Notes for Class: Analytic Center, Newton's Method, and Web-Based ACA

Robert M. Freund

April 8, 2004

1 The Analytic Center of a Polyhedral System

Given a polyhedral system of the form:

$$Ax \le b\ , \quad Mx = g\ ,$$

the *analytic center* is the solution of the following optimization problem:

$$\begin{array}{lll} \text{(ACP:)} & \text{maximize}_{x,s} & \prod_{i=1}^{m} s_i \\ & & \\ & \text{s.t.} & Ax + s = b \\ & & s \ge 0 \\ & & Mx = g\ . \end{array}$$

This is easily seen to be the same as:

$$\begin{array}{lll} \text{(ACP:)} & \text{minimize}_{x,s} & -\sum_{i=1}^{m} \ln(s_i) \\ & & \\ & \text{s.t.} & Ax + s = b \\ & & s \ge 0 \\ & & Mx = g\ . \end{array}$$

The analytic center possesses a very nice "centrality" property. Suppose that $(\hat{x}, \hat{s})$ is the analytic center. Define the following matrix:

$$\hat{S}^{-2} := \begin{pmatrix} (\hat{s}_1)^{-2} & 0 & \ldots & 0 \\ 0 & (\hat{s}_2)^{-2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & (\hat{s}_m)^{-2} \end{pmatrix}.$$

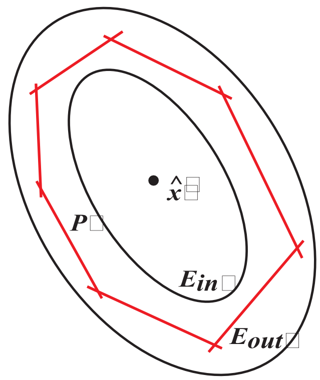

Figure 1: Illustration of the Ellipsoid construction at the analytic center.

Next define the following sets:

$$P := \{x \mid Mx = g,\ Ax \leq b\}$$

$$E_{\text{IN}} := \left\{x \mid Mx = g,\ \sqrt{(x-\hat{x})^T A^T \hat{S}^{-2} A(x-\hat{x})} \leq 1\right\}$$

$$E_{\text{OUT}} := \left\{x \mid Mx = g,\ \sqrt{(x-\hat{x})^T A^T \hat{S}^{-2} A(x-\hat{x})} \leq m\right\}$$

Theorem 1.1 *If $(\hat{x}, \hat{s})$ is the analytic center, then:*

$$E_{\text{IN}} \subset P \subset E_{\text{OUT}}\ .$$

This theorem is illustrated in Figure 1.

The theorem is actually pretty easy to prove.

Proof: Suppose that $x \in E_{\mathrm{IN}}$, and let $s = b - Ax$. Since $Mx = g$, we need only prove that $s \geq 0$ to show that $x \in P$. By construction of E_{IN}, s satisfies $(s-\hat{s})^T \hat{S}^{-2}(s-\hat{s}) \leq 1$, where $\hat{s} = b - A\hat{x}$. This in turn can be written as:

$$\sum_{i=1}^{m} \frac{(s_i - \hat{s}_i)^2}{\hat{s}_i^2} \leq 1 \ ,$$

whereby we see that each s_i must satisfy $s_i \geq 0$. Therefore $Ax \leq b$ and so $x \in P$.

We can write the optimality conditions (KKT conditions) for problem ACP as:

$$\begin{aligned} -\hat{S}^{-1}e + \lambda &= 0 \\ 0 + A^T\lambda + M^T u &= 0 \\ A\hat{x} + \hat{s} &= b \\ M\hat{x} &= g \ , \end{aligned}$$

where $e = (1, \ldots, 1)^T$, i.e., the e is the vector of ones.

