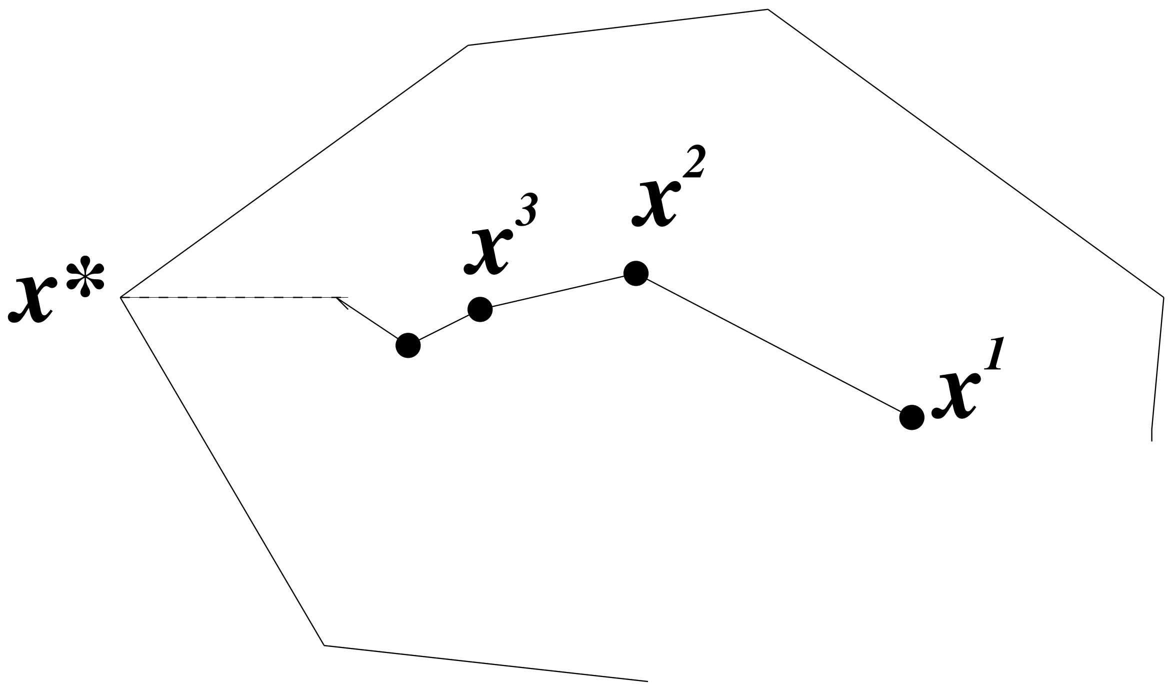
□ Interior Point Method

- Basic Idea:



Step 1: Start with an interior solution.

Step 2: If current solution is good enough, STOP.
Otherwise,

Step 3: Check all directions for improvement and
move to a better interior solution.
Go to Step 2.

- Performance:
  (1) It is a polynomial-time algorithm with complexity= $O(n^3L)$.
  (2) It outperformed the simplex method for large size LP.
  (3) 53 Netlib experiments.

□ Interior Movement

- Format:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{d}_{\mathbf{x}}^k$$

$$\begin{cases} \alpha \geq 0 : \text{ Step} - \text{length} \\ \mathbf{d}_{\mathbf{x}}^k \in R^n : \text{ moving direction} \end{cases}$$

- Questions:

1 How to find a "good" direction $\mathbf{d}_{\mathbf{x}}^k$?

2 How far should we move?

□ "Good" Directions

(A) Reduce the objective value

$$\underbrace{\mathbf{c}^T\mathbf{x}^{k+1}}_{\downarrow \; \mathbf{c}^T\mathbf{x}^k + \alpha\mathbf{c}^T\mathbf{d}_\mathbf{x}^k} \le \mathbf{c}^T\mathbf{x}^k$$

$$\Longrightarrow \mathbf{c}^T\mathbf{d}_\mathbf{x}^k \le 0$$

Candidate: $\mathbf{d}_\mathbf{x}^k = -\mathbf{c}$
(negative gradient)
(steepest decent)

(B) Keep feasibility

$$\begin{array}{ccc} \underline{\mathbf{A}\mathbf{x}^{k+1}} & = & \mathbf{b} \\ \downarrow & & \\ \mathbf{A}\mathbf{x}^k + \alpha \mathbf{A}\mathbf{d}_{\mathbf{x}}^k & & \\ \| & & \\ \mathbf{b} & & \end{array}$$

$$\Longrightarrow \mathbf{A}\mathbf{d}_{\mathbf{x}}^k = 0$$

i.e. $\mathbf{d}_{\mathbf{x}}^k \in \mathcal{N}(\mathbf{A})$ null space of $\mathbf{A}$.

Candidate: Projected negative gradient

$$\mathbf{d}_{\mathbf{x}}^k = (\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})(-\mathbf{c}).$$

- Valid Step-length

Fact: As long as $\mathbf{d}_{\mathbf{x}}^k \in \mathcal{N}(\mathbf{A})$,

$\mathbf{A}\mathbf{x}^{k+1} = \mathbf{b}$ no matter the value of α.
However

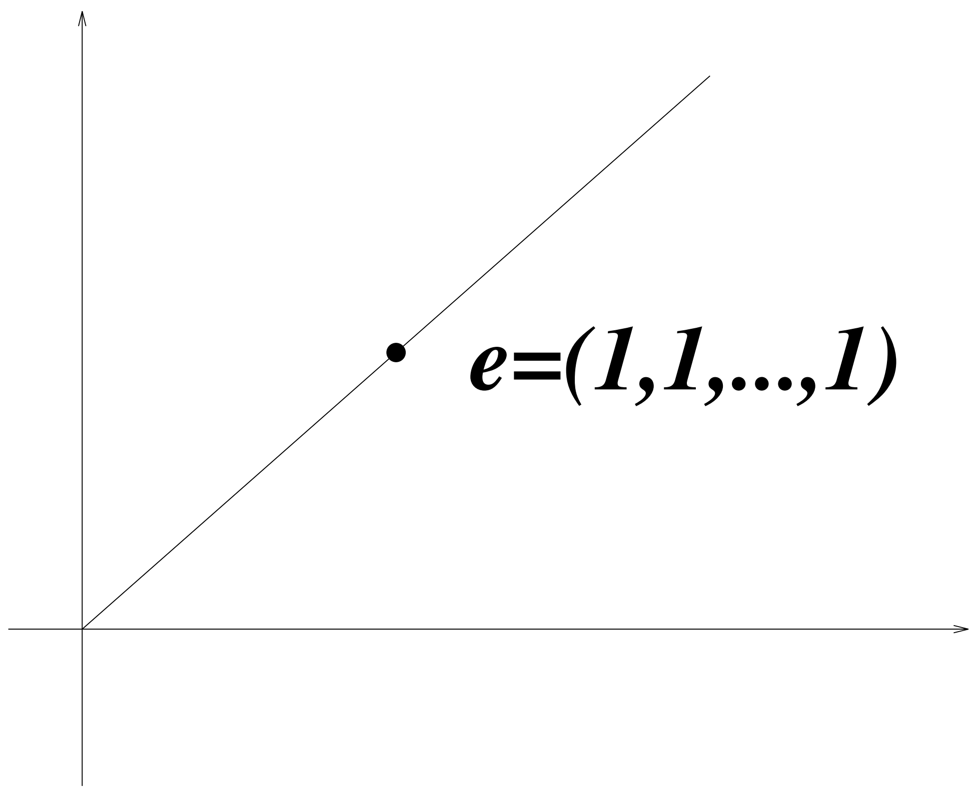
$$\mathbf{x}^{k+1} \geq 0$$

is required!

i.e., We have to know how far $\mathbf{x}^k$ is away from the boundary of non-negative orthant

$$\{\mathbf{x} \in R^n \mid \mathbf{x} \geq 0\}$$

(C) Scaling



If $\mathbf{x}^k = \mathbf{e}$, then

(1) $\mathbf{x}^k$ is one-unit away from the boundary.

(2) As long as $\alpha < 1$, $\mathbf{x}^{k+1} > 0$.

