Stack Complementarity

\[
\begin{align*}
\text{min} & \quad c^T z \\
\text{s.t.} & \quad Ax = b \quad (P) \quad \text{max} & \quad b^T y \\
& \quad x \geq 0 \\
& \quad \text{min} & \quad A^T y + s = c \quad (D) \\
& \quad s \geq 0.
\end{align*}
\]

Let \( \mathcal{P} \) be the set of optimal solutions to (P)

\[ \mathcal{P} = \text{set of optimal solutions to (P)} \]

\[ \mathcal{D} = \text{set of optimal solutions to (D)} \]

Either \( \mathcal{P} \) and \( \mathcal{D} \) are both nonempty or they are both empty.

Proof: 4 cases:

(i) \( (P) \) and \( (D) \) have some finite optimal value. Then \( \mathcal{P} \) and \( \mathcal{D} \) are both nonempty.

(ii) \( (P) \) has unbounded optimal value, \( (D) \) is infeasible. Then both \( \mathcal{P} \) and \( \mathcal{D} \) are empty.

(iii) \( (D) \) has unbounded optimal value, \( (P) \) is infeasible. Then both \( \mathcal{P} \) and \( \mathcal{D} \) are empty.

(iv) Both \( (P) \) and \( (D) \) infeasible. Then both \( \mathcal{P} \) and \( \mathcal{D} \) are empty.
A strictly feasible point for \((P)\) is \(\bar{x} > 0\) with \(A\bar{x} = b\).

A strictly feasible point for \((\bar{D})\) is \((\bar{y}, \bar{s})\) with \(\bar{s} > 0\) and \(A^T\bar{y} + \bar{s} = \bar{c}\).

**Lemma**: Suppose the primal and dual problems are feasible.

1. If the dual problem has a strictly feasible solution then \(\bar{D}\) is convex and bounded.
2. If \((P)\) has a strictly feasible solution then \(\bar{D}\) is convex and bounded.

**Proof**: See second, leave last as exercise:

Assume \((P)\) has a strictly feasible solution \(\bar{x}\).

Let \(z^*\) be the optimal value.

For any \(s^*\) in \(\mathbb{R}_+^n\) with corresponding \(y^*\), we have:

\[\bar{x}^T s^* = x^T (c - A^T y^*) = c^T \bar{x} - (A\bar{x})^T y^* = c^T \bar{x} - b^T y^* = c^T \bar{x} - z^*.\]

Since \(\bar{x} > 0\), we have

\[0 \leq s_i^* \leq \frac{c^T \bar{x} - z^*}{\bar{x}_i}\]

so \(\Omega_s\) is bounded.

**Note**: Converse is also true. (dual version of the theorem.)
Let \((x^*, y^*, s^*)\) be an optimal solution.

From complementary slackness, we have for each \(i\), \(x_i^* = 0\) and/or \(s_i^* = 0\).

Define \(B = \{i : x_i^* \neq 0 \text{ for some } x \in \mathcal{D}\}\)

and \(N = \{i : s_i^* \neq 0 \text{ for some } (y, s^*) \in \mathcal{D}\}\).

Example (i) \(B\) could be just the basic variables:

\[
\begin{align*}
\min & \quad 3x_1 + x_2 \\
\text{st.} & \quad x_1 + x_2 = 1 \\
& \quad x_i \geq 0
\end{align*}
\]

\[\text{new } B \uparrow y_1, \quad s_2\]

Unique optimal soln is \((0, 1, 1, 0)\),

\[x^* = (0, 1), \quad y^* = 1, \quad s^* = (2, 0)\]

So, \(B = \{2\}, \quad N = \emptyset\).
(ii) $B$ can be smaller than any basic variables:

\[ \begin{align*}
\min & \quad -x_1 + x_3 \\
\text{s.t.} & \quad x_1 + x_2 + x_3 = 1 \\
& \quad x_1 + x_4 = 1 \\
& \quad x_1 \geq 0
\end{align*} \] (P)

\[ \begin{align*}
\max & \quad y_1 + y_3 \\
\text{s.t.} & \quad y_1 + y_2 + s_1 = -1 \\
& \quad y_1 + s_2 = 1 \\
& \quad y_1 + s_3 = 0 \\
& \quad s_i \geq 0
\end{align*} \] (Q)

Unique optimal soln to (P), $x^* = (1, 0, 0, 0)$, so $\mathcal{B} = \{1\}$.

Optimal face.

Optimal solns are:

\[
\begin{align*}
\mathbf{y}^* = (0, 0), & \quad \mathbf{s}^* = (0, 1, 1, 0), \\
\mathbf{y}^* = (0, -1), & \quad \mathbf{s}^* = (0, 2, 0, 1)
\end{align*}
\]

and any convex combination:

\[
\begin{align*}
\mathbf{y}^* = (-\lambda, \lambda - 1), & \quad \mathbf{s}^* = (0, 2 - \lambda, \lambda, 1 - \lambda) \\
\text{for} & \quad 0 \leq \lambda \leq 1.
\end{align*}
\]

So $\mathcal{N} = \{2, 3, 4\}$.

