LECTURE 6: INTERIOR POINT METHOD

1. Motivation
2. Basic concepts
3. Primal affine scaling algorithm
4. Dual affine scaling algorithm

Motivation

- Simplex method works well in general, but suffers from exponential-time computational complexity.
- Klee-Minty example shows simplex method may have to visit every vertex to reach the optimal one.
- Total complexity of an iterative algorithm

  = # of iterations x # of operations in each iteration
- Simplex method

  - Simple operations: Only check adjacent extreme points

  - May take many iterations: Klee-Minty example

Question**:** **any fix?**

Complexity of the simplex method

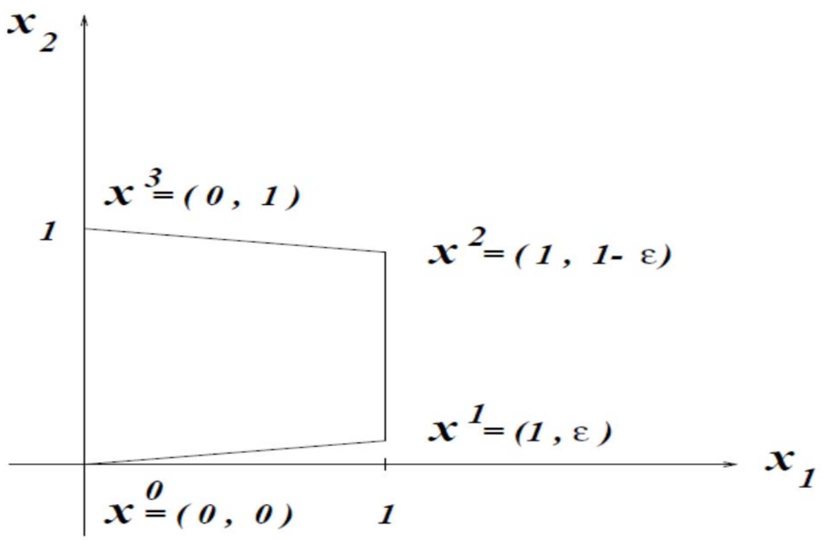
- Total # of elementary operations
  = (# of elementary operations at each iteration) $\times$ (# of iterations).
- # of elementary operations at each iteration of the revised simplex method $O(mn)$.
- From practical experience, the simplex method takes about (αm) iterations where $e^{\alpha} < \log_2(2 + n/m)$. Hence it is of $O(m^2 n)$.
- From the worst-case analysis, Klee and Minty [1972] showed a class of examples (in the d-dimensional space) which $2^d - 1$ iterations for the simplex method.

Worst case performance of the simplex method

Klee-Minty Example:

- Victor Klee, George J. Minty, "How good is the simplex algorithm?'' in (O. Shisha edited) Inequalities, Vol. III (1972), pp. 159-175.

$$
\begin{aligned}
(2 \text{ dim}) \quad & \min \quad -x_2 \\
& \text{s. t.} \quad x_1 \geq 0 \\
& \qquad x_1 \leq 1 \\
& \qquad x_2 \geq \epsilon x_1 \qquad (0 < \epsilon < \tfrac{1}{2}) \\
& \qquad x_2 \leq 1 - \epsilon x_1 \\
& \qquad x_1, x_2 \geq 0
\end{aligned}
$$

