Barrier Methods (Method of Centers)

The methods we've considered so far have used a potential function to analyze them, which is a useful tool and novel in linear programming. The method we are going to consider now are more traditionally motivated, being based on Newton's Method (1600s) and barrier methods (Fiacco & McCormick 1988).

Newton's Method

Unconstrained problems

\[
\min_{x \in \mathbb{R}^n} \, f(x).
\]

Having current point \( x \).

Use Taylor series expansion to approximate \( f(x) \)

Quadratic approximation of \( f(x) \)

\[
p(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k)^T \nabla^2 f(x_k) (x - x_k)
\]
In one dimension:

\[ f(x) \approx f(\bar{x}) + \frac{df}{dx}(x-\bar{x}) + \frac{1}{2} \frac{d^2f}{dx^2}(x-\bar{x})^2 = g(x) \]

\[ \text{min } g(x) = \text{Optimal soln: } \hat{x} = \bar{x} - \frac{df}{dx} \frac{d^2f}{dx^2} \]

provided \( \frac{d^2f}{dx^2} > 0 \),

\[ g(x) = \frac{df}{dx}(x) + \frac{d^2f}{dx^2}(x-\bar{x}) \]

In \( n \)-dimensions:

\[ f(x) \approx f(\bar{x}) + \frac{df}{dx}(x-\bar{x}) + \frac{1}{2} (x-\bar{x})^T \frac{d^2f}{dx^2}(x-\bar{x}) = g(x) \]

\( \text{min } g(x) = \text{x in } \mathbb{R}^n \)

\( \text{provided } \frac{d^2f}{dx^2}(\bar{x}) \text{ is positive definite, the optimal solution is } \hat{x} = \bar{x} - \left( \frac{d^2f}{dx^2}(\bar{x}) \right)^{-1} \nabla f(\bar{x}) \)

So, update \( \bar{x} \) to \( \hat{x} \) and repeat.

Note: If \( \hat{x} \) is optimal then \( \nabla f(\hat{x}) = 0 \), so \( \hat{x} = \bar{x} \) also.

Note: This converges quadratically, i.e. \( \frac{||x^{n+1} - \bar{x}||}{||x^n - \bar{x}||} \leq C \).
Consider problems
\[ \min \quad f(x) \]
\[ A x = b. \]

Assume there is a point \( x \) with \( A x = b \),
\[
 g(x) = f(x) + \nabla f(x)^T (x - x) + \frac{1}{2} (x - x)^T \nabla^2 f(x) (x - x)
\]
= second order Taylor series approximation of \( f \).

So update \( x \) by solving
\[
 w = J(x) \]
\[ A w = b. \]

Let \( p = x - x \). So equivalently, solve
\[
 \min_{p} \quad \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x) p
\]
\[ A^T p = 0. \]

Let \( h = \nabla f(x) \), \( H = \nabla^2 f(x) \). So equivalently, solve
\[
 \min \quad h^T p + \frac{1}{2} p^T H p
\]
\[ \text{sv.} \quad -A^T p = 0. \]

KKT Kuhn-Tucker conditions:
\[
 h + H p - A^T \pi = 0 \quad \pi = \text{Lagrange multipliers},
\]
\[ -A^T p = 0. \]

i.e.,
\[
 \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \pi \end{bmatrix} = \begin{bmatrix} 0 \\ h \end{bmatrix} \]
\[ \text{& solve this system.} \]
Basic Method & LP: All, et al. (1986) E253

\[
\begin{align*}
\min & & c^T x \\
& & A x = b \\
& & x \geq 0.
\end{align*}
\]

Consider instead \( \min \ c^T x - \mu \sum_i x_i \)
\( A x = b \),
\[
\text{with } \mu \geq 0.
\]

\[
\begin{array}{c}
\begin{array}{c}
\text{Graphical representation of the function.}
\end{array}
\end{array}
\]

Solution by Newton's Method:

\[
\begin{align*}
\hat{f}(x) &= c^T x - \mu \sum_i x_i \\
\hat{f}'(x) &= c - \mu x - e \\
\hat{f}''(x) &= \mu x^2.
\end{align*}
\]

So: Solve \( \hat{f}' x = 0 \) subject to

\[
\begin{bmatrix}
\mu x^{-2} & A^T \lambda \\
A & 0
\end{bmatrix}
\begin{bmatrix}
-\mu \\
\lambda
\end{bmatrix} =
\begin{bmatrix}
c - \mu x \cdot e \end{bmatrix}
\]

Update \( x \) by \( x + \mu \).
Equivalently, let \( r = X^{-1} \) (so \( r \) = direction in scaled space) and solve

\[
\begin{bmatrix}
\mu & I \\
AX & 0
\end{bmatrix}
\begin{bmatrix}
-r \\
-\pi
\end{bmatrix}
= \begin{bmatrix}
x c - \mu e \\
0
\end{bmatrix}
\]

Notice: \( r \) is in \( \mathbb{W}(AX) \)

\( AX^T r \) is in row space of \( AX \).

