.

Conic Optimization

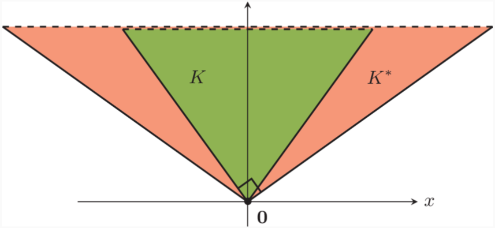
Definition
A **convex cone** is a subset C of $\mathbb{R}^n$ with 2 properties:

- If $x \in C$, and $\lambda \in \mathbb{R}$ with $\lambda \geq 0$, then $\lambda x \in C$.
- If $x, y \in C$, then $x + y \in C$.

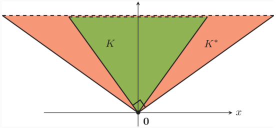
Conic Optimization

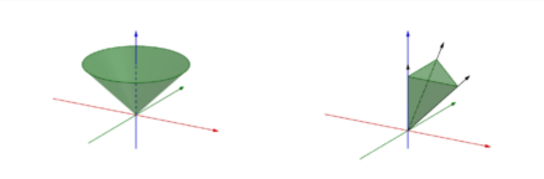
Definition
A **convex cone** is a subset C of $\mathbb{R}^n$ with 2 properties:

- If $x \in C$, and $\lambda \in \mathbb{R}$ with $\lambda \geq 0$, then $\lambda x \in C$.
- If $x, y \in C$, then $x + y \in C$.



Conic Optimization





Examples

- **Nonnegative Orthant:** $\{x \in \mathbb{R}^n : \forall i, x_i \geq 0\}$.
- **Second Order Cone:** $\{x \in \mathbb{R}^n : 0 \leq x_1 \leq \sqrt{\sum_{i=2}^{n} x_i^2}\}$.
- **Positive Semidefinite Cone**

Conic Optimization

Definition
Given a convex cone C, a **conic optimization problem** over C an optimization problem of the form

$$\begin{aligned} \text{minimize} \quad & h^\intercal x \\ \text{such that} \quad & Ax = b \\ & x \in C \end{aligned}$$

where $h \in \mathbb{R}^n$, $b \in \mathbb{R}^k$ and $A \in \mathbb{R}^{n \times k}$.

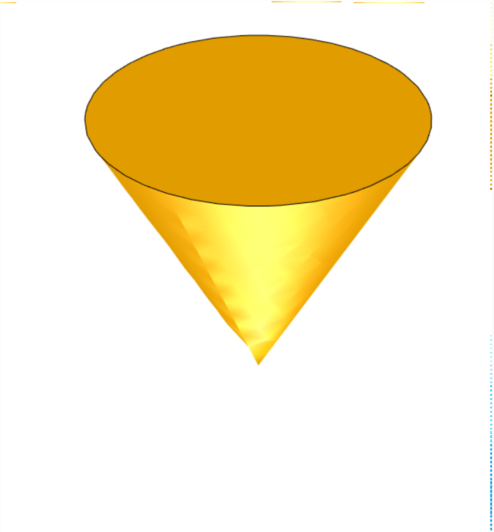


Figure 2: A convex cone C.

Conic Optimization

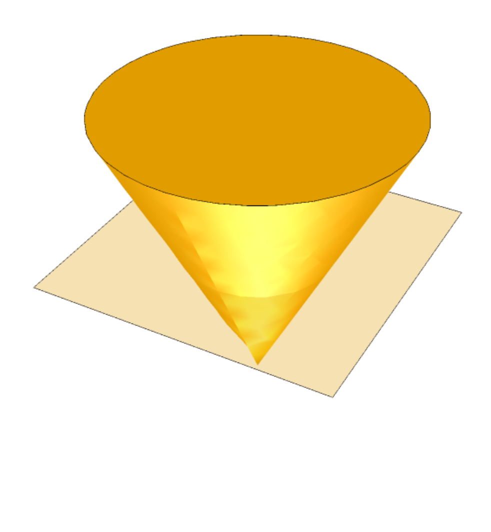


Figure 3: A linear subspace that intersects C.

