The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

LECTURE 19

Caramanis & Mokhtari Monday, April 4, 2022

Goal: In this lecture, we first discuss solving convex programs with equality constraints using Newton's method. Then, we study the interior point method to solve constrained convex programs.

1 Convex programs with equality constraints

Consider the following equality constrained optimization problem

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{1}$$

where the objective function $f_0 : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable. Further, note that $\mathbf{A} \in \mathbb{R}^{p \times n}$ and its rank is p where $p < n$. We further assume $\mathbf{b} \in \mathbb{R}^p$.
Note that the optimality conditions of the above problem are given by

$$\mathbf{A}\mathbf{x}^* = \mathbf{b}, \qquad \nabla f_0(\mathbf{x}^*) + \mathbf{A}^\top \mathbf{v}^* = \mathbf{0}$$

where $\mathbf{v}^* \in \mathbb{R}^p$ represents the optimal dual variable.

2 Eliminating equality constraints

Note that any solution of the $\mathbf{A}\mathbf{x} = \mathbf{b}$ can be represented as:

$$\{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}\} = \{\mathbf{F}\mathbf{z} + \hat{\mathbf{x}} \mid \mathbf{z} \in \mathbb{R}^{n-p}\}$$

where $\hat{\mathbf{x}}$ is a particular solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ and the range of $\mathbf{F} \in \mathbb{R}^{n \times (n-p)}$ is the null space of $\mathbf{A}$. Hence, our problem of interest boils down to the following unconstrained problem:

$$\min_{\mathbf{z} \in \mathbb{R}^{n-p}} \quad \tilde{f}(\mathbf{z}) := f_0(\mathbf{F}\mathbf{z} + \hat{\mathbf{x}})$$

By solving this problem we obtain $\mathbf{x}^* = \mathbf{F}\mathbf{z}^* + \hat{\mathbf{x}}$.

2.1 Example

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \; & f_1(x_1) + \cdots + f_n(x_n) \\ \text{s.t} \quad & \sum_{i=1}^{n} x_i = 1 \end{aligned}$$

This problem can be written as

$$\min_{\tilde{\mathbf{x}} \in \mathbb{R}^{n-1}} \quad f_1(\tilde{x}_1) + \cdots + f_{n-1}(\tilde{x}_{n-1}) + f_n(1 - \tilde{x}_1 - \tilde{x}_2 - \cdots - \tilde{x}_{n-1})$$

Indeed, by finding $\tilde{\mathbf{x}}^*$ we have $x_i^* = \tilde{x}_i^*$ for $i = 1, \ldots, n-1$ and $x_i^* = 1 - \sum_{i=1}^{n-1} \tilde{x}_i^*$. In this case, $\mathbf{F} = \begin{bmatrix} \mathbf{I} \\ \mathbf{1}^\top \end{bmatrix} \in \mathbb{R}^{n \times n-1}$ and $\hat{\mathbf{x}} = [0, \ldots, 0, 1]$.

3 The Newton step for equality constrained problems

Suppose $\mathbf{x} = \mathbf{x}_k$ is the current iterate and we need to find the next iterate $\mathbf{x}^+ = \mathbf{x} + \mathbf{v}$. In Newton step, we replace the objective function by its quadratic approximation around the current point $\mathbf{x}$ and ensure the updated variable is feasible:

$$\begin{aligned} \min \quad & \hat{f}_0(\mathbf{x}+\mathbf{v}) = f_0(\mathbf{x}) + \nabla f_0(\mathbf{x})^\top \mathbf{v} + (1/2)\mathbf{v}^\top \nabla^2 f_0(\mathbf{x})\mathbf{v} \\ \text{s.t} \quad & \mathbf{A}(\mathbf{x}+\mathbf{v}) = \mathbf{b} \end{aligned}$$

