

ÉCOLE POLYTECHNIQUE DE LOUVAIN

LINMA 2471 : COURSE NOTE

Optimization models and methods II :
Lecture 12 - S13

Authors :

BOUTRY Simon

1396 1000

ROMAINVILLE Sandrine

66051200

WERNER Jolan

58831100

Professor :

GLINEUR Francois

8 december 2015

1 Interior Point Method (IPM)

1.1 Newton's method

Remember we are trying to nullify a Newton step :

$$n(x) = -\nabla^2 f(x)^{-1} \nabla f(x) \quad \delta(x) = \dots$$

where $\delta(x)$ is a measure of accuracy.

2 cases are possible : (ϵ is the wanted accuracy of the optimum)

$$\bullet \delta(x) < 1 \rightarrow \delta(x + n(x)) \leq \left[\frac{\delta(x)}{1 - \delta(x)} \right]^2 \quad (1)$$

this case (1) gives ϵ solution after $\mathcal{O}(\log \log \frac{1}{\epsilon})$

The convergence is quadratic, we only need a few iterates to get the solution.

$$\bullet \forall \delta(x) \rightarrow f(x) - f\left(x + \frac{n(x)}{1 + \delta(x)}\right) \geq \delta(x) - \log(1 + \delta(x)) \quad (2)$$

in (2), if started with $\delta(x) > \tau$, we have $\delta(x) \leq \tau$ after $\lceil \frac{f(x) - f(x)^*}{\tau - \log(1 + \tau)} \rceil$ ($\lceil \cdot \rceil$ means rounded above) iterations ($\mathcal{O}(f(x) - f(x)^*)$).

Note that $\frac{n(x)}{1 + \delta(x)}$ is a damped Newton step.

In practise, we are starting with case 2 and moving to case 1 after some iterations.

1.2 Interior point method

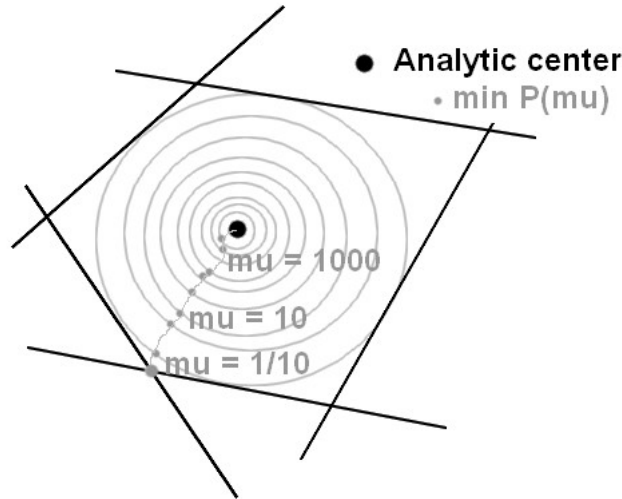
Consider

$$\min c^T x \quad x \in X \quad X \text{ convex}$$

Assume f is self concordant and defined in the interior of X ($\text{int } X$), then all the iterates will be in the set.

Take for example $f(y) = -\sum_{i=1}^m \log(c_i - a_i^T y)$

for domain $\{y | a_i^T y < c \quad \forall i = 1, \dots, m\} \iff \{y | A^T y < c\} = \text{interior of a polytope}$



Definition : analytic center : $x(\infty) = \underset{x \in \text{int}(X)}{\text{argmin}} f(x) \quad (*)$

We call this center "analytic" because it is defined by functions, by calculus. This corresponds to the value of x that minimizes $f(x)$

1.3 Barrier Problem

We are going to use the self-concordant function as a barrier to get to the border of the polytope.

$$(p_\mu) \quad \min_{x \in \text{int} X} \frac{c^T x}{\mu} + f(x) = \min_{x \in \text{int} X} f_\mu(x) \quad (3)$$

In (3) we saw that

- if $\mu = \infty$, we have the same form then (*), thus, the analytic center is a good approach for the minimum.
- if $\mu \rightarrow 0$, the point tends to the solution which is on the boundary segment.
- For every other value of μ , we minimize a balance between the barrier and the objective function

As we decrease the value of μ , we get closer to the effective solution of our problem.