Thus, \( \hat{r}_{ur} = P_{AX} (X c - \mu e) \)

so \( r \) is combination of \( \hat{r}_{ur} \) directions seen before.
The Primal-Dual Interior Point Method

(Kojima, Mizuno, Yoshire (27), Monteiro & Adler (37))

(Follow Narita et al. (31))

Have studied from linear programming problem:

\[
\begin{align*}
\min & \quad c^T x \\
\text{st.} & \quad A x = b \\
& \quad x \geq 0,
\end{align*}
\]

Introduce barrier function in objective function to replace the \( x \geq 0 \) constraints:

\[
\begin{align*}
\min & \quad c^T x - \mu \sum \ln x_i \\
\text{st.} & \quad b - A x = 0 \\
& \quad x \geq 0,
\end{align*}
\]

\((\text{Here, } \mu > 0)\)
Sketch of Algorithm

Given: \( x^0 > 0, y^0 > 0, z^0 > 0 \), with \( A^T x^0 = b, A^T y^0 + z^0 = c \). Given barrier parameter \( \mu^k > 0, k = 1 \).

While duality gap \( \rho = x^T z^k > 2 - L \)

Solve \( (P(\mu^k)) \)

Update \( x^k, y^k, z^k \) be solution to \( (P(\mu^k)) \) (approximately).

Reduce \( \mu^k \) if \( \mu^k L < \mu^k \), \( \mu^k > 0 \)

Let \( k = k + 1 \).

Repeat.

Could this be the central trajectory?

You do we reduce \( \mu^k \)? See later.
You do we solve \( P(\mu) \)?

Use KKT method for equations.

Find KKT conditions for \( P(\mu) \):

\[
\begin{align*}
  c - \mu^T x & = 0 \\
  b - A^T y & = 0 \\
  x & \geq 0 \\
  y & \geq 0
\end{align*}
\]

Let \( z \) be the dual slack.

Write \( z \) for \( \mu^T x \):

\[
\begin{align*}
  c - z - A^T y & = 0 \quad \text{Dual feasibility} \\
  b - A^T x & = 0 \quad \text{Primal feasibility} \\
  -\mu e - x^T z e & = 0 \quad \mu - \text{complementary slackness}
\end{align*}
\]

The third equation can be written \( x_i z_i = \mu \) for \( i = 1, \ldots, n \),

i.e., \( z_i = \frac{\mu}{x_i} \).

Thus, if \( \mu = 0 \), this is complementary slackness, so the optimality condition becomes exactly the standard optimality condition.

So, call the condition \( \mu e - x^T z e = 0 \)

\( \mu - \text{complementary slackness} \).
Can write these conditions as:
\[ F(x, y, z) = 0, \text{ when } F : \mathbb{R}^{2n+m} \to \mathbb{R}^{2n+m} \text{ is given by } \]
\[ F(x, y, z) = \begin{bmatrix} c - A^T y - z \\ b - A x \\ m/c - x \end{bmatrix}. \]

Newton's method for finding a solution to a system of equations \( F(x) = 0 \).

Given a current point \( \bar{w} \), find a direction \( dw \) by solving the equation
\[ J(\bar{w}) \cdot dw = -F(\bar{w}) \]
where \( J(\bar{w}) \) is the Jacobian of \( F \) evaluated at \( \bar{w} \), i.e., \( J(\bar{w}) \) is a matrix whose \((i,j)\)th entry is found by differentiating the \(i\)th element of the vector \( F(w) \) with respect to the \(j\)th component of \( w \).

Eg: Let \( F(w_1, w_2) = \begin{pmatrix} +2w_1^2 - \frac{4}{w_2} \\ 2w_1w_2 \\ 2w_1 \end{pmatrix} \) (Optimal point: \( \bar{w} = (0, 2) \)).

Then \( J(\bar{w}) = \begin{bmatrix} 0 & 2w_2 \\ 2w_1 & 2w_1 \\ 2 & 2 \end{bmatrix} \)
Let \( \bar{w} = (1) \) be initial point.

