AMSC 607 / CMSC 764 Advanced Numerical Optimization

Fall 2010

Renegar's LP IPM Convergence Analysis

Dianne P. O'Leary

Renegar's approach to IPMs

James Renegar's book, *A Mathematical View of Interior-Point Methods in Convex Optimization* (SIAM 2001) provides a 45 page development of the complexity theory for a particular IPM for LP.

Here we summarize the argument.

Our starting point

Our starting point in this discussion is your 10th homework problem.

---

10. In this problem, we establish various properties of the log function that will help us in studying its use as a barrier.

10a. (8) Let $f(x) = -\log(x)$ for $x > 0$. Let $\hat{f}(x) = -\log(ax + b)$ where $a, b > 0$.

Prove the following properties:

- $|f'''(x)| \le 2(f''(x))^{3/2}$.
- $|\hat{f}'''(x)| \le 2(\hat{f}''(x))^{3/2}$.

This is one way to characterize a "self-concordant function", but Renegar prefers a different definition.

10b. Let

$$f(\mathbf{x}) = -\sum_{j=1}^{n} \log(x_j).$$

As usual, let $\mathbf{H}(\mathbf{x})$ denote the Hessian of f.

The domain of f is $D_f = \{\mathbf{x} : \mathbf{x} > \mathbf{0}\}$.

Define the $\mathbf{x}$-inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle_x = \mathbf{u}^T \mathbf{H}(\mathbf{x}) \mathbf{v},$$

so that

$$\|\mathbf{u}\|_x^2 = \mathbf{u}^T \mathbf{H}(\mathbf{x}) \mathbf{u}.$$

Define the Diken ball of radius r with center $\mathbf{x}$:

$$B_x(\mathbf{x}, r) = \{\mathbf{w} : \|\mathbf{w} - \mathbf{x}\|_x < r\}.$$

Note that this is a very elongated ellipsoid as we approach the boundary.

Prove the following properties:

- P1: For any $\mathbf{x} \in D_f$, $B_x(\mathbf{x}, 1) \subset D_f$.
- P2: If $\mathbf{w} \in B_x(\mathbf{x}, 1)$ and if $\mathbf{v} \neq \mathbf{0}$, then

$$1 - \|\mathbf{w} - \mathbf{x}\|_x \leq \frac{\|\mathbf{v}\|_y}{\|\mathbf{v}\|_x} \leq \frac{1}{1 - \|\mathbf{w} - \mathbf{x}\|_x}.$$

  Hint: Show that $\|\mathbf{v}\|_y^2 \leq \|\mathbf{v}\|_x^2 \max_j (x_j/y_j)^2$. Then show that $y_j/x_j > 1 - |y_j/x_j - 1|$.

The set of functions that define norms so that P1 and P2 are satisfied is called the set of self-concordant functions, so you have shown that the log barrier is self-concordant.

The Newton step

Let's investigate Newton's method using our $\mathbf{x}$-inner product:

$$\begin{aligned} f(\mathbf{w}) &\approx f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{w} - \mathbf{x}) + \frac{1}{2}(\mathbf{w} - \mathbf{x})^T \mathbf{H}(\mathbf{x})(\mathbf{w} - \mathbf{x}) \\ &= f(\mathbf{x}) + \langle \mathbf{H}(\mathbf{x})^{-1}\mathbf{g}(\mathbf{x}), \mathbf{w} - \mathbf{x}\rangle_x + \frac{1}{2}\langle \mathbf{w} - \mathbf{x}, \mathbf{H}(\mathbf{x})(\mathbf{w} - \mathbf{x})\rangle \\ &\equiv f(\mathbf{x}) - \langle \mathbf{n}(\mathbf{x}), \mathbf{w} - \mathbf{x}\rangle_x + \frac{1}{2}\|\mathbf{w} - \mathbf{x}\|_x^2 \end{aligned}$$

Note that $\mathbf{n}(\mathbf{x}) = \mathbf{H}(\mathbf{x})^{-1}\mathbf{g}(\mathbf{x})$ is the Newton step for minimizing the function.

LP via minimizing the log barrier

We can formulate an IPM for linear programming by following the path of minimizers of the function

$$f_\eta(\mathbf{x}) = \eta \mathbf{c}^T \mathbf{x} + f(\mathbf{x})$$

as $\eta \to \infty$, where $f(\mathbf{x})$ is defined in 10b.

Note the relationship between Renegar's η and the parameter μ more commonly used.

