Interior-Point Methods

Stephen Wright

University of Wisconsin-Madison

Simons, Berkeley, August, 2017
Outline

- Introduction: Problems and Fundamentals
- Background: Newton’s Method, Central Path and Neighborhoods
- Primal-Dual Path Following for LP: Algorithm and Complexity
- Extensions: QP, LCP
- Primal and Dual Barrier methods
- Relationships between Dual Barrier and Primal-Dual

I give few citations (but note that several participants in this program have contributed hugely).

The following sources provide more background and bibliographies:

[Nesterov and Nemirovskii, 1994],
[Wright, 1997],
[Ye, 1997],
[Vandenberghe, 2016],
[Renegar, 2001]
Ethos

There’s extremely elegant theory underlying interior-point methods, that required the development of analysis: e.g. self-concordance, self-scaled barriers and cones, primal-dual potential functions. This theory was instrumental in extensions beyond LP, QP, monotone Linear Complementarity (LCP) to general conic programming, including semidefinite programming (SDP).

- [Todd, 2001] (SDP review)
- [Nesterov and Nemirovskii, 1994] (self-concordance, primal barriers, conic programming)
- [Vandenberghe, 2016] (major elements of theory in slides for class)
- [Renegar, 2001] (elegant treatment of main algorithmic ideas)

However, for LP / QP / LCP, algorithms can be developed and powerful results proved using only very elementary mathematical tools.

That’s the focus of this talk.
Linear Programming (LP)

Minimize a linear function in $\mathbb{R}^n$ over a polyhedron.

$$\min c^T x \text{ s.t. } Ax = b, \ x \geq 0.$$  \hspace{1cm} (P)

$A$ is $m \times n$ real matrix. WLOG assume full rank, so $m \leq n$. Dual:

$$\max b^T z \text{ s.t. } A^T z \leq c.$$  \hspace{1cm} \hspace{1cm}

Introduce dual slack variables for equivalent formulation:

$$\max b^T z \text{ s.t. } A^T z + s = c, \ s \geq 0.$$  \hspace{1cm} (D)

KKT conditions:

$$Ax = b, \ A^T z + s = c, \ 0 \leq x \perp s \geq 0,$$  \hspace{1cm} (KKT)

where $\perp$ means $x^T s = 0$. Thus $x_i \geq 0, s_i \geq 0$, and $x_is_i = 0$ for all $i$.

If $(x^*, z^*, s^*)$ satisfies KKT, then $x^*$ solves (P) and $(z^*, s^*)$ solves (D).
Weak Duality: If $x$ is feasible for (P) and $(z, s)$ is feasible for (D), then:

$$c^T x \geq b^T z.$$ 

Proof: You’ve seen it already this week. Twice!

Strong Duality: Exactly three possible scenarios for the primal-dual pair:

(i) (P) and (D) both have solutions $x^*$ and $(z^*, s^*)$, and their objectives are equal at optimality: $c^T x^* = b^T z^*$;
(ii) One of (P), (D) is unbounded and the other is infeasible;
(iii) Both (P) and (D) are infeasible.

Note: Don’t require a regularity condition (e.g. Slater).

Proof: Much trickier! Uses e.g. finite termination of simplex method with anti-cycling rules.
Duality and Complementarity

If \((x, z, s)\) is primal-dual feasible, all optimality conditions except complementarity are satisfied:

\[
Ax = b, \quad A^T z + s = c, \quad x \geq 0, \quad s \geq 0,
\]

All linear! Easy to place constraints on steps to ensure that all iterates \((x^k, z^k, s^k)\) remain feasible.

For feasible \((x, z, s)\), we have

\[
0 \leq \sum_{i=1}^{n} x_i s_i = x^T s = x^T (c - A^T z) = c^T x - b^T z,
\]

so that \(c^T x \geq b^T z\). (This is the famous one-line proof of weak duality.)

Notation: Measure near-optimality of a feasible triple \((x, z, s)\) by

\[
\mu = \frac{1}{n} x^T s = \frac{1}{n} \sum_{i=1}^{n} x_i s_i.
\]
KKT and Constrained Nonlinear Equations

Write KKT conditions as a system of constrained nonlinear equations: Define

\[ X = \text{diag}(x_1, x_2, \ldots, x_n), \quad S = \text{diag}(s_1, s_2, \ldots, s_n), \quad e = (1, 1, \ldots, 1)^T, \]

and rewrite KKT as follows:

\[
F_0(x, z, s) := \begin{bmatrix}
Ax - b \\
A^T z + s - c \\
X Se
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}, \quad (x, s) \geq 0.
\]

Note that \( F_0 \) is a square nonlinear system: \( F_0 : \mathbb{R}^{2n+m} \rightarrow \mathbb{R}^{2n+m} \).