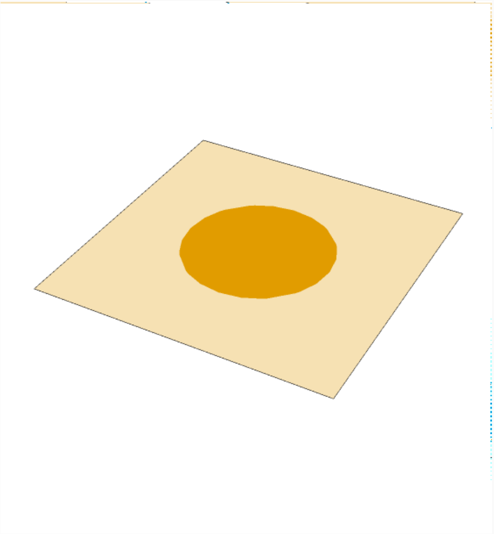


Figure 4: The intersection of a linear subspace and C

Conic Optimization

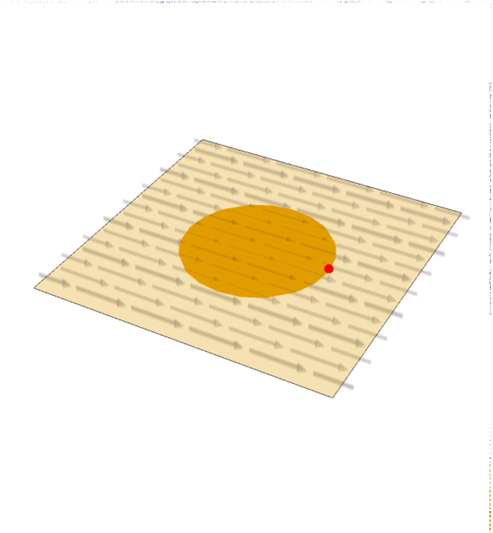


Figure 5: The optimal solution for an objective function.

Conic Optimization

When C is the nonnegative orthant, conic optimization problems over C are called **linear programming**.

Conic Optimization

When C is the nonnegative orthant, conic optimization problems over C are called **linear programming**.

Applications

- Approximating hard combinatorial problems like finding the maximum clique of a graph or finding a coloring of a graph.
- Used in economics and engineering to solve resource allocation problems.
- Used in machine learning for certain regression problems.

Conic Optimization

The more complicated C is, the more expressive the conic optimization problem is, and the more difficult they are to solve.

Meta-theorem

If you can efficiently determine whether or not a given point x is in C, then you can efficiently solve conic optimization problems over C.

Semidefinite Programming

Positive Semidefinite Matrices

Definition
An $n \times n$ matrix A where $A^\intercal = A$ is PSD if all of its eigenvalues are nonnegative.

Equivalent Conditions

- $X = \sum_i x_i x_i^\intercal$ where $x_i \in \mathbb{R}^n$.
- The determinants of all principal submatrices of A are nonnegative.
- A can be factored as $V^\intercal V$.

Positive Semidefinite Matrices

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{pmatrix}$$

Positive Semidefinite Matrices

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{pmatrix}$$

This is PSD: it can be factored

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}.$$

Positive Semidefinite Matrices

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

This is PSD: it can be factored

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}.$$

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{pmatrix}$$

This is not PSD, the submatrix

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

has determinant -1.

Positive Semidefinite Matrices

The PSD Cone
If X and Y are PSD, then $X + Y$ is PSD.

If X is PSD, then λX is PSD for $\lambda \geq 0$.

This makes the set of all PSD matrices a convex cone.

Semidefinite Programming

$$\begin{aligned} \text{minimize} \quad & \langle B^0, X \rangle \\ \text{such that} \quad & \langle B^\ell, X \rangle = b_\ell \quad \text{for } \ell \in \{1, \ldots, k\} \\ & X \succeq 0 \end{aligned}$$

- $X \succeq 0$ means that X is an $n \times n$ positive semidefinite matrix.
- A matrix is positive semidefinite if it is symmetric and all of its eigenvalues are nonnegative.
- The B^i are all $n \times n$ symmetric matrices.

Applications

- Approximates the NP-hard MAXCUT problem to a factor of 0.86.
- Computes clique numbers and chromatic numbers for perfect graphs.
- Approximates arbitrary polynomial optimization problems.
- Applications to learning mixtures of Gaussians.
- Many other engineering applications.