- Scale $\mathbf{x}^k$ to be $\mathbf{e}$

Define $\mathbf{X}_k = \mathrm{diag}(\mathbf{x}^k) = \begin{pmatrix} x_1^k & & \mathbf{0} & \\ & x_2^k & & \\ & & \ddots & \\ & \mathbf{0} & & x_n^k \end{pmatrix}$

then

$$\mathbf{X}_k^{-1}\mathbf{x}^k = \mathbf{e}.$$

Moreover,

$$\begin{array}{ccc} & \mathbf{X}_k^{-1} & \\ \mathbf{x} & \longrightarrow & \mathbf{y} = \mathbf{X}_k^{-1}\mathbf{x} \\ \boxed{R_+^n} & & \boxed{R_+^n} \\ \mathbf{x} = \mathbf{X}_k\mathbf{y} & \longleftarrow & \mathbf{y} \\ & \mathbf{X}_k & \end{array}$$

$$\begin{cases} \text{one} - \text{one} \\ \text{onto} \\ \text{boundary to boundary} \\ \text{interior to interior} \end{cases}$$

$$\mathbf{x} = \mathbf{X}_k \mathbf{y}$$

$$\boxed{\begin{array}{r} \text{Min } \mathbf{c}^T\mathbf{x} \\ \text{s. t. } \mathbf{A}\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq 0 \end{array}} \quad \longrightarrow \quad \boxed{\begin{array}{l} \text{Min } \mathbf{c}^T\mathbf{X}_k\mathbf{y} \\ \text{s. t. } \mathbf{A}\mathbf{X}_k\mathbf{y} = \mathbf{b} \\ \mathbf{y} \geq 0 \end{array}}$$

$$\mathbf{x}^k > 0 \qquad \mathbf{y}^k = \mathbf{e}$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha_k \frac{\mathbf{d}_{\mathbf{y}}^k}{\|\mathbf{d}_{\mathbf{y}}^k\|}$$

$$\mathbf{d}_{\mathbf{y}}^k = [\mathbf{I} - \mathbf{X}_k \mathbf{A}^T (\mathbf{A}\mathbf{X}_k^2 \mathbf{A}^T)^{-1} \mathbf{A}\mathbf{X}_k] \cdot (-\mathbf{X}_k \mathbf{c})$$

$$\begin{aligned} \mathbf{x}^{k+1} &= \mathbf{X}_k \mathbf{y}^{k+1} \\ &= \mathbf{X}_k \mathbf{y}^k + \alpha_k \mathbf{X}_k \frac{\mathbf{d}_{\mathbf{y}}^k}{\|\mathbf{d}_{\mathbf{y}}^k\|} \\ &= \mathbf{x}^k + \frac{\alpha_k}{\|\mathbf{d}_{\mathbf{y}}^k\|} \mathbf{d}_{\mathbf{x}}^k \end{aligned}$$

$$\mathbf{d}_{\mathbf{x}}^k = -\mathbf{X}_k [\mathbf{I} - \mathbf{X}_k \mathbf{A}^T (\mathbf{A}\mathbf{X}_k^2 \mathbf{A}^T)^{-1} \mathbf{A}\mathbf{X}_k] (\mathbf{X}_k \mathbf{c})$$

$\alpha_k = 0.99$ (say).

□ Recent Results

- Projective Scaling (N. Karmarkar 1984).
- Pimal Affine Scaling
  (Barnes/Vanderkei-Mcketon-Freedman 1985).
- Dual Affine Scaling
  (Adler-Karmarkar-Rosende-Veiga 1986).
- Pima-Dual (Kojima-Mizuno-Yoshise 1987).
- Pimal Affine with logarithmic barrier
  (Gonzaga 1989).
- Potential Reduction (Ye 1990).
- Unconstrained Convex Approach (Fang 1990).

  ⋮

□ New Trends

1. Integration of the Interior-point methods.
2. Develop hybrid algorithms
3. Solving very large size LP for real-world applications
   - exploring special structure.
   - sparsity.
   - decomposition.
   - parallel computations.

□ References

This talk is based on the book
Linear Optimization and Extensions: Theory and Algorithms, Prentice-Hall (1993),
authored by S.-C. Fang & S. Puthenpura.

□ Primal Affine Algorithm

Motivation

Total Complexity of iterative algorithm =
(# of iterations) $\times$ (operations in each iteration)

Simplex Method

- Simple operations
  : Only check adjacent extreme points.
- May take many iterations
  : Klee-Minty example.

Karmarkar's Algorithm

- Complicated iterations
  : Check all directions for the best one.
- Take fewer iterations

□ Primal Affine Algorithm

Is Simplex Method Good?

- In general, it visits about $0.7159\, m^{0.9522}\, n^{0.3109}$ vertices (linear in m, sub-linear in n).
- In the worst cases, Klee and Minty (1971) showed that it traverses $2^n - 1$ vertices (exponential-time algorithm)

Basic Strategy of Interior-Point Approach

- Stay inside of P
- Check more directions of movement
- Shorten travelling path
  i.e., Increase complexity at each iteration but reduce total number of iterations

Interior-Point Method

- Projective scaling method
- Affine scaling method
  - Primal Affine Scaling algorithm
  - Dual Affine Scaling algorithm
  - Primal-Dual algorithm
- Potential Reduction method
- Path-Following method

“Good” direction

(A). **Reduce the objective value**

$$\mathbf{c}^T\mathbf{x}^{k+1} \leq \mathbf{c}^T\mathbf{x}^k$$

$$\|$$

$$\mathbf{c}^T\mathbf{x}^k + \alpha_k \mathbf{c}^T\mathbf{d}_\mathbf{x}^k$$

$$\Rightarrow \ \mathbf{c}^T\mathbf{d}_\mathbf{x}^k \leq 0$$

Candidate: $\mathbf{d}_\mathbf{x}^k = -\mathbf{c}$

(negative gradient)

(Steepest descent)

(B). **Keep feasibility**

$$\mathbf{A}\mathbf{x}^{k+1} = \mathbf{b}$$

$$\|$$

$$\mathbf{A}\mathbf{x}^{k} + \alpha_k \mathbf{A}\mathbf{d}_{\mathbf{x}}^{k}$$

$$\|$$

$$\mathbf{b}$$

$$\Rightarrow \quad \mathbf{A}\mathbf{d}_{\mathbf{x}}^{k} = 0$$

i.e. $\mathbf{d}_{\mathbf{x}}^{k} \in \mathcal{N}(\mathbf{A})$: null space of $\mathbf{A}$

Candidate: projected negative gradient

$$\mathbf{d}_{\mathbf{x}}^{k} = (I - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})(-\mathbf{c})$$

Valid Step-length

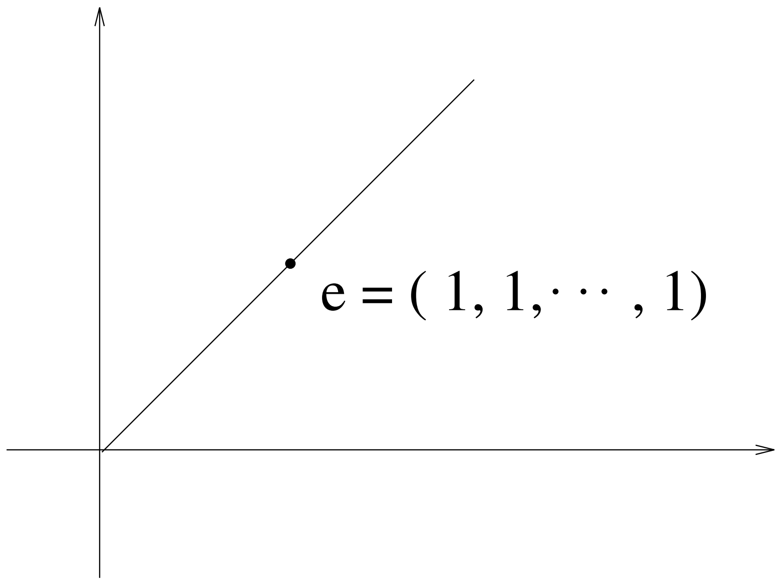
Fact: As long as $\mathbf{d}_{\mathbf{x}}^k \in \mathcal{N}(\mathbf{A})$ $\mathbf{A}\mathbf{x}^{k+1} = \mathbf{b}$ no matter the value of α.

