6.854J / 18.415J Advanced Algorithms
Fall 2008

18.415/6.854 Advanced Algorithms	October 29th, 2008

Lecture 15: Interior Point Algorithms for Conic Programming

Lecturer: Michel X. Goemans

1 Introduction

In this lecture, we continue our discussion of the interior-point algorithms for conic programming. In our study of conic programming, we will focus primarily on the following two spaces of possible solutions:

- Linear Programming: $K = \mathbb{R}_n^+$,
- Semi-definite Programming: $K =$PSD$_p$; that is, K is the cone of positive semi-definite matrices ($K = \{X : \; y^* X y \geq 0, \; \forall y \in \mathbb{R}^p, y \neq 0\}$).

We begin by revising the conic program for LP. Let K be a closed convex cone in $\mathbb{R}^n$. Then, K^* is the polar cone of K, which is defined as $K^* = \{s : \langle x, s \rangle \geq 0, \; \forall \; x \in K\}$.

The primal program for conic programming is

$$\begin{aligned} \min \; \langle c, x \rangle \text{ s.t. } & Ax = b, \\ & x \in K. \end{aligned}$$

The dual program for conic programming is

$$\begin{aligned} \max \; \langle b, y \rangle \text{ s.t. } & A^* y + s = c, \\ & s \in K^*, \end{aligned}$$

where A^* is the adjoint matrix of A. This generalizes both LP and SDP.

2 Barrier Functions

In the last lecture, we introduced barrier functions, which are useful in computing the optimal of the conic programs.

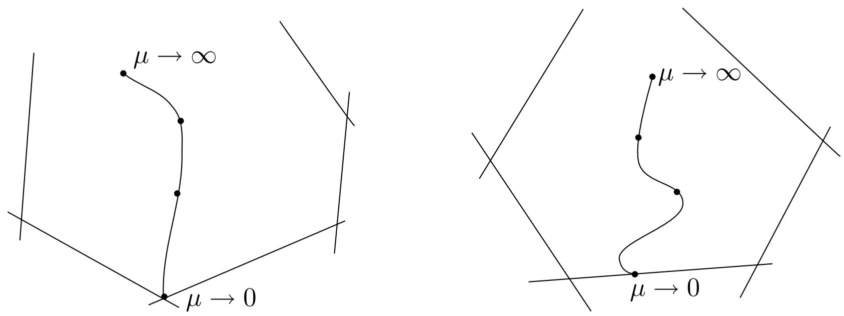
Definition 1 *A function $F : int(K) \to \mathbb{R}$ is a barrier function, if*

1. *F is strictly convex, and*
2. *$(x_k \to x \in \partial_K$ as $k \to \infty) \Rightarrow (F(x_k) \to \infty$ as $k \to \infty)$.*

The last property indicates that F approaches infinity as x_k moves closer to the boundary of K.

To compute the optimal of the conic programs, we first define the barrier primal (BP$_\mu$) and the barrier dual (BD$_\mu$) programs.

$$\begin{aligned} \text{BP}_\mu : \min \; \langle c, x \rangle + \mu F(x) \text{ s.t. } & Ax = b, \\ & (x \in \text{int}(K)). \end{aligned}$$



(a) Primal space. (b) Dual space.

Figure 1: The central path from $\mu = \infty$ to $\mu = 0$, the optimal solution, in the primal and dual space.

As we will start from an interior point, the condition $x \in \text{int}(K)$ will remain true because F becomes infinite closer to the boundary of K according to the definition of the barrier function (1).

$$\begin{aligned} \text{BD}_\mu : \max \ \langle b, y \rangle - \mu F_*(s) \text{ s.t. } & A^* y + s = c, \\ & (s \in \text{int}(K^*)), \end{aligned}$$

where F_* is a barrier function on K^*.

For the special cases of LP and SDP, we introduce two barrier functions, also known as *canonical barrier functions*, which will help us find an optimal for the initial conic programs efficiently.

$$\text{For } K = \mathbb{R}_n^+ : \quad F(X) = -\sum_{j=1}^n \ln(x_j) \tag{1}$$

$$\text{For } K = \text{PSD}_p : \quad F(X) = -\ln(\det(X)), \tag{2}$$

and similarly for F_*.

