EE 381V: Large Scale Optimization **Fall 2012**

Lecture 17 — October 25

Lecturer: Caramanis & Sanghavi *Scribe: Natalia Arzeno-Gonzalez, Joyce Ho*

17.1 Last time

In the previous lecture, we defined linear matrix inequalities (LMIs) as:

$$F(x) = F_0 + x_1F_1 + x_2F_2 + ... + x_nF_n \succeq 0, F_i \in S^n$$

We also defined the standard formulation of semidefinite programming (SDP):

$$\begin{aligned} \min \quad & c^{\mathsf{T}}x \\ \text{s.t} \quad & F(x) \succeq 0 \end{aligned} \tag{17.1}$$

We covered two applications of SDP, linear programming and quadratically constrainted quadratic programs.

17.2 Applications of SDP

SDP is useful for finding matrices with eigenvalues that satisfy certain properties. We will look at 4 more examples of SDP. Further applications are covered in Vandenberghe and Boyd's paper on semidefinite programming [2].

17.2.1 Matrix with largest eigenvalue

The optimization problem we'd like to solve is finding the matrix $X \in C$ with the largest eigenvalue. Though the constraint $X \in C$ is not a LMI, the problem can be formulated using SDP as:

$$\max \lambda_{max}(F(X)) \equiv \begin{cases} \min\limits_{X,t} & t \\ \text{s.t} & tI - F(X) \succeq 0 \end{cases}$$

17.2.2 Sum of r largest eigenvalues

Suppose we would like to minimize the sum of the r largest eigenvalues of A. Although the i_{th} eigenvalue, for $i \neq 1$ is neither convex or concave, the sum of the first r eigenvalues is a concave function. Thus the problem can be formulated using SDP [2]:

$$\begin{aligned} \min \quad & rt + \mathrm{Tr}(X) \\ \text{s.t} \quad & tI + X - A \succeq 0 \\ & X \succeq 0 \end{aligned}$$

17.2.3 Sum of singular values of a symmetric but not PSD matrix

Consider the problem of minimizing the sum of the singular values of a symmetric matrix. Note that for a symmetric matrix, the singular values are just the magnitude of the eigenvalues.

$$\begin{aligned} \min_X \quad & \sum_i |\lambda_i(X)| \\ \text{s.t} \quad & X \in C \\ & X = X^{\mathsf{T}} \end{aligned} \tag{17.2}$$

Fact Every symmetric matrix X can be described as the difference of two positive semidefinite matrices ($X = X_+ - X_-$ where $X_+, X_- \in S^n_+$) .

Proof: Using eigenvalue decomposition:

$$\begin{aligned} X &= \underbrace{U}_{\begin{bmatrix} U^+ & U_- \end{bmatrix}} \begin{bmatrix} \Lambda^+ & 0 \\ 0 & \Lambda_- \end{bmatrix} \underbrace{U^{\mathsf{T}}}_{\begin{bmatrix} U^+ & U_- \end{bmatrix}^{\mathsf{T}}} \\ X &= \underbrace{U^+\Lambda^+(U^+)^{\mathsf{T}}}_{\text{PSD}} + \underbrace{U_-\Lambda_-(U_-)^{\mathsf{T}}}_{NSD} \end{aligned}$$

□

Fact If $X \succeq 0$ then $\sum \lambda_i(X) = \mathrm{Tr}(X)$.

Then problem 17.2 can be rewritten as:

$$\begin{aligned} \min \quad & \mathrm{Tr}(X_+) + \mathrm{Tr}(X_-) \\ \text{s.t} \quad & X_+ - X_- \in C \\ & X_+ \succeq 0, X_- \succeq 0 \end{aligned}$$

17.2.4 Sum of Squares

Consider the following optimization problem (univariate polynomial of degree k):

$$\min f(x) = f_0 + f_1 x + f_2 x^2 + .. + f_k x^k \tag{17.3}$$

This is a non-convex optimization problem. However, we can reformulate problem 17.3 by introducing a variable γ:

$$\begin{aligned} \max_{\gamma} \quad & \gamma \\ \text{s.t.} \quad & f(x) - \gamma \geq 0 \quad \forall x \end{aligned} \tag{17.4}$$