Note that $\mathbf{B} \cup \mathcal{N} = \{1, 2, 3, 4\}$, i.e., all the indices.
(iii) $B$ could be bigger than any basic set. We consider

\[ \max \quad y, \]
\[ \text{st.} \quad y_1 + s_1 = 1 \]
\[ y_1 + s_2 = 1 \]
\[ y_2 + s_3 = 0 \]
\[ s_i \geq 0. \]

Optimal primal solution are

\[ x^* = (\lambda, 1-\lambda, 0) \text{ for any } \lambda \in [0, 1], \]

so, \[ B = \{1, 2\}. \]

Optimal dual solution is \[ y_1^* = 1, \quad s^* = (0, 0, 1), \] so, \[ \lambda \in \{3\}. \]

Again, \[ B \cup N = \{1, 2, 3\}, \] all the variables.
(iv) Could have both primal & dual multiple optimal solutions.

**N:** 

\[ 2x_1 + 3x_2 + x_3 = 4 \]
\[ 2x_1 + x_2 + 3x_3 - x_5 = 4 \]
\[ x_1 + x_2 + x_3 - x_6 = 2 \]
\[ x_6 \geq 0 \]

**Optimal set:** all three planes intersect.

**N:** 

\[ y^* = x(\frac{1}{3}, \frac{1}{3}, 0) + (1-x)(0, 0, 1) \]

\[ 5x = x(0, 0, 0, \frac{1}{3}, \frac{1}{3}, 0) + (1-x)(0, 0, 0, 0, 0, 1) \]

**N:** \( \{4, 5, 6\} \)
Lemma \( B \cap B^c \cap N = \emptyset \).

Proof. Follow immediately by complementary slackness.

If \( x_i > 0 \) then must have \( \pi^*_i = 0 \), any dual optimal rule.

So \( i \in B \) and \( i \notin N \).

\[ \square \]

Theorem (Goldman-Tucker)

\( B \cup N = \{1, 2, \ldots, n\} \). Thus, there exist at least one primal solution \( x^* \in \mathbb{R}_+^n \) and one dual solution \( (\mathbf{y}^*, \mathbf{s}^*) \in \mathbb{R}_+^m \times \mathbb{R}_+^n \).

Let \( F \) be the column \( \{1, 2, \ldots, n\} \setminus N \). Neither \( B \cup F \).

Let \( i \in F \). Consider the system:

\[ \begin{align*}
A_i^T \mathbf{w} &< 0 \\
(A_i^T)^\text{adj of } A &> 0 \\
-A_j^T \mathbf{w} &< 0 & j \in F \setminus \{i\} \\
A_B^T \mathbf{w} &> 0 & \text{when } A_B \text{ are only even } \& B.
\end{align*} \tag{I} \]

This system has a solution if and only if the following system does not have a solution:

\[ \begin{align*}
-\sum_{j \in F \setminus \{i\}} A_j^T \mu_j + A_B^T \mathbf{z} &= A_i \\
\mu_j &> 0, \quad \text{a price}
\end{align*} \tag{II} \] (No proof, similar)

\[ \text{to Farkas.} \]

Since \( i \in F \), we have \( x_i^* = 0 \) and \( s_i^* = 0 \) in every optimal solution.
Assume (I) holds.

Let $x^*$ be a dual solution with $s_i^* > 0 \text{ for } i \in N^-$
(E.g., $s_i^*$ is average of $N^-$ peaks, where $s_i > 0$ is least one for each peak; for each $i \in N^-$.

Let $y^*$ correspond to $x^*$, so $A^T y^* + s^* = 0$.

Let $z_i = y_i^*$ be a solution to (I).

Then, $y^* + z^* = 0$ is feasible, at least for $\epsilon$ small and positive.

Let $\bar{s} = A^T(y^* + \epsilon z^*)$.

Note that $A_{N^-}^T(y^* + \epsilon z^*) = A_{N^-} y^*$, so $\bar{s} = 0$ still.

Thus, $B^T x^* = 0$ for any optimal $x^*$, so $\bar{s}$ is optimal.

Also, $\bar{s} > 0$, so $i \in N^-$.

Assume (II) holds.

Let $x^*$, $y^*$ solve (II), let $x^*$ be optimal with $x_i^* > 0$.

Let $x_{N^-} = x_{N^-} + \epsilon z^*$, $x_{N^-} = x_{N^-}$, $\bar{x} = x$, $\bar{x}_i = \epsilon$, $\bar{x}_j = y_j$, for $j \in N \setminus N^-$. Then $x$ is feasible for $\epsilon$ small enough, and since $s_{N^-} = 0$ and $s_i > 0 \text{ for } i \in N$, we have that $i \in B$. $\bar{x}$
Lemma. Suppose there exist strictly feasible solutions for both (P) and (O).

Then, for any \( K \geq 0 \), the set

\[ \{ (x, s) : (x, y, s) \text{ is feasible and } x^Ts \leq K \} \]

is bounded.

Proof. Let \((x, y, s)\) be a strictly feasible solution.

Let \((x, s)\) be feasible with \( x^Tc \leq K \).

Now,

\[ (x - \bar{x})^T(s - \bar{s}) = (x - \bar{x})^TA^T(y - \bar{y}) \]
\[ = (Ax - A\bar{x})^T(y - \bar{y}) \]
\[ = 0 \quad \text{since } Ax = A\bar{x} = b. \]

Hence,

\[ 0 = x^Ts - \bar{x}^Ts - \bar{x}^Ts + \bar{s}^Tx \]
\[ \Rightarrow \bar{x}^Ts + \bar{s}^Tx = x^Ts - \bar{x}^Ts \leq K + \bar{s}^Tx. \]

Now, \( \bar{x} > 0 \) and \( \bar{s} > 0 \).