These functions are in fact very similar. In relation (2), X is a symmetric matrix, so $\det(X)$ is the product of its eigenvalues, say λ_i. Thus, $F(X) = -\sum_j \ln(\lambda_j)$ which is similar to the expression in relation (1).

Consider the set of optimal solutions as $\mu \to 0$. Note that when $\mu = 0$, the barrier programs become the initial conic programs and thus their optimal solutions are optimal solutions for the initial programs as well. The set of solutions as $\mu \to 0$ represents a path, called the *central path*. The path starts at $\mu = \infty$ and we show in the next section that as μ tends to 0, it converges to an optimal solution as illustrated in Figure 1.

3 Optimality Conditions

In this section we show a strong duality relation between the unique optimum solution to (BP_μ) and the unique optimum solution to (BD_μ). We focus on LP and SDP.

Claim 1 *If x is optimum in BP_μ for $K = \mathbb{R}_n^+$ (LP), then there exists y and s such that:*

1. $A^* y + s = c$,
2. $s - \mu x^{-1} = 0$.

Proof: The gradient of F at an optimum has to be normal to the region of feasible solutions because otherwise we would be able to improve on the optimum. Therefore, we have

$$\exists \ y, \ c + \mu \nabla F(x) = A^* y \tag{3}$$

Let us define

$$s = -\mu \nabla F(x). \tag{4}$$

By substituting for s in (3), we obtain $c - s = A^* y$ which proves the first relation.

Next, let us prove the second relation. We substiture $F(x)$ according to its definition (1) in (4) and obtain

$$s_j - \mu \frac{1}{x_j} = 0, j = 1 \dots n$$

and thus

$$s = \mu x^{-1}, \tag{5}$$

which proves the second relation. Note that since both x and μ are positive, it follows that $s \in K^*$. □

Claim 2 *If X is optimum in BP_μ for $K = \mathrm{PSD}_p$ (SDP) then there exists $y \in \mathbb{R}^m$ and $S \in \mathrm{PSD}_p$:*

1. $A^* y + S = C$,
2. $S - \mu X^{-1} = 0$.

Here $A^* y = \sum_i y_i A_i$ as we established last lecture.

Proof: Similarly to the proof of Claim 1, the gradient of F at X has to be normal on the region of feasible solutions because otherwise X would not be optimal. Therefore,

$$\exists\ y,\ c + \mu \nabla F(X) = A^* y, \tag{6}$$

where A^* is the adjoint. We claim that

$$\nabla F(X) = -X^{-1}. \tag{7}$$

Let us show this if X was not necessarily symmetric; our derivation thus will not be fully correct. Observe that

$$\frac{\partial F(X)}{\partial x_{ij}} = -\frac{1}{\det X} \frac{\partial \det X}{\partial x_{ij}} = -\frac{C_{ij}}{\det X},$$

where C_{ij} is the cofactor matrix of element (i, j). The last equality follows (for not necessarily symmetric matrices) from the fact that, for any i,

$$\det X = \sum_j x_{ij} C_{ij}.$$

(For symmetric matrices C_{ij} depends on $x_{ij} = x_{ji}$.) We can thus deduce our claim (7).

By substituting this relation in (6), we have

$$c - \mu X^{-1} = A^* y.$$

By substituting $S = \mu X^{-1}$ in this relation we obtain the desired relations. Furthermore, note that $S \in K^*$ (i.e. $S \succeq 0$) because X is positive definite and μ is positive. □

Note that the dual of the above claims are similar. That is, if y, s are optimal for BD_μ, we have $\exists\ x$, s.t. $Ax = b$ and $x + \mu \nabla F_*(s) = 0$. For LP, we would have $x - \mu s^{-1} = 0$ and thus $xs = \mu$ and for SDP, $X - \mu S^{-1} = 0$ and thus $XS = \mu I$.