Here $\mathbf{v} = \Delta\mathbf{x}_{nt}$. By writing the optimality condition for the above problem we obtain that

$$\nabla f_0(\mathbf{x}) + \nabla^2 f_0(\mathbf{x})\mathbf{v} + \mathbf{A}^\top \mathbf{w} = \mathbf{0}, \qquad \mathbf{A}(\mathbf{x}+\mathbf{v}) = \mathbf{b}$$

Hence the Newton step can be written as the solution of this **KKT system**

$$\begin{bmatrix} \nabla f_0^2(\mathbf{x}) & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_{nt} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} -\nabla f_0(\mathbf{x}) \\ \mathbf{0} \end{bmatrix}$$

It further can be verified that if we define

$$\lambda(\mathbf{x}) = \sqrt{\Delta\mathbf{x}_{nt}^\top \nabla^2 f_0(\mathbf{x}) \Delta\mathbf{x}_{nt}}$$

then we have

$$f(\mathbf{x}) - p* \;\le\; \frac{\lambda(\mathbf{x})^2}{2}$$

Newton's method with equality constraint is the following:

Start with a feasible point $\mathbf{x}_0$ satisfying $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$

- Newton Step: Solve the KKT condition system and find $\Delta\mathbf{x}_{nt}$. Then, compute $\lambda(\mathbf{x})$.
- If $\lambda(\mathbf{x})^2/2 \le \epsilon$ then stop.
- Line-search: choose step size η based on the backtracking line-search
- Update $\mathbf{x}^+ = \mathbf{x} + \eta\Delta\mathbf{x}_{nt}$

Important points:

- All iterates are feasible
- The algorithm is a descent method $f_0(\mathbf{x}_{k+1}) < f_0(\mathbf{x}_k)$

Remark 1. *If we run Newton's method for the eliminated problem $\tilde{f}(\mathbf{z}) := f_0(\mathbf{F}\mathbf{z} + \hat{\mathbf{x}})$, then the iterates coincide with the iterates of Newton method with equality constraints. In other words, if we started with $\mathbf{x}_0 = \mathbf{F}\mathbf{z}_0 + \hat{\mathbf{x}}$, then the iterates satisfy:*

$$\mathbf{x}_k = \mathbf{F}\mathbf{z}_k + \hat{\mathbf{x}}$$

Note that $\Delta\mathbf{z}_{nt} = -(\mathbf{F}^\top \nabla^2 f_0(\mathbf{x}_k)\mathbf{F})^{-1}\mathbf{F}^\top \nabla f_0(\mathbf{x}_k)$. Hence,

$$\mathbf{F}\Delta\mathbf{z}_{nt} = -\mathbf{F}(\mathbf{F}^\top \nabla^2 f_0(\mathbf{x}_k)\mathbf{F})^{-1}\mathbf{F}^\top \nabla f_0(\mathbf{x}_k)$$

Now if we set $\Delta\mathbf{x}_{nt} = \mathbf{F}\Delta\mathbf{z}_{nt}$ and $\mathbf{w} = -(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}(\nabla f_0(\mathbf{x}_k) + \nabla^2 f_0(\mathbf{x}_k)\Delta\mathbf{x}_{nt})$, we obtain that they satisfy the KKT system. Hence $\Delta\mathbf{x}_{nt} = \mathbf{F}\Delta\mathbf{z}_{nt}$.
Therefore, we don't need to develop a new analysis for Newton's method with equality constraint.

Remark 2. *If f_0 is self-concordant, then $\tilde{f}(\mathbf{z})$ is also self-concordant and hence the complexity of Newton's method with equality constraint is* $\log\log(1/\epsilon) + (f(\mathbf{x}_0) - f^*)/\gamma$

3.1 Solving the KKT systems

Approach 1. Since solving the KKT systems is equivalent to solving a linear system with $n+p$ equations and $n+p$ variables and the matrix is symmetric, we can simply use the $LDL^\top$ factorization which has a cost of $(1/3)(n+p)^3$.

