Lecture 5 — Jan 21, 2014

Inst. Mark Iwen *Scribe: Bosu Choi*

Contents

1 Overview

In the last lecture, we discussed how to cast problems as SDP. For this purpose, Schur Complement Theorem is introduced. In this lecture, we prove this theroem and discuss feasiblity and PCA.

2 SDP

2.1 Schur Complement

Theorem 1. *(Schur Complement) Suppose $M \in S^N$ has the block form,*

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}.$$

Then, the following properties must hold

i) $M > 0$ *iff* ($C > 0$ *and* $A - BC^{-1}B^{\mathrm{T}} > 0$).

ii) $C > 0 \Rightarrow$ ($M \geq 0$ *iff* $A - BC^{-1}B^{\mathrm{T}} \geq 0$).

iii) $A > 0 \Rightarrow$ ($M \geq 0$ *iff* $C - B^{\mathrm{T}}A^{-1}B \geq 0$).

iv) $M > 0$ *iff* ($A > 0$ *and* $C - B^{\mathrm{T}}A^{-1}B > 0$).

Proof: If C is invertible, then

$$M = \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BC^{-1}B^T & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} I & BC^{-1} \\ 0 & I \end{pmatrix}^T,$$

and if A is invertible, then

$$M = \begin{pmatrix} I & 0 \\ B^T A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix} \begin{pmatrix} I & 0 \\ B^T A^{-1} & I \end{pmatrix}^T.$$

If $C \in S^N$ is invertible, then

$$M \succ 0 \iff \begin{pmatrix} A - BC^{-1}B^T & 0 \\ 0 & C \end{pmatrix} \succ 0 \iff A - BC^{-1}B^T \succ 0 \;\&\; C \succ 0.$$

All four cases can be proven by similar considerations.

□

2.2 Homework Problem - due Jan 28th(Tue.)

Homework 6 Let $A \in \mathbb{R}^{N \times m}$, $\vec{b} \in \mathbb{R}^N$ and $\vec{d} \in \mathbb{R}^m$. Suppose that $\vec{d}^T \vec{x} > 0$ whenever $A\vec{x} + \vec{b} \geq 0$ (component-wise). Formulate the following problem as an SDP

$$Minimize \; \frac{(\vec{c}^T \vec{x})^2}{(\vec{d}^T \vec{x})} \;\; subject \; to \; A\vec{x} + \vec{b} \geq 0.$$

hint : look at the last example and Schur Complement.

3 Feasibility

Definition 1. *A point $\vec{x} \in \mathbb{R}^m$ is* **feasible** *for an SDP in standard form if it satisfies*

$$F(\vec{x}) = F_0 + \sum x_j F_j \succeq 0.$$

Note. Interior point methods use a sequence of "barrier function" that approximate this constraint. Here, however, we just want to demonstrate that the constraint can be tested "point-wise" in a fairly straightforward fashion. The constraint is not so unmanageable!

Theorem 2. *A matrix $A \in S^N$ is positive definite iff its pivots are all positive after it is reduced to upper triangular form via Gaussian elimination.*

Proof: Reduce $A \succ 0$ to upper triangular form (assuming no row swaps are necessary, which can be accounted for separately). This is equivalent to finding a lower triangular L such that $L^{-1}A = \tilde{U}$. Thus,

$$A = LDU \text{ where}$$

L : lower triangular matrix with ones on the diagonal
D : diagonal matrix with pivots on the diagonal
U : upper triangular matrix with ones on the diagonal

$$\Rightarrow \det(A) = \prod_{j=1}^{N} \lambda_j = \prod_{j=1}^{N} D_{j,j}$$

Here the λ_j are the eigenvalues of A.

Thus, $\det(A)$ is independent of $L's$ and $U's$ off-diagonal entires. We take advantage of this by defining

$$A_{1;2}(t) := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ tL_{2,1} & 1 & 0 & \vdots \\ \cdots & \cdots & \ddots & 0 \\ \cdots & \cdots & \cdots & 1 \end{pmatrix} D \begin{pmatrix} 1 & tU_{1,2} & \cdots & \cdots \\ 0 & 1 & \cdots & \cdots \\ \vdots & 0 & \ddots & \cdots \\ 0 & \cdots & 0 & 1 \end{pmatrix} \in S^N$$

for all $t \in [0, 1]$. Note that $L_{2,1} = U_{1,2}$ must hold.