However,

$\mathbf{x}^{k+1} > 0$ is required!

i.e., We have to know how far $\mathbf{x}^k$ is away from the boundary of the non-negative orthant $\{\mathbf{x} \in R^n \mid \mathbf{x} \geq 0\}$

(C) **Scaling**



If $\mathbf{x}^k = e$, then

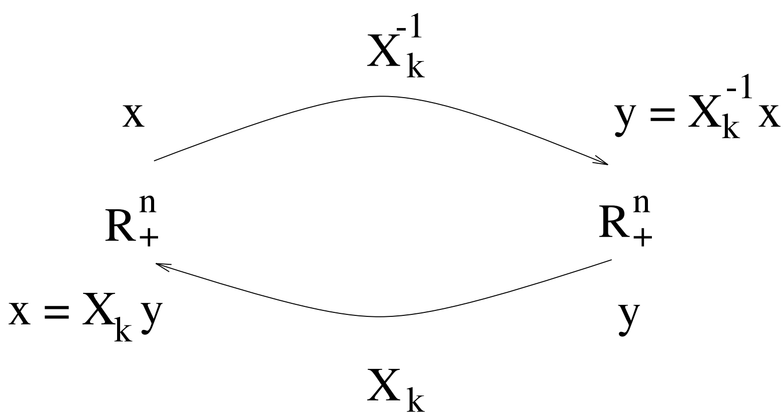
(1) $\mathbf{x}^k$ is one-unit away from the boundary

(2) As long as $\alpha < 1$, $\mathbf{x}^{k+1} > 0$

Scale $\mathbf{x}^k$ to be e

Define $X_k = \text{diag}(\mathbf{x}^k) = \begin{pmatrix} \mathbf{x}_1^k & & & \\ & \mathbf{x}_2^k & & 0 \\ 0 & & \ddots & \\ & & & \mathbf{x}_n^k \end{pmatrix}$

then $X_k^{-1}\mathbf{x}^k = e$



Moreover,

$$\begin{cases} \text{one-to-one} \\ \text{onto} \\ \text{boundary to boundary} \\ \text{interior to interior} \end{cases}$$

$x = X_k y$

$$\begin{array}{ll} \text{Min} & c^T x \\ \text{s.t.} & A x = b \\ & x \geqslant 0 \end{array} \qquad \longrightarrow \qquad \begin{array}{ll} \text{Min} & c^T X_k y \\ \text{s.t.} & A X_k y = b \\ & y \geqslant 0 \end{array}$$

$$\mathbf{x}^k > 0 \qquad\qquad \mathbf{y}^k = e$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha_k \frac{\mathbf{d}_\mathbf{y}^k}{\|\mathbf{d}_\mathbf{y}^k\|}$$

$$\mathbf{d}_\mathbf{y}^k = [I - X_k \mathbf{A}^T (\mathbf{A} X_k^2 \mathbf{A}^T)^{-1} \mathbf{A} X_k] \cdot (-X_k \mathbf{c})$$

$$\begin{aligned} \mathbf{x}^{k+1} &= X_k \mathbf{y}^{k+1} \\ &= X_k \mathbf{y}^k + \alpha_k X_k \tfrac{\mathbf{d}_\mathbf{y}^k}{\|\mathbf{d}_\mathbf{y}^k\|} \\ &= \mathbf{x}^k + \tfrac{\alpha_k}{\|\mathbf{d}_\mathbf{y}^k\|} \mathbf{d}_\mathbf{x}^k \end{aligned}$$

$$\therefore \ \underline{\mathbf{d}_\mathbf{x}^k = -X_k [I - X_k \mathbf{A}^T (\mathbf{A} X_k^2 \mathbf{A}^T)^{-1} \mathbf{A} X_k] X_k \mathbf{c}}$$

$$\underline{\alpha_k = 0.99 \text{ (say)}} \qquad 0 < \alpha_k < 1$$

Observations

(1) Another way to determine step-length α_k

Since $\mathbf{d}_{\mathbf{y}}^k = P_k(-X_k\mathbf{c})$

$\therefore\ \mathbf{A}X_k\mathbf{d}_{\mathbf{y}}^k = 0$ and

$\mathbf{A}X_k\mathbf{y}^{k+1} = \underline{\mathbf{A}X_k\mathbf{y}^k} + \underline{\alpha_k\mathbf{A}X_k\mathbf{d}_{\mathbf{y}}^k} = \mathbf{b}$

In order to make sure that $\mathbf{y}^{k+1} > 0$ we need

$$\mathbf{y}^k + \alpha_k\mathbf{d}_{\mathbf{y}}^k > 0$$

$$\|$$

$$e$$

Case 1: $\mathbf{d}_{\mathbf{y}}^k \geq 0$ then $\alpha_k \in (0, \infty)$

Case 2: $(\mathbf{d}_{\mathbf{y}}^k)_i < 0$ for some i

$\alpha_k = \min_i\{\frac{1}{-(\mathbf{d}_{\mathbf{y}}^k)_i} \mid (\mathbf{d}_{\mathbf{y}}^k)_i < 0\}$

or

$\alpha_k = \min_i\{\frac{\alpha}{-(\mathbf{d}_{\mathbf{y}}^k)_i} \mid (\mathbf{d}_{\mathbf{y}}^k)_i < 0\}$ for some

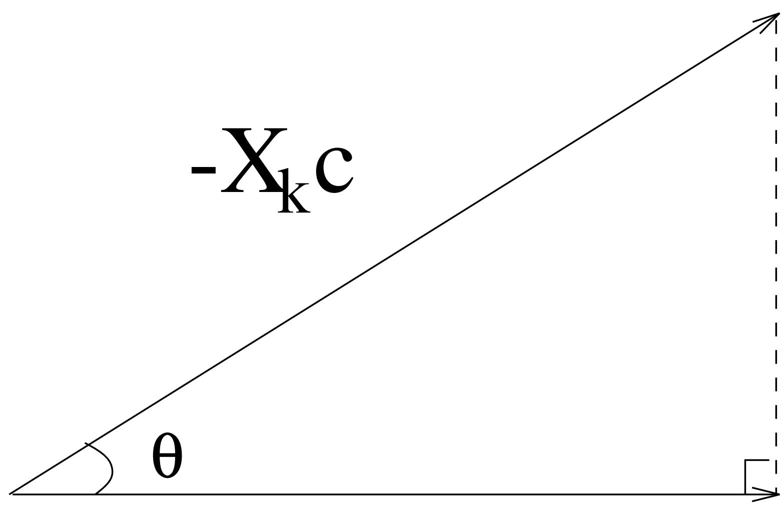
$\alpha \in (0, 1)$

(2) As in (1) $\mathbf{x}^{k+1}$

$$
\begin{aligned}
&= X_k \mathbf{y}^{k+1} \\
&= X_k(e + \alpha_k \mathbf{d}_{\mathbf{y}}^k) \\
&= \mathbf{x}^k + \alpha_k X_k \mathbf{d}_{\mathbf{y}}^k \\
&= \mathbf{x}^k + \alpha_k X_k(-P_k X_k \mathbf{c}) \\
&= \mathbf{x}^k - \alpha_k X_k[I - X_k \mathbf{A}^T(\mathbf{A}X_k^2\mathbf{A}^T)^{-1}\mathbf{A}X_k]X_k\mathbf{c} \\
&= \mathbf{x}^k - \alpha_k X_k^2[\mathbf{c} - \mathbf{A}^T \underbrace{(\mathbf{A}X_k^2\mathbf{A}^T)^{-1}\mathbf{A}X_k^2\mathbf{c}}_{\mathbf{w}^k}] \\
&= \mathbf{x}^k - \alpha_k \underbrace{X_k^2[\mathbf{c} - \mathbf{A}^T\mathbf{w}^k]}_{\mathbf{d}_{\mathbf{x}}^k} \\
&= \mathbf{x}^k - \alpha_k \mathbf{d}_{\mathbf{x}}^k
\end{aligned}
$$

(3)

$$\begin{aligned}\mathbf{c}^T\mathbf{x}^{k+1} &= \mathbf{c}^T(\mathbf{x}^k + \alpha_k X_k \mathbf{d}_{\mathbf{y}}^k) \\ &= \mathbf{c}^T\mathbf{x}^k + \alpha_k \mathbf{c}^T X_k(-P_k X_k \mathbf{c}) \\ &= \mathbf{c}^T\mathbf{x}^k - \alpha_k \| - P_k X_k \mathbf{c}\|^2 \\ &= \mathbf{c}^T\mathbf{x}^k - \alpha_k \|\mathbf{d}_{\mathbf{y}}^k\|^2\end{aligned}$$



$$d_y^k = -P_k X_k c$$

Hence, $\mathbf{c}^T\mathbf{x}^{k+1} \leq \mathbf{c}^T\mathbf{x}^k$

and $\mathbf{c}^T\mathbf{x}^{k+1} < \mathbf{c}^T\mathbf{x}^k$ if $\mathbf{d}_{\mathbf{y}}^k \neq 0$

Lemma 7.1 If $\exists\ \mathbf{x}^k \in P,\ \mathbf{x}^k > 0$ with $\mathbf{d}_{\mathbf{y}}^k > 0$, then the standard LP is unbounded below.