Approach 2. We can use the elimination technique. Consider the system:

$$\begin{bmatrix} \mathbf{H} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} -\mathbf{g} \\ \mathbf{0} \end{bmatrix}$$

Then we have

$$\mathbf{H}\mathbf{v} + \mathbf{A}^\top \mathbf{w} = -\mathbf{g} \qquad \Rightarrow \qquad \mathbf{v} = -\mathbf{H}^{-1}(\mathbf{A}^\top \mathbf{w} + \mathbf{g})$$

and we have

$$\mathbf{A}\mathbf{v} = 0$$

Therefore

$$\mathbf{A}(-\mathbf{H}^{-1}(\mathbf{A}^\top \mathbf{w} + \mathbf{g})) = \mathbf{0} \qquad \Longleftrightarrow \qquad \mathbf{w} = -(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{H}^{-1}\mathbf{g}$$

The overall cost of this procedure is $\mathcal{O}(p^3 + p^2 n)$.

4 Interior Point Methods

Consider the following convex program with equality and inequality constraints:

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) \\ \text{s.t} \quad & f_i(\mathbf{x}) \le 0, \qquad i = 1, \dots, m \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{2}$$

where here the functions are all convex and twice differentiable, and m is the number of inequality constraints. Moreover, $\mathbf{A} \in \mathbb{R}^{p \times n}$ whose rank is p and $p < n$. We further assume that the problem is strictly feasible: There exists $\hat{\mathbf{x}}$ such that

$$\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}, \qquad f_i(\hat{\mathbf{x}}) < 0, \qquad i = 1, \dots, m$$

hence, strong duality holds and dual optimum is attained.

If we define the indicator function $I_- : \mathbb{R} \to \mathbb{R}$ as

$$I_-(u) = 0 \quad u \le 0, \qquad\qquad I_-(u) = \infty \quad u > 0$$

then we can write the above problem as

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) + \sum_{i=1}^m I_-(f_i(\mathbf{x})) \\ \text{s.t} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Now note that the indicator function can be approximated using the function $-(1/t)\log(-u)$ for some $t > 0$. Indeed, this smooth approximation is more accurate if t is larger.

$$\begin{aligned} \min \quad & f_0(\mathbf{x}) - \frac{1}{t}\sum_{i=1}^m \log(-f_i(\mathbf{x})) \\ \text{s.t} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{3}$$

Remark 3. *Selecting a large value of t makes the approximation more accurate, but at the same time it makes the approximation problem harder for Newton's method.*

The following function

$$\phi(\mathbf{x}) := -\sum_{i=1}^{m} \log(-f_i(\mathbf{x}))$$

with domain $\{\mathbf{x} \mid f_i(\mathbf{x}) < 0),\ i = 1, \ldots, m\}$ is called the *log-barrier.* This function is indeed convex and twice differentiable with the following gradient and Hessian:

$$\nabla\phi(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}), \qquad \nabla^2\phi(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{f_i(\mathbf{x})^2} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^\top + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}).$$

Remark 4. *Note that (3) is an approximation of (2). Larger t makes the solutions closer, but at the same time it makes the problem $f_0(\mathbf{x})+(1/t)\phi(\mathbf{x})$ harder for Newton's method as Hessian is changing very rapidly. Hence, to address this issue, we solve a sequence of problems and geometrically increase the value t. In the meantime, we use the solution of the previous round with smaller t, as a the initial iterate of Newton's method with larger t.*

An equivalent version of Problem (3) is given by

$$\begin{aligned} \min \quad & t f_0(\mathbf{x}) + \phi(\mathbf{x})) \\ \text{s.t} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{4}$$

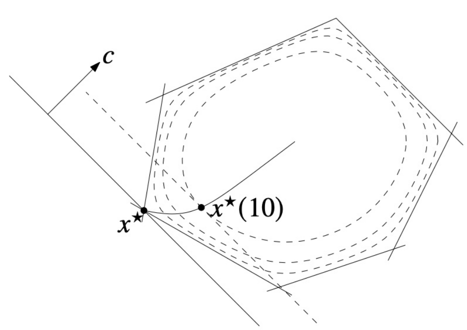
For now suppose the above problem has the required conditions that can be solved using Newton's method for properly selected t. We further assume that for any $t > 0$, the above problem has a unique solution denoted by $\mathbf{x}^*(t)$.