So, \( 0 \leq \bar{s} \leq K + \bar{s}^Tx \) and \( 0 \leq \bar{x} \leq K + \bar{s}^Tx \).

Corollary. If the set of optimal solutions to (P) (or (O)) is unbounded,

then (P) (or (O)) does not have a strictly feasible solution.

Proof. Directly from Lemma. Alternatively: \[ d \geq 0, d \neq 0 \text{ and } x^Td, c^Td = 0. \] Let \( y, x, s \) such \( A^Ty + s = 0, c^Tx = 0. \)
This problem is a linear programming problem.

Optimality conditions for

\[ \max \ c^T x \]
\[ Ax = b \quad (I) \]
\[ x \geq 0 \quad \text{or} \quad A^T y + s = c \quad (II) \]
\[ y, s \geq 0 \]

can be written:

\[
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S & 0 & X
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
s
\end{bmatrix}
= \begin{bmatrix}
0 \\
b \\
0
\end{bmatrix}
\]

\[ x, s \geq 0 \]

where \( X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \), \( S = \begin{bmatrix} s_1 & \cdots & s_n \end{bmatrix} \)

This is a nonlinear system of equations.
Can define a direction by using a Newton step:

\[
\begin{bmatrix}
0 & A^T & I
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta y \\
\Delta s
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
-5\times
\end{bmatrix}
\]

Note: we assume primal and dual feasibility, and strict feasibility, i.e., \( Ax = b, x > 0, A^Ty + s = c, s > 0 \).

Let \( \mathcal{F}^0 = \mathcal{F}(x, y, s) : Ax = b, x > 0, A^Ty + s = c, s > 0 \).

Primal-dual method, then, the search direction toward the center, in order to be able to take a larger step.

They keep the \( x \) and \( s \) away from the boundary, \( x \geq 0, s \geq 0 \).
Central part conditions:

\[ A^T x + s = c \]
\[ A x = b \]
\[ x \geq s = 1 \]
\[ c = (\ldots, a) \]
\[ (x, s) > 0. \]

For each \( t > 0 \), the solution to the system defines a point on the central part. Denote this point by \( T \).
Most powerful dual algorithms take Newton steps towards points on $C$ with $t > 0$, rather than solving for a point on $C$.

More target points when it comes near close.
Duality gap: 
\[ c^T x - b^T y = (A^T y + s)^T x - (A x)^T y \]
\[ = s^T x + y^T A x - y^T A x \]
\[ = s^T x. \]

Define the duality measure \( m = \frac{x_i^T s}{n} = \frac{1}{n} \sum x_i s. \)

Define the centring parameters \( \bar{c}, \bar{c} \in [0, 1]. \)

General step equation:
\[
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S & 0 & x
\end{bmatrix}
\begin{bmatrix}
\bar{c} x \\
\bar{c} y \\
\bar{c} s
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
-x S e + \bar{c} m e
\end{bmatrix}
\]

Taking \( \bar{c} = 0 \) gives back the earlier system, and gives the affine scaling direction.

Taking \( \bar{c} = 1 \) gives the centring direction. This is

The direction obtained if we try to quote solve the system of

where the root of the central point with \( z = m \).
Algorithm

Pinheir's dual framework

Given \((x^0, y^0, s^0) \in \mathbb{F}_0\).

For \(k = 0, 1, 2, \ldots\)

Solve

\[
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S^u & 0 & x^0 \\
S^l & 0 & x^0 \\
\end{bmatrix}
\begin{bmatrix}
\delta x^k \\
\delta y^k \\
\delta s^k \\
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
-x^k s^k + c^u s^k \\
-x^k s^k + c^u s^k \\
\end{bmatrix}
\]

where \(s^0 \in [0, 1]\) and \(\mu_k = \frac{(x^0)^T s^0}{n}\).

Set

\[
(x^{k+1}, y^{k+1}, s^{k+1}) \leftarrow (x^k, y^k, s^k) + \alpha_k (\delta x^k, \delta y^k, \delta s^k)
\]

choosing \(\alpha_k\) so that \((x^{k+1}, y^{k+1}, s^{k+1}) > 0\).

end for
Lemma: If $A$ is tall and full rank and if $x > 0, s > 0$ then

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & x \end{bmatrix}$$

is a square invertible matrix.

And $A^{-1}$ has

Block row below:

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & x \end{bmatrix} \xrightarrow{B_3 -XB_1} \begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & -XA^T & 0 \end{bmatrix}$$

$$\xrightarrow{B_2 -AS^{-1}B_3} \begin{bmatrix} 0 & A^T & I \\ 0 & AS^{-1}XA^T & 0 \\ S & -XA^T & 0 \end{bmatrix}$$

Look at $AS^{-1}XA^T$:

\[
\begin{bmatrix} & & \\
& & \\
& & \\
\end{bmatrix}
\]

Invertible by assumption. So we can find $y$. Can be found by and $\sigma$

\[\]
Other stuff to discuss

1) Expand on the lemma on page 99, to get \( d_1 = d_5 = 0 \).