Semidefinite Programs for Linear Regression

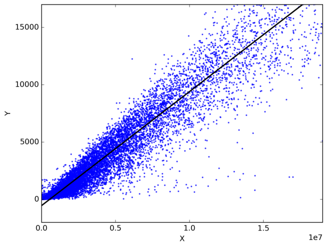
Linear Regression
Given a set of input variables $a_1, \ldots, a_k \in \mathbb{R}^n$, and a set of output variables $b_1, \ldots, b_k \in \mathbb{R}$, find $x \in \mathbb{R}^n$ that minimizes

$$\sum_{i=1}^{k} (x^\intercal a_i - b_i)^2$$

Semidefinite Programs for Linear Regression

Linear Regression
Given a set of input variables $a_1, \ldots, a_k \in \mathbb{R}^n$, and a set of output variables $b_1, \ldots, b_k \in \mathbb{R}$, find $x \in \mathbb{R}^n$ that minimizes

$$\sum_{i=1}^{k} (x^\intercal a_i - b_i)^2$$



Semidefinite Programs for Linear Regression

- $A = \begin{pmatrix} a_1 & a_2 & \dots & a_k \end{pmatrix}$
- $b = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_k \end{pmatrix}$
- Linear regression objective is

$$\min_{x \in \mathbb{R}^n} \|A^\intercal x - b\|^2.$$

Semidefinite Programs for Linear Regression

Problem: In conic optimization, the objective is supposed to be linear, but this is quadratic.

Solution: Introduce a new variable that absorbs the quadratic term.

Semidefinite Programs for Linear Regression

$$\begin{aligned} \|A^\intercal x - b\|^2 &= (A^\intercal x - b)^\intercal (A^\intercal x - b) \\ &= x^\intercal A A^\intercal x - 2b^\intercal A^\intercal x + \|b\|^2 \\ &= \mathrm{tr}(AA^\intercal x x^\intercal) - 2b^\intercal A^\intercal x + \|b\|^2 \end{aligned}$$

Semidefinite Programs for Linear Regression

$$\begin{aligned}
\|A^\intercal x - b\|^2 &= (A^\intercal x - b)^\intercal (A^\intercal x - b) \\
&= x^\intercal A A^\intercal x - 2b^\intercal A^\intercal x + \|b\|^2 \\
&= \operatorname{tr}(AA^\intercal \mathbf{x}\mathbf{x}^\intercal) - 2b^\intercal A^\intercal x + \|b\|^2
\end{aligned}$$

Semidefinite Programs for Linear Regression

$$\begin{aligned}\|A^\intercal x - b\|^2 &= (A^\intercal x - b)^\intercal (A^\intercal x - b) \\ &= x^\intercal A A^\intercal x - 2b^\intercal A^\intercal x + \|b\|^2 \\ &= \operatorname{tr}(AA^\intercal \mathbf{x}\mathbf{x}^\intercal) - 2b^\intercal A^\intercal x + \|b\|^2\end{aligned}$$

Notice that the matrix $xx^\intercal$ is PSD, so let's make $X = xx^\intercal$.

Semidefinite Programs for Linear Regression

$$\min_{x \in \mathbb{R}^n} \|A^\intercal x - b\|^2 = \min_{X = xx^\intercal} \operatorname{tr}(AA^\intercal X) - 2b^\intercal A^\intercal x + \|b\|^2$$

The right hand side is almost a semidefinite program. A little massaging let's us rewrite this as a semidefinite program.

Semidefinite Programs for Linear Regression

It turns out that the solution to the semidefinite program

$$\begin{aligned} \text{minimize} \quad & \|b\|^2 - \operatorname{tr}(XA^\intercal bb^\intercal A) \\ \text{such that} \quad & \operatorname{tr}(AA^\intercal X) = 1 \\ & X \succeq 0 \end{aligned}$$

is of the form $\alpha xx^\intercal$, where α is a scalar.

Part 2: The Locally PSD Cone and Its Relatives

Motivating Question

Sparse Linear Regression
Given a set of input variables $a_1, \ldots, a_k \in \mathbb{R}^n$, and a set of output variables $b_1, \ldots, b_k \in \mathbb{R}$, find $x \in \mathbb{R}^n$ that minimizes

$$\sum_{i=1}^{k} (x^\intercal a_i - b_i)^2,$$

and x has at most k nonzero entries.

Motivating Question

Sparse Linear Regression
Given a set of input variables $a_1, \ldots, a_k \in \mathbb{R}^n$, and a set of output variables $b_1, \ldots, b_k \in \mathbb{R}$, find $x \in \mathbb{R}^n$ that minimizes

$$\sum_{i=1}^{k} (x^\intercal a_i - b_i)^2,$$

and x has at most k nonzero entries.

NP-hard for some values of k.

Can use our earlier reformulation for linear regression to turn this into a conic optimization problem.