Only “slightly” nonlinear: just the last term, which is

\[
X Se = \begin{bmatrix}
x_1 s_1 \\
x_2 s_2 \\
\vdots \\
x_n s_n
\end{bmatrix} = 0.
\]
Primal-Dual Interior-Point Approach

Basic primal-dual approach generates a sequence \((x_k, z_k, s_k)\) of strictly feasible triples, satisfying (note strict inequalities):

\[
\mathcal{F}^\circ := \{(x, z, s) : Ax = b, \ A^T z + s = c, \ x > 0, \ s > 0\}.
\]

- Steps are Newton-like steps on \(F_0\):

\[
(x_{k+1}, z_{k+1}, s_{k+1}) = (x_k, z_k, s_k) + \alpha_k(\Delta x^k, \Delta z^k, \Delta s^k), \quad \text{some} \ \alpha_k > 0.
\]

- Iterate toward complementarity, hence optimality:

\[
\mu_k := \frac{1}{n} (x^k)^T s^k \to 0 \quad \text{as} \ k \to \infty.
\]

- Keep the pairwise products \(x_i^k s_i^k, \ i = 1, 2, \ldots, n\) more or less balanced — not too far from their average value \(\mu_k\). (Explicitly or implicity.)
Complexity

Polynomial complexity follows from a decrease per iteration of the form

$$\mu_{k+1} \leq (1 - Cn^{-\nu})\mu_k, \quad k = 0, 1, 2, \ldots,$$

where $C > 0$ is independent of $n$ and $\nu \geq .5$ for conventional methods. Such rates require delicate manipulation of the modified Newton strategy for computing steps, and also of the choice of step length $\alpha_k$.

They lead to complexity results via a standard argument:

$$\mu_K \leq \epsilon \mu_0 \iff (1 - Cn^{-\nu})^K \leq \epsilon$$

$$\iff K \log(1 - Cn^{-\nu}) \leq \log \epsilon$$

$$\iff K(1 - Cn^{-\nu}) \leq \log \epsilon \quad \text{since } \log(1 + t) \leq t$$

$$\iff K \geq n^{\nu}\log|\epsilon|/C.$$

Thus get $\mu_K/\mu_0 \leq \epsilon$ in $O(n^{\nu}\log|\epsilon|)$ iterations.
Newton’s Method for Nonlinear Algebraic Equations

Consider $F : \mathbb{R}^N \to \mathbb{R}^N$ smooth. We wish to solve $F(w) = 0$.

**Basic Idea:** At some iterate $w$, one form of Taylor’s theorem implies

$$F(w + d) = F(w) + J(w)d + o(\|d\|),$$

where $o(\|d\|)/\|d\| \to 0$ as $d \to 0$, and $J(w)$ is the Jacobian ($N \times N$ matrix of first partial derivatives):

$$J(w) = \left[ \frac{\partial F_i}{\partial w_j}(w) \right]_{i=1,2,\ldots,n}^{j=1,2,\ldots,n}.$$

If $J(w)$ is nonsingular, Newton step is $d = -J(w)^{-1}F(w)$. From expansion above we get for this $d$:

$$F(w + d) = o(\|J(w)^{-1}\| \|F(w)\|).$$

If the smallest singular value of $J(w)$ is uniformly bounded below on some region (as in the neighborhood of a nondegenerate solution $w^*$), have

$$F(w + d) = o(\|F(w)\|).$$
Newton’s Method in 1D

This suggests an iterative scheme: Starting from some $w^0 \in \mathbb{R}^N$:

$$d^k = -J(w^k)^{-1}F(w^k), \quad w^{k+1} = w^k + d^k, \quad k = 0, 1, 2, \ldots .$$
Newton’s Method for Minimization

There is also a Newton’s method for minimization:

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

where \( f : \mathbb{R}^N \to \mathbb{R} \) is smooth (typically twice Lipschitz continuously differentiable).

Motivated again by Taylor’s theorem, using the approximation

\[
f(x + d) \approx f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d.
\]

When \( \nabla^2 f(x) \) is positive definite, the minimizer is

\[
d = -\nabla^2 f(x)^{-1} \nabla f(x).
\]

... the same step we get by applying nonlinear-equations Newton’s method to \( F(x) = \nabla f(x) \).

Local quadratic convergence to nondegenerate minimizers \( x^* \), for which \( \nabla f(x^*) = 0 \) and \( \nabla^2 f(x^*) \) is positive definite.
Newton’s Method for Minimization

By substituting the Newton step \( d = -\nabla^2 f(x)^{-1} \nabla f(x) \) into the quadratic approximation, obtain a “Newton decrement”

\[
    f(x + d) \approx f(x) - \frac{1}{2} \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) \\
    = f(x) - \frac{1}{2} d^T \nabla^2 f(x) d.
\]

Introduce notation for the decrement:

\[
    \lambda(x) := \|\nabla f(x)\|_{x*} = \|d\|_x,
\]

where

\[
    \|v\|_{x*} := \left( v^T \nabla^2 f(x)^{-1} v \right)^{1/2}, \quad \|w\|_x := \left( d^T \nabla^2 f(x) d \right)^{1/2}
\]

are Hessian-weighted norms [Vandenberghe, 2016].

We can make stronger statements about the decrement when \( f \) is self-concordant. It plays an important role in the analysis of primal barrier methods.
Nonlocal Results? Do we need them?

Under certain assumptions, Newton’s method converges \textit{locally} at a \textit{quadratic} rate.

What happens outside the domain of quadraticness? Are global rates of convergence available? Leading to complexity bounds?

- In general, no. For some functions Newton is as slow as first-order methods.
- But for self-concordant functions, \textit{yes}! These have 3rd derivatives along any direction bounded in terms of 2nd derivatives, so the quadratic approximation has guaranteed quality everywhere, and a scaled Newton step makes significant progress.
- For \textit{primal-dual methods for LP}, we can specialize the analysis to the particular form of $F$ (perturbed KKT conditions).
Primal-Dual Path-Following

Path-following primal-dual interior-point methods generate search directions from perturbations of the KKT system $F_0$, and choose step lengths $\alpha_k$ with reference to central path neighborhoods.