2) Can handle infeasible starting points.

3) Extending to other convex problems. (Complementary equation is different).

4) Superlinear convergence.

5) Interior function relaxation method.

6) Mehrota predictor-corrector.
   Dynamic selection of \( T \) based on progress available with the affine direction.
Convergent Process (Watten, C1913)

String table data

A, b, c. A is mma.

If all entries integer: requires \( \lceil \log_2 |a| \rceil + 1 \) bit to store an integer a.

If all entries real:

multiply through by least common denominator.

So can exam have integer data.

Need also to indicate where each number starts and stops.

So total space required is

\[ L = L_0 + (m + m_1) \]

where

- \( L_0 \) is number of entries
- \( m \) space to store the numbers themselves.

Natural Number Model

Will work with reals. \( +, -, \div, \times \) all present reals, as do native version of these.

When we need a square root, we can approximate it closely, enough by a rational in this course.

Stated result: Since a solution is bounded by a poly in sin of original data,
Lemma. When the problem data are integers, with length $L$, the vertices of the primal and dual feasible polytopes defined by

$$\{ x | A x = b, x \geq 0 \}, \\ \{ y, s | A^T y + s = c, s \geq 0 \}$$

are rational. Moreover, the nonzero components of $x$ and $s$ for these vertices are bounded below by $2^{-L}$. \textit{Proof. See text.}$^{17}$

Corollary. (P(k,l)) have unimodular optimum value if

$$c^Tx - b^Ty < 2^{-(L+1)}$$

then the optimum value $x$ and $(y,s)$ is optimal, provided $(x,y,s)$ is feasible.

Let $(x^*, y^*, s^*)$ be optimal.

Then $x^* s^* = c^T x^* - b^T y^*$.

Corollary. Any non-optimal vertex pair $(x,y,s)$ has objective value at least $2^{-2L}$.

Proof. If $x$ since $(x,y,s)$ is non-optimal, it violates complementary slackness.

Thus, for some component $i$, $x_i \geq 2^{-L}$ and $s_i \geq 2^{-L}$. So dually $y_i s_i \geq 2^{-2L}$. $^{17}$
Polynomials, Strongly Polynomial Algorithms

A polynomial algorithm has runtime that depends polynomially on the size of the problem.

E.g.: \( 57^n^3 \).

An algorithm is strongly polynomial if the number of arithmetical operations it performs is polynomial in the dimension of the problem, independent of the size of the individual entries.

E.g.: Todd's algo. for LP poly in entries in \( A \) allow, not dependent on \( B, c \).

From this work:

If \( f(n) = O(g(n)) \) then there exists a constant \( c \) such that \( |f(n)| \leq cg(n) \) for all \( n \geq k \).

If \( f(n) = o(g(n)) \) then \( \frac{f(n)}{g(n)} \to 0 \) as \( n \to \infty \).
Let \( \epsilon \in (0,1) \) be given.

Let \( \mu = \frac{x^5}{n} \).

Assume our algorithm generates a sequence of feasible iterate satisfying

\[
\mu_{k+1} \leq \left( 1 - \frac{\epsilon}{\omega} \right) \mu_k, \quad k = 0, 1, 2, \ldots
\]

for some positive constants \( \epsilon \) and \( \omega \). Suppose the starting point \((x^0, y^0, z^0)\) satisfies\( \mu_0 \leq \frac{1}{\epsilon^p}, \)

for some positive constant \( p \). Then there exist a value \( K \)

\[
K = O(\sqrt[p]{\log \frac{1}{\epsilon}}),
\]

such that

\[
\mu_k \leq \epsilon \quad \text{for all } k \geq K.
\]

We have

\[
\log(\mu_{k+1}) \leq \log\left(1 - \frac{\epsilon}{\omega}\right) + \log \mu_k
\]

\[
\leq (k+1) \log\left(1 - \frac{\epsilon}{\omega}\right) + \log \mu_0
\]

\[
\leq (k+1) \log\left(1 - \frac{\epsilon}{\omega}\right) + \frac{1}{p} \log\left(\frac{1}{\epsilon}\right)
\]
We have:

\[ \ln(1 + \beta) \leq \beta \] for all \( \beta \geq 1 \).

Then, 

\[ (k+1) \left( \frac{-\sigma}{\sqrt{n}} \right) + \Theta_p(\log \frac{1}{n}) \leq \log \epsilon \]

so

\[ (k+1) \left( \frac{-\sigma}{\sqrt{n}} \right) \leq (k+1) \log \frac{1}{n} \]

Therefore, 

\[ k_{\alpha} \leq 3 \] if 

\[ (k+1) \frac{\sigma}{\sqrt{n}} \geq (k+1) \log \frac{1}{n} \]

Thus, 

\[ \mu_n \leq \epsilon \] for all \( n \geq N = n \log_3 \frac{p+1}{\sigma} \]
Will see:
Short-term pert. following methods use $w = \frac{1}{2}$, only need one main step to update $p_k$.
Long-term pert. following methods use $w = 0$, may need several inner steps to update $p_k$.

