# Barrier Method

Ryan Tibshirani
Convex Optimization 10-725

# Last time: Newton's method

Consider the problem

$$\min_x f(x)$$

for $f$ convex, twice differentiable, with $\mathrm{dom}(f) = \mathbb{R}^n$. Newton's method: choose initial $x^{(0)} \in \mathbb{R}^n$, repeat

$$x^{(k)} = x^{(k-1)} - t_k \big(\nabla^2 f(x^{(k-1)})\big)^{-1} \nabla f(x^{(k-1)}), \quad k = 1, 2, 3, \dots$$

Step sizes $t_k$ chosen by backtracking line search

If $\nabla f$ Lipschitz, $f$ strongly convex, $\nabla^2 f$ Lipschitz, then Newton's method has a local convergence rate $O(\log\log(1/\epsilon))$

Downsides:

- Requires solving systems in Hessian $\leftarrow$ quasi-Newton
- Can only handle equality constraints $\leftarrow$ this lecture

An important variant is equality-constrained Newton: start with $x^{(0)}$ such that $Ax^{(0)} = b$. Then we repeat the updates

$$x^+ = x + tv, \quad \text{where}$$

$$v = \underset{Az=0}{\operatorname{argmin}} \ \nabla f(x)^T(z-x) + \frac{1}{2}(z-x)^T \nabla^2 f(x)(z-x)$$

which keep $x^+$ in feasible set $Ax = b$

Here $v$ is characterized by KKT system

$$\left[ \begin{array}{cc} \nabla^2 f(x) & A^T \\ A & 0 \end{array} \right] \left[ \begin{array}{c} v \\ w \end{array} \right] = \left[ \begin{array}{c} -\nabla f(x) \\ 0 \end{array} \right]$$

for some $w$. Hence Newton direction $v$ is again given by solving a linear system in the Hessian (albeit a bigger one)

# Hierarchy of second-order methods

Assuming all problems are convex, you can think of the following hierarchy that we've worked through:

- Quadratic problems are the easiest: closed-form solution
- Equality-constrained quadratic problems are still easy: we use KKT conditions to derive closed-form solution
- Equality-constrained smooth problems are next: use Newton's method to reduce this to a sequence of equality-constrained quadratic problems
- Inequality-constrained and equality-constrained smooth problems are what we cover now: use interior-point methods to reduce this to a sequence of equality-constrained problems

# Log barrier function

Consider the convex optimization problem

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

We will assume that $f, h_1, \ldots h_m$ are convex, twice differentiable, each with domain $\mathbb{R}^n$. The function

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

is called the log barrier for the above problem. Its domain is the set of strictly feasible points, $\{x : h_i(x) < 0, \ i = 1, \ldots m\}$, which we assume is nonempty. (Note this implies strong duality holds)

Ignoring equality constraints for now, our problem can be written as

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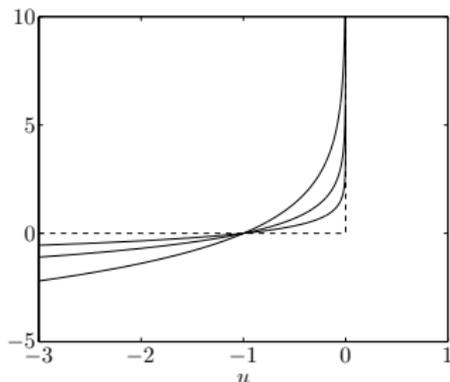


We can approximate the sum of indicators by the log barrier:

$$\min_x \; f(x) - (1/t) \cdot \sum_{i=1}^{m} \log(-h_i(x))$$

where $t > 0$ is a large number

This approximation is more accurate for larger $t$. But for any value of $t$, the log barrier approaches $\infty$ if any $h_i(x) \to 0$

# Outline

Today:
- Central path
- Properties and interpretations
- Barrier method
- Convergence analysis
- Feasibility methods

# Log barrier calculus

For the log barrier function

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

we have for its gradient:

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

and for its Hessian:

$$\nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{h_i(x)^2} \nabla h_i(x) \nabla h_i(x)^T - \sum_{i=1}^{m} \frac{1}{h_i(x)} \nabla^2 h_i(x)$$

# Central path

Consider barrier problem:

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

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

- Hope is that, as $t \to \infty$, we will have $x^\star(t) \to x^\star$, solution to our original problem
- Why don't we just set $t$ to be some huge value, and solve the above problem? Directly seek solution at end of central path?
- Problem is that this is seriously inefficient in practice
- Much more efficient to traverse the central path, as we will see

An important special case: barrier problem for a linear program:
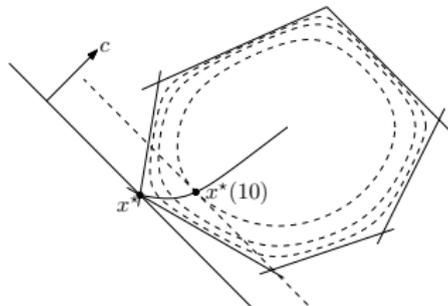
$$\min_x \ tc^T x - \sum_{i=1}^m \log(e_i - d_i^T x)$$

The barrier function corresponds to polyhedral constraint $Dx \le e$

Gradient optimality condition:

$$0 = tc - \sum_{i=1}^m \frac{1}{e_i - d_i^T x^\star(t)} d_i$$

This means that gradient $\nabla\phi(x^\star(t))$ must be parallel to $-c$, i.e., hyperplane $\{x : c^T x = c^T x^\star(t)\}$ lies tangent to contour of $\phi$ at $x^\star(t)$



(From B & V page 565)

# KKT conditions and duality

Central path is characterized by its KKT conditions:

$$t\nabla f(x^\star(t)) - \sum_{i=1}^{m} \frac{1}{h_i(x^\star(t))}\nabla h_i(x^\star(t)) + A^T w = 0,$$

$$Ax^\star(t) = b, \quad h_i(x^\star(t)) < 0, \quad i = 1, \ldots m$$

for some $w \in \mathbb{R}^m$. But we don't really care about dual variable for barrier problem ...

From central path points, we can derive feasible dual points for our original problem. Given $x^\star(t)$ and corresponding $w$, we define

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

We claim $u^\star(t), v^\star(t)$ are dual feasible for original problem, whose Lagrangian is

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

Why?

- Note that $u_i^\star(t) > 0$ since $h_i(x^\star(t)) < 0$ for all $i = 1, \ldots, m$
- Further, the point $(u^\star(t), v^\star(t))$ lies in domain of Lagrange dual function $g(u, v)$, since by definition

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

I.e., $x^\star(t)$ minimizes Lagrangian $L(x, u^\star(t), v^\star(t))$ over $x$, so $g(u^\star(t), v^\star(t)) > -\infty$

# Duality gap

This allows us to bound suboptimality of $f(x^\star(t))$, with respect to original problem, via the duality gap. We compute

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

That is, we know that $f(x^\star(t)) - f^\star \leq m/t$

This will be very useful as a stopping criterion; it also confirms the hope that $x^\star(t) \to x^\star$ as $t \to \infty$

# Perturbed KKT conditions

We can now reinterpret central path $(x^\star(t), u^\star(t), v^\star(t))$ as solving the perturbed KKT conditions:

$$\nabla f(x) + \sum_{i=1}^m u_i \nabla h_i(x) + A^T v = 0$$
$$u_i \cdot h_i(x) = -1/t, \quad i = 1, \dots m$$
$$h_i(x) \leq 0, \quad i = 1, \dots m, \quad Ax = b$$
$$u_i \geq 0, \quad i = 1, \dots m$$

Only difference between these and actual KKT conditions for our original problem is second line: these are replaced by

$$u_i \cdot h_i(x) = 0, \quad i = 1, \dots m$$

i.e., complementary slackness, in actual KKT conditions

# Barrier method

The barrier method solves a sequence of problems

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

for increasing values of $t > 0$, until duality gap satisfies $m/t \leq \epsilon$

We fix $t^{(0)} > 0$, $\mu > 1$. We use Newton to compute $x^{(0)} = x^\star(t)$, a solution to barrier problem at $t = t^{(0)}$. For $k = 1, 2, 3, \ldots$

- Solve the barrier problem at $t = t^{(k)}$, using Newton initialized at $x^{(k-1)}$, to yield $x^{(k)} = x^\star(t)$
- Stop if $m/t \leq \epsilon$, else update $t^{(k+1)} = \mu t$