The Central Path is a trajectory in primal-dual space defined by a scalar-parametrized variant of the KKT system: For any $\tau > 0$ let $(x_\tau, z_\tau, s_\tau)$ be the solution of

$$F_\tau(x, z, s) := \begin{bmatrix} Ax - b \\ A^T z + s - c \\ XSe - \tau e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad (x, s) > 0.$$  

Central path is

$$C := \{(x_\tau, z_\tau, s_\tau) : \tau > 0\}.$$

Strictly feasible set $\mathcal{F}^\circ$ nonempty $\Rightarrow (x_\tau, z_\tau, s_\tau)$ is defined uniquely for each $\tau > 0$. 

Central Path Neighborhoods

Require more than strict feasibility of iterates, which is $(x^k, s^k) > 0$. Require iterates to stay within certain neighborhoods of the central path.

Want the pairwise products $x_is_i$ to be not too different for $i = 1, 2, \ldots, n$.

\[ \mathcal{N}_2(\theta) := \{(x, z, s) \in \mathcal{F}^\circ : \|XSe - \mu e\|_2 \leq \theta \mu, \ \mu = x^T s/n\}, \]

for some $\theta \in (0, 1)$. Typically $\theta = .25$ or $.5$.

\[ \mathcal{N}_{-\infty}(\gamma) := \{(x, z, s) \in \mathcal{F}^\circ : x_is_i \geq \gamma \mu, \ \mu = x^T s/n, \ i = 1, 2, \ldots, n\}, \]

for some $\gamma \in (0, 1)$. Typically $\gamma = .001$.

The $\mathcal{N}_{-\infty}$ neighborhood is wider than $\mathcal{N}_2$.

Note that $\mathcal{N}_{-\infty}(\gamma) \to \mathcal{F}^\circ$ as $\gamma \to 0$. (But $\mathcal{N}_2(\theta) \not\to \mathcal{F}^\circ$ as $\theta \to 1$.)
Path-Following Strategy

- Define a target point on the central path: \((x(\tau), z(\tau), s(\tau))\) for some \(\tau > 0\) that depends on the current iterate;
- Calculate a Newton step for \(F_\tau\) from the current iterate;
- Choose step length \(\alpha_k\) along the Newton step to stay inside a central path neighborhood.

We describe a long-step path-following algorithm (LPF) that starts from a point \((x^0, z^0, s^0) \in \mathcal{N}_\infty(\gamma)\) for some \(\gamma \in (0, 1)\) and achieves a geometric decrease per iteration in \(\mu\):

\[
\mu_{k+1} \leq (1 - \delta n^{-1})\mu_k, \quad k = 0, 1, 2, \ldots,
\]

for \(\delta\) independent of \(n\). Thus we have

\[
\mu_k \leq \epsilon \mu_0 \quad \text{in } O(n \log \epsilon) \text{ iterations.}
\]

It’s also fast in practice.
Long-Step Path Following

Notation:

\[ \mu^k = (x^k)^T s^k / n, \]
\[ (x^k(\alpha), z^k(\alpha), s^k(\alpha)) = (x^k, z^k, s^k) + \alpha(\Delta x^k, \Delta z^k, \Delta s^k). \]

Algorithm LPF.

Choose \( \gamma \in (0, 1) \), \( \sigma_{\text{min}} \) and \( \sigma_{\text{max}} \) with \( 0 < \sigma_{\text{min}} < \sigma_{\text{max}} < 1 \);  
Choose \( (x^0, z^0, s^0) \in \mathcal{N}_{-\infty}(\gamma) \);  
for \( k = 0, 1, 2, \ldots \) do  
Choose \( \sigma_k \in [\sigma_{\text{min}}, \sigma_{\text{max}}] \);  
Let \( (\Delta x^k, \Delta z^k, \Delta s^k) \) be the Newton step for \( F_{\sigma_k \mu_k} \) at \( (x^k, z^k, s^k) \);  
Choose \( \alpha_k \) to be the largest value in \( (0, 1] \) for which  
\[ (x^k(\alpha), z^k(\alpha), s^k(\alpha)) \in \mathcal{N}_{-\infty}(\gamma); \]
Set \( (x^{k+1}, z^{k+1}, s^{k+1}) = (x^k(\alpha_k), z^k(\alpha_k), s^k(\alpha_k)) \).
The Search Direction

The Newton step for $F_{\sigma_k \mu_k}$ at $(x, z, s)$ satisfies (omitting superscripts):

$$
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S & 0 & X
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta z \\
\Delta s
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0 \\
-XXSe + \sigma_k \mu_ke
\end{bmatrix}.
$$

Decomposing the third block into individual components, we have

$$s_i \Delta x_i + x_i \Delta s_i = -x_i s_i + \sigma_k \mu_k.$$

Summing both sides over $i = 1, 2, \ldots, n$, obtain

$$s^T \Delta x + x^T \Delta s = -x^T s + \sigma_k n \mu_k = -(1 - \sigma_k) n \mu_k.$$

The difference between $(x + \Delta x, z + \Delta z, s + \Delta s)$ and the solution of $F_{\sigma_k \mu_k}(x, z, s) = 0$ is only the term $\Delta X \Delta Se$ in the third block, where

$$\Delta X = \text{diag}(\Delta x_1, \Delta x_2, \ldots, \Delta x_n), \quad \Delta S = \text{diag}(\Delta s_1, \Delta s_2, \ldots, \Delta s_n).$$

The analysis is mostly concerned with showing that this term is not too large, that is, the linear approximation to $F_{\sigma_k \mu_k}$ that is the basis of Newton’s method is good enough to make substantial progress.
Can plot progress of the algorithm in “xs” space:
Convergence and Complexity in Five Easy Steps

The result follow from five technical but elementary claims, each building on the one before:

1. If \( u, v \) are two vectors in \( \mathbb{R}^n \) with \( u^T v \geq 0 \), we have
   \[
   \| UVe \| \leq 2^{-3/2} \| u + v \|^2,
   \]
   where \( U = (u_1, u_2, \ldots, u_n) \), \( V = (v_1, v_2, \ldots, v_n) \).