4 Duality Gap

Recall that the duality gap between a primal feasible solution x and a dual feasible solution (y, s) is defined as the difference between their values. Furthermore, this expression simplifies to $\langle s, x\rangle$ for conic programs. Let $x(\mu)$ denote the (unique) optimum solution to BP_μ and $(y(\mu), s(\mu))$ the (unique) optimum solution to BD_μ. Since they are feasible for the original primal and dual conic programs, we have that the duality gap is $\langle x(\mu), s(\mu)\rangle$. This gives us an indication of how far we are from optimal. We will show now that this duality gap converges to 0 as μ tends to 0, and thus the central path converges to an optimum solution.

In the previous section, we found that optimality of the primal implies $s = \mu x^{-1}$. For LP, the duality gap will be

$$\langle x(\mu), s(\mu)\rangle = \sum_j x_j(\mu) \cdot \mu / x_j(\mu) = n\mu.$$

Therefore, as $\mu \to 0$, $\langle x(\mu), s(\mu)\rangle \to 0$.

For SDP, we have

$$\langle X(\mu), S(\mu)\rangle = X(\mu) \bullet S(\mu) = Tr(X(\mu) \cdot S(\mu)) = Tr(\mu I_p) = p\mu,$$

where I_p is the identity matrix of dimension of dimension p. Thus, as $\mu \to 0$, $\langle X(\mu), S(\mu)\rangle \to 0$.

5 Barrier Function Properties

Both canonical barrier functions we introduced, (1) and (2), are *self-concordant*. Let us mention the definition of self-concordance, but we shall not elaborate further on this property.

Definition 2 *Let $Q \subseteq \mathbb{R}^n$ be an open convex set. Function $F : Q \to \mathbb{R}^n$ is a self-concordant barrier function if it is at least three times differentiable, convex, and satisfies the properties:*

1. $|D^3F(x)[h, h, h]| \leq 2(D^2F(x)[h, h])^{3/2}$,
2. $|DF(x)[h]|^2 \leq \vartheta D^2F(x)[h, h]$, *and*
3. $F(x) \to \infty$ *as* $x \to \partial Q$.

Here $D^kF(x)[h, \ldots, h]$ is the k-th directional of F at x along the direction $h \in \mathbb{R}^n$, and the constant ϑ is called the parameter of the barrier function. The parameter ϑ determines the speed of the underlying interior point method.

Definition 3 *A function is ν-logarithmically homogenous if $\forall x, \forall \tau > 0, F(\tau x) = F(x) - \nu \ln(\tau)$.*

Remark 1 *The canonical barrier functions defined in (1) and (2) are ν-logarithmically homogenous.*

Proof: First, let us consider the case when $K = \mathbb{R}^n_+$.

$$\begin{aligned} F(\tau x) &= -\sum_{j=1}^n \ln(\tau x_j) \\ &= -n\ln(\tau) - \sum_{j=1}^n \ln(x_j) \\ &= -n\ln(\tau) + F(x) \end{aligned}$$

which proves the remark for $\nu = n$.

Let us consider next $K = \text{PSD}_p$. We have

$$\begin{aligned} F(\tau X) &= -\ln(\det(\tau X)) \\ &= -\ln(\tau^p \det(X)) \\ &= -p\ln(\tau) - \ln(\det(X)) \\ &= -p\ln(\tau) + F(X) \end{aligned}$$

which proves the remark for $\nu = p$. □

6 Interior-Point Algorithms

We begin with an overview of the algorithm.

1. Start with point x_0 for the primal and points y_0, s_0 for the dual and a value for μ of μ_0. These points should be close to the points on the central path: $x(\mu_0)$ for the primal and $s(\mu_0), y(\mu_0)$ for the dual for some definition of closeness that we will introduce.
2. At every step k, decrease μ and compute new points x_k, y_k, s_k close to the points $x(\mu_k)$ for the primal and $s(\mu_k), y(\mu_k)$ for the dual that are located on the central path.
3. As $\mu \to 0$, the solutions converge to the optimal solution.

We need to define a notion of distance and closeness to the central path.

7 Distance to Central Path

To define what we mean by being "close" to the central path, we need a distance function parametrized by μ that measures how close x is to s. This distance $d_\mu(x, s)$ should be equal for the primal and for the dual, and furthermore it should be obviously zero if we are on the central path:

$$d_\mu(x, s) = 0 \text{ if } s + \mu \nabla F(x) = 0.$$

Similar conditions must hold for the dual as well, hence $d_\mu(x, s) = 0$ if $x + \mu \nabla F_*(s) = 0$. Notice that at least for the canonical barrier functions these two conditions are equivalent and hence being on the dual central path implies being on the primal central path and viceversa.

