Convex optimization project

Annalle Ben Simon 205798846 bachelor, Evyatar Cohen 206121654 master.

June 17, 2021

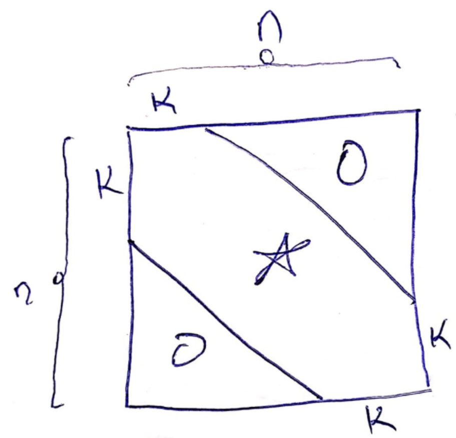
Abstract

We translated the original problem to an equavilent convex problem, that involves finding a global minimum over the set of k-upper-band martices. Then we introduced an algorithm that computes the solution in a close form way, that costs $O\left(nk^3\right)$. In addition, in the case of $k = n - 1$, the problem is much simpler, and we show faster algorithm for this cases.

1 Introduction

Def:

k**-band matrix** is a matrix A where $|i - j| > k \Rightarrow A_{ij} = 0$. We denote the set of all k**-band matrices** as $\mathcal{B}_k$. We can see that $\mathcal{B}_k$ is a linear sub-space and thus convex.



The problem

We wish to solve the following problem. Given $S \in S^n_{++}, k \in [n]$,

$$\begin{aligned} \min_{K} \quad & \mathrm{Tr}\,(SK) - \log |K| \\ \text{s.t.} \quad & K \in \mathcal{B}_k \\ & K \in S^n_{++} \end{aligned}$$

We denote this problem P.

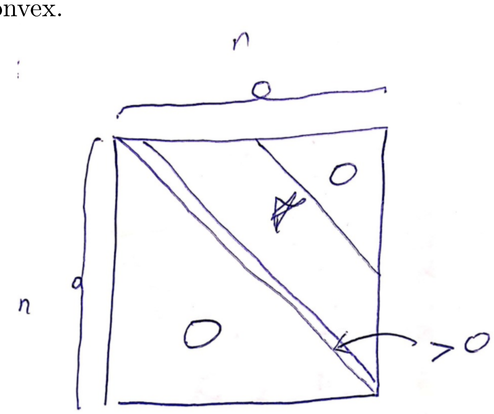
P is a convex problem

We can see that the domain is a convex set as it is an intersection of two convex sets. And that $\mathrm{Tr}\,(SK) - \log |K|$ is convex as $\mathrm{Tr}\,(SK)$ is linear in respect to K and $-\log |K|$ is convex as we saw in class.

2 k-upper-band

Def:

k-upper-band matrix is a matrix U that is both **k-band matrix** and **upper triangular** with real and **positive diagonal entries**. We denote the set of all **k-upper-band matrices** as $\mathcal{U}_k$. He can see that $\mathcal{U}_k$ is a polyhedron and thus convex.



Therom

Every **psd k-band matrix** K ($K \in \mathcal{B}_k \cap S^n_{++}$) can be represented as $K = U^T U$ s.t $U \in \mathcal{U}_k$.

proof

By Cholesky decomposition we know that $K = U^T U$ s.t U is **upper triangular** with real and **positive diagonal entries**. Now we need to show that for every $i, j \in [n]$ s.t. $j - i > k$, $U_{ij} = 0$ (We don't have to show the general $|i-j| > k$ as U is **upper triangular**)

$$\begin{aligned} 0 &= K_{ij} \\ &= \left(U^T U\right)_{ij} \\ &= \sum_{t=1}^{n} U_{it}^T U_{tj} \\ &= \sum_{t=1}^{n} U_{ti} U_{tj} \\ \text{upper triangular} &= \sum_{t=1}^{\min\{i,j\}} U_{ti} U_{tj} \\ [i<j] &= \sum_{t=1}^{i} U_{ti} U_{tj} \end{aligned}$$

Let $j \in \{i+k+1, ..., n\}$ (for $j \leq i+k$, we have $j - i \leq i + k - i = k$). Now by induction on i,

for $i = 1$ we have $0 = K_{1j} = U_{11} U_{1j}$ we know the diagonal entries of U are positive, so $U_{11} \neq 0$, and we are left with $U_{1j} = 0$.

For other is s.t. $j - i > k$, we have by the induction assumption $0 = K_{ij} = \sum_{t=1}^{i} U_{ti} U_{tj} = U_{ii} U_{ij} \Rightarrow U_{ij} = 0$.