(4) For $\mathbf{x}^k \in P^0 = \{\mathbf{x} \in R^n \mid \mathbf{Ax} = \mathbf{b},\ \mathbf{x} > 0\}$ if $\mathbf{d}_\mathbf{y}^k = -P_k X_k \mathbf{c} = 0$ then $X_k \mathbf{c}$ falls in the orthogonal space of $N(AX_k)$, *i.e.*

$X_k \mathbf{c} \in$ row space of $(\mathbf{A}X_k)$
$\Rightarrow \ \exists\, u^k$ s.t. $(\mathbf{A}X_k)^T u^k = X_k \mathbf{c}$
or $(u^k)^T \mathbf{A} X_k = \mathbf{c}^T X_k$
$\Rightarrow \ (u^k)^T A = \mathbf{c}^T$

For any $\mathbf{x} \in P$
$\mathbf{c}^T \mathbf{x} = (u^k)^T \mathbf{Ax} = \underline{(u^k)^T \mathbf{b}}$ (constant)
$\therefore$ Any feasible solution is optimal !! (Lemma 7.2) In particular, $\mathbf{x}^k$ is optimal !

(5) Combining (3) & (4), if the standard form LP is bounded below and $\mathbf{c}^T \mathbf{x}$ is not a constant, then $\{\mathbf{c}^T \mathbf{x}^k \mid k = 1, 2, \cdots\}$
is well-defined and strictly decreasing.
(Lemma 7.3)

(6) $\mathbf{w}^k \equiv (\mathbf{A}X_k^2\mathbf{A}^T)^{-1}AX_k^2\mathbf{c}$ dual estimate

$\mathbf{r}^k \equiv \mathbf{c} - A^T\mathbf{w}^k$ reduced cost

If $\mathbf{r}^k \geq 0$, then $\mathbf{w}^k$ is dual feasible

and $(\mathbf{x}^k)^T\mathbf{r}^k = e^T X_k \mathbf{r}^k$ becomes the duality gap, *i.e.*,

$\mathbf{c}^T\mathbf{x}^k - \mathbf{b}^T\mathbf{w}^k = e^T X_k \mathbf{r}^k$

Therefore, if $\underline{\mathbf{r}^k \geq 0 \text{ and } e^T X_k \mathbf{r}^k = 0}$

(Stopping rule) $\nearrow$

then $\mathbf{x}^k \leftarrow \mathbf{x}^*$, $\mathbf{w}^k \leftarrow \mathbf{w}^*$

(7) $\mathbf{d}_\mathbf{y}^k$

$$= [I - X_k\mathbf{A}^T(\mathbf{A}X_k^2\mathbf{A}^T)^{-1}\mathbf{A}X_k](-X_k\mathbf{c})$$
$$= -X_k(\mathbf{c} - \mathbf{A}^T(\mathbf{A}X_k^2\mathbf{A}^T)^{-1}\mathbf{A}X_k^2\mathbf{c})$$
$$= -X_k(\mathbf{c} - \mathbf{A}^T\mathbf{w}^k)$$
$$= -X_k\mathbf{r}^k$$

Primal Affine Scaling Algorithm

Step1 Set $k \leftarrow 0, \varepsilon > 0, 0 < \alpha < 1$

find $\mathbf{x}^0 > 0$ and $A\mathbf{x}^0 = \mathbf{b}$

Step2 Compute

$\mathbf{w}^k = (\mathbf{A}X_k^2\mathbf{A}^T)^{-1}\mathbf{A}X_k^2\mathbf{c}$

$\mathbf{r}^k = \mathbf{c} - \mathbf{A}^T\mathbf{w}^k$

If $\mathbf{r}^k \geq 0$, and $e^T X_k \mathbf{r}^k \leq \varepsilon$

then STOP! $\mathbf{x}^* \leftarrow \mathbf{x}^k$, $\mathbf{w}^* \leftarrow \mathbf{w}^k$

Otherwise,

Step3 Compute $\mathbf{d}_{\mathbf{y}}^k = -X_k\mathbf{r}^k$

If $\mathbf{d}_{\mathbf{y}}^k \overset{>}{\neq} 0$,then STOP! Unbounded.

If $\mathbf{d}_{\mathbf{y}}^k = 0$, then STOP! $\mathbf{x}^* \leftarrow \mathbf{x}^k$

Otherwise,

Step4 Find

$\alpha_k = \min_i\{\frac{\alpha}{-(\mathbf{d}_{\mathbf{y}}^k)_i} \mid (\mathbf{d}_{\mathbf{y}}^k)_i < 0\}$

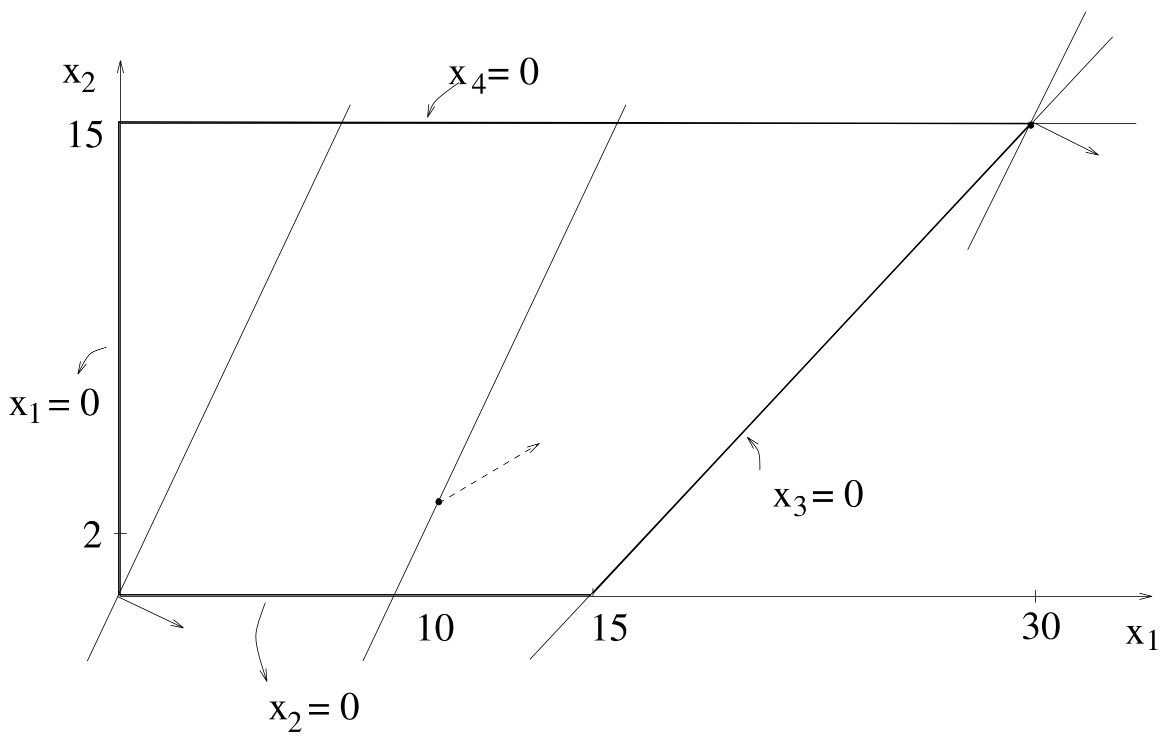
$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k X_k \mathbf{d}_{\mathbf{y}}^k$

$k \leftarrow k + 1$

Go to Step 2.

