10-725: Convex Optimization **Fall 2019**

Lecture 15: October 16

Lecturer: Ryan Tibshirani *Scribes: Avikalp Srivastava, Nicholas Trieu, Srinivas Ravishankar*

15.1 Barrier Method Motivation

Barrier method is the first of the two interior points methods that will be looked at in this class (the second being primal-dual interior point method).

15.1.1 Hierarchy of Second-Order Methods

We first look at a hierarchical overview of the kind of problems and methods we have seen thus far. We assume that all problems being considered here are convex.

- **Quadratic Problems**: Easiest, since for unconstrained convex quadratic problems, a closed-form solution can be obtained by solving the linear system given by differentiating and setting to zero.
- **Equality-constrained Quadratic Problems**: Also easy, since we can use KKT conditions to again get a closed form solution given by a linear system.
- **Equality-constrained Smooth Problems**: This is the previous category minus the quadratic assumption - thus we are minimizing a twice differentiable function subject to equality constraints. Newton's method reduces this to a sequence of equality-constrained quadratic minimization problems.
- **Inequality-constrained and Equality-constrained Smooth Problems**: Among the methods seen so far, only projected gradient descent (PGD) is feasible for solving this. However, the issues with using PGD are that projections can be hard and we lose the faster convergence rate of Newton's method. Therefore we will try to use interior-point methods, which reduce the problem to sequence of only equality-constrained problems - which brings us back to case 3 above.

15.1.2 Log Barrier Function

The intuition is to try to push the inequality constraints into the criterion in a smooth way and only be left with equality constraints.

Consider the convex optimization problem:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & h_i(x) \leq 0,\ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

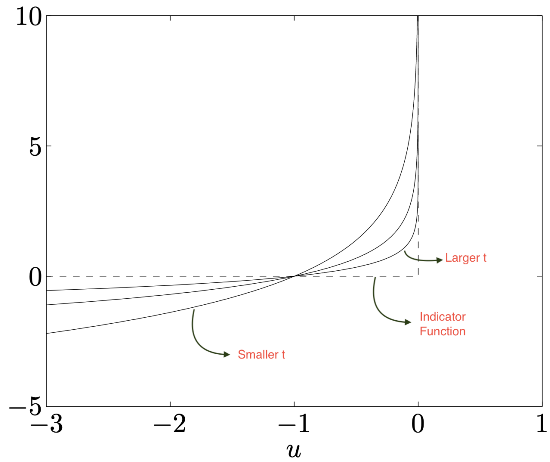

Figure 15.1: Log-barrier function as an approximation

Assume that f, $h_1, h_2, \ldots, h_m$ are convex, twice differentiable, each with domain $\mathbb{R}^n$ (this last condition is just for simplicity).

We define the *log-barrier* function for the above problem as:

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

The domain for the above function is the set of strictly feasible points $\{x : h_i(x) < 0, i = 1, \ldots, m\}$. We assume that this is non-empty - which, by using slater's condition, implies that strong-duality holds.

Now, consider the non-smooth exact transition of inequality constraints to the criterion:

$$\min_x \quad f(x) + \sum_{i=1}^{m} I_{h_i(x) \leq 0}(x)$$

The main idea is to try to approximate the indicator function using the log-barrier function as:

$$\min_x \quad f(x) + \frac{1}{t}\phi(x)$$

where $t > 0$ is a large number. Note that the approximation becomes more accurate as t becomes larger (see Fig. 15.1)

15.1.3 Convexity of Log Barrier Function

There are multiple ways of seeing this. One of them is to see that each term log-barrier function is composition of two functions, such that the inner function $h_i(x)$ is convex, and the outer function $-\log(-x)$ is convex and non-decreasing. Since log-barrier function involves summation of such compositions, it itself is convex as well.

15.1.4 Central Path

Using the log-barrier approximation, we have reduced our original problem to the barrier problem given as:

$$\begin{aligned} \min_x \quad & tf(x) + \phi(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

The *central path* is defined as the solution $x^*(t)$ as a function of $t > 0$.

We now discuss a bunch of observations about this formulation:

- We do this with the hope that as $t \to \infty$, $x^*(t) \to x^*$.
- We take a path instead of simply solving the problem for a very large value of t because of two issues with using a large valued t:
  - Numerical instability
  - Empirically, it leads to Newton method taking many iterations to reach the quadratic convergence phase.
- Therefore, in both practice and principle it is much more efficient to traverse the path by solving this problem for a bunch of intermediate values of t, and use the solution for one value of t as warm start for solution to next larger value of t.
- Informally, it is always keeping Newton's method in the quadratic convergence phase.

