
Lecture 17

In this lecture we introduce interior point methods for solving nonlinear convex optimization problems of the form

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{f}(\mathbf{x}) \leq \mathbf{0} \\ & && \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{1}$$

with $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{p \times d}$ and $\mathbf{f} = (f_1, \dots, f_m)^\top: \mathbb{R}^n \rightarrow \mathbb{R}^m$. The particular form of interior point methods we discuss is the *barrier method*, which differs slightly from the primal-dual method discussed for linear programming.

Recall the Karush-Kuhn-Tucker (KKT) conditions

$$\begin{aligned} \mathbf{f}(\mathbf{x}^*) &\leq \mathbf{0} \\ \mathbf{A}\mathbf{x}^* &= \mathbf{b} \\ \boldsymbol{\lambda}^* &\geq \mathbf{0} \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad 1 \leq i \leq m \\ \nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla_{\mathbf{x}} f_i(\mathbf{x}^*) + \mathbf{A}^\top \boldsymbol{\mu}^* &= \mathbf{0}, \end{aligned} \tag{2}$$

where $\nabla_{\mathbf{x}}$ denotes the gradient of a function with respect to \mathbf{x} . As discussed in the previous lecture, for convex problems these are necessary and sufficient conditions for $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^p$ being a set of optimal primal and dual solutions. By this we mean that \mathbf{x}^* is a minimizer of (1) and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a maximizer of the Lagrange dual function $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ subject to $\boldsymbol{\lambda} \geq \mathbf{0}$, which in turn was defined as the infimum

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}),$$

where $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x}) + \boldsymbol{\mu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b})$ is the Lagrangian. In this section we make the additional assumption that the functions f and f_i , $1 \leq i \leq m$, are two time continuously differentiable. The reason is that we want to apply variants of Newton's method to the KKT conditions. For later reference we list some of the more popular forms of convex optimization problems that fall into our scope, with the associated KKT conditions.

Example 17.1. (LP) For a linear programming problem of the form

$$\begin{aligned} & \text{minimize} && \langle c, x \rangle \\ & \text{subject to} && Ax = b \\ & && x \geq 0, \end{aligned}$$

the KKT conditions have the form

$$\begin{aligned} & Ax = b \\ & A^\top y + s = c \\ & s \geq 0 \\ & x \geq 0 \\ & x_i s_i = 0, \quad 1 \leq i. \end{aligned}$$

Example 17.2. (QP) A *quadratic programming problem* is a problem with quadratic objective and linear constraints of the form

$$\begin{aligned} & \text{minimize} && x^\top P x + q^\top x + r \\ & \text{subject to} && Ax = b \\ & && Bx \leq c. \end{aligned}$$

One typical example is portfolio optimization.

Many other problems can be cast in the required form, and we will see later that the framework also encompasses semidefinite programming.

17.1 The logarithmic barrier

The idea of barrier methods is to replace the constraints of the optimization problem (1) with a parametrized system of equality constraints, such that the solutions of the equality constraint system can be found, and converge to the solution of the original system. The first observation is that we can replace the inequality constrained system

$$\text{minimize } f(x) \quad \text{subject to } Ax = b, \quad f_i(x) \leq 0, \quad 1 \leq i \leq m,$$

with the equality constrained, but nonsmooth, problem

$$\text{minimize } f(x) + \sum_{i=1}^m I_-(f_i(x)), \quad \text{subject to } Ax = b,$$

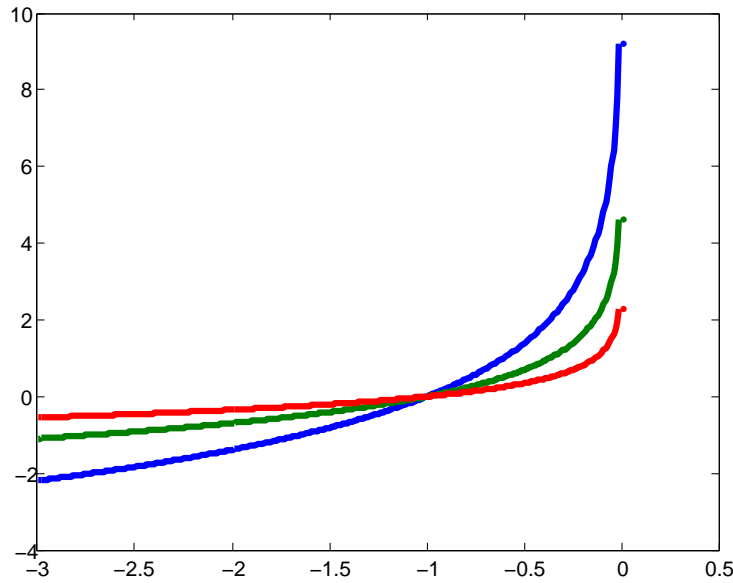
where

$$I_-(t) = \begin{cases} 0 & t \leq 0, \\ \infty & t > 0. \end{cases}$$

Clearly, the objective is $f(\mathbf{x})$ whenever the inequality constraints $f_i(\mathbf{x}) \leq 0$ are satisfied, and ∞ if not, so that points that do not satisfy the constraints are forbidden. Since this function is not easy to deal with analytically, we *approximate* it with the function