Therom

$U \in \mathcal{U}_k \Longrightarrow U^T U \in \mathcal{B}_k \cap S_{++}^n$

Proof

Let $U \in \mathcal{U}_k$, first of all, $U^T U$ is simetric so we only have to check cases where $j - i > k$. Now let $i, j \in [n]$ s.t. $j - i > k$:

$$\left(U^T U\right)_{ij} = \sum_{t=1}^{n} U_{it}^T U_{tj} = \sum_{t=1}^{n} U_{ti} U_{tj}$$

$U_{ti} U_{tj}$ if alwase equal to 0:

1. If $t \leq i$ we have $j - t \geq j - i > k$ and thus $U_{tj} = 0$
2. If $t > i$ we have U_{ti} is under the diagonal and thus $U_{tj} = 0$

So overall,

$$j - i > k \Longrightarrow \left(U^T U\right)_{ij} = 0$$

Now we have to show that $U^T U \in S_{++}$. It is by clear that $U^T U \in S_+$, so we only have to show that $\left|U^T U\right| \neq 0$. $|U| > 0$ as the determinant of a **triangular matrix** in the product of the diagonal entries and the diagonal entries of U are all positive. So

$$\left|U^T U\right| = |U|^2 > 0$$

3 Translating the problem P and relaxation

We know that P is equivalent to

$$\begin{aligned} \min_{U} \quad & \mathrm{Tr}\left(SU^T U\right) - \log\left|U^T U\right| \\ \text{s.t.} \quad & U \in \mathcal{U}_k \end{aligned}$$

As $U^T U \in S_{++} \cap \mathcal{B}_k$ and every $K \in S_{++} \cap \mathcal{B}_k$ can be represented as such U.

We can see that $\log\left|U^T U\right| = \log\left|U^T\right| |U| = \log|U|^2 = 2\log|U|$ (without absolute-value becouse the the derminant of a **upper triangular** matrix in the product of the diagonal entries and U have positive diagonal entries). So we are left with the P' form.

$$\begin{aligned} \min_{U} \quad & \mathrm{Tr}\left(SU^T U\right) - 2\log|U| \\ \text{s.t.} \quad & U \in \mathcal{U}_k \end{aligned}$$

And if we want K^* that solves P, we take $K^* = U^{*T} U^*$

The problem P' is convex

$\mathcal{U}_k$ is convex as it is a polyhedron. Now we will show that $f(U) = \mathrm{Tr}\left(SU^T U\right) - 2\log|U|$ is convex.

$$\mathrm{Tr}\left(SU^T U\right) = \sum_{ij} \left(SU^T\right)_{ij} U_{ij} = \sum_{ij} \langle S_i, U_j\rangle U_{ij}$$

$$\begin{aligned} \frac{\partial \mathrm{Tr}\left(SU^T U\right)}{\partial U_{xy}} &= \sum_{ij} \frac{\partial \langle S_i, U_j\rangle}{\partial U_{xy}} U_{ij} + \sum_{ij} \frac{\partial U_{ij}}{\partial U_{xy}} \langle S_i, U_j\rangle \\ &= \sum_{i} S_{xi} U_{iy} + \langle S_x, U_y\rangle \\ &= 2\langle S_x, U_y\rangle \end{aligned}$$

Thus

$$\frac{\partial \mathrm{Tr}\left(SU^TU\right)}{\partial U}=2SU^T$$

Now to show strong convexity by 1st order, we will show that $\forall X, Y \in \mathbb{R}^{n\times n}$, $X\neq Y$

$$g\left(Y\right)>g\left(X\right)+\left\langle \nabla g\left(X\right),\left(Y-X\right)\right\rangle$$

In our case, $g\left(X\right)=\mathrm{Tr}\left(SX^TX\right)$, $\left\langle A,B\right\rangle=\mathrm{Tr}\left(AB\right)$. So

$$\mathrm{Tr}\left(SY^TY\right)\overset{?}{<}\mathrm{Tr}\left(SX^TX\right)+\mathrm{Tr}\left(2SX^T\left(Y-X\right)\right)$$

$$\begin{aligned}
0 &\overset{?}{<} \mathrm{Tr}\left(SY^TY\right)-\mathrm{Tr}\left(SX^TX\right)-\mathrm{Tr}\left(2SX^T\left(Y-X\right)\right)\\
&=_1 \mathrm{Tr}\left(S\left(Y^TY-X^TX-2X^T\left(Y-X\right)\right)\right)\\
&= \mathrm{Tr}\left(S\left(Y^TY-X^TX-2X^TY+2X^TX\right)\right)\\
&= \mathrm{Tr}\left(S\left(Y^TY-2X^TY+X^TX\right)\right)\\
&= \mathrm{Tr}\left(S\left(Y-X\right)^T\left(Y-X\right)\right)>_2 0
\end{aligned}$$

$=_1$ linarity of trace.