Algorithm will use neighborhoods:

$N(0, \varepsilon) \subseteq \mathbb{R}^3$ feasible:
$E_j = \sum_{i=1}^{N} x_i - \mu e_i^2 \leq \Theta \mu^3$

$N_{\infty} (y) = \mathbb{F}(x, y, t) \text{ feasible: } x_i \leq y \mu \text{ for } i = 1, 2, \ldots, N$. 
**Interior Point Methods**

**Affine Methods**

Rescale problem and go in direction of steepest descent.


\[
\text{min } \mathbf{c}^T \mathbf{x} \\
\mathbf{A} \mathbf{x} = \mathbf{b} \\
\mathbf{x} \geq 0.
\]

- Normally: Have initial point \( x^0 > 0 \).
- Move in steepest descent direction.
- Cook inside taking shortest length.
- \( \mathbf{c} \) rescale so all coefficients differ.

Now can take longer.
Algebraically:

\[
\min_{x \geq 0} c^T x \\
\text{s.t. } A x = b \\
X^o \geq 0 \\
x^o i
\]

Have feasible point \( x^o > 0 \), \( x^o = \left[ \begin{array}{c} x_i^o \\ \vdots \\ x_n^o \end{array} \right] \)

Construct diagonal \( X^o = \left[ \begin{array}{c} x_i^o \\ \vdots \\ x_n^o \end{array} \right] \).

Let \( e = \left[ \begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \right] \). Then \( X^o e = x^o \) so \( A X^o e = b \) and \( (X^o c)^T e = c^T x \).

So consider the problem \( \min_{x \geq 0} \frac{c^T x}{x^T X^o} \leq \frac{c^T x}{x_i^o} \), \( \text{s.t. } A X^o x = b \).

where \( c = X^o c \), \( A = A X^o \).

\( e \) is feasible in \( \left( P \right) \), \( c^T e = c^T x \).

If \( \bar{x} \) is feasible in \( \left( P \right) \), then \( A X^o \bar{x} = b \), so \( A (X^o \bar{x}) = b \), so \( X^o \bar{x} \) is feasible in \( \left( P \right) \); also, \( \bar{c}^T e = c^T (X^o \bar{x}) \).

Thus, \( \left( P \right) \) is equivalent to \( \left( P \right) \), and \( \bar{x} \) is optimal for \( \left( P \right) \) with optimal value \( \bar{c} \) \( \Rightarrow \) \( X^o \bar{x} \) is optimal for \( \left( P \right) \) with optimal value \( \bar{c} \).

When we take a step in \( \left( \Pi \right) \), we need to worry about violating nonnegativity. So we'll correspond to nonnegativity. Since current point is \( \left( \Pi \right) \), \( \bar{e} \), maximum possible step length can always take a step of length at least 1.
Choosing a direction in $(P)$:

Ideally, move in direction $e$.

But this may move us off the affine span $Ax = b$.

So pick a direction in the null space of $A$; choose $\bar{d}$ with $\bar{A}d = 0$.

![Diagram showing null space and $\bar{A}d = 0$]

Project $e$ onto nullspace of $\bar{A}$:

Call this direction $\bar{d} = \frac{1}{\alpha} \bar{e}$.

New point: $x' = e\bar{e} - \alpha \bar{d}$ for some step length $\alpha > 0$.

Now $\bar{e}^T x' = \bar{e}^T x - \alpha \bar{e}^T \bar{d} = \bar{e}^T x - \alpha \bar{e}^T p = 0$.

Projection matrix is idempotent, i.e., $p = p^T = p^2$.

So $\bar{e}^T p\bar{e} = \bar{e}^T p^{2} \bar{e} = \bar{e}^T p\bar{e}^{2} \bar{e} = \|p_{\bar{A}}\bar{e}\|^2$.

So, projected $p_{\bar{A}}\bar{e} \neq 0$, $\bar{e}^T x' < \bar{e}^T x$. 
The new point in the problem \((p)\) is \(x' = x^o + \bar{x}'\)

**Newton's algorithm**

1. Initial \(x^o > 0\) for \(x' > 0\)

**Trust region algorithm**

1. Given \(c^T x = 0, \ A x^o = b, \ k = 0\)
2. Compute direction \(\bar{x} = \mathbf{A} x^o, \ c = x^o c\)
3. Calculate direction \(\bar{x} = \mathbf{A} x^o, \ c = x^o c\)
4. Scale back \(x^k+1 = \frac{x^k}{x' x'}^T x^k\)
5. Terminate loop: If termination criterion satisfied, STOP.

**Step length:**

1. Take \(\alpha = \frac{1}{||d||} \) more to edge of acceptability ellipsoid.
2. Take a fraction of \(\alpha \) to boundary:
   \[
   \alpha = 0.95 \min \left\{ \frac{1}{d_i} : \bar{d}_i > 0 \right\}
   \]

**Termination criteria:** Since algorithm is monotone, can block improve
\[c^T x < c^T x^o\] if this becomes small, STOP.
Dual variables:

\[ w = (\tilde{A} \tilde{A}^T)^{-1} \tilde{A} \tilde{z} \] is a possible dual vector.

Convergence proof:

Dikin's. Assume primal and dual degeneracy.

The Step is boundary of inscribed ellipsoid.

The \( (\tilde{A} \tilde{A}^T)^{-1} \tilde{A} \tilde{z} \) is a feasible solution to (P).