Conical Optimization for Sparse Linear Regression

Let

$$\mathcal{FW}_n^k = \{X \in \mathbb{R}^{n\times n} : X = \sum_i x_i x_i^\intercal, \|x_i\|_0 \leq k\}.$$

Here, $\|x\|_0$ is the number of nonzero entries of x.

$$\begin{aligned} \text{minimize} \quad & \|b\|^2 - \text{tr}(XA^\intercal bb^\intercal A) \\ \text{such that} \quad & \text{tr}(AA^\intercal X) = 1 \\ & X \in \mathcal{FW}_n^k \end{aligned}$$

is equivalent to sparse linear regression.

How can we approximate solutions to conical programs over $\mathcal{FW}_n^k$?

Two aspects:

- It is clear that nonsparse linear regression will result in a better objective. How much better can it be?
- Can we produce a solution to this program that is close to the optimal solution?

How can we approximate solutions to conical programs over $\mathcal{FW}_n^k$?

Two aspects:

- It is clear that nonsparse linear regression will result in a better objective. How much better can it be?
- Can we produce a solution to this program that is close to the optimal solution?

We will focus on the dual cone of $\mathcal{FW}_n^k$, which we call the k-locally PSD cone.

Locally PSD Matrices

k-Locally Positive Semidefinite Cones

For $S \subseteq [n]$, denote by X_S the submatrix of X indexed by elements of S.

A matrix $M \in \mathbb{R}^{n \times n}$ is k-locally PSD if every $k \times k$ submatrix of M is PSD.

$$\mathcal{S}^{n,k} = \{M : M \text{ is } k\text{-locally PSD}\}$$

is a convex cone.

E.g. the following 4×4 matrix is 2-locally PSD:

$$\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$$

k-Locally Positive Semidefinite Cones

For $S \subseteq [n]$, denote by X_S the principle submatrix of X indexed by elements of S.

A matrix $M \in \mathbb{R}^{n \times n}$ is k-locally PSD if every $k \times k$ submatrix of M is PSD. The k-locally PSD matrices form a convex cone: $\mathcal{S}^{n,k}$.

E.g. the following 4×4 matrix is 2-locally PSD:

$$\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$$

All 2×2 matrices are rank 1 PSD.

The whole matrix has eigenvalues

$$2, 2, 2, -2$$

k-Locally Positive Semidefinite Cones

NP hard to optimize over $\mathcal{S}^{n,k}$ if k is an input.

These are connected to ‘Restricted Isometry Matrices’, which are norm preserving transformations for sparse vectors.

Comparison Between These Cones

If X is a PSD matrix, then any submatrix of X is PSD.

The k-locally PSD matrices form a chain under inclusion:

$$\mathcal{S}^{n,2} \supseteq \mathcal{S}^{n,3} \supseteq \cdots \supseteq \mathcal{S}^{n,n} = \Sigma_n$$

We can think of these as being successive relaxations of the PSD cone Σ_n, and we might ask **how far away from Σ_n is $\mathcal{S}^{n,k}$?**

We can formulate this eigenvalue question as an optimization problem.

$$\begin{aligned} \text{Minimize } & \lambda_1(X) \\ \text{subject to } & X \in S^{n,k} \\ & \operatorname{tr}(X) = 1 \end{aligned}$$

$\lambda_1(X)$, the minimum eigenvalue of X, is a **concave function** in X.

We can formulate this eigenvalue question as an optimization problem.

$$\begin{aligned} \text{Minimize } & \lambda_1(X) \\ \text{subject to } & X \in S^{n,k} \\ & \operatorname{tr}(X) = 1 \end{aligned}$$

$\lambda_1(X)$, the minimum eigenvalue of X, is a **concave function** in X.

Concave minimization is hard!

Theorem
If $X \in S^{n,k}$, then $\lambda_1(X) \geq \frac{k-n}{(k-1)n}\operatorname{tr}(X)$. This bound is tight.

This implies that if $k = \alpha n$, then $\lambda_1(X) \geq \frac{\alpha' \operatorname{tr}(X)}{n}$.

Hyperbolicity Cones

Hyperbolic Polynomials

We say that a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ is *hyperbolic* with respect to a vector v if for each $x \in \mathbb{R}^n$, the univariate polynomial

$$g(t) = f(x + tv)$$

is **real rooted**, in the sense that all complex roots of g are in fact real.

