COMS 4995-8: Advanced Algorithms (Spring'21) Apr 6, 2021

Lecture 23: Interior Point Method

Instructor: *Alex Andoni* Scribes: *Shrihan Pasikanti, Nikhil Cherukupalli*

1 Introduction

In today's lecture, we dive into the concept of the interior point method by applying the previously visited principles of convexity and Newton's Method.

2 Interior Point Method for Linear Programs

In this section, we try to solve the following problem:

$$\min c^T x$$

$$\text{s.t. } Ax \leq b$$

We take K to be the set of acceptable values. That is, $K := \{x : Ax \leq b\}$.

Definition 1.

$$f(x) = \begin{cases} c^T x & \text{if } x \in K \\ +\infty & \text{if } x \notin K \end{cases}$$

The problem is that f may not be sufficiently smooth on the boundary ∂K.

Definition 2.

$$F(x) = \begin{cases} < \infty & \text{if } x \in K \\ +\infty & \text{if } x \notin K \end{cases}$$

We want $F(x) \to \infty$ as $x \to \partial K$. We will define this function in detail later so that it has the aforementioned property.

Definition 3 (Barrier Function)**.** *Let's define:*

$$f_\eta(x) = \eta c^T x + F(x)$$

where η is a scalar s.t. $\eta > 0$

Using our new definitions, we can restate the original problem.
New goal: Optimize $f_\eta(x)$ (i.e find $\min_{x \epsilon \mathbb{R}} f_\eta(x)$)

Now, we take

$$
\begin{aligned}
F(x) :&= \lg \prod_{i=1}^{m} \frac{1}{b_i - A_i x} \\
&= \sum_{i=1}^{m} \lg \frac{1}{b_i - A_i x} \\
&= \sum_{i=1}^{m} - \lg(b_i - A_i x)
\end{aligned}
$$

where A_i denotes the i-th row of matrix i.

Definition 4. $x_\eta^* = \arg\min_x f_\eta(x) = \arg\min_x \{\eta c^T x + F(x)\}$

Claim 5. *$f_\eta(x)$ is convex*

Proof. It is sufficient to show that the Hessian matrix ($\nabla^2 f_\eta$) is positive semi-definite, which implies that all its eigenvalues ($\lambda_i \geq 0$).
We know:

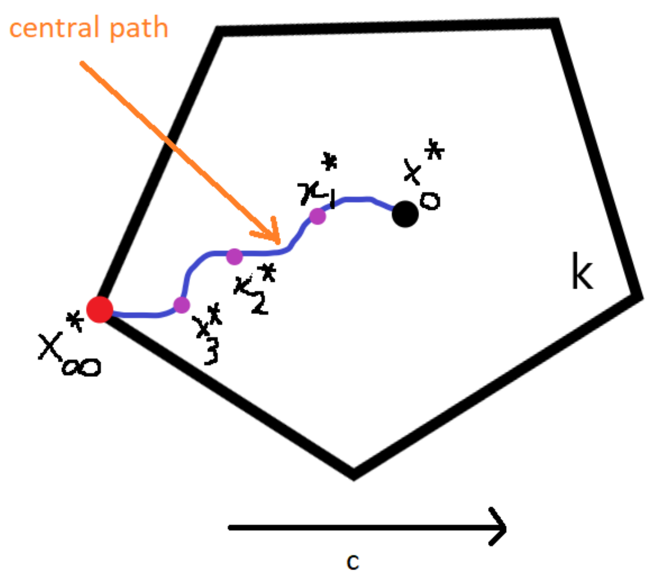
$$
\begin{aligned}
f_\eta(x) &= \eta c^T x + F(x) \\
\nabla f_\eta(x) &= \eta c^T + \sum \frac{A_i}{b_i - A_i x} \\
\nabla^2 f_\eta(x) &= \sum \frac{A_i^T A_i}{(b_i - A_i x)^2}
\end{aligned}
$$

Now, consider an arbitrary vector y ,

$$
\begin{aligned}
y^T \nabla^2 f_\eta(x).y &= \sum \frac{y^T A_i^T A_i.y}{(b_i - A_i x)^2} \\
&= \sum \frac{||A_i y||_2^2}{(b_i - A_i x)^2} \\
&\geq 0
\end{aligned}
$$

Hence, the Hessian is positive semi-definite and therefore, $f_\eta(x)$ is convex. □

Definition 6 (Slack Variable)**.** $\xi := (b_i - A - ix)^2$



Remark 7. *If A is a full-rank matrix ($vol(k) > 0$), then $\lambda_{min}(\nabla f_\eta(x)) > 0$, implying that $f_\eta(x)$ is strongly convex.*

Definition 8 (Analytic Center)**.** *x_0^* is the analytic center of K if $x_0^* = arg\, min_{x \in K} f_o(x)$*

We **observe** that x_η^* is a continuous function w.r.t. η.

Definition 9 (Central Path)**.** *The set $\{x_\eta^* : \eta > 0\}$ is the central path of f.*

Algorithm 0: We solve x_η^* when η is a very large number

1. Recall that Gradient Descent depends on condition number (κ) of F, which could be very large.
2. Newton's method is much faster but requires a "warm" start.