If \( w = (\tilde{A} \tilde{A}^T)^{-1} \tilde{A} \tilde{z} \) tends to an optimal dual solution.

Notice:

Dual to (P) is \( \max \{ b^T y \mid A y \leq c \} \) (D)

This is exactly the same as \( \max \{ b^T y \mid A y \leq c \} \) exceptslackare rescaled.

The slack, in (D) are \( 2 - A^T \tilde{w} = \tilde{d} \), and slack in (D) are \( (x^k)^{-1} \tilde{d} \)

So complementary slackness tends to \( (x^k)^{-1} [(x^k)^{-1} \tilde{d}] = \tilde{d} \).

Since \( w \) is becoming dual feasible, it follows that \( \tilde{d} \) tends to something nonnegative. Thus, since complementary slackness, must have \( \tilde{d} \to 0 \).

So, \[ \exists (x^k) \to 0 \text{ if } x \text{ is optimal point.} \]
Phase I: Need initial feasible point $b = a^T x$

So set up problem: $\max x_0$

s.t. $(b-Ax)x_0 + Ax = b$

$x_0, x \geq 0.$

Notice that $x_0 = 1, x = e$ is feasible in this problem.

Optimal solution gives feasible solution to $(P)$ if $(P)$ is feasible.
Dual Affine Method. (Adler, Karahan, Rescic, Vega.)

Instead of rescaling primal variables, we rescale the dual slacks.

\[ \begin{align*}
\text{min} & \quad c^T y \\
A x &= b \quad (P) \\
A y + z &= c \quad (D). \\
y & \geq 0, \\
z & \geq 0.
\end{align*} \]

Have dual feasible point \((y^*, z^*)\) with \(z^* > 0\).
Rescale \((D)\) so all slacks \(z_i = z^*_i = 1\):

Multiply by \((2^*)^{-1}\):

\[ \begin{align*}
\text{max} & \quad b^T \tilde{y} \\
\text{st.} & \quad A^T \tilde{y} + (2) = c. \\
\text{Equivalently:} & \quad \text{max} \quad b^T \tilde{y} \\
\text{st.} & \quad (2)^{-1} A^T \tilde{y} + z = (2)^{-1} c. \quad (D).
\end{align*} \]

\[ \begin{align*}
\text{Let rows of } H \quad (H \text{ is } (m \times n) \text{ full rank}) \text{ be basis of span.} \\
\text{Then } \begin{pmatrix} A(2)^{-1} \\ H \end{pmatrix} \text{ is non-invertible matrix.}
\end{align*} \]

So \((D)\) is equivalent to

\[ \begin{align*}
\text{max} & \quad b^T \tilde{y} \\
\text{st.} & \quad A(2)^{-1} A^T \tilde{y} + (2)^{-1} z = A(2)^{-1} c \\
& \quad H z = H c. \\
z & \geq 0.
\end{align*} \]

So \(\tilde{y} = (A(2)^{-1} A^T)^{-1} (A(2)^{-1} c - A(2)^{-1} z). \)
So (D) is equivalent to

\[ \min b^T (A(2)^{-1} A^T)^{-1} H (2)^{-1} c - b^T (A(2)^{-1} A^T)^{-1} A(2)^{-1} z \]

s.t. \[ H z = H c \]
\[ z \geq 0. \]

Steep descent direction:

\[ \Delta z = -P_H \left( (2)^{-1} A^T (A(2)^{-1} A^T)^{-1} b \right) \]

Now, \[-P_H \left( (2)^{-1} A^T (A(2)^{-1} A^T)^{-1} b \right) \]

\[ = \left( I - H (H^T H)^{-1} H \right) (2)^{-1} A^T (A(2)^{-1} A^T)^{-1} b \]

\[ = (2)^{-1} A^T (A(2)^{-1} A^T)^{-1} b, \text{ since } H (2)^{-1} A^T = 0, \]

by choice of \( H \).

Thus, \[ \Delta y = -\left( A(2)^{-1} A^T \right)^{-1} A(2)^{-1} \Delta z \]

\[ = + \left( A(2)^{-1} A^T \right)^{-1} A(2)^{-1} A^T (A(2)^{-1} A^T)^{-1} b \]

\[ = + \left( A(2)^{-1} A^T \right)^{-1} b. \]

and \[ \Delta z = -(2)^{-1} A^T (A(2)^{-1} A^T)^{-1} b. \]

Thus, \[ \Delta z = -(2)^{-1} A^T (A(2)^{-1} A^T)^{-1} b. \]

Final estimate: \[ -\Delta z, \text{ since } A\Delta z = -b. \]
**Projective Methods**

**The Centering Direction**

\[
\begin{align*}
\text{E.g.} \quad & \min \quad -x_1 - 7x_2 \\
\text{s.t.} \quad & x_1 + 10x_2 + 100x_3 = 111 \\
& x_i \geq 0,
\end{align*}
\]

Optimal point: \(x = [111 \ 0 \ 0]^T\).

Current feasible point: \(x = [1 \ 0 \ 0]^T\), so don't need to rescale.