Example of Hyperbolic Polynomial:

$$f(x, y, z) = x^2 + y^2 - z^2$$

This is a quadratic polynomial, which is hyperbolic in the direction $(0, 0, 1)$.

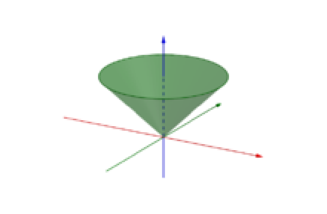
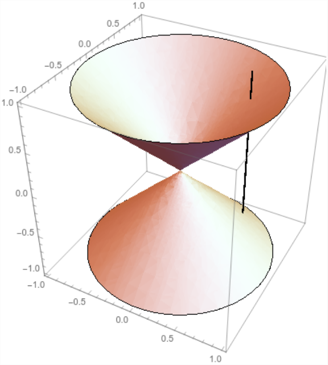


Figure 8: The zeros of a hyperbolic polynomial.

Example of Hyperbolic Polynomial (Cont): If we pick the line through $(1,1,0)$ going in the $(0,0,1)$ direction, we see it hits the zero set of f in 2 places, the right number for a quadratic.



Hyperbolicity Polynomials

Example: $f(x) = \prod_{i=1}^{n} x_i$, and the all ones vector $v = (1, \ldots, 1)$.

For any fixed $x \in \mathbb{R}^n$, look at

$$g(t) = f(x + tv) = (x_1 - t)(x_2 - t)\ldots(x_n - t)$$

The roots of this polynomial are exactly $x_1, \ldots, x_n$, and are therefore real.

Example: The **spectral theorem** says that if X is a Hermitian matrix, then X has real eigenvalues.

The eigenvalues of X are exactly the roots of $\det(X + tI)$, so the determinant on symmetric matrices is a hyperbolic polynomial in $\binom{n+1}{2}$ variables with respect to the vector I.

Hyperbolicity Cones

If f is a hyperbolic polynomial which is hyperbolic with respect to a vector v, we can associate a convex cone called its **hyperbolicity cone, $\Lambda_{f,v}$.**

Hyperbolicity cones can be defined in a number of different ways:

1. $\Lambda_{f,v}$ is the set of $x \in R^n$ so that $f(x + tv) \geq 0$ for all $t \geq 0$.
2. Let $\mathcal{V}_f$ be the set of zeros of f as a subset of $\mathbb{R}^n$. $\Lambda_{f,v}$ is the connected component of $\mathbb{R}^n - \mathcal{V}_f$ containing v.
3. $\Lambda_{f,v}$ is the set of $x \in \mathbb{R}^n$ so that all roots of $f(x + tv)$ are nonnegative.

Examples: If f is the polynomial $\prod_{i=1}^{n} x_i$, then f is hyperbolic with respect to the all ones vector, and $\Lambda_{f,v}$ is the nonnegative orthant in $\mathbb{R}^n$.

If f is the determinant, then f is hyperbolic with respect to the identity, and $\Lambda_{f,v}$ is the PSD cone.

Derivative Relaxations

Lemma
If f is hyperbolic with respect to v, then the directional derivative $D_v f$ is hyperbolic with respect to v. Also, $\Lambda_{f,v} \subseteq \Lambda_{D_v f,v}$.

Examples:

$$e_{n,k} = \sum_{S \subseteq [n], |S| = k} \prod_{i \in S} x_i$$

Let $L^{n,k}$ be the hyperbolicity cone of $e_{n,k}$.

Examples:

$$c_{n,k}(X) = \sum_{S \subseteq [n], |S| = k} \det(X_S)$$

Let $H^{n,k}$ be the hyperbolicity cone of $c_{n,k}$.

Connecting Things Together

Connections between Cones

$c_{n,k}$ is basis invariant!

If X is any symmetric matrix, and λ is any ordering of the eigenvalues of X, then

$$c_{n,k}(X) = e_{n,k}(\lambda)$$

So, $H^{n,k}$ is basis invariant, and is exactly those matrices whose eigenvalues are in $L^{n,k}$.

Connection between $H^{n,k}$ and $\mathcal{S}^{n,k}$

$c_{n,k}(X)$ is the sum of $k \times k$ minors of X, so it is nonnegative on $\mathcal{S}^{n,k}$. That implies that

Theorem

$$\mathcal{S}^{n,k} \subseteq H^{n,k}$$

Theorem
If $X \in H^{n,k}$, then $\lambda_1(X) \geq \frac{k-n}{(k-1)n} \operatorname{tr}(X)$. This bound is tight.