2. \[
   \| \Delta X \Delta Se \| \leq 2^{-3/2} \| (XS)^{-1/2}(-XSe + \sigma \mu e) \|^2.
   \]

3. \[
   \| \Delta X \Delta Se \| \leq 2^{-3/2}(1 + \gamma^{-1})n\mu.
   \]

4. \[
   (x^k(\alpha), z^k(\alpha), s^k(\alpha)) \in \mathcal{N}_{-\infty}(\gamma) \quad \text{for all} \quad \alpha \in \left[0, 2^{3/2} \gamma \frac{1 - \gamma \sigma}{1 + \gamma \frac{n}{n}}\right].
   \]

5. \[
   \mu_{k+1} \leq (1 - \delta n^{-1})\mu_k \quad \text{for some} \ \delta \ \text{independent of} \ n \ \text{and all} \ k \geq 0.
   \]
Proof of \(1\)

For scalars \(\alpha\) and \(\beta\) with \(\alpha \beta \geq 0\), we have \(|\alpha \beta| = \alpha \beta \leq (\alpha + \beta)^2/4\).

Define \(\mathcal{P} = \{i : u_i v_i \geq 0\}\), \(\mathcal{N} := \{i : u_i v_i < 0\}\). Since \(u^T v \geq 0\), we have

\[
0 \leq u^T v = \sum_{i \in \mathcal{P}} u_i v_i + \sum_{i \in \mathcal{N}} u_i v_i = \|[u_i v_i]_{i \in \mathcal{P}}\|_1 - \|[u_i v_i]_{i \in \mathcal{N}}\|_1.
\]

Thus

\[
\|UVe\| = \left(\|[u_i v_i]_{i \in \mathcal{P}}\|^2 + \|[u_i v_i]_{i \in \mathcal{N}}\|^2\right)^{1/2}
\leq \left(\|[u_i v_i]_{i \in \mathcal{P}}\|^2_1 + \|[u_i v_i]_{i \in \mathcal{N}}\|^2_1\right)^{1/2}
\leq \left(2 \|[u_i v_i]_{i \in \mathcal{P}}\|^2_1\right)^{1/2}
\leq \sqrt{2} \left\|\left[\frac{1}{4} (u_i + v_i)^2\right]_{i \in \mathcal{P}}\right\|_1
\leq 2^{-3/2} \sum_{i \in \mathcal{P}} (u_i + v_i)^2 \leq 2^{-3/2} \sum_{i=1}^n (u_i + v_i)^2 = 2^{-3/2} \|u + v\|^2.
\]
Proof of 2

Multiply last block of Newton equations by \((XS)^{-1/2}\), and define \(D = X^{1/2}S^{-1/2}\) to obtain

\[
S \Delta x + X \Delta s = -XSe + \sigma \mu e
\]

\[
\Rightarrow D^{-1} \Delta x + D \Delta s = (XS)^{-1/2}(-XSe + \sigma \mu e).
\]

Set \(u = D^{-1} \Delta x\) and \(v = D \Delta s\) and note from first two blocks of Newton equations that

\[
u^T v = \Delta x^T \Delta s = -\Delta x^T A^T \Delta z = -(A \Delta x)^T \Delta z = 0.\]

Thus can apply 1 to deduce that

\[
\|\Delta X \Delta Se\| \leq 2^{-3/2} \|(XS)^{-1/2}(-XSe + \sigma \mu e)\|^2.
\]
Proof of 3

Expand the right-hand side of 2:

\[
\|\Delta X \Delta S e\| \leq 2^{-3/2} \left\| - (XS)^{1/2} e + \sigma \mu (XS)^{-1/2} e \right\|^2 \\
\leq 2^{-3/2} \left[ x^T s - 2\sigma \mu e^T e + \sigma^2 \mu^2 \sum_{i=1}^{n} \frac{1}{x_i s_i} \right] \\
\leq 2^{-3/2} \left[ x^T s - 2\sigma \mu e^T e + \sigma^2 \mu^2 \frac{n}{\gamma \mu} \right] \quad \text{since } x_i s_i \geq \gamma \mu \\
\leq 2^{-3/2} \left[ 1 - 2\sigma + \frac{\sigma^2}{\gamma} \right] n \mu \quad \text{since } e^T e = n \\
\leq 2^{-3/2} (1 + \gamma^{-1}) n \mu \quad \text{since } \sigma \in (0, 1).
Proof of 4

From 3 we have

\[ |\Delta x_i \Delta s_i| \leq \|\Delta X \Delta S e\|_2 \leq 2^{-3/2} (1 + \gamma^{-1}) n \mu. \]

Thus from \( x_i s_i \geq \gamma \mu \) and the third block of the Newton equations, we have

\[
\begin{align*}
x_i(\alpha) s_i(\alpha) &= (x_i + \alpha \Delta x_i)(s_i + \alpha \Delta s_i) \\
&= x_i s_i + \alpha(x_i \Delta s_i + s_i \Delta x_i) + \alpha^2 \Delta x_i \Delta s_i \\
&\geq x_i s_i (1 - \alpha) + \alpha \sigma \mu - \alpha^2 |\Delta x_i \Delta s_i| \\
&\geq \gamma (1 - \alpha) \mu + \alpha \sigma \mu - \alpha^2 2^{-3/2} (1 + \gamma^{-1}) n \mu.
\end{align*}
\]