The first step above is called a centering step (since it brings $x^{(k)}$ onto the central path)

Considerations:

- Choice of $\mu$: if $\mu$ is too small, then many outer iterations might be needed; if $\mu$ is too big, then Newton's method (each centering step) might take many iterations
- Choice of $t^{(0)}$: if $t^{(0)}$ is too small, then many outer iterations might be needed; if $t^{(0)}$ is too big, then the first Newton solve (first centering step) might require many iterations

Fortunately, the performance of the barrier method is often quite robust to the choice of $\mu$ and $t^{(0)}$ in practice

(However, note that the appropriate range for these parameters is scale dependent)

Example of a small LP in $n = 50$ dimensions, $m = 100$ inequality constraints (from B & V page 571):
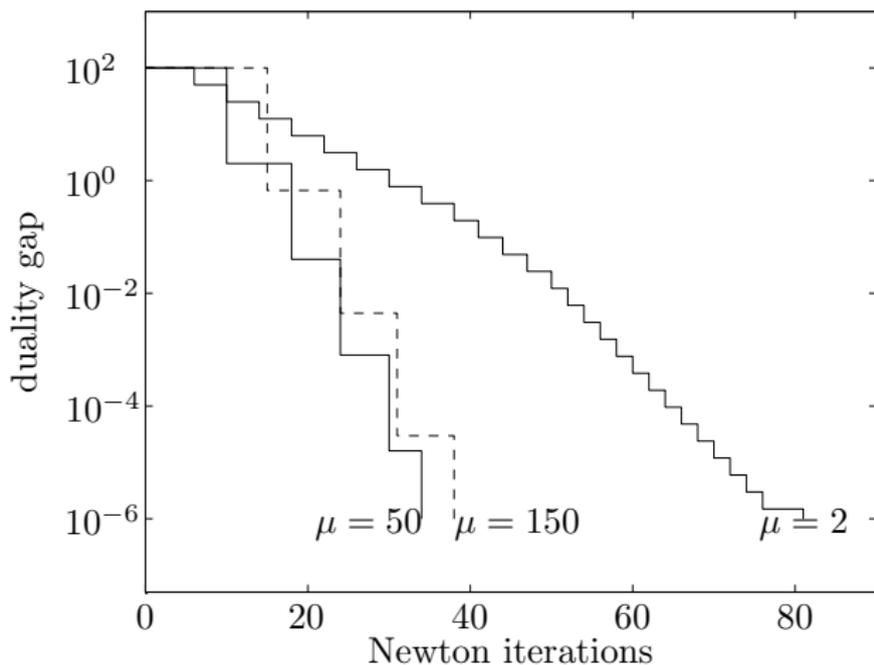
# Convergence analysis

Assume that we solve the centering steps exactly. The following result is immediate

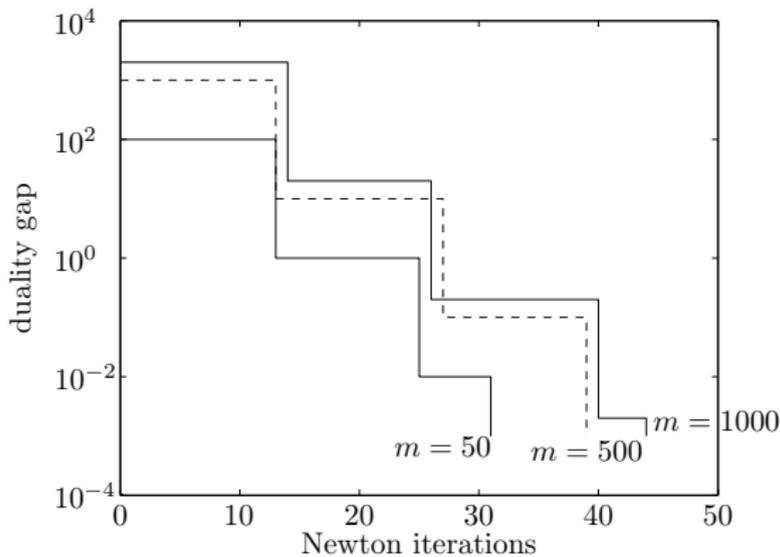**Theorem:** The barrier method after $k$ centering steps satisfies
$$f(x^{(k)}) - f^\star \leq \frac{m}{\mu^k t^{(0)}}$$