15.1.5 Special Case: Linear Program

Consider the linear program:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Dx \le e \quad \text{or} \quad d_i^T x \le e_i,\ i = 1, \dots, m \end{aligned}$$

The log-barrier function corresponding to this problem is:

$$\phi(x) = -\sum_{i=1}^{m} log(e_i - d_i^T x)$$

So, the barrier criterion becomes:

$$\min_x \quad tc^T x - \sum_{i=1}^{m} log(e_i - d_i^T x)$$

Since this is a convex program, we can use gradient optimality to see that:

$$0 = tc + \nabla\phi(x^*(t)) \implies 0 = tc - \sum_{i=1}^{m} \frac{1}{e_i - d_i^T x^*(t)} d_i$$

So the gradient $\nabla\phi(x^*(t))$ must be parallel to -c. This tells us that at the central path solution $x^*(t)$, the gradient of log-barrier function points in a direction that is parallel to the negative gradient of the criterion - this direction maximally reduces the criterion. See Fig. 15.2 for more details.

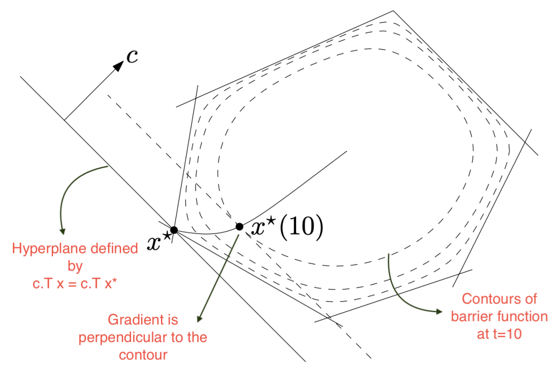

Figure 15.2: The central point methods arrives at the minimizer (x^*) by minimizing a sequence of smooth problems each of which approximates the constraint polyhedron by a smooth constraint region lying inside the polyhedron - hence the name interior point method.

15.1.6 KKT Conditions and Duality

The central path is characterized by its KKT conditions:

$$t\nabla f(x^*(t)) - \sum_{i=1}^{m} \frac{1}{h_i(x^*(t))} \nabla h_i(x^*(t)) + A^T w = 0,$$
$$Ax^*(t) = b, \quad h_i(x^*(t)) < 0, \quad i = 1, \dots, m$$

for some $w \in \mathbb{R}^m$.

We may derive feasible dual points for our original problem

$$\min_x \quad f(x)$$
$$Ax = b$$

as follows. Given $x^*(t)$ and corresponding w, define

$$u_i^*(t) = -\frac{1}{th_i(x^*(t))}, \quad i = 1, \dots, m$$
$$v^*(t) = w/t.$$

It follows that $u^*(t), v^*(t)$ are dual feasible for the original problem, whose Lagrangian is

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + v^T(Ax - b).$$

Proof sketch:
Note that $u_i^*(t) > 0$ since $h_i(x^*(t)) < 0$ for all $i, \dots, m$. Furthermore, the point $(u^*(t), v^*(t))$ lies in the domain of the Lagrange dual function $g(u, v)$ because

$$\nabla f(x^*(t)) + \sum_{i=1}^{m} u_i(x^*(t)) \nabla h_i(x^*(t)) + A^T v^*(t) = 0.$$

So $x^*(t)$ minimizes the Lagrangian $L(x, u^*(t), v^*(t))$ over x, so $g(u^*(t), v^*(t)) > -\infty$.

15.1.7 Duality Gap

Using the dual feasible points, we may bound the suboptimality of $f(x^*(t))$ with respect to the original problem via the duality gap. We compute

$$\begin{aligned} g(u^*(t), v^*(t)) &= f(x^*(t)) + \sum_{i=1}^m u_i^*(t) h_i(x^*(t)) + v^*(t)^T (Ax^*(t) - b) \\ &= f(x^*(t)) - m/t \end{aligned}$$

so $f(x^*(t)) - f^* \le m/t$. This bound may be used as a stopping criterion, and it confirms that $x^*(t) \to x^*$ as $t \to \infty$.

15.1.8 Perturbed KKT Conditions

We can now interpret the central path $(x^*(t), u^*(t), v^*(t))$ as solving the *perturbed* KKT conditions

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^m u_i \nabla h_i(x) + A^T v = 0 \\ u_i \cdot h_i(x) = 1/t, \quad i = 1, \dots, m \quad \text{(perturbation)} \\ h_i(x) \le 0, \quad i = 1, \dots, m, \quad Ax = b \\ u_i \ge 0, \quad i = 1, \dots, m \end{aligned}$$

where the only difference between these and the actual KKT conditions is the second line, i.e. complementary slackness. Note that as $t \to \infty$, $1/t \to 0$ which brings us closer to the actual KKT condition $u_i \cdot h_i(x) = 0$.