Meanwhile, again using the third block of Newton equations, we have

\[
\begin{align*}
n \mu(\alpha) &= x(\alpha)^T s(\alpha) = (x + \alpha \Delta x)^T (s + \alpha \Delta s) \\
&= x^T s + \alpha(s^T \Delta x + x^T \Delta s) + \alpha^2 \Delta x^T \Delta s \\
&= (1 - \alpha + \alpha \sigma) n \mu.
\end{align*}
\]
So for $x_i(\alpha)s_i(\alpha) \geq \gamma \mu(\alpha)$, it suffices to have

$$\gamma(1 - \alpha)\mu + \alpha \sigma \mu - \alpha^2 2^{-3/2}(1 + \gamma^{-1})n\mu \geq \gamma(1 - \alpha + \alpha \sigma)\mu,$$

which is equivalent to

$$\alpha \leq 2^{3/2} \gamma \frac{1 - \gamma \sigma}{1 + \gamma \frac{n}{n}},$$

proving 4.
Proof of 5

We showed already that

\[ \mu_k(\alpha) = (1 - \alpha(1 - \sigma_k))\mu_k, \]

so that \( \mu_k(\alpha) \) is decreasing in \( \alpha \) over \( \alpha \in [0, 1] \).

From 4, we have for step actually taken that

\[ \alpha_k \geq 2^{3/2} \gamma \frac{1 - \gamma \sigma}{1 + \gamma n}. \]

Thus

\[ \mu_{k+1} = \mu_k(\alpha_k) \leq \mu_k \left( 1 - \left( 2^{3/2} \gamma \frac{1 - \gamma \sigma_k}{1 + \gamma \frac{1}{n}} (1 - \sigma_k) \right) \right) \mu_k. \]

Recalling that \( \sigma_k \in [\sigma_{\min}, \sigma_{\max}] \) for all \( k \), we have that

\[ \mu_{k+1} \leq (1 - \delta n^{-1})\mu_k, \]

for

\[ \delta = 2^{3/2} \gamma \frac{1 - \gamma}{n \left( 1 + \gamma \right)} \min(\sigma_{\min}(1 - \sigma_{\min}), \sigma_{\max}(1 - \sigma_{\max})). \]
Other Path-Following Algorithms: SPF

Short-step Path-Following (SPF) uses the more restricted neighborhood $\mathcal{N}_2(0.4)$, fixed $\sigma = 1 - 0.4/\sqrt{n}$, and full modified Newton steps.

Algorithm SPF.

Set $\sigma = 1 - 0.4/\sqrt{n}$;
Choose $(x^0, z^0, s^0) \in \mathcal{N}_2(0.4)$;
for $k = 0, 1, 2, \ldots$ do
   Let $(\Delta x^k, \Delta z^k, \Delta s^k)$ be the Newton step for $F_{\sigma \mu_k}$ at $(x^k, z^k, s^k)$;
   Set $(x^{k+1}, z^{k+1}, s^{k+1}) = (x^k, z^k, s^k) + (\Delta x^k, \Delta z^k, \Delta s^k)$;
end for

Analysis is similar to LPF, leading to a better dependence on $n$:

$$\mu_{k+1} = (1 - 0.4/\sqrt{n})\mu_k,$$

thus achieving $\mu_k \leq \epsilon \mu_0$ in $O(\sqrt{n} \log \epsilon)$ iterations.

Better complexity bound, but slower in practice. (That’s a theme!)
Other Path-Following Algorithms: Predictor-Corrector

Use two neighborhoods $\mathcal{N}_2(.25)$ and $\mathcal{N}_2(.5)$. Alternate between
- Predictor steps with $\sigma = 0$ (unmodified Newton on KKT) that start in $\mathcal{N}_2(.25)$ and stop at the boundary of $\mathcal{N}_2(.5)$;
- Corrector steps with $\sigma = 1$ that return to $\mathcal{N}_2(.25)$.

**Algorithm PC.**

Choose $(x^0, z^0, s^0) \in \mathcal{N}_2(0.25)$;

for $k = 0, 1, 2, \ldots$ do

if $k$ even then

Let $(\Delta x^k, \Delta z^k, \Delta s^k)$ be the Newton step for $F_0$ at $(x^k, z^k, s^k)$;

Choose $\alpha_k$ to be the largest value in $(0, 1]$ for which

$$(x^k(\alpha), z^k(\alpha), s^k(\alpha)) \in \mathcal{N}_2(0.5);$$

Set $(x^{k+1}, z^{k+1}, s^{k+1}) = (x^k(\alpha_k), z^k(\alpha_k), s^k(\alpha_k))$;

else

Let $(\Delta x^k, \Delta z^k, \Delta s^k)$ be the Newton step for $F_{\mu_k}$ at $(x^k, z^k, s^k)$;

Set $(x^{k+1}, z^{k+1}, s^{k+1}) = (x^k, z^k, s^k) + (\Delta x^k, \Delta z^k, \Delta s^k)$;

end if

end for
Analysis and complexity is similar to SPF: $O(\sqrt{n \log \epsilon})$ iterations.
Primal-Dual Potential Reduction

An alternative way to balance centrality of \((x^k, z^k, s^k)\) while steadily reducing duality gap \(\mu_k\) to zero is by using a primal-dual logarithmic potential function:

\[
\Phi_{\rho}(x, s) := \rho \log x^T s - \sum_{i=1}^{n} \log x_i s_i,
\]

for \(\rho > n\). Note that

\[
\Phi_{\rho}(x, s) := (\rho - n) \log \mu - \sum_{i=1}^{n} \log \frac{x_i s_i}{\mu} + \rho \log n, \quad \text{where} \quad \mu = \frac{x^T s}{n},
\]

so “reducing \(\rho\)” and “maintaining centrality” are both represented.