Definition 1. *The sequence of solutions $\mathbf{x}^*(t)$ for different values of $t > 0$, is called the **central path** of Problem (4).*

Central Path for Linear Programs. Consider the following linear program defined as

$$\min \mathbf{c}^\top \mathbf{x}, \qquad s.t. \quad \mathbf{a}_i^\top \mathbf{x} \le b_i \quad i = 1, \ldots, 6$$

Basically approximate a polyhedron by a smooth convex set that is within the polytope, when we follow the problem in (4).



Note that the hyperplane $\mathbf{c}^\top(\mathbf{x} - \mathbf{x}^*(t)) = 0$ is tangent to the level curve of ϕ at $\mathbf{x}^*(t)$.

4.1 Dual points from central path

As we assume that strict feasibility holds and duality gap is zero we have

$$\mathbf{A}\mathbf{x}^*(t) = \mathbf{b}, \qquad f_i(\mathbf{x}^*(t)) < 0, \qquad i = 1, \dots, m.$$

and there exists a $\hat{\mathbf{v}}$ such that

$$t\nabla f_0(\mathbf{x}^*(t)) + \nabla\phi(\mathbf{x}^*(t)) + \mathbf{A}^\top \hat{\mathbf{v}} = \mathbf{0}.$$

which implies that

$$t\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x}^*(t))} \nabla f_i(\mathbf{x}^*(t)) + \mathbf{A}^\top \hat{\mathbf{v}} = \mathbf{0}.$$

Now we claim that

$$\lambda_i^*(t) = -\frac{1}{tf_i(\mathbf{x}^*(t))}, \qquad i = 1, \dots, m \qquad \boldsymbol{\nu}^*(t) = \mathbf{w}/t$$

are dual feasible and $\mathbf{x}^*(t)$ minimizes their corresponding Lagrangian. Why? cause it can be verified that

$$\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) \nabla f_i(\mathbf{x}^*(t)) + \mathbf{A}^\top \boldsymbol{\nu}^*(t) = \mathbf{0}.$$

Now considering the fact that the Lagrangian is given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \boldsymbol{\nu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b})$$

we have

$$\begin{aligned} g(\lambda_i^*(t), \boldsymbol{\nu}^*(t)) &= f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(\mathbf{x}^*(t)) + \boldsymbol{\nu}^*(t)^\top (\mathbf{A}\mathbf{x}^*(t) - \mathbf{b}) \\ &= f_0(\mathbf{x}^*(t)) - \frac{m}{t} \end{aligned}$$

Now using the fact that $p^* \geq g(\lambda_i^*(t), \boldsymbol{\nu}^*(t))$, we have

$$f_0(\mathbf{x}^*(t)) - p^* \leq \frac{m}{t}$$

The Barrier Method is given by

Start with a strictly feasible point $\mathbf{x}_0$ and select $\mu > 1$, $t = t_0 > 0$, and $\epsilon > 0$.

- Centering Step: Solve the centering problem in (4) (using Newton's method) and find $\mathbf{x}^*(t)$
- Update $\mathbf{x} = \mathbf{x}^*(t)$
- If $m/t < \epsilon$ stop
- Increase the barrier parameter $t = \mu t$

It can be easily verified that this method terminates after running for the following number of iterations

$$\left\lceil \frac{\log \frac{m}{\epsilon t_0}}{\log \mu} \right\rceil$$

4.2 Complexity analysis for self-concordance functions

Newton's method can be used for the barrier method we need the following conditions:

- The sublevel sets of f_0 (on its domain) are bounded.
- The function $tf_0(\mathbf{x}) + \phi(\mathbf{x})$ should be self-concordant.
- The function $tf_0(\mathbf{x}) + \phi(\mathbf{x})$ must have closed sub-level sets for $t \geq t_0$.