AN EXAMPLE

$$\begin{aligned} \min \quad & -2x_1 + x_2 \\ \text{s.t.} \quad & x_1 - x_2 \le 15 \\ & x_2 \le 15 \\ & x_1, x_2 \ge 0 \end{aligned}$$



Reformulate to standard form

$$\begin{aligned} \min \quad & -2x_1 + x_2 \\ \text{s.t.} \quad & x_1 - x_2 + x_3 &= 15 \\ & x_2 + x_4 &= 15 \\ & x_1, x_2, x_3, x_4 &\geq 0 \end{aligned}$$

and $\mathbf{x}^0 = \begin{pmatrix} 10 \\ 2 \\ 7 \\ 13 \end{pmatrix}$ is feasible

MATRIX FORMAT

$$
\begin{aligned}
\min \quad & \mathbf{c}^T\mathbf{x} \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{x} > \mathbf{0}
\end{aligned}
$$

$$
\text{where } \mathbf{A} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 15 \\ 15 \end{bmatrix}
$$

$$
\mathbf{c} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \; \mathbf{x}^0 = \begin{bmatrix} 10 \\ 2 \\ 7 \\ 13 \end{bmatrix}
$$

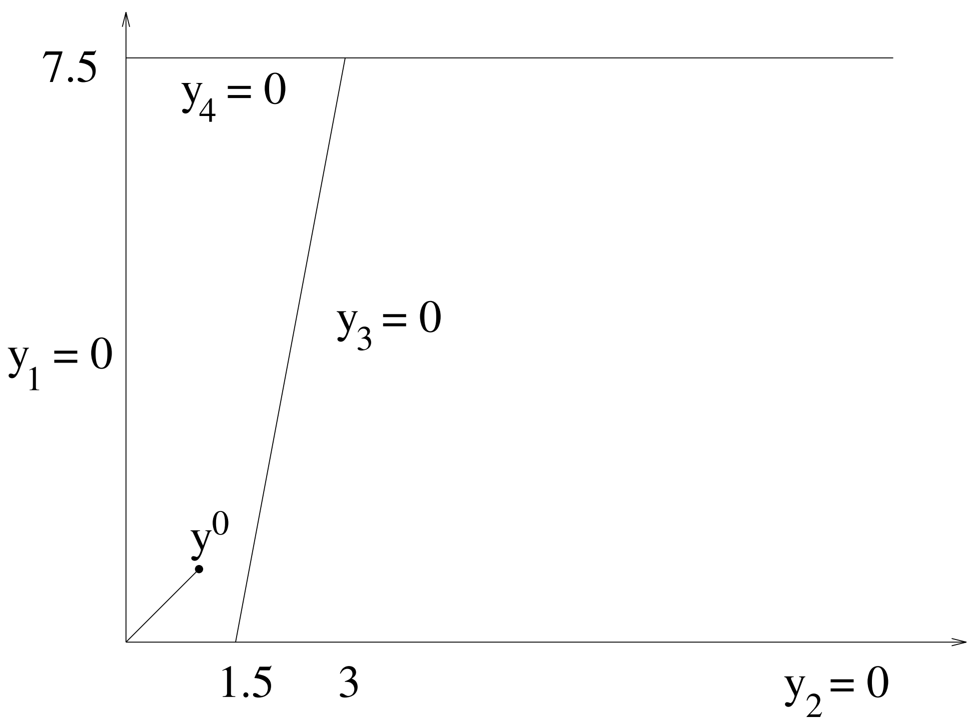
$$
\mathbf{X}_0 = \begin{bmatrix} 10 & & & 0 \\ & 2 & & \\ 0 & & 7 & \\ & & & 13 \end{bmatrix}
$$

SCALING

$$\mathbf{y} = X_0^{-1}\mathbf{x} = \begin{bmatrix} \frac{1}{10} & & & \\ & \frac{1}{2} & & 0 \\ 0 & & \frac{1}{7} & \\ & & & \frac{1}{13} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \qquad \begin{array}{l} y_1 = x_1/10 \\ y_2 = x_2/2 \\ y_3 = x_3/7 \\ y_4 = x_4/13 \end{array}$$

The problem is transformed to

$$\begin{array}{lrr} \min & -2(10y_1) + (2y_2) & = -20y_1 + 2y_2 \\ \text{s.t.} & 10y_1 - 2y_2 & +7y_3 = 15 \\ & 2y_2 & +13y_4 = 15 \\ & y_1, y_2, y_3, y_4 & \geq 0 \end{array}$$



The new matrix form

$$\begin{aligned} \min \quad & \bar{\mathbf{c}}^T \mathbf{y} \\ \text{s.t.} \quad & \bar{\mathbf{A}}\mathbf{y} = \mathbf{b} \\ & \mathbf{y} > \mathbf{0} \end{aligned}$$

where

$$\bar{\mathbf{A}} = \begin{bmatrix} 10 & -2 & 7 & 0 \\ 0 & 2 & 0 & 13 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 15 \\ 15 \end{bmatrix}$$

$$\bar{\mathbf{c}} = \begin{bmatrix} -20 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \ \text{and } \mathbf{y}^0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{cases} \bar{A} = AX_0 \\ \bar{\mathbf{c}} = X_0 \mathbf{c} \\ \mathbf{y}^0 = X_0^{-1} \mathbf{x}_0 \end{cases}$$

Step direction in transformed space

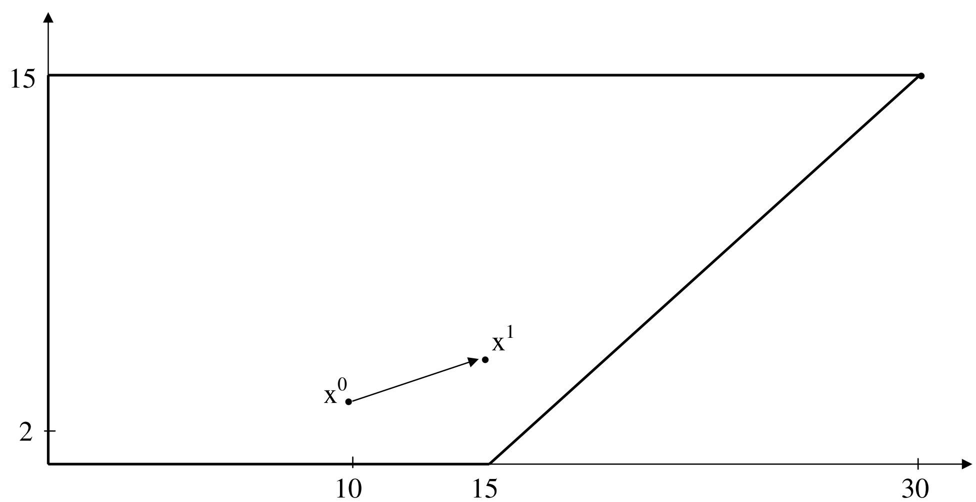
$$\mathbf{d}_{\mathbf{y}}^{0} = -PX_0\mathbf{c} = \begin{bmatrix} +6.66 \\ +0.68 \\ -9.33 \\ -0.10 \end{bmatrix}$$

$$\mathbf{y}^1 = \mathbf{y}^0 + \frac{0.95}{11.48} \begin{bmatrix} +6.66 \\ +0.68 \\ -9.33 \\ -0.10 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 0.083 \begin{bmatrix} +6.66 \\ +0.68 \\ -9.33 \\ -0.10 \end{bmatrix} = \begin{bmatrix} 1.55 \\ 1.05 \\ 0.23 \\ 0.99 \end{bmatrix}$$

Scale back

$$\mathbf{x}^1 = X_0\mathbf{y}^1 \Rightarrow \begin{cases} x_1 = 10 \cdot y_1 = 15.5 \\ x_2 = 2 \cdot y_2 = 2.10 \\ x_3 = 7 \cdot y_3 = 1.61 \\ x_4 = 13 \cdot y_4 = 12.87 \end{cases}$$



How to Start ?

1. Big-M method

$$(LP)\begin{cases} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{cases}$$

Objective: to make $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ be feasible, i.e.,

$\mathbf{A}e = \mathbf{b}$?

Method: Adding an artificial variable x^a with a large positive number M for

$$(LP')\begin{cases} \min & \mathbf{c}^T\mathbf{x} + Mx^a \\ \text{s.t.} & \mathbf{A}\mathbf{x} + (\mathbf{b} - \mathbf{A}e)x^a = \mathbf{b} \\ & \mathbf{x} \geq 0,\ x^a \geq 0 \end{cases}$$

Properties:

(1) (LP$'$) is a standard form LP with $n+1$ variables and m constraints.

(2) $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in R^{n+1}$ is an interior feasible solution of (LP$'$).

(3) If $x^{a^*} > 0$ in $(\mathbf{x}^*, x^{a^*})$ then (LP) is infeasible. Otherwise, either (LP) is unbounded or $\mathbf{x}^*$ is optimal to (LP).