- All iterates are strictly feasible: \((x^k, z^k, s^k) \in \mathcal{F}^\circ\).
- Search directions are modified Newton steps for \(F_{\sigma \mu_k}\), with \(\sigma \equiv n/\rho\).
- Choose step length \(\alpha\) to minimize \(\Phi_{\rho}\) along
\( (x^k, z^k, s^k) + \alpha(\Delta x^k, \Delta z^k, \Delta s^k) \);
Algorithm PDPR

Algorithm PDPR.

Choose $\rho > n$, $\sigma = n/\rho$, and $(x^0, z^0, s^0) \in \mathcal{F}^\circ$;

for $k = 0, 1, 2, \ldots$ do

Let $(\Delta x^k, \Delta z^k, \Delta s^k)$ be the Newton step for $F_{\sigma \mu_k}$ at $(x^k, z^k, s^k)$;

Set $\alpha_{\max} =$ largest value of $\alpha$ s.t. $(x^k(\alpha), z^k(\alpha), s^k(\alpha)) \in \mathcal{F}^\circ$;

Set

$$\alpha_k = \arg \min_{\alpha \in (0, \alpha_{\max})} \Phi_\rho(x^k(\alpha), s^k(\alpha));$$

Set $(x^{k+1}, z^{k+1}, s^{k+1}) = (x^k(\alpha_k), z^k(\alpha_k), s^k(\alpha_k))$;

end for
For $\rho \geq n + \sqrt{n}$, we have

$$
\Phi_\rho(x^{k+1}, s^{k+1}) \leq \Phi_\rho(x^k, s^k) - 0.15, \quad k = 0, 1, 2, \ldots,
$$

so that $\Phi_\rho(x^k, s^k) \downarrow -\infty$ at a steady rate.

Can show that

$$
\mu \leq \exp \left( \Phi_\rho(x, s) / (\rho - n) \right),
$$

so we have $\mu_k \leq \epsilon$ in

$$
\frac{\Phi_\rho(x^0, s^0) + (\rho - n) |\log \epsilon|}{\delta} \text{ iterations}.
$$
Initialization

The description of LPF assumes that we can find an initial point \((x^0, z^0, s^0) \in \mathcal{N}_{-\infty}(\gamma)\). This is nontrivial in general.

Two approaches:

- Modify the algorithm to \textit{not} require feasibility of the equality constraints \(Ax = b\) and \(A^T z + s = c\).
  - \textbf{Infeasible-interior-point} algorithms.
- Modify the LP formulation so that a feasible point is obvious.
  - \textbf{Homogeneous self-dual (HSD)} formulations.
  - Solved using interior-point methods for monotone linear complementarity (see later).

In implementations, infeasible-interior-point is preferred, and the analysis above can be extended to show convergence and complexity, at slightly worse rates: \(O(n^2 \log \epsilon)\) instead of \(O(n \log \epsilon)\).

But HSD is fairly practical, and has elegant theory.
Infeasible-Interior-Point

Start from any point \((x^0, z^0, s^0)\) with \((x^0, s^0) > 0\).

For any iterate \((x^k, z^k, s^k)\), define residuals

\[
r_b^k := Ax^k - b, \quad r_c^k := A^T z^k + s^k - c.
\]

Search directions are still Newton steps for \(F_{\sigma_k \mu_k}\) at \((x^k, z^k, s^k)\), now defined by these Newton equations:

\[
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S^k & 0 & X^k
\end{bmatrix}
\begin{bmatrix}
\Delta x^k \\
\Delta z^k \\
\Delta s^k
\end{bmatrix}
= 
\begin{bmatrix}
-r_c^k \\
-r_b^k \\
-X^k S^k e + \sigma_k \mu_k e
\end{bmatrix}.
\]

No longer have the nice relation \((\Delta x^k)^T \Delta s^k = 0\), so \(\mu\) does not decrease linearly in steplength \(\alpha\) along \((\Delta x^k, \Delta z^k, \Delta s^k)\).

The residual norms \(r_b\) and \(r_c\) do however decrease linearly in \(\alpha\), since the first two blocks of the Newton system are linear.
Algorithm IPF

Infeasible extension of the $\mathcal{N}_{-\infty}$ central path neighborhood:

$$
\mathcal{N}_{-\infty}(\gamma, \beta) := \left\{ (x, z, s) : \| (r_b, r_c) \| \leq \beta \| (r_b^0, r_c^0) \| (\mu/\mu_0), \ x_i s_i \geq \gamma \mu \right\},
$$

where $\mu = x^T s / n$ as usual, and $\gamma \in (0, 1)$ and $\beta \geq 1$.

Algorithm IPF.