So, find \( dw \) by solving
\[ J(1,1) \cdot \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} = -F(1,1) \]
\[ \Rightarrow \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \cdot \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \Rightarrow \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}. \]
\[ \Rightarrow dw_1 = \frac{3}{2}, dw_2 = -\frac{1}{2}. \]

New point: \( \bar{w} = \bar{w} + dw = \left( \frac{5}{2} \right) + \left( \frac{3}{2} \right) = \left( \frac{5}{2}, \frac{3}{2} \right). \)
Return to \( F(x,y,z) = \begin{pmatrix} c-A^Ty-z \\ b-Ax \\ \mu e - \lambda e \end{pmatrix} \).

Then \( J(x,y,z) = \begin{bmatrix} 0 & -A^T & -I \\ -A & 0 & 0 \\ -2 & 0 & -X \end{bmatrix} \).

Let \( x, y, z \) be current point with \( x > 0, \bar{x} > 0, Ax = b, A^Ty + z = c, \) but \( \mu e - \lambda e \neq 0 \).

5. Find direction by solving

\[
\begin{bmatrix}
0 & -A^T & -I \\
-A & 0 & 0 \\
-2 & 0 & -X
\end{bmatrix}
\begin{bmatrix}
dx \\
dy \\
dz
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
-\mu e + \lambda e
\end{bmatrix}
\]

i.e.,

\[
\begin{align*}
A^Tdy + dz &= 0 \\
A dx + Z dx + X dx &= -\mu e - \lambda e
\end{align*}
\]

Substitute in (3) from (1):

\[
\frac{\lambda}{2} dx \times \bar{X} A^T dy = \mu e - \lambda e
\]

Multiply by \( \bar{X} A^{-1} \):

\[
A dx \times \bar{X} A^{-1} \times A^T dy = \bar{X} A^{-1} (\mu e - \lambda e)
\]

From (1), \( Ax = 0 \), we get

\[
\int A^{-1} \times \bar{X} A^T dy = -\mu \bar{X} A^{-1} e + X \bar{X} e = b - \mu \bar{X} A^{-1} e.
\]
Let $D^2 = \bar{Z}^{-1}X$.

So solve $(AD^2A^T)dy = b - \mu A\bar{Z}^{-1}e$.

Then

\[
\begin{align*}
  dz &= -A^Tdy \\
  dx &= \bar{Z}^{-1}(\mu e - X\bar{Z}e - Xdz)
\end{align*}
\]

Can show that

\[
dx = D\bar{P}_{AD}D(\mu X^{-1}e - Ze)
\]
In $O \left( \frac{1}{|L|} \right)$ iterations, we can find the following algorithm (Will follow Monteiro & Adler (37)). Also (Bieler, p. 83-91).

By choosing $\mu$ appropriately, we can show that we only need to spend an iteration on each subproblem.

Can maintain $\|X z e - p e\| \leq \Theta \mu$, where $0 \leq \Theta < 1$.

Case $\Theta = 0.35$ works.

We have neighborhood $N_x$.

For each $x \geq 0$, $Ax = b$ and $z \geq 0$, $Ay + z = c$ for some $y^2$.

are nonempty, then the set of optimal solutions to (P) and (D)

are nonempty and bounded.

Proof: (L) exercise

Case $\Theta > 0$.

If $x^* \geq 0$, $Ax = b$ and $z^* \geq 0$, $Ay + z = c$ for some $y^2$.

are nonempty and bounded.

Then, $(P^*(\mu))$ has a unique optimal solution, for each $\mu > 0$.

Proof: $P^*(\mu)$ is convex (p. 130, 131).

The objective function is convex (p. 130, 132, 134, 135, 136).

Unif. convex open set.

The feasible region is bounded, so $x^* \in P^*(\mu)$ has a unique minimizer.

Let $\ell = \inf_{x \in X} \{ c^T x + d \}$.

Then, there is some $d, d > 0$ such that $c^T x + d$ is decreasing monotonically as $c^T x + d \rightarrow c^T x_0 + d_0$.

for all $(x, d)$. But

$c^T x + d \rightarrow c^T x_0 + d_0$ as $d \rightarrow 0$

if $\mu$ is a regularity condition.
Corollary. Under the assumptions, $P(\mu)$ has a unique optimal solution for each $\mu > 0$.

Proof. Follows immediately from corresponding theory.

Note: Even if optimal solution to $P(\mu)$ is not unique, the optimal solution to $P(\mu)$ is.

Proposition. Under the assumptions, as $\mu \to 0$, $x(\mu)$ and $(y(\mu), z(\mu))$ converge to optimal solutions to (P) and (D) respectively.

No proof, but follows from fact that the duality gap is $x(\mu)^T z(\mu) = \mu$. 

\[\square\]
Let \( W = \{ (x, y, z) : Ax = 0, A^Ty + z = c, x > 0, z > 0 \} \).