$$\hat{I}(u) := -\frac{1}{t} \ln(-u)$$

for suitable values t . Figure ?? shows the approximating functions for $t = 0.5, 1, 2$. As is easily seen, as t increases the approximation becomes better. An approximation



to the nonsmooth problem is given by the following equality constrained problem,

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) - \sum_{i=1}^m \frac{1}{t} \log(f_i(\mathbf{x})) \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

The domain \mathcal{D} of this problem is the set of points such that the inequalities $f_i(\mathbf{x}) < 0$ are strictly satisfied. One can also verify that the objective function is convex, giving rise to a convex optimization problem. The function

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

is called the *logarithmic barrier function* of the system of inequalities $f_i(\mathbf{x}) < 0$: the reason is that it prevents candidates \mathbf{x} from becoming too close to the boundary of the feasible set by becoming very large near it. With this notation, we can write the

approximate problem as

$$\begin{aligned} & \text{minimize} && t f(\mathbf{x}) + \varphi(\mathbf{x}) \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{17.1}$$

Note that here we multiplied the objective with t , which does not change the optimal points (but the values). For reference, we record the gradient and Hessian of φ ,

$$\begin{aligned} \nabla \varphi &= \sum_{i=1}^m \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x}) \\ \nabla^2 \varphi &= \sum_{i=1}^m \frac{1}{f_i^2(\mathbf{x})} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^\top + \sum_{i=1}^m \frac{1}{-\nabla f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x}). \end{aligned}$$

Two questions are of importance:

1. How well does a solution of (17.1) approximate one of (1), and
2. How do we solve (17.1)?

17.2 The central path

The set of solutions $\mathbf{x}^*(t)$ of (17.1) for $t > 0$ is called the *central path*. Points on the central path are strictly feasible: they satisfy $\mathbf{A}\mathbf{x}^*(t) = \mathbf{b}$ and $f_i(\mathbf{x}^*(t)) < 0$. Moreover, by the Lagrange multiplier theorem, they satisfy the optimality condition

$$t \nabla f(\mathbf{x}) + \nabla \varphi(\mathbf{x}) + \mathbf{A}^\top \boldsymbol{\mu}^* = \mathbf{0}$$

for some suitable Lagrange multipliers $\boldsymbol{\mu} \in \mathbb{R}^p$. An important property of the central path is that we can derive dual points $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, and therefore (by Lagrange duality) lower bounds on the optimal value p^* of (1), from any point \mathbf{x}^* on the central path. In fact, the central path can be derived from the solutions of a variant of the KKT conditions.

Lemma 17.3. *A point \mathbf{x} is equal to a point $\mathbf{x}^*(t)$ on the central path if and only if there exist dual multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ such that the following conditions are satisfied.*

$$\begin{aligned} f(\mathbf{x}^*) &\leq \mathbf{0} \\ \mathbf{A}\mathbf{x}^* &= \mathbf{b} \\ \boldsymbol{\lambda}^* &\geq \mathbf{0} \\ -\lambda_i^* f_i(\mathbf{x}^*) &= \frac{1}{t}, \quad 1 \leq i \leq m \\ \nabla_x f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla_x f_i(\mathbf{x}^*) + \mathbf{A}^\top \boldsymbol{\mu}^* &= \mathbf{0}, \end{aligned}$$

The proof is left as an exercise. Note the analogy with the definition of the central path in the case of linear programming! Accordingly, the solution methods discussed in the context of linear programming carry over to the nonlinear case. We conclude by sketching the *barrier method*.

1. Start with a strictly feasible \mathbf{x} and $t := t^{(0)} > 0$, $\mu > 1$ and tolerance $\varepsilon > 0$;
2. Compute $\mathbf{x}^*(t)$ by minimizing $tf + \varphi$ subject to $\mathbf{Ax} = \mathbf{b}$, starting with \mathbf{x} ;
3. Update $\mathbf{x} = \mathbf{x}^*(t)$;
4. Stop if $m/t < \varepsilon$, otherwise set $t := \mu t$ (increase t).

The only unexplained part is how to solve the equality constrained minimization problem. One way is to use Newton's method on the optimality conditions of this problem.