Fact A univariate polynomial is positive if and only if it is a sum of squares.

$$f(x) \geq 0 \Leftrightarrow \exists\ h_1(x), h_2(x) \text{ s.t. } f(x) = \sum_i h_i(x)^2$$

The fact can then be used to rewrite the constraint in problem formulation 17.4. Define $x = \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^k \end{bmatrix}$ and $h(x) = \underbrace{\begin{bmatrix} h_0 & h_1 & \dots & h_k \end{bmatrix}}_{h^\mathsf{T}} x$. Then we know that $h(x)^2 = x^\mathsf{T} h h^\mathsf{T} x = x^\mathsf{T} F(\gamma) x$ where $F(\gamma) \succeq 0$. So the constraint:

$$f(x) - \gamma \geq 0 \ \ \forall x \Leftrightarrow \exists\ F(\gamma) \succeq 0 \text{ s.t. } x^\mathsf{T} F(\gamma) x \ \ \forall x$$

The problem formulation now is:

$$\begin{aligned} \max_{\gamma, F} \quad & \gamma \\ \text{s.t.} \quad & f(x) - \gamma = x^\mathsf{T} F x \ \ \forall x \\ & F \succeq 0 \end{aligned}$$

Note that the first constraint is still not a linear matrix inequality, but we can replace it with linear constraints.

$$\begin{aligned} f_0 - \gamma &= F_{00} \\ f_1 &= F_{01} + F_{10} \\ f_2 &= F_{02} + F_{11} + F_{20} \\ &\vdots \end{aligned}$$

17.3 Log Determinant Optimization

So far we have studied SDP when the objective is a linear function. In log determinant optimization, the objective can be a particular nonlinear function. The general form of the problem is

$$\begin{aligned} \min_{x} \quad & c^\mathsf{T} - \log\det G(x) \\ \text{s.t.} \quad & G(x) \succeq 0 \\ & F(x) \succeq 0 \end{aligned} \tag{17.5}$$

Fact $-\log\det G(X)$ is a convex function of $G(X)$.

Proof: We follow the proof in section 3.1 of Boyd and Vandenberghe [1]. To show that $-\log\det X$ is convex in X over $X \succeq 0$, it is sufficient to show that, for $t \in \mathbb{R}$, $-\log\det(Z+tY)$ is convex in t for all $Z \succeq 0$, and any symmetric Y.
Define $f(X) = -\log\det X$ and $g(t) = f(Z+tY)$, restricting g such that $Z+tY \succ 0$. Without loss of generality, assume $t=0$ is inside this interval so that $Z \succ 0$. Then we have

$$\begin{aligned} g(t) = & - \log\det(Z+tY) \\ = & - \log\det(Z^{1/2}(I+tZ^{-1/2}YZ^{-1/2})Z^{1/2}) \\ = & - \sum_{i=0}^{n} \log(1+t\lambda_i) + \log\det Z \end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $Z^{-1/2}YZ^{-1/2}$. The first and second derivatives of the function are:

$$g'(t) = -\sum_{i=1}^{n} \frac{\lambda_i}{1+t\lambda_i}$$

$$g''(t) = \sum_{i=1}^{n} \frac{\lambda_i^2}{(1+t\lambda_i)^2}$$

Notice that $g''(t)$ is always nonnegative, so f is convex. $\square$

One advantage of writing an optimization problem as a log determinant optimization is that algorithms such as Newton's method and gradient descent can be made more efficient for this type of problem.