15.1.9 Barrier Method Algorithm and Considerations

We solve a sequence of optimization problems:

$$\min_x tf(x) + \phi(x)$$

for increasing values of t. We stop when $m/t \le \epsilon$, since duality gap is upper bounded by m/t.

Concretely:

1. Initialize $t^{(0)} > 0, \mu > 0$. Solve the minimization problem to get $x^{(0)} = x^*(t^{(0)})$.
2. For k = 1,2,3...
   (a) Compute $t^{(k)} = \mu t^{(k-1)}$.
   (b) Solve minimization problem using Newton's method initialized at $x^{(k-1)}$ to get $x^{(k)} = x^*(t^{(k)})$
   (c) if $m/t \le \epsilon$ break

Considerations:

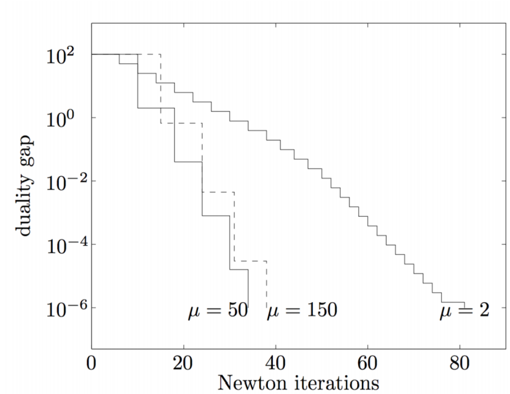

Figure 15.3: Convergence for different values of μ

1. Choice of μ: If small, then we make only a small step on the centering path. Many such steps needed. If large, Newton's method might meander around before reaching $x^{(k)}$. Note that Newton's method's individual iterations need not generally follow the centering path, only its solutions lie on it.
2. Choice of $t^{(0)}$: If too small, many steps to take on the centering path. If too large, then the first centering step will require many inner iterations of the Newton method.

The barrier method is quite robust to the choice of μ and $t^{(0)}$ in practice. However, note that the appropriate range for these parameters is scale dependent. We can see the effect of different choices of μ for an LP of $n = 50$ and $m = 100$ in Fig 15.3.

15.1.10 Convergence Analysis

We have $t^{(k)} = \mu^k t^{(0)}$. Stopping criterion is $m/t^{(k)} \leq \epsilon$. Number of steps is smallest k such that

$$\frac{m}{\mu^k t^{(0)}} \leq \epsilon$$

$$\mu^k \geq \frac{m}{\epsilon t^{(0)}}$$

$$k \log \mu \geq \log \frac{m}{\epsilon t^{(0)}}$$

$$k \geq \frac{\log m/\epsilon t^{(0)}}{\log \mu}$$

Implications: Scales logarithmically with number of constraints, which is good. Roughly linear convergence for a given m. We can see an example of this for an LP in figure 15.4

Now, how many Newton iterations for each of the above steps? For a self-concordant function, small changes in x lead to small changes in its Hessian. This leads to 'good' behaviour of the Newton's method, in the sense that it requires roughly constant number of iterations per centering step. This would then cause the total number of steps to still be $O(\log m/\epsilon t^{(0)})$.

$tf + \phi$ is self-concordant when f,h_i are all linear or quadratic. This covers LPs, QPs, QCQPs.

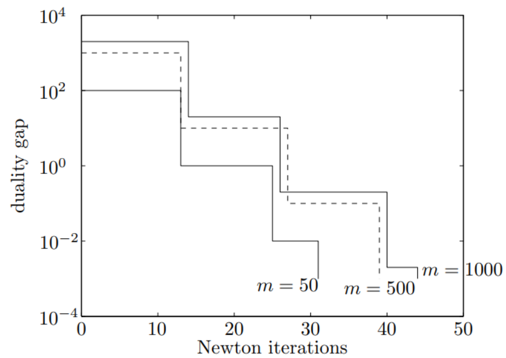

Figure 15.4: Convergence scales only with log of constraints, very slow scaling, good behaviour

15.1.11 Feasibility Criterion

Observe that for the analysis of showing dual feasibility, we require the first centering step $x^{(0)}$ to be strictly feasible.

We can get such a point by solving the following LP using the barrier method:

$$\min_{x,s} s$$

$$\text{s.t } h_i(x) \leq s \text{ for i = 1,2,...m}$$

$$Ax = b$$

Note that we don't require an optimal solution, we can stop as soon as $s < 0$ since this gives us a feasible point $x^{(0)}$. This is known as the feasibility method.

An alternative to get a strictly feasible point is to choose different upper bounds on each $h_i(x)$:

$$\min_{x,s} 1^T s$$

$$\text{s.t } h_i(x) \leq s_i \text{ for i = 1,2,...m}$$

$$Ax = b, s \geq 0$$

An advantage over the previous LP is that this is more informative. If the original system is infeasible, the non-zero entries of s tell us which inequalities cannot be satisfied.