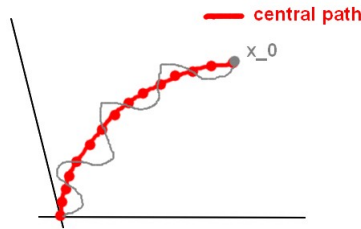
Theorem 1. Assuming that X contains no lines, for $\mu > 0$, $x(\mu) = \min_{x \in \text{int}X} \frac{c^T x}{\mu} + f(x)$ is well defined and unique (simply proved)

Consider $\{x(\mu); \mu > 0\}$ and call it the *central path*. It tries to be as far as possible from the boundary and, in the same time, tends to the minimum.

Property 1. $\lim_{\mu \rightarrow 0} x(\mu) = x^*$ and it is well defined and optimal for the original problem

A possible method would be to target a point $x(\mu)$ and after reaching it jump a bit further on the path and target another point until reaching the true minimum of the objective function. However, this method is too slow (the complexity is not good).

So we will introduce a method which will skip some values of μ but without be too far from the central path. The path is the set of points minimizing P_μ for different values of μ



1.4 Proximity measure

Definition : Proximity measure to the central path point $x(\mu)$:

$$\delta_\mu(x) = \|n_\mu(x)\|_x$$

with $n_\mu(x) = -\nabla^2 f_\mu(x)^{-1} \nabla f_\mu(x) = -\nabla^2 f(x)^{-1} [\frac{c}{\mu} + \nabla f(x)]$

This corresponds to a Newton step for the min function $f_\mu(x)$.

Therefore, we successively find

$$\delta_\mu(x) = \|\nabla f_\mu(x)\|_x^* = \|\frac{c}{\mu} + \nabla f(x)\|_x^*$$

δ_μ is the space between iter points and central path. Since always requiring to get to the targeted point is too slow, we propose to start from a starting point close to the central path and make only one Newton step per target. The algorithm is the following :

Algorithm

Given $x_0, \mu_0, 0 < \tau < 1$ and $0 < \theta < 1$, such that $\delta_{\mu_0}(x_0) \leq \tau$

While $\mu_k > \mu_{final}$

1. $\mu_{k+1} = (1 - \theta)\mu_k$
2. $x_{k+1} = x_k + n_{\mu_{k+1}}(x_k)$
3. $k \leftarrow k + 1$

Definition : Assume that f is a self-concordant function and $\delta(x) \leq \sqrt{\nu}$ for some $\nu \geq 1$. Then, f is a ν -self-concordant barrier.

Example $-\log(x)$ is a 1-self-concordant barrier

Property 1. if f_1 is ν_1 -self-concordant barrier, f_2 is ν_2 -self-concordant barrier then $f_1 + f_2$ is a $(\nu_1 + \nu_2)$ -self-concordant barrier

Property 2. $x \rightarrow f(x)$ is θ -self-concordant barrier then $y \rightarrow f(c - A^T y)$ is also θ -self-concordant barrier

Exercise

1. $-\log(c_i - a_i^T y)$ is 1-self-concordant barrier

2. $-\sum_{i=1}^m \log(c_i - a_i^T y)$ is m -self-concordant barrier

Assuming f is a ν -self-concordant barrier. We have $c^T x(\mu) - c^T x^* \leq \nu\mu$.

For example, if we have a 6-self-concordant barrier,

- $x(\frac{1}{10})$ has accuracy $\frac{6}{10}$
- $x(\frac{1}{1000})$ has accuracy $\frac{6}{1000}$

But in practise we never compute $x(\mu)$.

Theorem 2. Assume $\delta_\mu(x) \leq \tau < 1$. Then

$$c^T x(\mu) - c^T x^* < \frac{\nu\mu}{1 - \tau}$$

This means that μ_{final} is the solution to

$$\frac{\nu\mu}{1 - \tau} = \epsilon$$

1.5 Short-step path following method

Theorem 3. Choosing $\tau = \frac{1}{4}$, $\theta = \frac{1}{16\sqrt{\nu}}$, the method

1. maintains $\delta_{\mu_k}(x_k) \leq \tau \quad \forall k$
2. stops after $\mathcal{O}(\sqrt{\nu} \log(\frac{1}{\epsilon}))$ steps

This method is known as the *Short-step path following method*. The term $\theta = \frac{1}{16\sqrt{\nu}}$ is too small and the step between two μ is also too small. Thus, the complexity is too bad.