To simplify the calculations we can scale these vectors by $\frac{1}{\mu}$, thus we have that $d_\mu(x, s) = 0$ if $\frac{s}{\mu} + \nabla F(x) = 0 \iff \frac{x}{\mu} + \nabla F_*(s) = 0$.

Finally we define the distance function as the norm of these vectors, and since we are free to choose what norm to use we define a norm respect to x and a norm with respect to s, such that:

$$d_\mu(x, s) = \left\| \frac{s}{\mu} + \nabla F(x) \right\|_x = \left\| \frac{x}{\mu} + \nabla F_*(s) \right\|_s.$$

The norm $\|a\|_b$ is defined as $\|a\|_b = \sqrt{\langle (\nabla^2 F(b))^{-1} a, a \rangle}$ where $\nabla^2 F(b)$ represents the Hessian matrix.

7.1 Distance Function for LP

To compute the Hessian matrix for LP, we plug in the expression for $F(x)$ defined in (1).

$$\nabla F(x) = -x^{-1} = \begin{bmatrix} -\frac{1}{x_1} \\ \vdots \\ -\frac{1}{x_n} \end{bmatrix}.$$

Hence the matrix of second derivatives is,

$$\nabla^2 F(x) = \begin{bmatrix} \frac{1}{x_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{x_n^2} \end{bmatrix}.$$

Finally since the Hessian matrix is a diagonal matrix we can calculate its inverse by taking the inverse of each element in its diagonal,

$$(\nabla^2 F(x))^{-1} = \begin{bmatrix} x_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n^2 \end{bmatrix}.$$

Therefore $\|a\|_b = \sqrt{a^T \begin{bmatrix} b_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_n^2 \end{bmatrix} a}$. We can now evaluate this norm on the vector $\frac{s}{\mu} + x^{-1}$ to obtain

$$\left\| \frac{s}{\mu} + x^{-1} \right\|_x = \sqrt{\sum_j x_j^2 \left(\frac{s_j}{\mu} - \frac{1}{x_j} \right)^2} = \sqrt{\sum_j \left(\frac{x_j s_j}{\mu} - 1 \right)^2}.$$

The same computation can be performed for the dual, which will result in a similar expression.

7.2 Distance Function for SDP

Similarly, for SDP we have

$$\left\| \frac{1}{\mu} X - S^{-1} \right\|_x = \sqrt{Tr\left(\frac{1}{\mu} X^{\frac{1}{2}} S X^{\frac{1}{2}} - I \right)^2} = \sqrt{Tr\left(\frac{1}{\mu} S^{\frac{1}{2}} X S^{\frac{1}{2}} - I \right)^2} = \left\| \frac{1}{\mu} S - X^{-1} \right\|_s.$$

Here, the the last equality holds since $Tr(AB) = Tr(BA)$ even if A and B do not commute.

We can also write this expression in the more compact (but less symmetric) form $\sqrt{Tr(\frac{1}{\mu} SX - I)^2}$.

Finally, the following lemma concerning this metric has been proved (and we will not go over the proof in this lecture).

Lemma 3 *If $d_\mu(x, s) \leq 1$ then $\langle x, s \rangle \leq 2\nu\mu$.*

The lemma suggests that if we keep a distance of at most 1 from the central path, as $\mu \to 0$, the duality gap will become 0 as well which means that we will reach the optimal solution.

8 Follow the Central Path

Suppose that at iteration k we have some value μ_k and x_k, which is close to $x(\mu_k)$; we want to compute $\mu_{k+1} < \mu_k$ and x_{k+1}, which should be close to $x(\mu_{k+1})$.

More concretely at iteration k we have x_k, s_k, y_k, μ_k which are close to the central path, and we want to obtain values $x_{k+1}, s_{k+1}, y_{k+1}, \mu_{k+1}$ which are still close to the central path.