$>_2$ $\left(Y-X\right)^T\left(Y-X\right)\in S_+$, trace of the product of psd and pd matrices in positive.

Thus $\mathrm{Tr}\left(SU^TU\right)$ is convex for all $U\in\mathbb{R}^{n\times n}$. As it is convex in every subset of the indices of the matrix in particular, convex is the k-upper-band indices $i\leq j\leq i+k$.

We know that $\left(-\log x\right)$ is convex so

$$-\log\left|U\right|=\sum_{i=1}^{n}-\log U_{ii}$$

is a sum convex functions. So we got that $f\left(U\right)=\mathrm{Tr}\left(SU^TU\right)-2\log\left|U\right|$ is convex as it is a sum convex functions. Thus P' is a convex problem.

4 Finding the minimum of P'

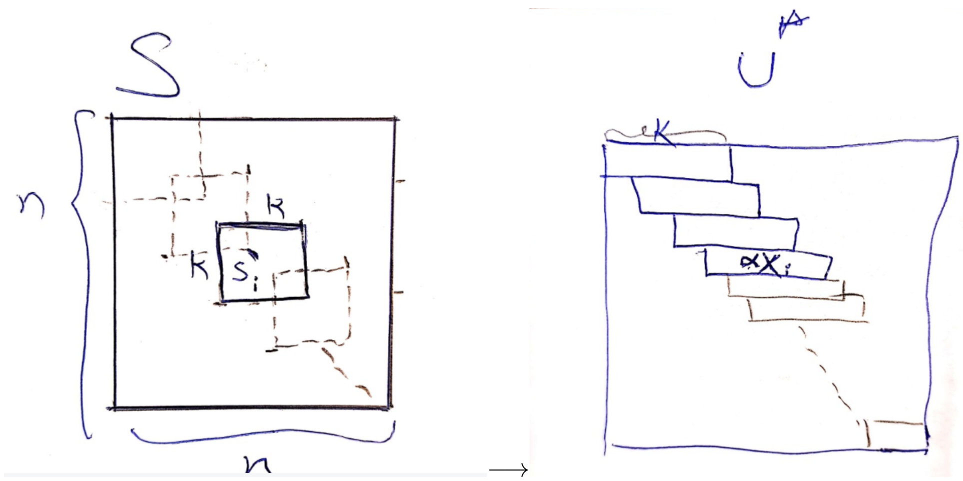
The algorithm to find U is:

1. Initiate U to $n\times n$ zeros matrix.
2. For every i:
   (a) find $x_i\in\mathbb{R}^k$ s.t. $S_{[i:i+k,i:i+k]}x_i=\left[1,0,...,0\right]^T$ (S is $k\times k$ matrix)
   (b) calculate $\alpha=\frac{1}{\sqrt{x_i[0]}}$ ($x_i[0]$ is the first entry of x_i)

(c) $U_{[i,i:i+k]} \leftarrow \alpha x_i$

3. return U

   (a) (To solve for P, return $K = U^T U$)



Proof that the algorithm is properly defined

We have 2 fishey steps in the algorithm: (a) and (b).

Lemma $S \in S_{++} \implies S_{[i:i+k,i:i+k]} \in S_{++}$.

Assume it doesn't, It means that there is a $z \in \mathbb{R}^k$ such that $z^T S_{[i:i+k,i:i+k]} z \leq 0$. We take $z' \in \mathbb{R}^n$, s.t. $z'_{[i:i+k]} = z$ and 0 otherwise.

$$z'^T S z' = z^T S_{[i:i+k,i:i+k]} z \leq 0$$

Contradiction to the fact that $S \in S_{++}$

(a) proof that the system $S_{[i:i+k,i:i+k]} x_i = [1, 0, ..., 0]^T$ has a solution.

$S_{[i:i+k,i:i+k]} \in S_{++}$ (Lemma) and thus invertable and thered is a unique solution $x = S^{-1}_{[i:i+k,i:i+k]} [1, 0, ..., 0]^T$.

(b) proof that $x_i [0] > 0$

$$0 < x_i^T S_{[i:i+k,i:i+k]} x_i = x_i^T [1, 0, ..., 0]^T = x_i [0]$$

The inequlity comes from the fact that $S_{[i:i+k,i:i+k]} \in S_{++}$.