17.4 Applications of Log Determinant Optimization

17.4.1 Minimum volume ellipsoid

Given points $x_1, \ldots, x_n$ we want to find the minimum volume ellipsoid containing all the points. Recall that an ellipsoid can be thought of as a linear transformation on a sphere, where, in order to preserve volume, the transformation has to be invertible. We can also characterize an ellipsoid by its transformation back into a ball (inverse transform parametrization), as:

$$\varepsilon = \{x : ||Ax+b|| \leq 1\} \tag{17.6}$$

With this parametrization of an ellipsoid, we have vol $\propto \det A^{-1}$. Then, in order to find the ellipsoid with minimum value, we want $\min \det A^{-1}$, which is equivalent to $\min \log\det A^{-1}$, and thus we can write the problem as

$$\begin{aligned} \min_{A} \quad & -\log\det A \\ \text{s.t.} \quad & ||Ax_i - b|| \leq 1 \quad \text{for } i = 1, \ldots, n \end{aligned} \tag{17.7}$$

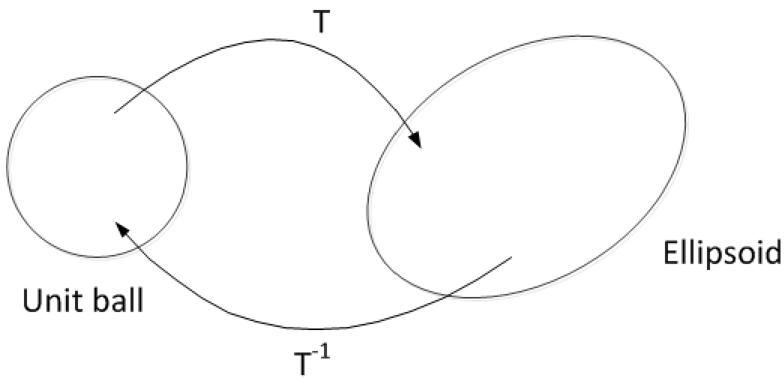

Figure 17.1. The linear transformation from the unit ball to an ellipsoid and the inverse transformation back to a unit ball.

However, the constraints are not written as in linear matrix inequality form. We then use the Schur complement to rewrite the problem as

$$\begin{aligned} \min \quad & -\log\det A \\ \text{s.t.} \quad & \begin{bmatrix} \mathrm{I} & Ax_i - b \\ (Ax_i - b)^\mathsf{T} & 1 \end{bmatrix} \quad \text{for } i = 1, \dots, n \\ & A \succeq 0 \end{aligned} \tag{17.8}$$

17.4.2 Maximum volume ellipsoid inside a polyhedron

Consider the problem of finding the ellipsoid with maximum volume that fits inside the polyhedron $P = \{a_i^\mathrm{T} x \leq b_i \text{ for } i = 1, \dots, m\}$ (Figure 17.2). For this problem, we will parametrize the ellipsoid in terms of the transform from sphere to ellipsoid:

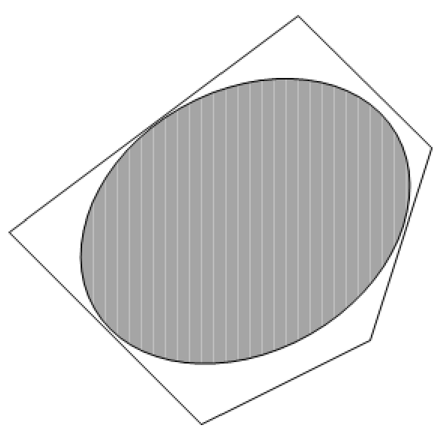

Figure 17.2. An ellipsoid inside a polyhedron.

$$\varepsilon = \{Cy + d : \ ||y|| \leq 1\} \tag{17.9}$$

With this parametrization of an ellipsoid, we have vol $\propto \det C$. The problem can then be written as

$$\begin{array}{ll} \max\limits_{C,d} & \log\det C \\ \text{s.t.} & \left(\sup\limits_{||y||\leq 1} a_i^\top C y\right) + a_i^\top d \leq b_i \end{array} \tag{17.10}$$

where the constraint can be rewritten as

$$||a_i^\top C|| + a_i^\top d \leq b_i$$

$$||a_i^\top C||^2 \leq \left(b_i - a_i^\top d\right)^2$$

As with the minimum volume ellipsoid problem, we have to rewrite our constraints as linear matrix inequalities. The final problem then becomes

$$\begin{array}{ll} \max\limits_{C,d} & \log\det C \\ \text{s.t.} & \begin{bmatrix} (b_i - a_i^\top d)\mathrm{I} & Ca_i \\ (Ca_i)^\top & (b_i - a_i^\top d) \end{bmatrix} \succeq 0 \end{array} \tag{17.11}$$

The form of equation (17.10) is similar to what we saw for robust optimization, with the constraint written as $a_i^\top Cy + a_i^\top \leq b_i \ \forall y$ such that $||y|| \leq 1$.