From this we can derive the following fact: if (x, s) is feasible for problem ACP, then

$$e^T \hat{S}^{-1} s = e^T \hat{S}^{-1}(b - Ax) = \lambda^T b + u^T Mx = \lambda^T b + u^T g \ .$$

Since this is true for any (x, s) feasible for ACP, it will also be true for $(\hat{x}, \hat{s})$ (where $\hat{s} = b - A\hat{x}$), and so

$$e^T \hat{S}^{-1} s = \lambda^T b + u^T g = e^T \hat{S}^{-1} \hat{s} = m \ .$$

This means that s must lie in the set

$$T := \left\{ s \mid s \geq 0, e^T \hat{S}^{-1} s = m \right\} \ .$$

Now the extreme points of T are simply the vectors:

$$v^1 := m\hat{s}_1 e^1, \ldots, v^m := m\hat{s}_m e^m \ ,$$

where e^i is the i^{th} unit vector. Notice that each of these extreme points v^i satisfies:

$$(v^i - \hat{s})^T \hat{S}^{-2} (v^i - \hat{s}) = (me^i - e)^T (me^i - e) = m^2 - m \le m^2 \ ,$$

and so any $s \in T$ will satisfy

$$(s - \hat{s})^T \hat{S}^{-2} (s - \hat{s}) \le m^2 \ .$$

Therefore for $x \in P$, $s = b - Ax$ will satisfy $\sqrt{(s-\hat{s})^T \hat{S}^{-2}(s-\hat{s})} \le m$, which is equivalent to $\sqrt{(x-\hat{x})^T A^T \hat{S}^{-2} A(x-\hat{x})} \le m$. This in turn implies that $x \in E_{\text{OUT}}$.
q.e.d.

2 Newton's Method

Suppose we want to solve:

$$\text{(P:)} \qquad \min f(x)$$

$$x \in \Re^n.$$

At $x = \bar{x}$, $f(x)$ can be approximated by:

$$f(x) \approx h(x) := f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^t H(\bar{x})(x - \bar{x}),$$

which is the quadratic Taylor expansion of $f(x)$ at $x = \bar{x}$. Here $\nabla f(x)$ is the gradient of $f(x)$ and $H(x)$ is the Hessian of $f(x)$.

Notice that $h(x)$ is a quadratic function, which is minimized by solving $\nabla h(x) = 0$. Since the gradient of $h(x)$ is:

$$\nabla h(x) = \nabla f(\bar{x}) + H(\bar{x})(x - \bar{x}) \ ,$$

we therefore are motivated to solve:

$$\nabla f(\bar{x}) + H(\bar{x})(x - \bar{x}) = 0 \ ,$$

which yields

$$x - \bar{x} = -H(\bar{x})^{-1}\nabla f(\bar{x}).$$

The direction $-H(\bar{x})^{-1}\nabla f(\bar{x})$ is called the *Newton direction*, or the *Newton step* at $x = \bar{x}$.

This leads to the following algorithm for solving (P):

Newton's Method:

Step 0 Given x^0, set $k \leftarrow 0$

Step 1 $d^k = -H(x^k)^{-1}\nabla f(x^k)$. If $d^k = 0$, then stop.

Step 2 Choose stepsize $\alpha^k = 1$.

Step 3 Set $x^{k+1} \leftarrow x^k + \alpha^k d^k, \ k \leftarrow k + 1$. Go to **Step 1**.

Note the following:

- The method assumes $H(x^k)$ is nonsingular at each iteration.
- There is no guarantee that $f(x^{k+1}) \leq f(x^k)$.
- Step 2 could be augmented by a linesearch of $f(x^k + \alpha d^k)$ to find an optimal value of the stepsize parameter α.