Proof of the correctness of the algorithm - U^* is the minimum of P'

First we can see that U^* is a feasible solution ($U \in \mathcal{U}_k$) as we change only the proper indices. Also note that $x_0 > 0$, thus $U_{ii} = \frac{x_0}{\sqrt{x_0}} > 0$. Now we look on f and it's gradient:

$$\begin{aligned} f(U) &= \mathrm{Tr}\left(SU^TU\right) - 2\log|U| \\ &= \mathrm{Tr}\left(SU^TU\right) - 2\sum_{i=1}^{n}\log U_{ii} \\ \frac{\partial f}{\partial U_{ij}} &= 2\left\langle S_i, U_j\right\rangle - \frac{2}{U_{ij}}\cdot\mathbf{1}_{\boldsymbol{i=j}} \end{aligned}$$

So we need to show that for i, j s.t. $i \le j \le i+k$

$$\left\langle S_i, U_j^*\right\rangle = \frac{1}{U_{ij}^*}\cdot\mathbf{1}_{\boldsymbol{i=j}}\quad {}_{ij}$$

In other words

$$\left\langle S_i, U_j^*\right\rangle = \begin{cases} \frac{1}{U_{ii}^*} & i=j \\ 0 & i<j\le i+k \end{cases}$$

Note that as $U^* \in \mathcal{U}_k$, and S is simetric

$$\left\langle U_{[i]}^*, S_{[j]}\right\rangle = \left\langle U_{[i,i:i+k]}^*, S_{[j,i:i+k]}\right\rangle = \left\langle \frac{x_i}{\sqrt{x_i[0]}}, S_{[j,i:i+k]}\right\rangle$$

where x_i was chosen to satisfy

$$S_{[i:i+k,i:i+k]}x_i = [1,0,...,0]^T$$

so

$$S_{[i:i+k,i:i+k]}\left(ax_i\right) = [a,0,...,0]^T$$

Note that

$$\begin{aligned} \left\langle U_{[i]}^*, S_{[j]}\right\rangle &= \left\langle \alpha x_i, S_{[j,i:i+k]}\right\rangle = \begin{cases} 0 & i<j<i+k \\ \alpha & i=j \end{cases} \\ &= \left\langle \frac{1}{\sqrt{x_i[0]}}x_i, S_{[j,i:i+k]}\right\rangle = \begin{cases} 0 & i<j<i+k \\ \frac{1}{\sqrt{x_i[0]}} & i=j \end{cases} \end{aligned}$$

And this is indeed what we want

$$\left\langle U_{[i]}^*, S_{[i]}\right\rangle = \frac{1}{\sqrt{x_i[0]}} = \frac{\sqrt{x_i[0]}}{x_i[0]} = \frac{1}{U_{ii}^*}$$

To conclude we have

1. $U^* \in \mathcal{U}_k$

2. $\left\langle U^*_{[i]}, S_{[j]} \right\rangle = \begin{cases} \frac{1}{U^*_{ii}} & i = j \\ 0 & i < j \leq i+k \end{cases}$, It meants that $\frac{\partial f}{\partial U_{ij}} (U^*) = 0$ for $i \leq j \leq i+k$. This means that this U^* is the minimum of P'.

5 Run time analysis

The algorithm to find U is:

1. Initiate U to $n \times n$ zeros matrix.

2. For every i:

   (a) find $x_i \in \mathbb{R}^k$ s.t. $S_{[i:i+k,i:i+k]} x_i = [1, 0, ..., 0]^T$ (S is $k \times k$ matrix)

   (b) calculate $\alpha = \frac{1}{\sqrt{x_i[0]}}$ ($x_i[0]$ is the first entry of x_i)

   (c) $U_{[i,i:i+k]} \leftarrow \alpha x_i$

solving a $k \times k$ linear equation system costs $O(k^3)$, we do that n times. Overall the algorithm costs $O\left(nk^3\right)$.

parallization:

Notice that for every i, the computation in independent of the others. So we can use parallization to make to comutation faster.

6 Algorithm for $k = n - 1$

Lemma $S \in S_{++} \Longrightarrow S^{-1} \in S_{++}$

proof: Let $y \in \mathbb{R}^n$,denote $x = S^{-1}y$. This means that $y = Sx$.

$$y^T S^{-1} y = y^T x = x^T S^T x = x^T S x > 0$$

Algorithm for $k = n - 1$

In this case, $\mathcal{B}_k = \mathbb{R}^{n \times n}$.

$$f(K) = \mathrm{Tr}(SK) - \log |K|$$
$$f'(K) = S - K^{-1}$$

$$S = K^{-1}$$
$$K = S^{-1}$$

So S^{-1} is the global minimum of f over all $K \in \mathbb{R}^{n\times n}$. notice that as $S \in S_{++}$,$S^{-1} \in S_{++}$ as well and thus S^{-1} is fisible. Run time of comuting S^{-1} iS $O\left(n^3\right)$ which is better then the general case.