**Alone direction:**

\[
A = \begin{bmatrix} 1 & 10 & 100 \end{bmatrix} \quad AA^T = 10101, \quad (AA^T)^{-1} = \frac{1}{10101} A_c = 91
\]

\[
P_{A} c = -c + A^T(1) = \begin{bmatrix} 1 \ 0 \ 0 \end{bmatrix} c - \begin{bmatrix} 91 \ 0 \ 0 \end{bmatrix} c \begin{bmatrix} 0.99 \ 0.99 \ 0.99 \end{bmatrix}^T
\]

\[
\approx 0.9 \begin{bmatrix} 1 \ 0 \ 0 \end{bmatrix}
\]

\[
x \approx \begin{bmatrix} 1 \ 0 \ 0 \end{bmatrix} + \alpha \begin{bmatrix} 0.99 \ -1 \ -1 \end{bmatrix}^T
\]

Taking \(\alpha = 1\) gives the point \(x \approx [2.1 \ 10.89 \ 0]^T\) on the boundary.

So the direction does not move us appreciably close to the optimal point.

**Direction:** Draw a sphere around current point, find best point in sphere in future.
In general:

Draw an ellipsoid around the current point, and find the best point on that ellipsoid. That gives the direction.

Can define an uninterpreted version of the algorithm:
Calculate direction, move to $x + \varepsilon d$, find new direction.

Step 1: Find the vertices
In the limit as $\varepsilon \to 0$, the path of iteration can appear every vertex of the polyhedron on the way to the optimal value (Megiddo & Schu, 1989, 1989)

Polyhedral is optimal solution.
Problems arise when the algorithm starts from near a vertex. So it is useful to try to "centralize the iterate.

Return to example: To try to move toward the "far-away" boundaries:

\[ x_2 \quad \text{look at } P_{Ae}. \]

We know movement in this direction.

\[
P_A e = e - A^T(AA^T)^{-1}Ae = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.111 \\ 1001 \\ 100 \end{bmatrix}
\]

\[ \approx \begin{bmatrix} 0.99 \\ 0.9 \\ -0.1 \end{bmatrix}. \]

This is better job of increasing \( x_1 \),\( \text{ let } -P_Ae \approx \begin{bmatrix} 11 \\ 10 \\ -1 \end{bmatrix}. \)

In general:

Define lower L.P. \( \min c^T x \) and the feasible point \( \bar{x}, \bar{x} > 0 \).

The affine scaling direction is \( -X_P A^{-1} c \), and

the centering direction is \( \bar{x} P_A e \).\]

Centropic direction is useful when close to a bad vertex, not so useful when close to a good vertex. So the combination of the directions
\[ d = -\bar{X}x - e + \bar{X}x e \]

should reflect this: \( b \) should get smaller as we approach the optimal vertex.

E.g. could have \( b = \) current duality gap.

One way to get a balance between these directions is to use a potential function (assum optimal value = 0 for now).
\[ f(x) = \frac{1}{2} \ln c^Tx - 2 \ln x^c \]

\[ \frac{df}{dx_i} = \frac{1}{c^Tx} - \frac{1}{x_i} \]

So at \( x = e \):
\[ Df = \frac{1}{c^Tx} (c - e) \]

So going in steepest descent direction for \( Df \) is going in the direction \[ -\frac{1}{c^Tx} P e + P e . \]
A Primal-Dual Potential Reduction Algorithm.

Un potential function \( \Phi(x,s) = \rho \log x^S - \sum \log x_i s_i \) for \( s_i > 0 \). Typical choice: \( \rho = n + \sqrt{n} \).

We have \( \Phi(x,s) = (\rho - n) \log x^S + n \log x^S - \sum \log x_i s_i \)

\[ = (\rho - n) \log x^S - \sum \log \left( \frac{x^S_i}{n} \right) + n \log n \]

\[ = \left( n - \frac{\rho - n}{n} \right) \log x^S - \sum \log x_i s_i + n \log n \]

Thus, if \( \Phi(x,s) \to -\infty \) then we must have \( x^S \to 0 \).

So try to decrease the potential function.

Define diagonal matrices \( X, S \) and then \( D = X^S S^{-1} \).

Take direction \( \Delta x = -\nabla D x \Phi(x,s) \) \hspace{1cm} (1)

Now, \( D x \Phi(x,s) = \frac{\rho}{x^S} S - S^{-1} e \)

Hence, 6.26, show that this can be found by solving

\[
\begin{bmatrix}
0 & A^T & I \\
A & 0 & 0 \\
S & 0 & X
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta y \\
\Delta g
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
-XS e + \frac{x^S}{\rho} e
\end{bmatrix}
\]
It can also be shown that

\[ \Delta s = -DA \tilde{\mathbf{T}}(DA^2 \tilde{\mathbf{T}}^{-1} \mathbf{A}) \tilde{\mathbf{T}}_s \tilde{\mathbf{I}}, \quad (x,s) \]

with \( \tilde{\mathbf{D}}_s \tilde{\mathbf{I}}_s(x,s) = \frac{x^T}{\pi^s} x + S^{-1} e \)

\[ \Delta y = (DA^2 \tilde{\mathbf{T}}^{-1} \mathbf{A}) \tilde{\mathbf{D}}_s \tilde{\mathbf{I}}, \quad (x,s) \]

Note: expressions given by (\(x\)) are scaled those given in (1), (2), (3), multiplied by \( \frac{x^T}{\pi^s} \). In what follows we use the directions given by (\(x\))

