Renegar’s approach to IPMs


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:

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

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

10b. Let

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

As usual, let \( H(x) \) denote the Hessian of \( f \).

The domain of \( f \) is \( D_f = \{ x : x > 0 \} \).

Define the \( x \)-inner product

\[
\langle u, v \rangle_x = u^T H(x)v,
\]

so that

\[
\|u\|_x^2 = u^T H(x)u.
\]
Define the Diken ball of radius $r$ with center $x$:

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

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

Prove the following properties:

- **P1**: For any $x \in D_f$, $B_{x}(x, 1) \subset D_f$.
- **P2**: If $w \in B_{x}(x, 1)$ and if $v \neq 0$, then
  
  $$1 - \|w - x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq 1 - \|w - x\|_x.$$  

Hint: Show that $\|v\|_y^2 \leq \|v\|_x^2 \max_j (x_j/y_j)^2$. 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 $x$-inner product:

$$f(w) \approx f(x) + g(x)^T (w - x) + \frac{1}{2} (w - x)^T H(x) (w - x)$$

$$= f(x) + \langle H(x)^{-1} g(x), w - x \rangle_x + \frac{1}{2} \langle w - x, H(x) (w - x) \rangle$$

$$= f(x) - \langle n(x), w - x \rangle_x + \frac{1}{2} \|w - x\|_x^2$$

Note that $n(x) = H(x)^{-1} g(x)$ 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

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

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

Note the relationship between Renegar’s $\eta$ and the parameter $\mu$ more commonly used.

The Newton step is

$$n_{\eta}(x) = -H(x)^{-1} (\eta c + g(x)) \equiv - (\eta c_x + g_x(x)).$$
Unquiz: Show that
\[ n_{\eta}(x) = \frac{\eta}{\eta} n_{\eta}(x) + (\frac{\eta}{\eta} - 1) g_x(x), \]
and therefore
\[ \|n_{\eta}(x)\|_x \leq \frac{\eta}{\eta} \|n_{\eta}(x)\|_x + (\frac{\eta}{\eta} - 1) \|g_x(x)\|_x. \]
We will call this Equation 1.

Unquiz: Show that the \( j \)th component of the vector \( g_x(x) = H(x)^{-1} g(x) \) is \( x_j \), and therefore
\[ \|g_x(x)\|_2^2 = g_x(x)^T H(x) g_x(x) = n. \]
This is Equation 2.

We will call \( \|g_x(x)\|_2^2 = n \) the complexity value of \( f \).

A key convergence result for Newton’s method

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

Renegar shows (Thm 2.2.4) Equation 3:
\[ \|n(w)\|_w \leq \left( \frac{\|n(x)\|_x}{1 - \|n(x)\|_x} \right)^2. \]
This doesn’t look very interesting until we realize that the size of the Newton step is related to the size of the error \( \|x - x(\eta)\|_x \), where \( x(\eta) \) is the minimizer of \( f_\eta \). This is Equation 4:
\[ \|w - x(\eta)\|_x \leq \frac{3\|n(x)\|_x^2}{(1 - \|n(x)\|_x)^3}, \]
valid when \( \|n(x)\|_x < 1/4 \). (Thm 2.2.5)

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

Example: Suppose we start with a point \( x_1 \) for which \( \|n(x_1)\|_x < 1/4 \). Then
\[ \|n(x_2)\|_x < \left( \frac{1/4}{1 - 1/4} \right)^2 = \left( \frac{4}{3} \cdot \frac{4}{3} \right)^2 = \frac{1}{9}, \]
\[ \|n(x_3)\|_x < \left( \frac{1/9}{1 - 1/9} \right)^2 = \left( \frac{9}{8} \cdot \frac{9}{8} \right)^2 = \frac{1}{64}, \]
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 \( \eta \) fixed, can be bootstrapped into a result for a sequence of self-concordant functions, with a sequence of \( \eta \) values.