Furthermore, notice that (i) $A_{1;2}(1) = A$, (ii) $A_{1;2}(0)$ has zeros in its associated $U_{1,2}$ and $L_{2,1}$ entries, (iii) the eigenvalues of $A_{1;2}(t) \in S^N$ are always real numbers, and (iv) the eigenvalues change continuously with t by Weyl's perturbation bounds (Lecture 3, Theorem 2). Thus, $\det(A_{1;2}(t)) = \prod_{j=1}^{N} D_{j,j} = \prod_{j=1}^{N} \lambda_j > 0$ for all $t \in [0, 1]$.

As $t \to 0$, none of the eigenvalues of $A_{1;2}(t)$ can change sign since, if they did, they would have to become zero for some t (and then we'd have $\det(A_{1;2}(t)) = 0$, which can't happen). Thus, sending $U_{1,2}$ and $L_{2,1}$ to zero does not change the sign of any of the eigenvalues. Similarly, we can do this for all off-diagonal entries of L and U (one row/column at a time). In the process, all eigenvalues of A change continuously into the pivots $D_{j,j}$, and can't change sign in the process. Thus, all the pivots must be positive. $\square$

For more details on this test, read up on the Cholesky decomposition.

4 PCA (i.e., using the SVD to approximate a point cloud with a hyperplane)

- Given $P = \{\vec{x}_1, \cdots, \vec{x}_N\} \subseteq \mathbb{R}^D$.

- Our fitness measure for an affine subspace H is $R_\tau(H, P) = (\sum_{\vec{x}_j \in P} d(\vec{x}_j, H)^\tau)^{1/\tau}$ for some $\tau \in \mathbb{R}^+$. Here $d(\cdot, \cdot)$ is Hausdorff distance.

- Assume that P has mean $\frac{1}{N} \sum_{j=1}^{N} \vec{x}_j = \vec{0}$

- For $\tau = 2$ we get a least squares approximation to P.

- Review $\tau = 2$: This can be solved exactly in $O(ND\min\{N, D\})$-time

Goal : Minimize $(R_2(H,P))^2 = \sum_{\vec{x}\in P}^{N} d(\vec{x}_j, H)^2$ over all $n < D$ dimensional subspace H.

- Let $X_P = (\vec{x}_1, \cdots, \vec{x}_N) \in \mathbb{R}^{D\times N}$
- Represent an n-dimensional H(subspace) by a projection matrix $\Pi_H \in \mathbb{R}^{D\times D}$(rank n) that projects onto H.

$$
\begin{aligned}
(R_2(H,P))^2 &= \sum_{\vec{x}_j \in P} d(\vec{x}_j, H)^2 \\
&= \sum_{\vec{x}_j \in P} \|\vec{x}_j - \Pi_H \vec{x}_j\|_2^2 \\
&= \|X_P - \Pi_H X_P\|_F^2 \;\; (Recall\ \|A\|_F^2 = \sum\sum |a_{ij}|^2) \\
&= \|(I - \Pi_H)X_P\|_F^2
\end{aligned}
$$

- We want to minimize this $\|\cdot\|_F$ over all H. Recall that

$$
\begin{aligned}
\|A\|_F &= \sqrt{trace(A^T A)} \\
&= \sqrt{trace(V\Sigma^2 V^T))} \\
&= \sqrt{trace(\Sigma^2)} \;\; (when\ A = U\Sigma V^T, the\ SVD\ of A) \\
&= \sqrt{\sum_{j=1}^{N} \sigma_j(A)^2}
\end{aligned}
$$

  So we want to minimize $\sum_{j=1}^{min(N,D)} \sigma_j((I - \Pi_H)X_P)^2$ over all H.

- If H is n-dimensional, $(I - \Pi_H)$ is $(D - n)$-dimensional projection.
- Let $X_P = U\Sigma V^T$ (SVD of X_P).
- We should let $I - \Pi_H$ project onto the subspace spanned by $D - n$ columns of U associated with $\sigma_D, \cdots, \sigma_{n+1}$.

To minimize $R_2(P, H)$ over H, we want to

1. calculate SVD of X_P, $X_P = U\Sigma V^T$.

2. set $\Pi_H = U_n U_n^T$ where $U_n = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_n & \vec{0} & \cdots \vec{0} \end{pmatrix}$; $U = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_D \end{pmatrix}$.