There are several schemes to achieve this goal; one way is to focus on the primal program and on the conditions that we derived in Section 3:

$$
\begin{aligned}
Ax_{k+1} &= b \\
A^* y_{k+1} + s_{k+1} &= c \\
s_{k+1} + \mu_{k+1}\nabla F(x_{k+1}) &= 0
\end{aligned}
$$

Since we do not know the value of x_{k+1}, we can use the Taylor expansion on $\nabla F(x_{k+1})$:

$$\nabla F(x_{k+1}) \sim \nabla F(x_k) + (x_{k+1} - x_k)\nabla^2 F(x_k)$$

Now we have a system of linear equations on x_{k+1}. To solve the system, consider the following definitions:

$$\Delta x = x_{k+1} - x_k \tag{8}$$

$$\Delta y = y_{k+1} - y_k \tag{9}$$

$$\Delta s = s_{k+1} - s_k \tag{10}$$

One can prove that if at iteration k you are "close" to the central path (as defined by the distance function) there is a way to decrease μ by a constant fraction and still remain "close" to the central path. This is formalized by the following theorem, which we will not prove in class.

Theorem 4 *If* $d_{\mu_k}(x_k, s_k) \leq 0.1$ *and*

$$\mu_{k+1} = \frac{\mu_k}{1 + \frac{0.1}{\sqrt{\nu}}}$$

then $d_{\mu_k+1}(x_{k+1}, s_{k+1}) \leq 0.1$.

The method described so far is referred to as **primal path following** since we considered the conditions of $x(\mu_k)$ as defined by the primal, but we could have done the same thing on the dual which leads to a different set of linearized equations.

9 Number of Iterations

We require $\sqrt{\nu}$ iterations to decrease μ by a constant factor, and μ is equal to the duality gap. Therefore decreasing the duality gap to some constant ε starting from x_0, y_0, s_0 requires $O(\sqrt{\nu}\log\frac{<x_0,s_0>}{\varepsilon})$ iterations.

It is interesting to observe that SDP with n^2 variables requires the same number of iterations as LP with n variables. However the resulting system of linear equations that need to be solved per iteration for SDP involves n^2 variables as opposed to n variables as in the case of LP.

10 How Do We Start?

Let us assume that we have a point x that is inside the primal and inside the dual, but it could be very far away from the central path. This point is not suitable to start the algorithm since we need to be close to the central path to be guaranteed to stay close to the central path.

However, there is a nice trick that works when the region is bounded; here we will sketch the informal intuition behind it. Observe that as $\mu \to \infty$ the position of the point $x(\mu)$ does not depend on the objective function. This means that the central paths of all objective functions can be traced back to a common "origin".

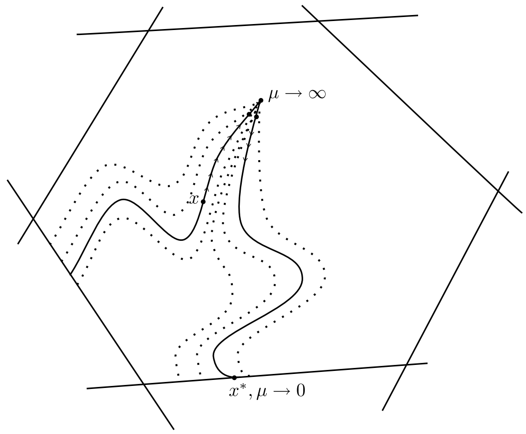


Figure 2: Tracing back from a path that starts at x, jumping to the "correct" central path and starting the algorithm from there.

There is a continuum of central paths and it is easy to find the one that passes through x. Then we can trace this path back towards $\mu = \infty$ until we are close enough to the desired central path. Next, we can follow the desired central path described by the objective function of interest and start the algorithm from there. Figure 2 illustrates this procedure.

For more information regarding interior-point methods for conic programming and its special cases, the reader is referred to the references below.

References

[1] A.S. Nemirovski, Lectures notes on Modern Convex Optimization, 2005. Available at `http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf`.

[2] A.S. Nemirovski and M.J. Todd, "Interior-point methods for optimization", Acta Numerica 17 (2008), 191–234.

[3] J. Renegar, "A mathematical view of interior-point methods in convex optimization", SIAM, 2001.