All of the above conditions are satisfied for LP, QP, QCQP!

Suppose $\mathbf{x}$ is the solution for problem with t and we use it as initial point for problem with parameter μt. Then. the number of Newton iterations for solving the problem with parameter μt is

$$N_{newton} \leq \frac{\mu t f_0(\mathbf{x}) + \phi(\mathbf{x}) - \mu t f_0(\mathbf{x}^+) - \phi(\mathbf{x}^+)}{\gamma} + \log\log\left(\frac{m}{\mu t}\right)$$

Now note we have $\mathbf{x} = \mathbf{x}^*(t)$ and $\mathbf{x}^+ = \mathbf{x}^*(\mu t)$. Moreover,

$$\lambda_i^*(t) = -\frac{1}{tf_i(\mathbf{x}^*(t))}, \qquad i = 1, \ldots, m \qquad\qquad \boldsymbol{\nu}^*(t) = \mathbf{w}/t$$

Hence,

$$\begin{aligned}
&\mu t f_0(\mathbf{x}) + \phi(\mathbf{x}) - \mu t f_0(\mathbf{x}^+) - \phi(\mathbf{x}^+) \\
&= \mu t f_0(\mathbf{x}^*(t)) - \mu t f_0(\mathbf{x}^*(\mu t)) - \sum_{i=1}^m \log(-f_i(\mathbf{x}^*(t))) + \sum_{i=1}^m \log(-f_i(\mathbf{x}^*(\mu t))) \\
&= \mu t f_0(\mathbf{x}^*(t)) - \mu t f_0(\mathbf{x}^*(\mu t)) - \sum_{i=1}^m \log\left(\frac{1}{t\lambda_i^*(t)}\right) + \sum_{i=1}^m \log(-f_i(\mathbf{x}^*(\mu t))) \\
&= \mu t f_0(\mathbf{x}^*(t)) - \mu t f_0(\mathbf{x}^*(\mu t)) + \sum_{i=1}^m \log(-t\lambda_i^*(t) f_i(\mathbf{x}^*(\mu t))) \\
&= \mu t f_0(\mathbf{x}^*(t)) - \mu t f_0(\mathbf{x}^*(\mu t)) + \sum_{i=1}^m \log(-\mu t\lambda_i^*(t) f_i(\mathbf{x}^*(\mu t))) - m\log\mu \\
&\leq \mu t f_0(\mathbf{x}^*(t)) - \mu t f_0(\mathbf{x}^*(\mu t)) - \mu t \sum_{i=1}^m (\lambda_i^*(t) f_i(\mathbf{x}^*(\mu t))) - m - m\log\mu \\
&= \mu t f_0(\mathbf{x}^*(t)) - \mu t \left(f_0(\mathbf{x}^*(\mu t)) + \sum_{i=1}^m \lambda_i^*(t) f_i(\mathbf{x}^*(\mu t)) + (\boldsymbol{\nu}^*(t))^\top (\mathbf{A}\mathbf{x}^*(\mu t) - \mathbf{b}) \right) - m - m\log\mu \\
&\leq \mu t f_0(\mathbf{x}^*(t)) - \mu t g(\boldsymbol{\lambda}^*(t), \boldsymbol{\nu}^*(t)) - m - m\log\mu \\
&\leq \mu m - m - m\log\mu \\
&= m(\mu - 1 - \log\mu)
\end{aligned}$$

Hence, the total number of Newton iterations will be

$$N = \left(\frac{m(\mu - 1 - \log\mu)}{\gamma} + \log\log(1/\epsilon)\right) \left\lceil \frac{\log\frac{m}{\epsilon t_0}}{\log\mu} \right\rceil \qquad \text{where} \qquad \frac{1}{\gamma} = M_f^2 \frac{20 - 8\alpha}{\alpha\beta(1-2\alpha)^2}$$

If we set $\mu = 1 + 1/\sqrt{m}$ then we have

$$N = \mathcal{O}\left(\sqrt{m}\log\frac{m}{t_0\epsilon}\right)$$