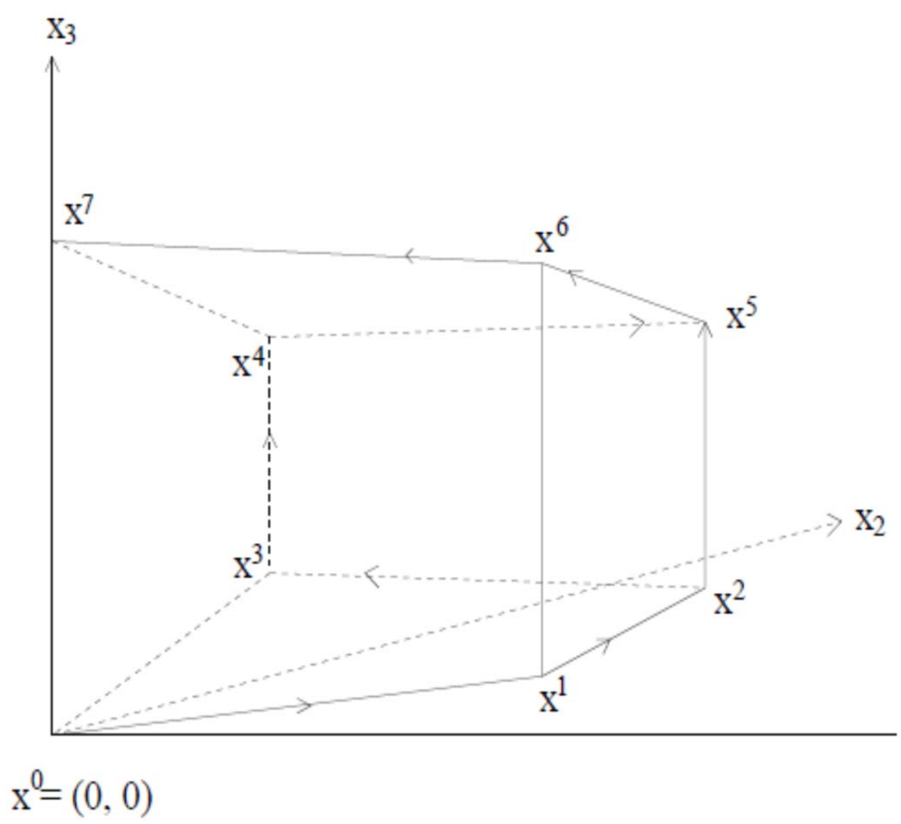


$$\mathbf{x}^0 \rightarrow \mathbf{x}^1 \rightarrow \mathbf{x}^1 \rightarrow \mathbf{x}^3 \text{ (optimal)}$$

$$2^2 - 1 = 3 \text{ iterations}$$

Klee-Minty Example

$$
\begin{aligned}
(3\ \text{dim})\quad & \min \quad -x_3 \\
& \text{s. t.} \quad x_1 \geq 0 \\
& \qquad x_1 \leq 1 \\
& \qquad x_2 \geq \epsilon x_1 \\
& \qquad x_2 \leq 1 - \epsilon x_1 \\
& \qquad x_3 \geq \epsilon x_2 \\
& \qquad x_3 \leq 1 - \epsilon x_2 \\
& \qquad x_1, x_2, x_3 \geq 0
\end{aligned}
$$

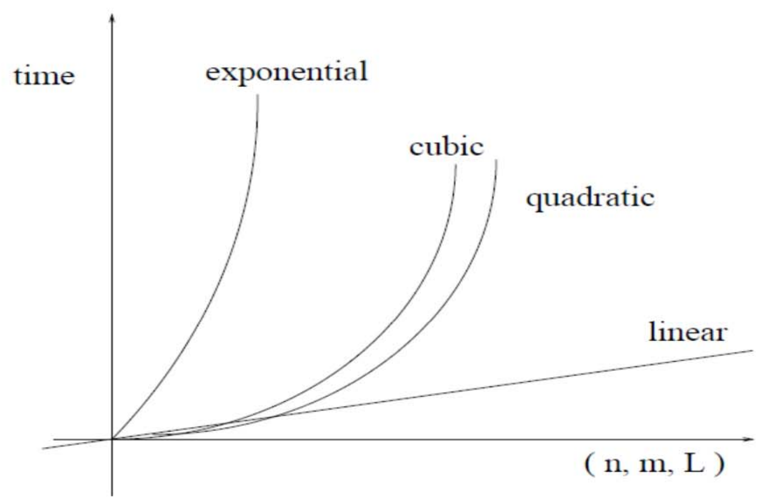


$2^3 - 1 = 7$ iterations

Klee-Minty Example

$$
\begin{aligned}
(d \ \text{dim}) \quad & \min \quad -x_d \\
& \text{s. t.} \quad x_1 \geq 0 \\
& \qquad x_1 \leq 1 \\
& \qquad x_2 \geq \epsilon x_1 \\
& \qquad x_2 \leq 1 - \epsilon x_1 \\
& \qquad \vdots \\
& \qquad x_d \geq \epsilon x_{d-1} \\
& \qquad x_d \leq 1 - \epsilon x_{d-1} \\
& \qquad x_i \geq 0
\end{aligned}
$$