Proposition 2.1 *If $H(x)$ is SPD, then $d = -H(x)^{-1}\nabla f(x)$ is a descent direction, i.e. $f(x + \alpha d) < f(x)$ for all sufficiently small values of α.*

Proof: It is sufficient to show that $\nabla f(x)^t d = -\nabla f(x)^t H(x)^{-1}\nabla f(x) < 0$. This will clearly be the case if $H(x)^{-1}$ is SPD. Since $H(x)$ is SPD, if $v \neq 0$,

$$0 < (H(x)^{-1}v)^t H(x)(H(x)^{-1}v) = v^t H(x)^{-1}v,$$

thereby showing that $H(x)^{-1}$ is SPD. ∎

2.1 Rates of convergence

A sequence of numbers $\{s_i\}$ exhibits *linear* convergence if $\lim_{i\to\infty} s_i = \bar{s}$ and

$$\lim_{i\to\infty} \frac{|s_{i+1} - \bar{s}|}{|s_i - \bar{s}|} = \delta < 1.$$

If $\delta = 0$ in the above expression, the sequence exhibits *superlinear* convergence.

A sequence of numbers $\{s_i\}$ exhibits *quadratic* convergence if $\lim_{i\to\infty} s_i = \bar{s}$ and

$$\lim_{i\to\infty} \frac{|s_{i+1} - \bar{s}|}{|s_i - \bar{s}|^2} = \delta < \infty.$$

Examples:

Linear convergence: $s_i = \left(\frac{1}{10}\right)^i$: 0.1, 0.01, 0.001, etc. $\bar{s} = 0$.

$$\frac{|s_{i+1} - \bar{s}|}{|s_i - \bar{s}|} = 0.1.$$

Superlinear convergence: $s_i = \frac{1}{i!}$: 1, $\frac{1}{2}$, $\frac{1}{6}$, $\frac{1}{24}$, $\frac{1}{125}$, etc. $\bar{s} = 0$.

$$\frac{|s_{i+1} - \bar{s}|}{|s_i - \bar{s}|} = \frac{i!}{(i+1)!} = \frac{1}{i+1} \to 0 \text{ as } i \to \infty.$$

Quadratic convergence: $s_i = \left(\frac{1}{10}\right)^{(2^i)}$: 0.1, 0.01, 0.0001, 0.00000001, etc. $\bar{s} = 0$.

$$\frac{|s_{i+1} - \bar{s}|}{|s_i - \bar{s}|^2} = \frac{(10^{2^i})^2}{10^{2^{i+1}}} = 1.$$

Theorem 2.1 (Quadratic Convergence Theorem) *Suppose $f(x) \in C^3$ on $\Re^n$ (i.e., $f(x)$ is three times continuously differentiable) and x^* is a point that satisfies:*

$$\nabla f(x^*) = 0 \quad \textit{and} \quad H(x^*) \;\; \textit{is} \;\; \textit{nonsingular.}$$

If Newton's method is started sufficiently close to x^*, the sequence of iterates converges quadratically to x^*.

Example 1: Let $f(x) = 7x - \ln(x)$. Then $\nabla f(x) = f'(x) = 7 - \frac{1}{x}$ and $H(x) = f''(x) = \frac{1}{x^2}$. It is not hard to check that $x^* = \frac{1}{7} = 0.142857143$ is the unique global minimum. The Newton direction at x is

$$d = -H(x)^{-1}\nabla f(x) = -\frac{f'(x)}{f''(x)} = -x^2\left(7 - \frac{1}{x}\right) = x - 7x^2.$$

Newton's method will generate the sequence of iterates $\{x^k\}$ satisfying:

$$x^{k+1} = x^k + (x^k - 7(x^k)^2) = 2x^k - 7(x^k)^2 \ .$$

Below are some examples of the sequences generated by this method for different starting points.

k	x^k	x^k	x^k	x^k
0	1.0	0	0.1	0.01
1	-5.0	0	0.13	0.0193
2	-185.0	0	0.1417	0.03599257
3	$-239,945.0$	0	0.14284777	0.062916884
4	-4.0×10^{11}	0	0.142857142	0.098124028
5			0.142857143	0.128849782
6				0.1414837
7				0.142843938
8				0.142857142
9				0.142857143
10				0.142857143

By the way, the "range of convergence" for Newton's method for this function happens to be

$$x \in (0.0 \ , \ 0.2857143) \ .$$

Example 2: $f(x) = -\ln(1 - x_1 - x_2) - \ln x_1 - \ln x_2$.