The Newton step is

$$\mathbf{n}_\eta(\mathbf{x}) = -\mathbf{H}(\mathbf{x})^{-1}(\eta \mathbf{c} + \mathbf{g}(\mathbf{x})) \equiv -(\eta \mathbf{c}_x + \mathbf{g}_x(\mathbf{x})).$$

Unquiz: Show that

$$\mathbf{n}_{\hat{\eta}}(\mathbf{x}) = \frac{\hat{\eta}}{\eta}\mathbf{n}_{\eta}(\mathbf{x}) + (\frac{\hat{\eta}}{\eta} - 1)\mathbf{g}_x(\mathbf{x}),$$

and therefore

$$\|\mathbf{n}_{\hat{\eta}}(\mathbf{x})\|_x \le \frac{\hat{\eta}}{\eta}\|\mathbf{n}_{\eta}(\mathbf{x})\|_x + (\frac{\hat{\eta}}{\eta} - 1)\|\mathbf{g}_x(\mathbf{x})\|_x.$$

We will call this Equation 1.

Unquiz: Show that the jth component of the vector $\mathbf{g}_x(\mathbf{x}) = \mathbf{H}(\mathbf{x})^{-1}\mathbf{g}(\mathbf{x})$ is x_j, and therefore

$$\|\mathbf{g}_x(\mathbf{x})\|_x^2 = \mathbf{g}_x(\mathbf{x})^T\mathbf{H}(\mathbf{x})\mathbf{g}_x(\mathbf{x}) = n.$$

This is Equation 2.

We will call $\|\mathbf{g}_x(\mathbf{x})\|_x^2 = n$ the complexity value of f.

A key convergence result for Newton's method

Suppose $\|\mathbf{n}(\mathbf{x})\|_x < 1$. This means that $\mathbf{w} = \mathbf{x} + \mathbf{n}(\mathbf{x})$ is inside $B_x(\mathbf{x}, 1) \subset D_f$ and therefore $\mathbf{w} > \mathbf{0}$. We can take the full Newton step without hitting the boundary of the feasible region.

Renegar shows (Thm 2.2.4) Equation 3:

$$\|\mathbf{n}(\mathbf{w})\|_w \le \left(\frac{\|\mathbf{n}(\mathbf{x})\|_x}{1 - \|\mathbf{n}(\mathbf{x})\|_x}\right)^2.$$

This doesn't look very interesting until we realize that the size of the Newton step is related to the size of the error $\|\mathbf{x} - \mathbf{x}(\eta)\|_x$, where $\mathbf{x}(\eta)$ is the minimizer of f_η. This is Equation 4:

$$\|\mathbf{w} - \mathbf{x}(\eta)\|_x \le \frac{3\|\mathbf{n}(\mathbf{x})\|_x^2}{(1 - \|\mathbf{n}(\mathbf{x})\|_x)^3},$$

valid when $\|\mathbf{n}(\mathbf{x})\|_x < 1/4$. (Thm 2.2.5)

So if we can force $\|\mathbf{n}(\mathbf{x})\|_x$ to go to zero for our sequence of iterates, we force convergence to the solution $\mathbf{x}(\eta)$.

Example: Suppose we start with a point $\mathbf{x}_1$ for which $\|\mathbf{n}(\mathbf{x}_1)\|_{x_1} < 1/4$. Then

$$\begin{aligned}
\|\mathbf{n}(\mathbf{x}_2)\|_{x_2} &< \left(\frac{1/4}{1-1/4}\right)^2 = \left(\frac{1}{4} * \frac{4}{3}\right)^2 = \frac{1}{9}, \\
\|\mathbf{n}(\mathbf{x}_3)\|_{x_3} &< \left(\frac{1/9}{1-1/9}\right)^2 = \left(\frac{1}{9} * \frac{9}{8}\right)^2 = \frac{1}{64},
\end{aligned}$$

etc.

Stop and think how amazing Equation 4 is. We have a computable bound on the error! Too bad this only works for self-concordant functions....

But luckily,

... in this IPM, we are minimizing a sequence of self-concordant functions.

So we have to see if this convergence result for a single self-concordant function, with η fixed, can be bootstrapped into a result for a sequence of self concordant functions, with a sequence of η values.