Connection between $H^{n,k}$ and $\mathcal{S}^{n,k}$

We can relax the program above

$$\begin{aligned} &\text{Minimize } \lambda_1(X) \\ &\text{subject to } X \in \mathcal{S}^{n,k} \\ &\qquad tr(X) = 1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} &\text{Minimize } \lambda_1(X) \\ &\text{subject to } X \in H^{n,k} \\ &\qquad tr(X) = 1 \end{aligned}$$

Connection between $H^{n,k}$ and $L^{n,k}$

Questions about eigenvalues of matrices in $H^{n,k}$ descend to questions about $L^{n,k}$, so our program becomes.

$$\begin{aligned} &\text{Minimize } \lambda_1(X) \\ &\text{subject to } X \in \mathcal{S}^{n,k} \\ &\qquad tr(X) = 1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} &\text{Minimize } \lambda_1 \\ &\text{subject to } \lambda \in L^{n,k} \\ &\qquad \sum_{i=1}^{n} \lambda_i = 1 \end{aligned}$$

This is now a linear conical optimization problem!

Connection between $\mathcal{S}^{n,k}$ and $L^{n,k}$

Using symmetry, this last program reduces to a 1 dimensional convex optimization problem that we can exactly solve.

$$\mathrm{OPT} = \frac{k-n}{(k-1)n}$$

It happens that this is also the minimum eigenvalue of the matrix

$$G(n,k) = \begin{pmatrix} \alpha & \beta & \beta & \beta & \dots \\ \beta & \alpha & \beta & \beta & \dots \\ \beta & \beta & \alpha & \beta & \dots \\ \beta & \beta & \beta & \alpha & \dots \\ & & \ddots & & \end{pmatrix},$$

where $\alpha = \frac{1}{n}$ and $\beta = -\frac{1}{(k-1)n}$. This is in $\mathcal{S}^{n,k}$.

Approximation Guarantees

Approximation Guarantees

We learned that $\Sigma_n \subseteq S^{n,k} \subseteq H_n^k$.

Moreover, for any $X \in H_n^k$, $X + \frac{n-k}{(k-1)n} \operatorname{tr}(X) I \in \Sigma_n$.

Dually, $\Sigma_n \supseteq \mathcal{FW}^{n,k} \supseteq (H_n^k)^*$.

Moreover, for any $X \in \Sigma_n$, $X + \frac{n-k}{(k-1)n} \operatorname{tr}(X) I \in (H_n^k)^*$.

Consider the conical optimization problem

$$\begin{aligned} \text{minimize} \quad & \|b\|^2 - \operatorname{tr}(XA^\intercal bb^\intercal A) \\ \text{such that} \quad & \operatorname{tr}(AA^\intercal X) = 1 \\ & X \in C \end{aligned}$$

Let α be the value of this program for when $C = \mathcal{FW}^{n,k}$; let α_ℓ be the value of this program when $C = \Sigma_n$, and α_H be the value when $C = (H_n^k)^*$.

We know $\alpha_\ell \leq \alpha \leq \alpha_H$, but also that if X is the optimal solution when $C = \Sigma_n$, then

$$\frac{(k-1)n}{(k-1)n + (n-k)\operatorname{tr}(AA^\intercal)\operatorname{tr}(X)}\left(X + \frac{(n-k)\operatorname{tr}(X)}{(k-1)n} I\right)$$

is a feasible point when $C = \mathcal{FW}_n^k$.

Plugging this into the program, we get

$$\alpha_\ell \geq \frac{(k-1)n}{(k-1)n + (n-k)\operatorname{tr}(AA^\intercal)\operatorname{tr}(X)}\alpha + \frac{(k-1)n\operatorname{tr}(A^\intercal bb^\intercal A))}{(k-1)n + (n-k)\operatorname{tr}(AA^\intercal)}$$

This can be lower bounded* by

$$\alpha_\ell \geq \frac{(k-1)n}{(k-1)n + (n-k)\chi(AA^\intercal)}\alpha$$

where $\chi(AA^\intercal)$ is the condition number of $AA^\intercal$ (this has not been checked for correctness).

Similarly, we get that

$$\alpha \geq \frac{(k-1)n}{(k-1)n + (n-k)\chi(AA^\intercal)} \alpha_H$$

where $\chi(AA^\intercal)$ is the condition number of $AA^\intercal$ (this has not been checked for correctness).

Details of the proof can be found at
`https://arxiv.org/abs/2012.04031`.