How to find x_0 and μ_0 ?

Pick whatever μ_0 you like, depends on the problem (could be heuristic or the central point,...). The chosen x_0 must be close to the central path.

Init : Pick a reasonable μ_0 and compute x_0 by minimization of $(P_{\mu_0}) : \min \frac{c^T x}{\mu_0} + f(x)$ using damped Newton steps. Stop the procedure when distance to the path is less or equal to $\frac{1}{4}$ ($= \tau$).

Let's consider a similar method : long step method

1.6 Long step method

Given $x_0, \mu_0, 0 < \tau < 1, 0 < \theta < 1$ such that $\delta_{\mu_0}(x_0) \leq \tau$
while $(\mu_k > \mu_{final})$

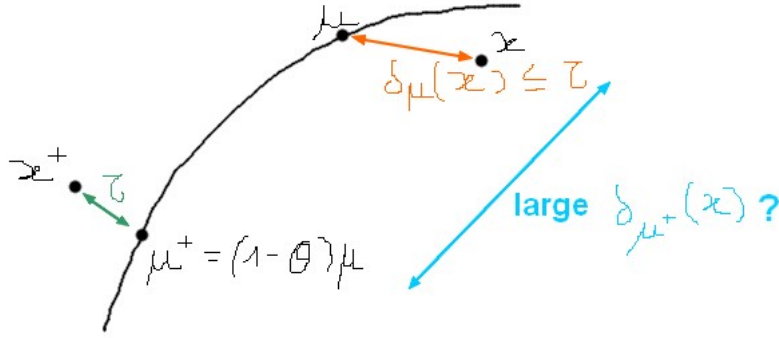
1. $\mu_{k+1} = (1 - \theta)\mu_k$
2. Compute x_{k+1} such that $\delta_{\mu_{k+1}}(x_{k+1}) \leq \tau$ (\star)
3. $k \leftarrow k + 1$

(\star) using possibly several damped Newton steps.

Theorem 4. For any $0 < \tau < 1$ and $0 < \theta < 1$, this method computes a ν -self-concordant barrier and method stops after $\mathcal{O}(\nu \log(\frac{1}{\epsilon}))$ steps (The number of steps means the number of times the method re-enters the while loop (k)).

Remark that this method works good for any τ, θ , which make it flexible. Theoretically, this method is worst than the short step method, but in practice, it rarely needs more than 20 - 50 iterations.

1.7 Proof that short step method works



Now, $\tau = \frac{1}{4}$

$$\delta_{\mu_{k+1}}(x_k) = \left\| \frac{c}{\mu_{k+1}} + \nabla f(x_k) \right\|_{x_k}^*$$

$$\begin{aligned} \text{let } x_k &\rightsquigarrow x \\ x_{k+1} &\rightsquigarrow x^+ \\ \mu_k &\rightsquigarrow \mu \\ \mu_{k+1} &\rightsquigarrow \mu^+ = (1-\theta)\mu \end{aligned}$$

$$\begin{aligned} \delta_{\mu_{k+1}}(x_k) &= \left\| \frac{c}{\mu^+} + \nabla f(x) \right\|_x^* \\ &= \left\| \frac{\mu}{\mu^+} \left(\frac{c}{\mu} + \nabla f(x) \right) + \left(1 - \frac{\mu}{\mu^+} \right) \nabla f(x) \right\|_x^* \\ &\leq \frac{1}{1-\theta} \left\| \frac{c}{\mu} + \nabla f(x) \right\|_x^* + \frac{\theta}{1-\theta} \left\| \nabla f(x) \right\|_x^* \\ &\leq \frac{1/4 + 1/(16\sqrt{\nu}) \sqrt{\nu}}{1-\theta} = \frac{5/16}{1-\theta} \\ &\leq \frac{5/16}{15/16} = \frac{1}{3} \quad \text{using } \nu \geq 1 \end{aligned}$$

Then $\delta_{\mu^+}(x^+) = \left(\frac{1/3}{1-1/3} \right)^2 = \frac{1}{4}$ Starting from a point at a distance lower than $\frac{1}{4}$ from the central path, we move forward to another point at a distance lower than $\frac{1}{4}$ from the central path.

■