2. **Two-Phase method**

$$(LP) \begin{cases} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{cases}$$

Choose any $\mathbf{x}^0 > 0$, calculate

$$\mathbf{v} = \mathbf{b} - \mathbf{A}\mathbf{x}^0$$

If $\mathbf{v} = 0$, then $\mathbf{x}^0$ is interior feasible.
Otherwise, consider

$$(Phase - I) \begin{cases} \min & u \\ \text{s.t.} & \mathbf{A}\mathbf{x} + \mathbf{v}u = \mathbf{b} \\ & \mathbf{x} \geq 0, \ u \geq 0 \end{cases}$$

Properties:

(1) (Phase-I) is a standard form LP with $n+1$ variables and m constraints.

(2) $\hat{\mathbf{x}}^0 = \begin{pmatrix} \mathbf{x}^0 \\ u^0 \end{pmatrix} = \begin{pmatrix} \mathbf{x}^0 \\ 1 \end{pmatrix}$ is interior feasible for (Phase-I).

(3) (Phase-I) is bounded below by 0.

(4) Apply primal-affine scaling to (Phase-I) will generate $\begin{pmatrix} \mathbf{x}^* \\ u^* \end{pmatrix}$ for (Phase-I).
If $u^* > 0$, (LP) is infeasible.
Otherwise, $\mathbf{x}^* > 0$ for (Phase-II) as an initial feasible solution.

Properties of Primal Affine Scaling

(1) The convergence proof, *i.e.*,

$$\{\mathbf{x}^k\} \to \mathbf{x}^*$$

under Non-degeneracy assumption (Theorem 7.2) is given by Vanderbei/Meketon/ Freedman in (1985).

(2) Convergence proof without Non-degeneracy assumption,
T. Tsuchiya (1991)
P. Tseng/ Z. Luo (1992)

(3) The computational bottleneck is to find

$$(AX_k^2A^T)^{-1}$$

(4) No polynomial-time proof
- J. Lagarias showed primal affine scaling is only of super-linear rate.
- N. Megiddo/ M. Shub showed that primal affine scaling might visit all vertices if it moves too close to the boundary.

(5) In practice, VMF reported

	# iterations
Simplex	$0.7159\ m^{0.9522}\ n^{0.3109}$
Affline Scaling	$7.3385\ m^{-0.0187}\ n^{0.1694}$

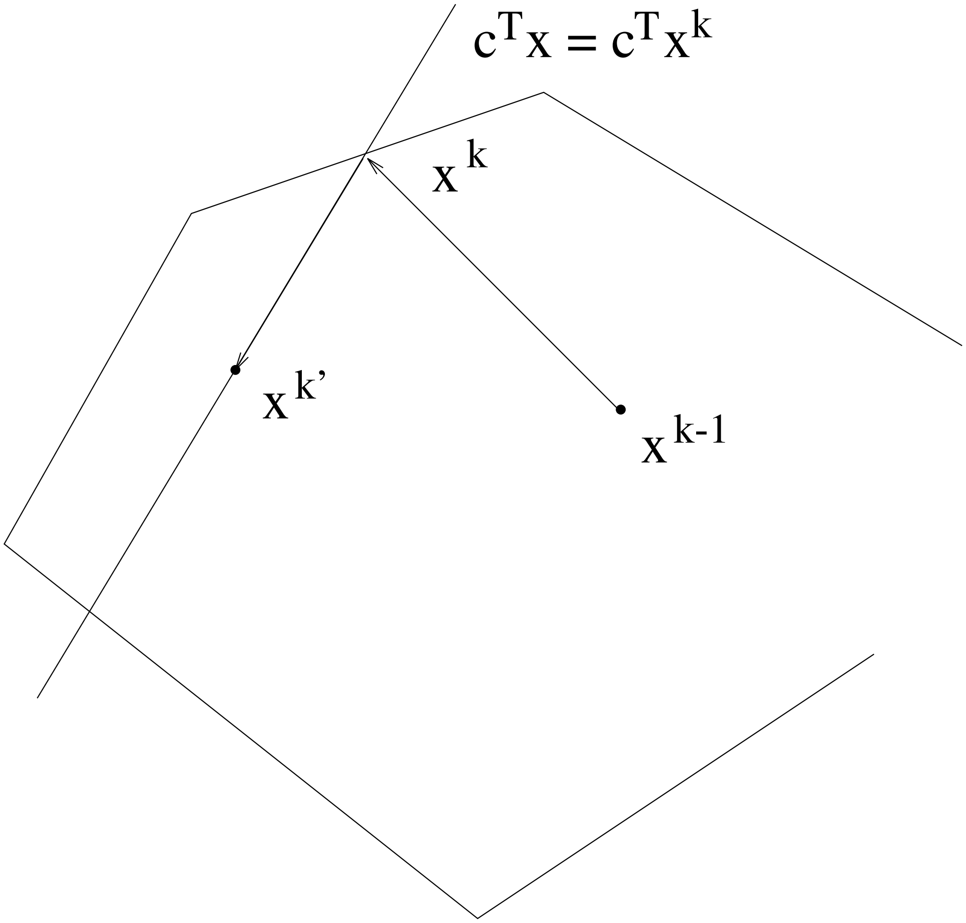
(6) It may lose primal feasibility due to machine accuracy (Phase-I again).

(7) May be sensitive to primal degeneracy.

Improving Primal-Affine Scaling

Objective: Stay away from the boundary !

1. Potential Push Method



$$\begin{aligned} \min \quad & -\sum_{j=1}^{n} \log_e x_j \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} > 0 \\ & \mathbf{c}^T\mathbf{x} = \mathbf{c}^T\mathbf{x}^k \end{aligned}$$

$$\Downarrow$$

$(\mathbf{x}^k)'$ to replace $\mathbf{x}^k$

2. Logarithmic Barrier Function Method

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} - \mu \textstyle\sum_{j=1}^n \log_e x_j \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} > 0 \end{aligned}$$

(1) $\{\mathbf{x}^*(\mu) \mid \mu > 0\} \longrightarrow \mathbf{x}^*$

(2)

$$\begin{aligned} \mathbf{d}_\mu^k &= X_k[I - X_k\mathbf{A}^T(\mathbf{A}X_k^2\mathbf{A}^T)^{-1}\mathbf{A}X_k](-X_k\mathbf{c} + \mu e) \\ &= X_kP_k(-X_k\mathbf{c}) + \mu X_kP_ke \\ &= \mathbf{d}_\mathbf{x}^k + \underbrace{\mu X_kP_ke}_{\text{centering force}} \end{aligned}$$

(3) Polynomial-time proof, *i.e.*,
terminates in $O(\sqrt{n}L)$ iterations.

C. Gonzaga (1989) (Problems in Proof !!)
C. Roos/ J. Vial (1990)
- Total complexity $O(n^3L)$!

Dual Affine Scaling

$$
(D) \quad \begin{array}{ll} \max & \mathbf{b}^T\mathbf{w} \\ \text{s.t.} & \mathbf{A}^T\mathbf{w} + \mathbf{s} = \mathbf{c} \\ & \mathbf{s} \geq 0 \end{array}
$$

Given $(\mathbf{w}^k, \mathbf{s}^k)$ dual interior feasible, *i.e.*,

$$
\begin{aligned} \mathbf{A}^T\mathbf{w}^k + \mathbf{s}^k &= \mathbf{c} \\ \mathbf{s}^k &> 0 \end{aligned}
$$

Objective find $(\mathbf{d}_{\mathbf{w}}^k, \mathbf{d}_s^k)$ and $\beta_k > 0$ such that

$$
\begin{aligned} \mathbf{w}^{k+1} &= \mathbf{w}^k + \beta_k \mathbf{d}_{\mathbf{w}}^k \\ \mathbf{s}^{k+1} &= \mathbf{s}^k + \beta_k \mathbf{d}_{\mathbf{s}}^k \end{aligned}
$$

is still dual interior feasible, and

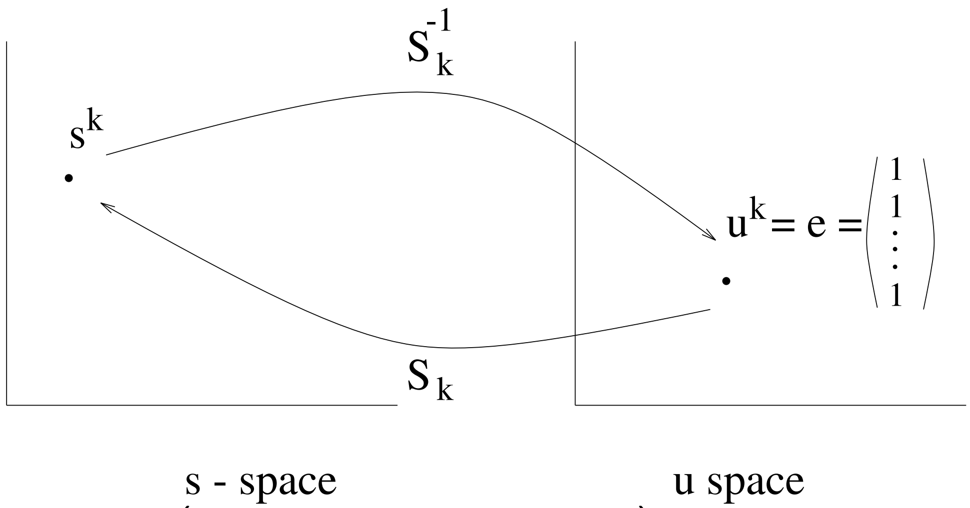
$$
\mathbf{b}^T\mathbf{w}^{k+1} \geq \mathbf{b}^T\mathbf{w}^k
$$