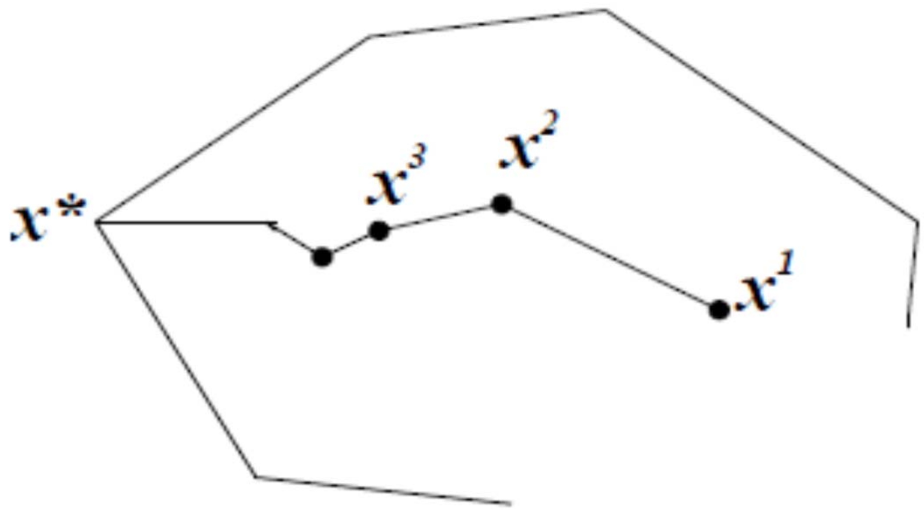
Hence, in theory, the simplex method is not a polynomial-time algorithm. It is an *exponential time* algorithm!



$2^d - 1$ iterations

Karmarkar's (interior point) approach

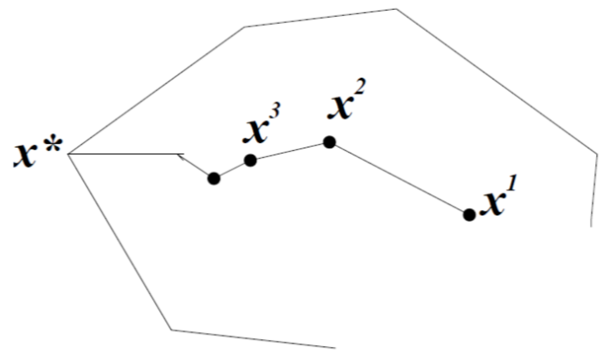
- Basic idea: approach optimal solutions from the interior of the feasible domain



- Take more complicated operations in each iteration to find a better moving direction
- Require much fewer iterations

General scheme of an interior point method

- An iterative method that moves in the interior of the feasible domain



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.

Interior movement (iteration)

- Given a current interior feasible solution $\mathbf{x}^k$, we have

$$\begin{aligned} \mathbf{A}\mathbf{x}^k &= \mathbf{b} \\ \mathbf{x}^k &> 0 \end{aligned}$$

An interior movement has a general format

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

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

Key knowledge

- 1. Who is in the interior?
  - Initial solution
- 2. How do we know a current solution is optimal?
  - Optimality condition
- 3. How to move to a new solution?
  - Which direction to move? (good feasible direction)
  - How far to go? (step-length)

Q1 - Who is in the interior?

- Standard for LP

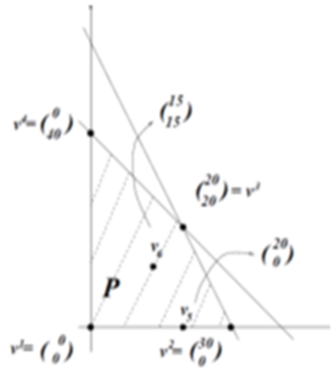
$$
\text{(LP)} \qquad \begin{aligned} & \text{Min } \mathbf{c}^T\mathbf{x} \\ \text{s. t. } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{aligned}
$$

- Who is at the vertex?
- Who is on the edge?
- Who is on the boundary?
- Who is in the interior?