Choose $\gamma \in (0, 1)$, $\beta \geq 1$, $\sigma_{\text{min}}$ and $\sigma_{\text{max}}$ with $0 < \sigma_{\text{min}} < \sigma_{\text{max}} < 1$;

Choose $(x_0, z_0, s_0)$ with $(x_0, s_0) > 0$;

for $k = 0, 1, 2, \ldots$ do

Choose $\sigma_k \in [\sigma_{\text{min}}, \sigma_{\text{max}}]$;

Let $(\Delta x^k, \Delta z^k, \Delta s^k)$ be the Newton step for $F_{\sigma_k \mu_k}$ at $(x^k, z^k, s^k)$;

Choose $\alpha_k$ to be the largest value in $(0, 1]$ for which

$$(x^k(\alpha), z^k(\alpha), s^k(\alpha)) \in \mathcal{N}_{-\infty}(\gamma, \beta);$$

and

$$
\mu_k(\alpha) \leq (1 - .01\alpha) \mu_k;
$$

Set $(x^{k+1}, z^{k+1}, s^{k+1}) = (x^k(\alpha_k), z^k(\alpha_k), s^k(\alpha_k))$.

end for
Infeasibility complicates the analysis considerably, but still requires only elementary tools.

Ultimately, show that $\alpha_k \geq \bar{\delta}/n^2$ for some $\bar{\delta}$ independent of $n$ and all $k$. Thus get $\mu_k \leq \epsilon \mu_0$ and $\|(r_b^k, r_c^k)\| \leq \beta \epsilon \|(r_b^0, r_c^0)\|$ in $O(n^2 \log \epsilon)$ iterations.
Extension: Monotone Linear Complementarity

Given $M \in \mathbb{R}^{n \times n}$ positive semidefinite and $q \in \mathbb{R}^n$, seek $(x, s)$ such that

$$s = Mx + q, \quad (x, s) \geq 0, \quad x^T s = 0. \quad \text{(LCP)}$$

Can add extra unconstrained variables and equality constraints to get a mixed monotone LCP.

KKT conditions for LP and convex QP are mixed monotone LCP. For LP:

$$
\begin{bmatrix}
  s \\
  0
\end{bmatrix} = 
\begin{bmatrix}
  0 & -A^T \\
  A & 0
\end{bmatrix}
\begin{bmatrix}
  x \\
  z
\end{bmatrix} + 
\begin{bmatrix}
  c \\
  -b
\end{bmatrix}, \quad (x, s) \geq 0, \quad x^T s = 0.
$$

For convex QP problem

$$
\min \frac{1}{2} x^T Qx + c^T x \quad \text{s.t.} \quad Ax = b, \quad x \geq 0,
$$

the corresponding LCP is

$$
\begin{bmatrix}
  s \\
  0
\end{bmatrix} = 
\begin{bmatrix}
  Q & -A^T \\
  A & 0
\end{bmatrix}
\begin{bmatrix}
  x \\
  z
\end{bmatrix} + 
\begin{bmatrix}
  c \\
  -b
\end{bmatrix}, \quad (x, s) \geq 0, \quad x^T s = 0.
$$
Algorithms for monotone LCP are almost identical to those for LP. It is an appealing framework because

- notation is simpler — just \((x, s)\);
- can be extended immediately to mixed monotone LCP;
- apply to convex QP as well as LP, and also other LCP applications e.g. from games.

The analysis is a little different from LP because we have only \(\Delta x^T \Delta s \geq 0\) instead of \(\Delta x^T \Delta s = 0\), but otherwise algorithms, convergence analysis, complexity results are much the same.
LP: Homogeneous Self-Dual Formulation

Returning to LP: We can derive a mixed monotone LCP from the KKT conditions for LP, such that

- It’s easy to find a strictly feasible initial point for the mixed LCP;
- The mixed LCP always has a solution;
- This solution yields either a primal-dual solution pair for the LP, or a certificate of infeasibility.

Given \((x^0, z^0, s^0)\) such that \((x^0, s^0) > 0\) define

\[
\bar{b} := b - Ax^0, \quad \bar{c} := c - A^Tz^0 - s^0, \quad \bar{t} := c^Tx^0 + 1 - b^Tz^0.
\]

(The first two represent initial infeasibilities.) The mixed LCP is then

\[
\begin{bmatrix}
\kappa \\
s \\
0 \\
0
\end{bmatrix}
= \begin{bmatrix}
0 & -c^T & -b^T & \bar{t} \\
c & 0 & -A^T & -\bar{c} \\
-b & A & 0 & \bar{b} \\
-\bar{t} & \bar{c}^T & -b^T & 0
\end{bmatrix}
\begin{bmatrix}
\tau \\
x \\
z \\
\theta
\end{bmatrix}
+ \begin{bmatrix}
0 \\
0 \\
0 \\
(x^0)^Ts^0 + 1
\end{bmatrix},
\]

\[0 \leq \tau \perp \kappa \geq 0, \quad 0 \leq x \perp s \geq 0.\]
HSD Results

Note that the original LP variables appear alongside scalar auxiliary variables \( \kappa, \tau, \theta \).

A strictly feasible initial point for the mixed LCP is

\[
(\tau, x, z, \theta, \kappa, s) = (1, x^0, z^0, 1, 1, s^0).
\]

The term “homogeneous self-dual” (HSD) arises from the mixed LCP being reduced KKT conditions for a self-dual LP formulation that looks quite similar — and that is homogeneous except for the term \((x^0)^T s^0 + 1\).

**Result 0:** *The mixed LCP has a strictly complementary solution.* This follows from

- the close relationship of the mixed LCP to the HSD linear program
- the fact that the HSD LP is feasible
- when a primal-dual pair of LPs has a solution, it has a strictly complementary solution.
HSD Results

Result 1: Any solution \((\tau^*, x^*, z^*, \theta^*, \kappa^*, s^*)\) of the mixed LCP has \(\theta^* = 0\). (Proof: Elementary)

Result 2: The original LP has primal-dual solutions if and only if all strictly complementary solutions of the mixed LCP have \(\kappa^* = 0\) and \(\tau^* > 0\). These solutions are \(x^*/\tau^*\) for the primal and \((z^*/\tau^*, s^*/\tau^*)\) for the dual.