Putting it all together

Equation 1:

$$\|\mathbf{n}_{\hat{\eta}}(\mathbf{x})\|_x \leq \frac{\hat{\eta}}{\eta}\|\mathbf{n}_\eta(\mathbf{x})\|_x + (\frac{\hat{\eta}}{\eta} - 1)\|\mathbf{g}_x(\mathbf{x})\|_x.$$

Equation 2 tells us that $\|\mathbf{g}_x(\mathbf{x})\|_x = \sqrt{n}$. Suppose

- $\|\mathbf{n}_\eta(\mathbf{x})\|_x \leq 1/9$.
- $\frac{\hat{\eta}}{\eta} = 1 + 1/(8\sqrt{n})$.

Then

$$\|\mathbf{n}_{\hat{\eta}}(\mathbf{x})\|_x \leq \left(1 + \frac{1}{8\sqrt{n}}\right) * \frac{1}{9} + \frac{1}{8\sqrt{n}}\sqrt{n} = \frac{1}{9}\left(1 + \frac{1}{8\sqrt{n}}\right) + \frac{1}{8} \equiv \gamma.$$

Now use Equation 3: if $\mathbf{w} = \mathbf{x} + \mathbf{n}(\mathbf{x})$, then

$$\|\mathbf{n}(\mathbf{w})\|_w \leq \left(\frac{\|\mathbf{n}(\mathbf{x})\|_x}{1 - \|\mathbf{n}(\mathbf{x})\|_x}\right)^2.$$

We write this for $\hat{\eta}$:

$$\begin{aligned} \|\mathbf{n}_{\hat{\eta}}(\mathbf{w})\|_w &\leq \left(\frac{\|\mathbf{n}_{\hat{\eta}}(\mathbf{x})\|_x}{1 - \|\mathbf{n}_{\hat{\eta}}(\mathbf{x})\|_x}\right)^2 \\ &\leq \left(\frac{\gamma}{1-\gamma}\right)^2 \qquad \text{Unquiz:} \\ &\leq \frac{1}{9}. \end{aligned}$$

Thus, $\mathbf{w}$ is also close to the central path, and we can continue the process, again increasing η by $1 + 1/(8\sqrt{n})$.

Now we can calculate how many steps it takes to increase η from some initial value η_0 to some target value η^*:

$$\left(1 + \frac{1}{8\sqrt{n}}\right)^k \eta_0 \geq \eta^*$$

when

$$k \log \frac{8\sqrt{n}+1}{8\sqrt{n}} \geq \log(\eta^*/\eta_0),$$

or

$$k \geq \frac{\log(\eta^*/\eta_0)}{\log \frac{8\sqrt{n}+1}{8\sqrt{n}}}.$$

It suffices to take $k = \lceil 10\sqrt{n} \log(\eta^*/\eta_0) \rceil$. (This can be verified by direct computation for $n \geq 1$.)

How big does η^* need to be?

Renegar's goal is to find a feasible point $\mathbf{x}$ whose objective function value $\mathbf{c}^T\mathbf{x}$ is close to optimal. Let $\mathbf{x}_{opt}$ be a solution to the LP and let $v_{opt} = \mathbf{c}^T\mathbf{x}_{opt}$.

Step 1: Get a bound for $\|\mathbf{c}_x\|_x$, where $\mathbf{c}_x = \mathbf{H}(\mathbf{x})^{-1}\mathbf{c}$, and $\mathbf{x}$ is any feasible point.

Since $B_x(\mathbf{x}, 1) \subset D_f$, $\mathbf{x} - t\mathbf{c}_x \in D_f$ if $0 \leq t < 1/\|\mathbf{c}_x\|_x$. We note that

$$\begin{aligned} \mathbf{c}^T(\mathbf{x} - t\mathbf{c}_x) &= \mathbf{c}^T\mathbf{x} - t\mathbf{c}^T\mathbf{c}_x \\ &= \mathbf{c}^T\mathbf{x} - t\mathbf{c}_x^T\mathbf{H}(\mathbf{x})^{-1}\mathbf{c}_x \\ &= \mathbf{c}^T\mathbf{x} - t\|\mathbf{c}_x\|_x^2, \end{aligned}$$

so

$$t\|\mathbf{c}_x\|_x^2 = \mathbf{c}^T\mathbf{x} - \mathbf{c}^T(\mathbf{x} - t\mathbf{c}_{\mathbf{x}}) \leq \mathbf{c}^T\mathbf{x} - v_{opt}.$$