$$\nabla f(x) = \begin{bmatrix} \frac{1}{1-x_1-x_2} - \frac{1}{x_1} \\ \\ \frac{1}{1-x_1-x_2} - \frac{1}{x_2} \end{bmatrix},$$

$$H(x) = \begin{bmatrix} \left(\frac{1}{1-x_1-x_2}\right)^2 + \left(\frac{1}{x_1}\right)^2 & \left(\frac{1}{1-x_1-x_2}\right)^2 \\ \left(\frac{1}{1-x_1-x_2}\right)^2 & \left(\frac{1}{1-x_1-x_2}\right)^2 + \left(\frac{1}{x_2}\right)^2 \end{bmatrix}.$$

$x^* = \left(\frac{1}{3}, \frac{1}{3}\right)$, $f(x^*) = 3.295836866$.

k	x_1^k	x_2^k	$\|x^k - x^*\|$
0	0.85	0.05	0.58925565098879
1	0.717006802721088	0.0965986394557823	0.450831061926011
2	0.512975199133209	0.176479706723556	0.238483249157462
3	0.352478577567272	0.273248784105084	0.0630610294297446
4	0.338449016006352	0.32623807005996	0.00874716926379655
5	0.333337722134802	0.333259330511655	$7.41328482837195e^{-5}$
6	0.333333343617612	0.33333332724128	$1.19532211855443e^{-8}$
7	0.333333333333333	0.333333333333333	$1.57009245868378e^{-16}$

Some remarks:

- Note from the statement of the convergence theorem that the iterates of Newton's method are equally attracted to local minima and local maxima. Indeed, the method is just trying to solve $\nabla f(x) = 0$.
- What if $H(x^k)$ becomes increasingly singular (or not positive definite)? In this case, one way to "fix" this is to use $H(x^k) + \epsilon I$.
- The work per iteration of Newton's method is $O(n^3)$
- So-called "quasi-Newton methods" use approximations of $H(x^k)$ at each iteration in an attempt to do less work per iteration.

3 Modification of Newton's Method with Linear Equality Constraints

Here we consider the following problem:

$$\text{(P:)} \qquad \text{minimize}_x \quad f(x)$$

$$\text{s.t.} \qquad Ax = b.$$

Just as in the regular version of Newton's method, we approximate the objective with the quadratic expansion of $f(x)$ at $x = \bar{x}$:

$$(\tilde{\text{P}}:) \qquad \text{minimize}_x \quad h(x) := f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \tfrac{1}{2}(x - \bar{x})^t H(\bar{x})(x - \bar{x})$$

$$\text{s.t.} \qquad Ax = b.$$

Now we solve this problem by applying the KKT conditions, and so we solve the following system for (x, u):

$$\begin{array}{rcl} Ax & & = & b \\ \nabla h(x) & +A^T u & = & 0 \ . \end{array}$$

Now let us substitute the fact that:

$$\nabla h(x) = \nabla f(\bar{x}) + H(\bar{x})(x - \bar{x}) \quad \text{and} \quad A\bar{x} = b.$$

Substituting this and replacing $d = x - \bar{x}$, we have the system:

$$\begin{array}{rcl} Ad & & = & 0 \\ H(\bar{x})d & +A^T u & = & -\nabla f(\bar{x}) \ . \end{array}$$

The solution (d, u) to this system yields the Newton direction d at $\bar{x}$.

Notice that there is actually a closed form solution to this system, if we want to pursue this route. It is:

$$
\begin{aligned}
d &= -H(\bar{x})^{-1}\nabla f(\bar{x}) + H(\bar{x})^{-1}A^T\left(AH(\bar{x})^{-1}A^T\right)^{-1}AH(\bar{x})^{-1}\nabla f(\bar{x}) \\
u &= -\left(AH(\bar{x})^{-1}A^T\right)^{-1}AH(\bar{x})^{-1}\nabla f(\bar{x}) \ .
\end{aligned}
$$

4 Web-Based ACA (Adaptive Conjoint Analysis)