17.4.3 Gaussian maximum likelihood

For a zero-mean m-dimensional multivariate Gaussian: $\mathcal{N}(0, \Sigma)$,

$$p(x;\Sigma) = \left(\frac{1}{2\pi\det(\Sigma)}\right)^{m/2} \exp\left(\frac{-x^\top\Sigma^{-1}x}{2}\right)$$

$$\log p(x;\Sigma) \propto -\frac{m}{2}\log\det(\Sigma) - \frac{x^\top\Sigma^{-1}x}{2}$$

If we perform a change of variables: $\Theta = \Sigma^{-1}$:

$$\log p(x;\Theta) = \frac{m}{2}\log\det(\Theta) - \frac{1}{2}\langle\Theta, xx^\top\rangle$$

In the maximum likelihood problem, given $x_1, \ldots, x_n$ we want to find

$$\begin{array}{ll} \arg\max\limits_{\Theta} & \prod\limits_{i=1}^{n} p(x_i;\Theta) = \sum\limits_{i} \log p(x_i;\Theta) \\ \text{s.t.} & \Theta \succeq 0 \end{array}$$

The problem in log determinant form then becomes

$$\arg\max_{\Theta} \quad \frac{m}{2}\log\det\Theta - \langle\Theta,\hat{\Sigma}\rangle \tag{17.12}$$

Where $\hat{\Sigma}$ is the empirical covariance matrix of X. This formulation doesn't seem interesting from a probability standpoint, since we would just have $\Sigma = \hat{\Sigma}$. However, this formulation becomes useful in cases where the number of samples is much less than the number of dimensions or when we have a constraint on what the Gaussian should be, such as imposing a prior on Θ.

17.4.4 Gaussian channel capacity

Given a channel $y = x + v$, with input $x \sim \mathcal{N}(0,\Sigma)$ and additive noise $v \sim \mathcal{N}(0,R)$, where R is known,we want to find the Σ that maximizes the capacity of the channel. When $y, x, v \in \mathbb{R}^n$, the model can represent n parallel channels or one channel at n time instants or n frequencies. The channel capacity is the maximum mutual information over Σ subject to power constraints. Mutual information between x and y is given by [3]

$$\frac{1}{2}\left(\log\det(\Sigma+R) - \log\det(R)\right) = \frac{1}{2}\log\det(I + R^{-1/2}\Sigma R^{-1/2})$$

When our constraint is a limit on the average total power, $\mathbb{E}(x^\mathsf{T}x/n) = \mathrm{Tr}(\Sigma)/n \le P$, the problem can be written as

$$\begin{array}{ll} \max & \frac{1}{2}\log\det(I + R^{-1/2}\Sigma R^{-1/2}) \\ \text{s.t.} & \mathrm{Tr}(\Sigma) \le nP \\ & \Sigma \succeq 0 \end{array} \tag{17.13}$$

This is similar in form to the Waterfilling problem we saw in Lecture 9. Let the eigenvalue decomposition of R be $R = V\Lambda V^\mathsf{T}$, we can introduce a new variable $\tilde{\Sigma} = V^\mathsf{T}\Sigma V$ and rewrite the problem as

$$\begin{array}{ll} \max & \frac{1}{2}\log\det(I + \Lambda^{-1/2}\tilde{\Sigma}\Lambda^{-1/2}) \\ \text{s.t.} & \mathrm{Tr}(\tilde{\Sigma}) \le nP \\ & \tilde{\Sigma} \succeq 0 \end{array} \tag{17.14}$$

The solution will have $\tilde{\Sigma}$ as a diagonal matrix since the off-diagonal elements do not appear in the constraints but would decrease the objective.

Bibliography

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar. 2004.

[2] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[3] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(499-533), 1998.