Algorithm 1: The main idea here is to walk along the central path as η increases from $\eta \approx 0$ to a very large value of η.

1. Start at $x_{\eta_0}^*$for $\eta_0 > 0$ such that $[x_{\eta_0}^* \approx x_0^*]$. Note here we **assume** that we know x_0^*
2. For each iteration t : let $\eta_{t+1} = \eta_t.(1 + \alpha)$, for some small value $\alpha > 0$. Then, compute $x_{\eta_{t+1}}^*$ using Newton's method with a warm start setting the initial value to $x_{\eta_t}^*$
3. Terminate once η_t is "sufficiently large enough". Return $x_{\eta_{t_f}}^*$ for where η_{t_f} is the stopping point.

Fact 10. *The value $x_{\eta_{i+1}}^*$ is within the convergence radius of $x_{\eta_i}^*$ (i.e there is only a minor perturbation between them).*

Algorithm 2: Here we present an optimization of *Algorithm 1*.
For each iteration t : $\eta_{t+1} = \eta_t.(1 + \alpha)$, we don't need to compute the optimal value of $x_{\eta_{t+1}}^*$'s for each intermediate η_{t+1}. In fact, each time we take a Newton's step, we are already in the radius of convergence. So, we are merely trying to approximate $x_{\eta_{t+1}}^*$ in this algorithm, which results in a cruder approximation of the central path but also uses fewer steps.

3 Analysis of the Terminal Condition

Here we examine some properties of the terminal condition η_T.

Claim 11.

$$c^T x_\eta^* - c^T x^* \leq m/\eta$$

Before proving the claim, let us make a brief aside.
First, set $\epsilon = \frac{m}{\eta_T}$. Then, $(1+\alpha)\eta_0^T = \frac{m}{\epsilon}$ So, $T = \mathcal{O}(\frac{\lg \frac{m}{\epsilon \eta_0}}{\alpha})$. Note that it suffices to take $\alpha = \frac{1}{Poly(n,m)}$.

Now, let us prove the claim.

Proof. By definition of x_η^*: the gradient convex function $\nabla f_\eta(x^*) = 0$. Then,

$$\iff \eta c + \nabla F(x_\eta^*) = 0$$
$$\frac{-\nabla F(x_\eta^*)}{\eta} = c$$

So, we need to show that:

$$\frac{\nabla F(x_\eta^*)^T}{\eta}\left(x^* - x_\eta^*\right) \leq \frac{m}{\eta}$$

Take any $x, y \in K$. Then,

$$\begin{aligned}
\nabla F(x)^T(y-x) &= \sum_{i=1}^{m} \frac{A_i}{b_i - A_i x}(y-x) \\
&= \sum_{i=1}^{m} \frac{A_i y - A_i x}{b_i - A_i x} \\
&= \sum_{i=1}^{m} \frac{b_i - A_i x - (b_i - A_i y)}{b_i - A_i x} \\
&= m - \sum_{i=1}^{m} \frac{b_i - A_i y}{b_i - A_i x} \\
&\leq m
\end{aligned}$$

Then, if we set $x = x_\eta^*, y = x^*$ and divide the expression above by η, our result is proven. □
Therefore, the assumed value of $T = \mathcal{O}(\frac{\lg \frac{m}{\epsilon \eta_0}}{\alpha})$ is correct.

4 Analysis of Starting Point

Here we analyze the process to compute the true analytical center.

We can obviously compute $x^*_{\eta_0}$ from x^*_0 by Newton's method as long as η_0 is sufficiently small. Suppose we have any $x' \in K$:

Claim 12. *$\forall\, x' \in K \setminus \partial K$ (i.e. x' strictly inside the boundary of K), $\exists\, \eta, c'$ such that:*

$$x' = \arg\min_x \left(\eta c' x + F(x)\right)$$

Proof. The gradient at point $x' = 0$. So,

$$\begin{aligned} \nabla \left(\eta c' x + F(x)\right)\left(x'\right) &= 0 \\ \implies \eta c' + \nabla F(x') &= 0 \\ \implies -\frac{\nabla F(x')}{\eta} &= c' \end{aligned}$$

So, we can merely fix $\eta = 1$ and find c'. □

Algorithm:

1. Given x' , we compute $c' = -\nabla F(x')$ with $\eta = 1$.
2. Walk the central path *back*, and decrease $\eta_{t+1} = \eta_t.(1 - \alpha)$ by taking Newton's step. This roughly approximates x^*_0 (i.e. the true analytical center).
3. Stop at $t = T$ large enough so that we are close enough to x^*_0

Remark 13. *To find a feasible x', we solve a different linear program LP' to* $\min t$ *subject to the constraints $A_i x \leq b_i + t$.*
For this LP', a feasible solution can be: $x = 0$, $t = \max_i(-b_i)$

Remark 14. *It is enough to set $\alpha := \Theta\left(\frac{1}{\sqrt{m}}\right)$*

In the next lecture, we will prove Remark 14, and proceed to talk about *Multiplicative Weights Update* and then switch to *Large Scale Models.*