Result 3: If the mixed LCP has a strictly complementary solution for which \(\kappa^* > 0\), then at least one of \(c^T x^*\) and \(-b^T z^*\) is negative, and

- if \(c^T x^* < 0\), then the dual LP is infeasible;
- if \(-b^T z^* < 0\), then the primal LP is infeasible.
Log Barrier Functions for LP

We turn now to a more traditional approach based on log barrier functions [Frisch, 1955].

Deal with an algebraic constraint $c_i(x) \geq 0$ by adding a term $-\log c_i(x)$ into the objective, weighted by a parameter.

- Log is defined only when the constraint is strictly feasible: $c_i(x) > 0$;
- Goes to $\infty$ as $c_i(x) \to 0$.

It is also self-concordant, which allows non-local results to be proved about convergence of Newton’s method applied to the barrier formulation.

For LP, we have barrier formulations of P and D, parametrized by $\tau > 0$:

\[
\min_x \frac{1}{\mu} c^T x - \sum_{i=1}^n \log x_i \text{ s.t. } Ax = b, \quad (P-\mu)
\]

\[
\max_z \frac{1}{\mu} b^T z + \sum_{i=1}^m \log (c_i - A_i^T z), \quad (D-\mu)
\]
Optimality Conditions, Central Path

KKT conditions for \((P-\mu)\) are:

\[
c - \mu X^{-1}e - A^T z = 0, \quad Ax = b,
\]

where as before \(X = \text{diag}(x_1, x_2, \ldots, x_n)\) and \(e = (1, 1, \ldots, 1)^T\). Defining \(s_i = \mu / x_i, \ i = 1, 2, \ldots, n\), we can rewrite these conditions as:

\[
c - s - A^T z = 0, \quad Ax = b, \quad XSe = \mu e,
\]

which are exactly the conditions that define the point \((x_\mu, z_\mu, s_\mu)\) on the central path \(C\)!

Thus the solutions of \((P-\mu)\) for \(\mu > 0\) are exactly the projections of primal-dual central path onto \(x\)-space.

A similar derivation works for the dual: The solutions of \((D-\mu)\) for \(\mu > 0\) are exactly the projection of \(C\) into \(z\)-space. (Here we set \(s := c - A^T z\) and \(x_i := \mu / s_i\).)
Path-Following

Thus we can solve (D) by following the path of solutions of (D-µ) as µ ↓ 0. The strategy is basically the same as in primal-dual methods:

- Take Newton steps for (D-µ) at current µ, with steplengths chosen to maintain strict feasibility of iterates, until a near-solution for (D-µ) is found;
- Decrease µ by some factor σ < 1.

The same strategy can be applied to (P-µ), except that here we take Lagrange-Newton steps, because of the constraint Ax = b.

Complexity results follow from:

- Only one or a few Newton steps needed for each µ;
- \((1 - \sigma)^{-1} \log \epsilon\) outer iterations to achieve \(\mu/\mu_0 < \epsilon\).
Newton Directions: Dual Barrier and Primal-Dual

Defining $s := c - A^T z$, Newton equations for (D-$\tau$) are

$$-AS^{-2}A^T \Delta z = -\frac{1}{\tau} b + AS^{-1} e.$$

We can “unpack” these equations to make a direct comparison with the primal-dual equations. Suppose that the current $s$ is the exact solution of (D-$\mu$). Then for $x := \mu S^{-1} e$, we have

$$b = Ax = \mu AS^{-1} e,$$

so the system above is

$$-AS^{-2}A^T \Delta z = \left(-\frac{\mu}{\tau} + 1\right) AS^{-1} e.$$

Choosing $\tau = \sigma \mu$ as in primal-dual, we find that $\Delta z$ is part of the solution of the unpacked system

$$
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S & 0 & X
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta z \\
\Delta s
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0 \\
\mu \left(1 - \frac{1}{\sigma}\right) e
\end{bmatrix}.$$
Primal-Dual vs Dual Barrier

Compare the dual system

$$
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S & 0 & X
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta z \\
\Delta s
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
\mu \left(1 - \frac{1}{\sigma}\right) e
\end{bmatrix}.
$$

with the primal-dual system from the point \((x_\mu, z_\mu, s_\mu)\):

$$
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S & 0 & X
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta z \\
\Delta s
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
-\lambda Se + \sigma \mu e
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
\mu(-1 + \sigma)e
\end{bmatrix}.
$$

The RHS coefficients are:

- **Dual Barrier:** \(\frac{\sigma - 1}{\sigma}\),
- **Primal-Dual:** \(\sigma - 1\).
Primal-Dual vs Dual Barrier

When $\sigma$ is close to 1, there is little difference. But when $\sigma$ is significantly less than 1 (as in long-step methods), the dual barrier step is much longer. It overshoots the solution of $(D-\sigma\mu)$.

However, the directions $\Delta z$ and $\tilde{\Delta} z$ are the same, modulo a scaling. Thus we can recover good behavior of dual barrier for long step methods by

- ensuring accurate enough solution of each subproblem $(D-\mu)$;
- scaling the first step taken after each resetting of the target to $\sigma \mu$.

[Wright and Jarre, 1998]
References