Let \( \theta, \delta \) be constants satisfying \( 0 \leq \theta < \frac{1}{4}, 0 < \delta < \frac{1}{4} \), and
\[
\theta^2 + \delta^2 \leq \frac{\theta(1 - \delta)}{\delta(1 - \theta)}.
\]

\( \theta = \delta = 0.15, \theta = \delta = 0.4 \) (Wright, p. 91)

Criterion of closeness to central part:
\[ \| XZe - \mu \| \leq \theta \mu. \]

\[ (x_2, x_2e) = \left( (1 - \theta) \mu, \mu \right) \]
\[ (x_2, x_2e) = \left( (1 - \theta) \mu, \mu \right) \]
\[ (x_2, x_2e) = \left( (1 + \theta) \frac{\mu}{\mu}, (1 - \frac{\theta}{\delta}) \mu \right) \]

All close to \( \mu \).

"Complementary" space.

\[ \| XZe - \mu \| \leq \theta \mu. \]

Each value of \( \mu \) gives a circle.

The union of all these circles is a pointed cone.

For a given \( x, z \), the "best" \( \mu \) is
\[ \mu = \frac{x_2}{v} = \mu \text{ which minimizes } \| XZe - \mu \|. \]
Algorithm

Step 0: Assume we have an initial point \( w^0 = (x^0, y^0, z^0) \in \mathcal{W} \) satisfying \( \| x^0 - \mu^0 e \| \leq \Theta \mu^0 \). Let \( \mu^0 = \frac{x^0 e}{n} \).

Step 1: If \( x^0 e \leq \varepsilon \), stop.

Step 2: Set \( w_{-1} = \mu^0 \left( 1 - \frac{\varepsilon}{\sqrt{n}} \right) \). Calculate direction \( d_{-1} = \nabla f(w_{-1}) \mu^0 \).

Step 3: Set \( w_{0} = w_{-1} + d_{-1} \).

Set \( k = k + 1 \), return to Step 1.

Theorem. Assume that \( w \in \mathcal{W} \) satisfies \( \| x^2 - \mu e \| \leq \Theta \mu^0 \), where \( \mu = \frac{x^2}{n} \). Let \( \hat{\mu} = \mu \left( 1 - \frac{\varepsilon}{\sqrt{n}} \right) \). Let \( \Delta = w + d \). Then

(a) \( \| x^2 - \hat{\mu} e \| \leq \Theta \hat{\mu} \)  
(b) \( \hat{\mu} \) is in \( \mathcal{W} \).  
(c) \( x^2 = \hat{\mu} e \).

Still close  
Still in \( \mathcal{W} \)  
Still have the best \( \mu \).
Corollary: The sequence of points $w^n$ generated by the algorithm satisfies

(a) $\|x^n 2^k e^{-\mu w^n}\| \leq 2^n \mu w^n$
(b) $w^n$ is $\mathcal{L}$-close.
(c) $x^n 2^k = np_w w^n$

where $\mu_w = \frac{\mu_w}{\sqrt{n}} (1 - \frac{\sigma}{\sqrt{n}})$.

Proof: Follows directly from the Theorem.

Prop: The total number of iterations is at most greater than $\mathcal{O} = \frac{\ln(np_0)}{\ln\delta}$.

Proof: Since $x^n 2^k \leq x$, i.e. $np_w \leq x$.

Now, $np_w = np_0 (1 - \frac{\sigma}{\sqrt{n}})^k$.

Thus, we need $\mu_w$ to satisfy $\mu_w (1 - \frac{\sigma}{\sqrt{n}})^k \leq \epsilon$, i.e.

$\ln \frac{\epsilon}{\mu_w} \geq \ln np_0 + \ln (1 - \frac{\sigma}{\sqrt{n}}) = \ln np_0 + \ln \frac{\sqrt{n}}{\sqrt{n} - \sigma} \geq \ln np_0 + \ln \frac{\sqrt{n}}{\sqrt{n} - \sigma}$.

Since $\ln (1 - \alpha) \leq -\alpha$ for $\alpha < 1$, we get

$\ln \epsilon \geq \ln np_0 - \ln \frac{\sigma}{\sqrt{n}}$, i.e. $\ln \epsilon \geq \ln np_0 \frac{\sqrt{n}}{\sigma}$. \[1\]

So, provided $p_0$ is $O(\frac{\ln n}{\ln \delta})$, all operations are $O(n^2 \ln n \ln(n \ln L))$.