Observations:

(1) Scaling

$\mathbf{w}^k \in R^m$ no scaling needed

$\mathbf{s}^k > 0$ scale to $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$



$$S_k = \begin{pmatrix} s_1^k & & & \\ & s_2^k & & 0 \\ & 0 & \ddots & \\ & & & s_n^k \end{pmatrix} = \text{diag } (\mathbf{s}^k)$$

$\mathbf{u} = S_k^{-1}\mathbf{s} \quad \mathbf{d_u} = S_k^{-1}\mathbf{d_s}$

$\mathbf{s} = S_k\mathbf{u} \quad \mathbf{d_s} = S_k\mathbf{d_u}$

(2) Dual Feasibility

$$\underbrace{\mathbf{A}^T\mathbf{w}^{k+1} + \mathbf{s}^{k+1}}_{\mathbf{c}} = \mathbf{A}^T(\mathbf{w}^k + \beta_k \mathbf{d}_{\mathbf{w}}^k) + (\mathbf{s}^k + \beta_k \mathbf{d}_{\mathbf{s}}^k)$$

$$= \underbrace{(\mathbf{A}^T\mathbf{w}^k + \mathbf{s}^k)}_{\mathbf{c}}$$

$$+ \underbrace{\beta_k}_{>0}(\mathbf{A}^T\mathbf{d}_{\mathbf{w}}^k + \mathbf{d}_{\mathbf{s}}^k)$$

$$\Rightarrow \mathbf{A}^T\mathbf{d}_{\mathbf{w}}^k + \mathbf{d}_{\mathbf{s}}^k = 0 \text{ is required !}$$

$$\Leftrightarrow S_k^{-1}\mathbf{A}^T\mathbf{d}_{\mathbf{w}}^k + \underbrace{S_k^{-1}\mathbf{d}_{\mathbf{s}}^k}_{\mathbf{d}_{\mathbf{u}}^k} = 0$$

$$\Leftrightarrow \mathbf{A}S_k^{-1}(S_k^{-1}\mathbf{A}^T\mathbf{d}_{\mathbf{w}}^k + \mathbf{d}_{\mathbf{u}}^k) = 0$$

$$\Leftrightarrow (\mathbf{A}S_k^{-2}\mathbf{A}^T)\mathbf{d}_{\mathbf{w}}^k + \mathbf{A}S_k^{-1}\mathbf{d}_{\mathbf{u}}^k = 0$$

$$\Leftrightarrow \mathbf{d}_{\mathbf{w}}^k = -\underbrace{(\mathbf{A}S_k^{-2}\mathbf{A}^T)^{-1}\mathbf{A}S_k^{-1}}_{Q}\mathbf{d}_{\mathbf{u}}^k$$

(3) Increase Objective value

$$\mathbf{b}^T \mathbf{d}_{\mathbf{w}}^k = -\mathbf{b}^T Q \mathbf{d}_{\mathbf{u}}^k \geq 0$$

We can choose

$$\mathbf{d}_{\mathbf{u}}^k = -Q^T \mathbf{b}$$

then $\mathbf{b}^T \mathbf{d}_{\mathbf{w}}^k = \mathbf{b}^T Q Q^T \mathbf{b} = \|Q^T \mathbf{b}\|^2 \geq 0$!!
Thus

$$\begin{aligned}
\mathbf{d}_{\mathbf{w}}^k &= -Q \mathbf{d}_{\mathbf{u}}^k \\
&= Q Q^T \mathbf{b} \\
&= \underbrace{(\mathbf{A} S_k^{-2} \mathbf{A}^T)^{-1} \mathbf{A} S_k^{-1}}_{Q} \underbrace{S_k^{-1} \mathbf{A}^T (\mathbf{A} S_k^{-2} \mathbf{A}^T)^{-1}}_{Q^T} \mathbf{b} \\
&= (\mathbf{A} S_k^{-2} \mathbf{A}^T)^{-1} \mathbf{b}
\end{aligned}$$

and $\mathbf{d}_{\mathbf{s}}^k = -\mathbf{A}^T \mathbf{d}_{\mathbf{w}}^k = -\mathbf{A}^T (\mathbf{A} S_k^{-2} \mathbf{A}^T)^{-1} \mathbf{b}$

(4) Step-size β_k

$\mathbf{s}^{k+1} = \underbrace{\mathbf{s}^k}_{>0} + \beta_k \mathbf{d}_\mathbf{s}^k > 0$

(i) $\mathbf{d}_\mathbf{s}^k = 0$, problem (D) has a constant objective value and $(\mathbf{w}^k, \mathbf{s}^k)$ optimal.

(ii) $\mathbf{d}_\mathbf{s}^k \overset{>}{\neq} 0$, $\beta_k \in (0, \infty)$
problem (D) is unbounded

(iii) some $(\mathbf{d}_\mathbf{s}^k)_i < 0$

$$\beta_k = \min_i \{ \frac{\alpha s_i^k}{-(d_s^k)_i} | (d_s^k)_i < 0 \}$$

for $\alpha \in (0, 1)$

(5) Primal estimate

$$\mathbf{x}^k \triangleq -S_k^{-2}\mathbf{d}_{\mathbf{s}}^k$$

then

$$\begin{aligned}
\mathbf{A}\mathbf{x}^k &= -\mathbf{A}S_k^{-2}(-\mathbf{A}^T\mathbf{d}_{\mathbf{w}}^k) \\
&= \mathbf{A}S_k^{-2}\mathbf{A}^T\mathbf{d}_{\mathbf{w}}^k \\
&= (\mathbf{A}S_k^{-2}\mathbf{A}^T)(\mathbf{A}S_k^{-2}\mathbf{A}^T)^{-1}\mathbf{b} \\
&= \mathbf{b}
\end{aligned}$$

Hence $\mathbf{x}^k$ is a primal estimate,

once $\mathbf{x}^k \geq 0$, then $\mathbf{x}^k$ is primal feasible.

If $\mathbf{c}^T\mathbf{x}^k - \mathbf{b}^T\mathbf{w}^k = 0$, then

$$\begin{aligned}
\mathbf{x}^k &\leftarrow \mathbf{x}^* \\
\mathbf{w}^k &\leftarrow \mathbf{w}^* \\
\mathbf{s}^k &\leftarrow \mathbf{s}^*
\end{aligned}$$

(6) Dual Affine Scaling Algorithm

Step 1: Set $k = 0$ and find $(\mathbf{w}^0, \mathbf{s}^0)$ s.t.

$$\mathbf{A}^T\mathbf{w}^0 + \mathbf{s}^0 = \mathbf{c},\ \mathbf{s}^0 > 0$$

Step 2: Set $S_k = \text{diag}\,(\mathbf{s}^k)$

$$\text{Compute} \quad \mathbf{d}_{\mathbf{w}}^k = (\mathbf{A}S_k^{-2}\mathbf{A}^T)^{-1}\mathbf{b}$$

$$\mathbf{d}_{\mathbf{s}}^k = -\mathbf{A}^T\mathbf{d}_{\mathbf{w}}^k$$

Step 3: If $\mathbf{d}_{\mathbf{s}}^k = 0$, STOP! $\mathbf{w}^k \leftarrow \mathbf{w}^*,\ \mathbf{s}^k \leftarrow \mathbf{s}^*$

If $\mathbf{d}_{\mathbf{s}}^k \overset{>}{\neq} 0$, STOP ! (D) is unbounded

Step 4: Compute

$$\mathbf{x}^k = -S_k^{-2}\mathbf{d}_{\mathbf{s}}^k$$

If $\mathbf{x}^k \geq 0$ and $\mathbf{c}^T\mathbf{x}^k - \mathbf{b}^T\mathbf{w}^k \leq \varepsilon$

STOP !