What have learned before

Learning from example

$$\begin{aligned}\text{Minimize}\quad & x_1 - 2x_2 \\ \text{subject to}\quad & x_1 + x_2 + x_3 \quad\quad = 40 \\ & 2x_1 + x_2 \quad\quad + x_4 = 60 \\ & x_1, x_2, x_3, x_4 \geq 0.\end{aligned}$$



What's special?

- Vertices

$$v^1 = \begin{pmatrix} 0 \\ 0 \\ 40 \\ 60 \end{pmatrix}, v^2 = \begin{pmatrix} 30 \\ 0 \\ 10 \\ 0 \end{pmatrix}, v^3 = \begin{pmatrix} 20 \\ 20 \\ 0 \\ 0 \end{pmatrix}, v^4 = \begin{pmatrix} 0 \\ 40 \\ 0 \\ 20 \end{pmatrix}.$$

- Edge

$$v^5 = \begin{pmatrix} 20 \\ 0 \\ 20 \\ 20 \end{pmatrix} \leftarrow \text{one zero } x_i$$

Interior

$$v^6 = \begin{pmatrix} 15 \\ 15 \\ 10 \\ 15 \end{pmatrix} \leftarrow \text{no zero } x_i$$

$$n = 4,\ m = 2,\ n - m = 2$$

Who is in the interior?

- Two criteria for a point **x** to be an interior feasible solution:

  1. A**x** = **b** (every linear constraint is satisfied)
  2. **x** > **0** (every component is positive)

- Comments:
  1. On a hyperplane $H = \{\mathbf{x} \in \mathbf{R}^n | \mathbf{a}^T\mathbf{x} = \beta\}$,
     every point is interior relative to H.
  2. For the first orthant K = $\{\mathbf{x} \in \mathbf{R}^n |\ \mathbf{x} \geq \mathbf{0}\}$,
     only those **x** > 0 are interior relative to K.