Theorem: Given a strictly feasible starting point \((x^0, y^0, s^0)\), suppose that an algorithm generates a sequence \((x^k, y^k, s^k)\) of strictly feasible iterates satisfying

\[ \forall (x^{k+1}, y^{k+1}) \leq (x^k, s^k) - \delta \quad \text{for all } k = 0, 1, 2, \ldots \]

for some fixed \( \delta > 0 \). Then for any \( \epsilon > 0 \) we have an index \( k_0 \),

\[ k_0 = \left[ \frac{\overline{\pi}(x^0, s^0) + C^{-n} \log \epsilon}{\delta} \right] \]

such that \( \mu_k \leq \epsilon \) for all \( k \geq k_0 \). \( \text{Proof} \) See text.
Now do we reason \((\tau), \text{ it,}\)

$$\bar{\mu} \left( x^{4t+1}, \sum s_{i} \right) \leq \bar{\mu} \left( x, s_{i} \right) - \tau.$$ 

Let $$\alpha_{\text{var}} = \min \left\{ \alpha : x_{t} + \alpha x_{t} \geq 0, \sum s_{i} \alpha s_{i} \geq 0 \right\}.$$ 

Now, $$\bar{\mu} \left( x + \alpha x, \sum s_{i} \right) - \bar{\mu} \left( x, s_{i} \right)$$

$$= \mu \log \left( \frac{(x + \alpha x)^{T} (\sum s_{i})}{x^{T} x} \right) - \sum \log \frac{x_{t} + \alpha x_{t}}{x_{t}} - \sum \log \frac{s_{i} \alpha s_{i}}{s_{i}^{T} s_{i}}$$

$$= \mu \log \left( 1 + \alpha s_{T} (x + x_{T}) \frac{x_{T}}{x_{T}} \right) - \sum \log \left( 1 + \frac{\alpha x_{T}}{x_{T}} \right) - \sum \log \left( 1 + \frac{\alpha s_{i}}{s_{i}} \right),$$

since $$\alpha x_{T} \sum s_{i} = 0,$$ because $$\alpha x$$ a nullspace of $$A$$

$$= 0$$ is a nullspace of $$A.$$ 

Choose a constant $$\tau \in (0, 1).$$ Let $$\tau = 0.5.$$ 

Define $$\alpha_{T}$$ so that

$$\alpha_{T} = \max \left\{ \frac{\alpha x}{x_{T}}, \frac{s_{T}}{s_{T}} \right\} = \tau.$$ 

This lets us bound the \(\sum\) term above, i.e., take $$\alpha \in (0, \tau]$$

Note: $$\alpha_{T} < \alpha_{\text{max}}.$$
We can then set up a quadratic approximation to $\phi(x + \alpha x, s + \alpha s)$. 

$q(\alpha)$ is bigger than $\phi(x + \alpha x, s + \alpha s)$ for $\alpha \in (0, \infty)$. 

Also, minimum value of $q(\alpha)$ is smaller than $\phi(x, s) - \delta$, 

$p(\alpha \in (0, \infty))$. 

So: Find minimum of $q(\alpha)$ in $(0, \infty)$ analytically. 

This step will guarantee sufficient decrease in $\phi(x, s)$. 

(For details, see text.) 

Let $v_c = \sqrt{x_c s_c}$, let $r_c = -v_c + \frac{x_c}{s_c} \cdot \frac{1}{v_c x_c}$. 

We can choose $\alpha = \frac{\text{V_dist}}{211.11}$ to guarantee a decrease of at least 0.15.
Relationship to centrality:

If iterate is centered then \( x_i S_i = \frac{x^T s}{n} = \mu \) (defined earlier).

Then, \( V_{\text{min}} = \sqrt{\nu} \) and \( c = -\sqrt{\nu} c + \frac{x^T s}{\sqrt{\nu}} \Rightarrow x = \left( \frac{x^T s}{\sqrt{\nu}} - \mu \right) c \)

so \( \|r\|^2 = n \left( \frac{x^T s}{\sqrt{\nu}} - \mu \right)^2 = n \left( \frac{(x^T s)^2}{\nu} + \mu - 2 \frac{x^T s}{\sqrt{\nu}} \right) \)

\[ = n \left( \frac{n x^T s}{\nu} + \frac{x^T s}{\sqrt{\nu}} - 2 \frac{x^T s}{\sqrt{\nu}} \right) \]

\[ = x^T s \left( \frac{n}{\sqrt{\nu}} - 1 \right)^2 \]

\[ = x^T s \left( \frac{\rho - n}{\rho} \right)^2 = \lambda \mu \left( \frac{\rho - n}{\rho} \right)^2 \].

Thus, \( \frac{V_{\text{min}}}{\|r\|^2} = \frac{1}{n} \frac{\rho}{\sqrt{\nu}} \sim 1 \) with \( \rho = n + \delta n \).

If poorly centered, so \( \frac{V_{\text{min}}}{\|r\|^2} \ll \mu \), get \( \alpha \ll 1 \).

So can take larger steps if iterate is better centered.

Note that we can’t get too poorly centered, at least Cond. \( x_i S_i \geq \varepsilon \).

In practice:

Try long steps, e.g. \( 0.95 \alpha \). If get sufficient decrease in potential function, accept the step. Else, reduce the step and try again.