$\mathbf{w}^k \leftarrow \mathbf{w}^*,\ \mathbf{s}^k \leftarrow \mathbf{x}^k \leftarrow \mathbf{x}^*$

Step 5: Compute

$$\beta_k = \min_i \{ \frac{\alpha s_i^k}{-(\mathbf{d}_s^k)_i} | (\mathbf{d}_s^k)_i < 0 \}$$

Step 6: $\mathbf{w}^{k+1} = \mathbf{w}^k + \beta_k \mathbf{d}_{\mathbf{w}}^k$

$\mathbf{s}^{k+1} = \mathbf{s}^k + \beta_k \mathbf{d}_{\mathbf{s}}^k$

Set $k \leftarrow k + 1$ Go to Step 2.

(7) Starting Dual Affine Scaling

Find $(\mathbf{w}^0, \mathbf{s}^0)$ s.t.

$$\mathbf{A}^T\mathbf{w}^0 + \mathbf{s}^0 = \mathbf{c}$$

$$\mathbf{s}^0 > 0$$

If $\mathbf{c} > 0$, then $\mathbf{w}^0 = 0,\ \mathbf{s}^0 = \mathbf{c}$ will do.

(Big - M Method)

Define $\mathbf{p} \in R^n,\ p_i = \begin{cases} 1 & \text{if } c_i \leq 0 \\ 0 & \text{if } c_i > 0 \end{cases}$

Consider, for a large $M > 0$,

(Big-M Problem)

$$\begin{aligned} \max \quad & \mathbf{b}^T\mathbf{w} + Mw^a \\ \text{s.t.} \quad & \mathbf{A}^T\mathbf{w} + \mathbf{p}w^a + \mathbf{s} = \mathbf{c} \\ & \mathbf{w}, w^a \ \text{ unrestricted} \\ & \mathbf{s} \geq 0 \end{aligned}$$

(a) (Big-M) is a standard LP with n constraints and $m + 1 + n$ variables.

(b) Define $\bar{c} = \max_i |c_i|$ and $\theta > 1$ then

$$\mathbf{w} = 0$$
$$w^a = -\theta\bar{\mathbf{c}}$$
$$\mathbf{s} = \mathbf{c} + \theta\bar{\mathbf{c}}\mathbf{p} > 0$$

is an intial interior feasible solution for problem (D).

(c) $(w^a)^0 = -\theta\bar{c} < 0$
Since $M > 0$ is large
$(w^a)^k \nearrow 0$ as $k \nearrow +\infty$
if $(w^a)^k$ does not approach or cross zero, then problem (D) is infeasible.

(8) Performance

(i) No polynomial-time proof.

(ii) Computational bottleneck

$$(AS_k^{-2}A^T)^{-1}.$$

(iii) Less sensitive to primal degeneracy and numerical errors, but sensitive to dual degeneracy.

(iv) Improves dual objective function very fast, but attaining primal feasibility is slow.

(9) Improvement

(i) Logarithmic Barrier Function Method

($\mu > 0$)

$$\begin{cases} \max & \mathbf{b}^T\mathbf{w} + \mu \sum_{j=1}^{n} \ln[c_j - \mathbf{A}_j^T \mathbf{w}] \\ \text{s.t.} & \mathbf{A}^T\mathbf{w} < \mathbf{c} \end{cases}$$

$$\triangle \mathbf{w} = \frac{1}{\mu}\underbrace{(\mathbf{A}S_K^{-2}\mathbf{A}^T)^{-1}\mathbf{b}}_{\mathbf{d}_{\mathbf{w}}^k} \quad \underbrace{-(\mathbf{A}S_K^{-2}\mathbf{A}^T)\mathbf{A}S_k^{-1}e}_{\text{centering force}}$$

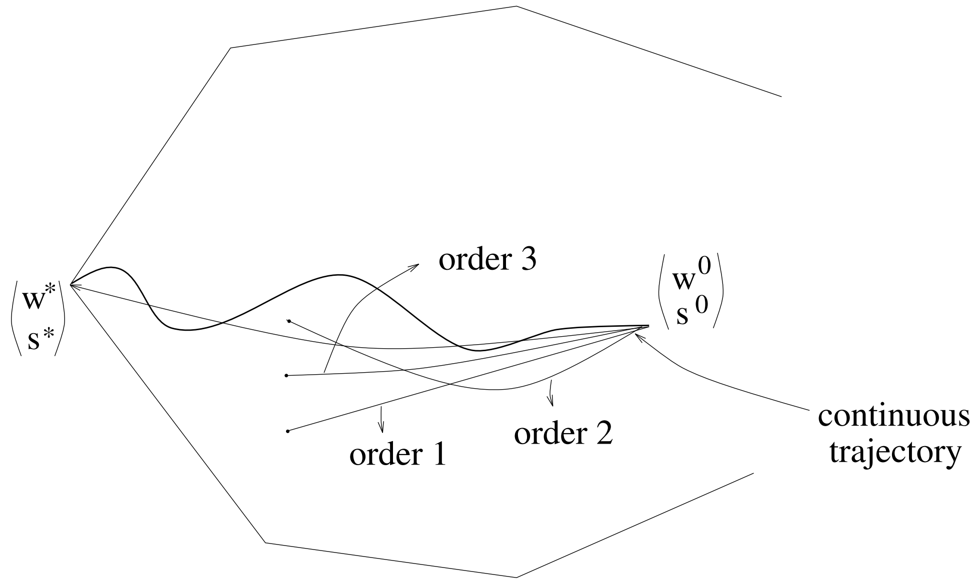
as $\mu \to 0, \quad \mathbf{w}^k(\mu) \to \mathbf{w}^*$

J. Renegar $O(n^{3.5}L)$

P. Vaidya $O(n^3L)$

C. Roos/ J. Vial $O(n^3L)$

(ii) Power Series Method



$$\text{O.D.E.} \begin{cases} \frac{d\ \mathbf{w}(\beta)}{d\ \beta} & = \lim_{\ \beta_k \to 0} \frac{\mathbf{w}^{k+1} - \mathbf{w}^k}{\beta_k} \\ & = [\mathbf{A} S(\beta)^{-2} \mathbf{A}^T]^{-1} \mathbf{b} \\ \frac{d\ \mathbf{s}(\beta)}{d\ \beta} & = -\mathbf{A}^T \frac{d\ \mathbf{w}(\beta)}{d\ \beta} \end{cases}$$

Initial condition

$$\mathbf{w}(0) = \mathbf{w}^0,\ \mathbf{s}(0) = \mathbf{s}^0$$

where

$$S(\beta) = \text{diag}(\mathbf{s}^0 + \beta \mathbf{d}_s)$$

Power-Series Expansion:

$$\mathbf{w}(\beta) = \mathbf{w}^0 + \sum_{i=1}^{\infty} \beta^j \ [\frac{1}{j!}] \ [\frac{d^j \ \mathbf{w}(\beta)}{d \ \beta^j}]_{\beta=0}$$

$$\mathbf{s}(\beta) = \mathbf{s}^0 + \sum_{i=1}^{\infty} \beta^j \ [\frac{1}{j!}][\frac{d^j \ \mathbf{s}(\beta)}{d \ \beta^j}]_{\beta=0}$$

(a) As long as $[\frac{d^j \ \mathbf{w}(\beta)}{d \ \beta^j}]_{\beta=0}$ and $[\frac{d^j \ \mathbf{s}(\beta)}{d \ \beta^j}]_{\beta=0}, \ j = 1, 2, \cdots n$ are known, $\mathbf{w}(\beta), \ \mathbf{s}(\beta)$ are known.

(b) Dual Affine Scaling is the case of first-order approximation

$$\mathbf{w}(\beta) = \mathbf{w}^0 + \beta[\frac{d \ \mathbf{w}(\beta)}{d \ \beta}]_{\beta=0}$$

$$\mathbf{s}(\beta) = \mathbf{s}^0 + \beta[\frac{d \ \mathbf{s}(\beta)}{d \ \beta}]_{\beta=0}$$

(c) A power-series approximation of order 4 or 5 cuts total # of iterations by 1/2.