In other words, to reach a desired accuracy level of $\epsilon$, we require

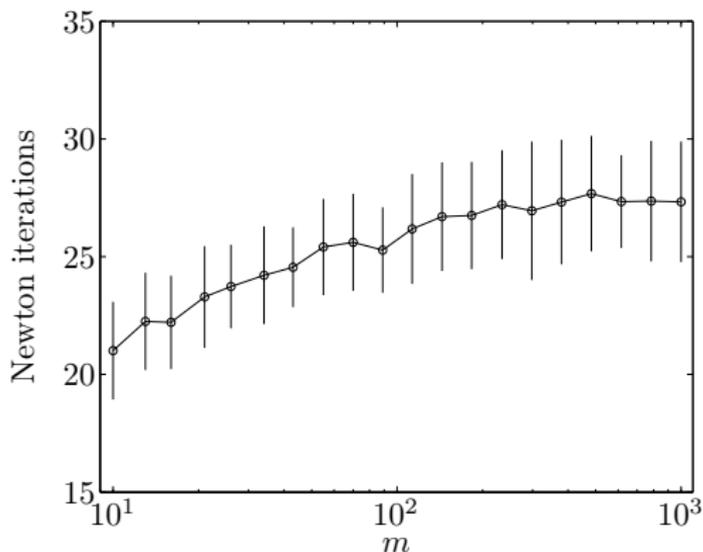$$\frac{\log(m/(t^{(0)}\epsilon))}{\log \mu}$$

centering steps with the barrier method (plus initial centering step)

Example of barrier method progress for an LP with $m$ constraints (from B & V page 575):



Can see roughly linear convergence in each case, and logarithmic scaling with $m$

Seen differently, the number of Newton steps needed (to decrease initial duality gap by factor of $10^4$) grows very slowly with $m$:



Note that cost of a single Newton step does depends on $m$ (and on problem dimension $n$)

# How many Newton iterations?

Informally, due to careful central path traversal, in each centering step, Newton is already in <span style="color:red">quadratic convergence phase</span>, so takes nearly constant number of iterations

This can be formalized under self-concordance. Suppose:

- The function $tf + \phi$ is self-concordant
- Our original problem has bounded sublevel sets

Then we can terminate each Newton solve at appropriate accuracy, and the <span style="color:red">total number of Newton iterations is</span> still $O(\log(m/(t^{(0)}\epsilon))$ (where constants do not depend on problem-specific conditioning). See Chapter 11.5 of B & V

Importantly, $tf + \phi = tf - \sum_{i=1}^{m} \log(-h_i)$ is self-concordant when $f, h_i$ are all linear or quadratic. So this covers LPs, QPs, QCQPs

# Feasibility methods

We have implicitly assumed that we have a strictly feasible point for the first centering step, i.e., for computing $x^{(0)} = x^\star$, solution of barrier problem at $t = t^{(0)}$. This is $x$ such that

$$h_i(x) < 0, \quad i = 1, \ldots m, \quad Ax = b$$

How to find such a feasible $x$? By solving

$$\min_{x,s} \quad s$$
$$\text{subject to} \quad h_i(x) \leq s, \ i = 1, \ldots m$$
$$Ax = b$$

The goal is for $s$ to be negative at the solution. This is known as a feasibility method. We can apply the barrier method to the above problem, since it is easy to find a strictly feasible starting point

Note that we do not need to solve this problem to high accuracy.
Once we find a feasible $(x, s)$ with $s < 0$, we can terminate early

An alternative is to solve the problem

$$\min_{x,s} \quad 1^T s$$
$$\text{subject to} \quad h_i(x) \leq s_i, \ i = 1, \ldots m$$
$$Ax = b, \ s \geq 0$$

Previously $s$ was the maximum infeasibility across all inequalities.
Now each inequality has own infeasibility variable $s_i, \ i = 1, \ldots m$

One advantage: when the original system is infeasible, the solution
of the above problem will be informative. The nonzero entries of $s$
will tell us which of the constraints cannot be satisfied

# References and further reading

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapter 11
- A. Nemirovski (2004), "Interior-point polynomial time methods in convex programming", Chapter 4
- J. Nocedal and S. Wright (2006), "Numerical optimization", Chapters 14 and 19