Since each iteration $O(\frac{\ln n}{\ln \delta})$ and fewer ops, all ops are $O(\frac{n^2 \ln n \ln(n \ln L)}{\ln\delta})$. \[2\]

Sou.
(i) \( \dot{x} = \dot{z} = \mu + \Delta x \Delta z \)

(ii) \( (\Delta x)^T (\Delta z) = 0 \)

(iii) \( x^T z = \nu \mu \)

Proof: (i) \( x \dot{z} = (x \Delta x)(\Delta z) = x \Delta z + \Delta x \Delta z + z \Delta x \Delta z + \Delta x \Delta z \)

\[ = \mu + \Delta x \Delta z \quad \text{by from the system \( J \Delta w = -F(w) \).} \]

(ii) \( \Delta x = 0 \), \( A^T \Delta y + \Delta z = 0 \)

i.e. \( \Delta x^T \Delta z = -\Delta x^T A^T \Delta y = 0 \).

(iii) \( x^T z = \sum x_i \dot{z}_i = \nu \mu + \sum \Delta x_i \Delta z_i = \nu \mu \). \( \square \)

(i) and (ii) show that \( \mu \) does not vary very far from the central trajectory. Second order change \( \Delta x^T \Delta z = 0 \), so second order change keeps us on the plane \( x^T z = \mu \), i.e. first order change never moves us down to the plane \( x^T z = \mu \) in the space, as shown by (iii) in the "confining" space.
Then the proof shows that \( DX \Delta Z \in \text{has small norm} \) so step are short, so can't move too far away, (a) will follow.

(b) follow from (a) and the fact that \( DX \Delta Z \in \text{has small norm} \). Only need to worry about non-negativity, (c) \( \Rightarrow \) if \( X < 0 \) then \( X < 0 \), but can't do this if \( DX \Delta Z \in \text{has small norm} \).

(c) follow directly from (iii) of Proposition.

1. First, \( \|DX \Delta Z e\| = \|X^2 e - \mu e\| \leq \frac{2^{14/\mu}}{\mu} \) for \( \mu \geq 54 \) (eq. (5.11))

But here term can be bounded from \( \|X^2 e - \mu e\| \leq \theta \mu \)
Consider the problem
\[ \min \quad c^T x + \frac{1}{2} x^T Q x \]
\[ A x = b \quad (P), \text{ where } Q \text{ is symmetric positive semi-definite matrix.} \]

The method we used for CP was a nonlinear method. So apply it to the problem. Set up \( P(\mu) \):

\[ \min \quad c^T x + \frac{1}{2} x^T Q x - \mu \sum (a_i x_i) \]
\[ A x = b \]
\[ x \geq 0. \]

Gradient of objective is now \( c + Q x - \mu x - \varepsilon \), so optimality conditions are:
\[ c + Q x - A^T y - z = 0 \]
\[ b - A x = 0 \]
\[ \mu e - x - Z e = 0. \]

Again, solve the system by Newton's method:

Jacobian is:
\[
\begin{bmatrix}
Q & -A^T & -I \\
-A & 0 & 0 \\
-Z & 0 & -x
\end{bmatrix}
\]
Solve
\[
\begin{bmatrix}
Q & -A^T - I \\
-A & 0 & 0 \\
-2 & 0 & -x
\end{bmatrix}
\begin{bmatrix}
dx \\
dy \\
dz
\end{bmatrix} =
\begin{bmatrix}
0 \\
\frac{c}{x + \bar{y} - \bar{z}} \\
\bar{x} \bar{z} - \mu \bar{x}
\end{bmatrix},
\tag{6}
\]
where \(\bar{x}, \bar{y}, \bar{z}\) is current point and \(dx, dy, dz\) is direction.

(again, assume satisfy primal and dual feasibility, with toward, complementary slackness.)

Can set up a dual problem to (P):

\[
\begin{align*}
\max \quad b^T y - \frac{1}{2} x^T Q x \\
\text{s.t.} \quad A^T y - Q x + z = c \\
\quad x \geq 0.
\end{align*}
\tag{D}
\]

Then we get weak duality: \(b^T y - \frac{1}{2} x^T Q x \leq c^T x + \frac{1}{2} x^T Q x\) for any feasible \(x, y, z\).

In fact, \((c^T x + \frac{1}{2} x^T Q x) - (b^T y - \frac{1}{2} x^T Q x) = x^T z\).

So get strong duality: \(\exists \quad (x, y) \in (P)\) has an optimal solution \(\bar{x}\), the \(\bar{x}\) has an optimal solution with \(c^T \bar{x} + \frac{1}{2} \bar{x}^T Q \bar{x} = b^T \bar{y} - \frac{1}{2} \bar{x}^T Q \bar{x}\).