Taking a limit as $t \to 1/\|\mathbf{c}_x\|_x$ gives

$$\|\mathbf{c}_x\|_x \leq \mathbf{c}^T\mathbf{x} - v_{opt}.$$

Step 2: Notice that

$$\mathbf{c}^T\mathbf{w} - v_{opt} = \mathbf{c}^T\mathbf{x} - v_{opt} + \langle \mathbf{c}_x, \mathbf{w} - \mathbf{x} \rangle_x,$$

so if $\mathbf{x} \in D_f$,

$$\begin{aligned} \frac{\mathbf{c}^T\mathbf{w} - v_{opt}}{\mathbf{c}^T\mathbf{x} - v_{opt}} &= 1 + \frac{\langle \mathbf{c}_x, \mathbf{w} - \mathbf{x} \rangle_x}{\mathbf{c}^T\mathbf{x} - v_{opt}} \\ &\leq 1 + \frac{\langle \mathbf{c}_x, \mathbf{w} - \mathbf{x} \rangle_x}{\|\mathbf{c}_x\|_x} \\ &\leq 1 + \frac{\|\mathbf{c}_x\|_x \|\mathbf{w} - \mathbf{x}\|_x}{\|\mathbf{c}_x\|_x} \\ &= 1 + \|\mathbf{w} - \mathbf{x}\|_x. \end{aligned}$$

Step 3: On the central path, $\mathbf{X}\mathbf{s} = (1/\eta)\mathbf{e}$, so the duality gap is n/η. We need one further fact about the central path (equation 2.12 in Renegar), and I wish I knew a short proof for it:

$$\mathbf{c}^T\mathbf{x}(\eta) - v_{opt} \leq \frac{n}{\eta}.$$

But maybe I do. Let $\mathbf{y}(\eta)$ and $\mathbf{z}(\eta)$ be the dual variables corresponding to $\mathbf{x}(\eta)$. Then

$$\mathbf{A}^T\mathbf{y}(\eta) + \mathbf{z}(\eta) = \mathbf{c},$$

so

$$\mathbf{x}(\eta)^T\mathbf{A}^T\mathbf{y}(\eta) + \mathbf{x}(\eta)^T\mathbf{z}(\eta) = \mathbf{x}(\eta)^T\mathbf{c},$$

or

$$\mathbf{y}(\eta)^T\mathbf{b} + \frac{n}{\eta} = \mathbf{x}(\eta)^T\mathbf{c}.$$

Also recall:

- The duality gap at the optimal solution is zero, so $\mathbf{y}_{opt}^T\mathbf{b} = \mathbf{x}_{opt}^T\mathbf{c}$,
- $\mathbf{y}_{opt}$ is the maximizer among all dual feasible points.

Therefore

$$\mathbf{x}(\eta)^T\mathbf{c} - \mathbf{x}_{opt}^T\mathbf{c} = \mathbf{y}(\eta)^T\mathbf{b} - \mathbf{y}_{opt}^T\mathbf{b} + \frac{n}{\eta} \le \frac{n}{\eta}.$$

and we have our fact.

Step 4: From Step 2 we now see that

$$\begin{aligned} \mathbf{c}^T\mathbf{w} - v_{opt} &\le (\mathbf{c}^T\mathbf{x}(\eta) - v_{opt})(1 + \|\mathbf{w} - \mathbf{x}(\eta)\|_{x(\eta)}) \\ &\le \frac{n}{\eta}(1 + \|\mathbf{w} - \mathbf{x}(\eta)\|_{x(\eta)}). \end{aligned}$$

Now use Equation 4 and the fact that $\|\mathbf{n}(\mathbf{x})\|_x \le 1/9$:

$$\|\mathbf{w} - \mathbf{x}(\eta)\|_x \le \frac{3\|\mathbf{n}(\mathbf{x})\|_x^2}{(1 - \|\mathbf{n}(\mathbf{x})\|_x)^3} \le \frac{3/81}{(1 - 1/9)^3} \equiv C.$$

Therefore,

$$\mathbf{c}^T\mathbf{w} - v_{opt} \le \frac{n(1 + C)}{\eta} < \epsilon$$

when $\eta > n(1 + C)/\epsilon$.

Conclusion: The algorithm takes $O(\sqrt{n}\log(n/\epsilon))$ iterations.

What algorithm?

Unquiz: Write the resulting algorithm: given a point close enough to the central path, take a Newton step, increase η as defined above, repeat until convergence.

The most onerous requirement: Finding an initial point close enough to the central path!