Putting it all together

Equation 1:

\[
|\mathbf{n}_\eta(x)\|_x \leq \frac{\tilde{\eta}}{\eta} |\mathbf{n}_{\tilde{\eta}}(x)\|_x + \left( \frac{\tilde{\eta}}{\eta} - 1 \right) |\mathbf{g}_x(x)\|_x.
\]

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

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

Then

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

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 \( \tilde{\eta} \):

\[
|\mathbf{n}_{\tilde{\eta}}(\mathbf{w})\|_w \leq \left( \frac{|\mathbf{n}_{\tilde{\eta}}(\mathbf{x})\|_x}{1-|\mathbf{n}_{\tilde{\eta}}(\mathbf{x})\|_x} \right)^2 \\
\leq \left( \frac{\gamma}{1 - \gamma} \right)^2 \quad \text{Unquiz:} \\
\leq \frac{1}{9}.
\]

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

Now we can calculate how many steps it takes to increase \( \eta \) from some initial value \( \eta_0 \) to some target value \( \eta^* \):

\[
\left( 1 + \frac{1}{8\sqrt{n}} \right)^k \eta_0 \geq \eta^*
\]
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 \( \eta^* \) need to be?

Renegar’s goal is to find a feasible point \( x \) whose objective function value \( c^T x \) is close to optimal. Let \( x_{\text{opt}} \) be a solution to the LP and let \( v_{\text{opt}} = c^T x_{\text{opt}} \).

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

Since \( B_x(x, 1) \subset D_f \), \( x - tc_x \in D_f \) if \( 0 \leq t < 1/\|c_x\|_x \). We note that
\[
\begin{align*}
    c^T(x - tc_x) &= c^T x - tc^T c_x \\
                 &= c^T x - t c^T H(x)^{-1} c_x \\
                 &= c^T x - t \|c_x\|_x^2,
\end{align*}
\]
so
\[
    t \|c_x\|_x^2 = c^T x - c^T(x - tc_x) \leq c^T x - v_{\text{opt}}.
\]
Taking a limit as \( t \to 1/\|c_x\|_x \) gives
\[
    \|c_x\|_x \leq c^T x - v_{\text{opt}}.
\]

**Step 2:** Notice that
\[
    c^T w - v_{\text{opt}} = c^T x - v_{\text{opt}} + \langle c_x, w - x \rangle_x,
\]
so if \( x \in D_f \),
\[
\begin{align*}
    \frac{c^T w - v_{\text{opt}}}{c^T x - v_{\text{opt}}} &= 1 + \frac{\langle c_x, w - x \rangle_x}{c^T x - v_{\text{opt}}} \\
    &\leq 1 + \frac{\|c_x\|_x \|w - x\|_x}{\|c_x\|_x} \\
    &\leq 1 + \frac{\|w - x\|_x}{\|c_x\|_x}.
\end{align*}
\]

**Step 3:** On the central path, \( Xs = (1/\eta)e \), so the duality gap is \( n/\eta \). We need one further fact about the central path (equation 2.12 in Renegar), and I wish I knew a short proof for it:
\[
    c^T x(\eta) - v_{\text{opt}} \leq \frac{n}{\eta}.
\]
But maybe I do. Let \( y(\eta) \) and \( z(\eta) \) be the dual variables corresponding to \( x(\eta) \). Then
\[
A^T y(\eta) + z(\eta) = c,
\]
so
\[
x(\eta)^T A^T y(\eta) + x(\eta)^T z(\eta) = x(\eta)^T c.
\]
or
\[
y(\eta)^T b + \frac{n}{\eta} = x(\eta)^T c.
\]
Also recall:

- The duality gap at the optimal solution is zero, so \( y_{opt}^T b = x_{opt}^T c \),
- \( y_{opt} \) is the maximizer among all dual feasible points.

Therefore
\[
x(\eta)^T c - x_{opt}^T c = y(\eta)^T b - y_{opt}^T b + \frac{n}{\eta} \leq \frac{n}{\eta},
\]
and we have our fact.

**Step 4:** From Step 2 we now see that
\[
c^T w - v_{opt} \leq (c^T x(\eta) - v_{opt})(1 + \|w - x(\eta)\|_{x(\eta)})
\leq \frac{n}{\eta} (1 + \|w - x(\eta)\|_{x(\eta)}),
\]
Now use Equation 4 and the fact that \( \|n(x)\|_x \leq 1/9 \):
\[
\|w - x(\eta)\|_x \leq \frac{3\|n(x)\|_x^2}{(1 - \|n(x)\|_x)^3} \leq \frac{3/81}{(1 - 1/9)^3} \equiv C.
\]
Therefore,
\[
c^T w - v_{opt} \leq \frac{n(1 + C)}{\eta} < \epsilon
\]
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 \( \eta \) as defined above, repeat until convergence.

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