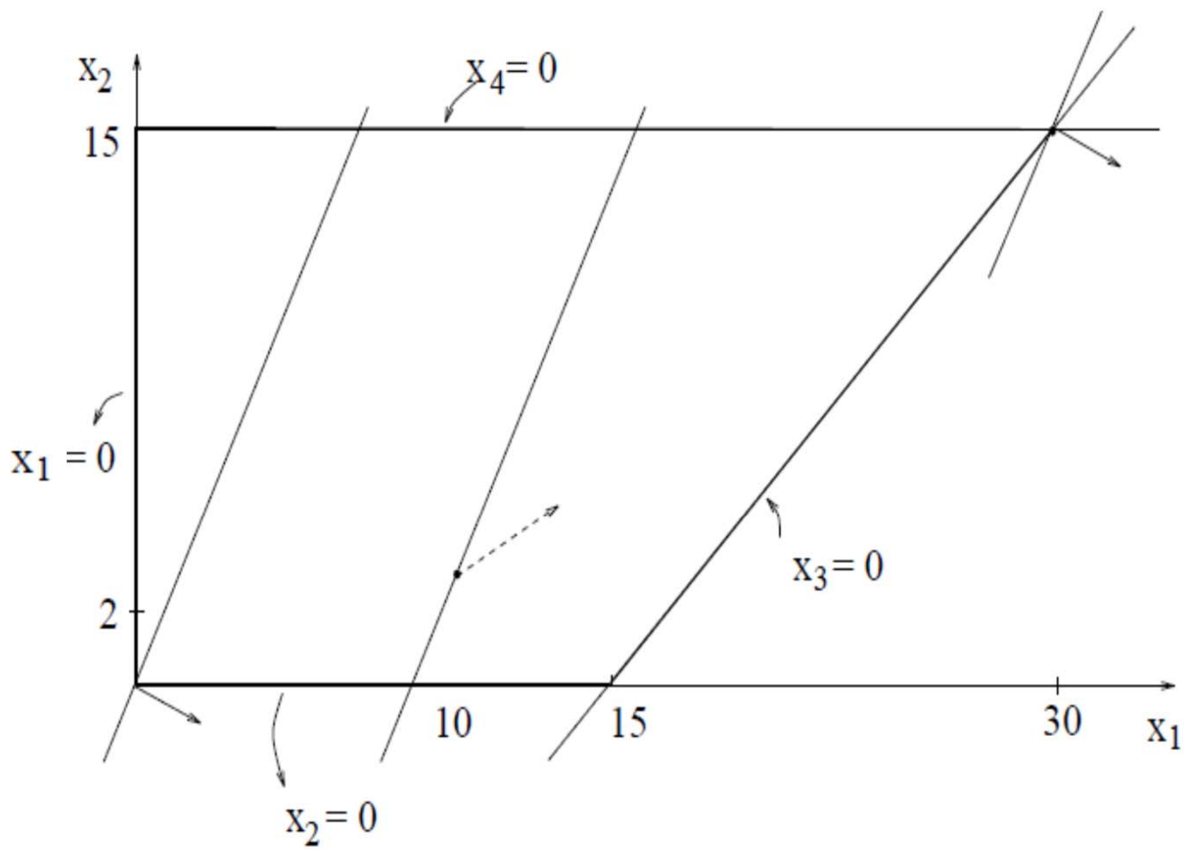
Example

$$\min \quad -2x_1 + x_2$$

$$\text{s.t.} \quad x_1 - x_2 \leq 15$$

$$x_2 \leq 15$$

$$x_1, x_2 \geq 0$$



How to find an initial interior solution?

- Like the simplex method, we have

  - Big M method
  - Two-phase method

  (to be discussed later!)

Key knowledge

- 1. Who is in the interior?
  - Initial solution

- 2. How do we know a current solution is optimal?
  - Optimality condition

- 3. How to move to a new solution?
  - Which direction to move? (good feasible direction)
  - How far to go? (step-length)

Q2 - How do we know a current solution is optimal?

- Basic concept of optimality:

  A current feasible solution is optimal if and only if

  "no feasible direction at this point is a good direction."

- In other words, "every feasible direction is not a good direction to move!"

Feasible direction

- In an interior-point method, a feasible direction at a current solution is a direction that allows it to take a small movement while staying to be interior feasible.
- Observations:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{d}_{\mathbf{x}}^k \qquad \begin{aligned} \mathbf{A}\mathbf{x}^k &= \mathbf{b} \\ \mathbf{x}^k &> 0 \end{aligned}$$

- There is no problem to stay interior if the step-length is small enough.
- To maintain feasibility, we need

$$\begin{aligned} \underline{\mathbf{A}\mathbf{x}^{k+1}} &= \mathbf{b} \\ \mathbf{A}\mathbf{x}^k + \alpha \mathbf{A}\mathbf{d}_{\mathbf{x}}^k &= \mathbf{b} \end{aligned} \implies \mathbf{A}\mathbf{d}_{\mathbf{x}}^k = 0$$

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

Good direction

- In an interior-point method, a good direction at a current solution is a direction that leads it to a new solution with a lower objective value.

- Observations:

$$\begin{aligned} \underline{\mathbf{c}^T\mathbf{x}^{k+1}} &\leq \mathbf{c}^T\mathbf{x}^k \\ \mathbf{c}^T\mathbf{x}^k + \alpha\mathbf{c}^T\mathbf{d}_{\mathbf{x}}^k &\leq \mathbf{c}^T\mathbf{x}^k \end{aligned} \implies \mathbf{c}^T\mathbf{d}_{\mathbf{x}}^k \leq 0$$

Optimality check

- Principle:

  "no feasible direction at this point is a good direction."

  - At a current solution, we check that

    No $\mathbf{d}_{\mathbf{x}}^{k} \in R^{n}$ with $\mathbf{A}\mathbf{d}_{\mathbf{x}}^{k} = 0$

    can make

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

Key knowledge

- 1. Who is in the interior?
  - Initial solution

- 2. How do we know a current solution is optimal?
  - Optimality condition

- 3. How to move to a new solution?
  - Which direction to move? (good feasible direction)
  - How far to go? (step-length)

Q3 – How to move to a new solution?

1. Which direction to move?

   - a good, feasible direction

     "Good" requires $\mathbf{c}^T \mathbf{d}_{\mathbf{x}}^k \leq 0$

     "Feasible" requires

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

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

Question: any suggestion?

A good feasible direction

- Reduce the objective value

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

<u>**Candidate**</u>: $\mathbf{d}_{\mathbf{x}}^k = -\mathbf{c}$

(negative gradient)

(Steepest descent)

- Maintain feasibility

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

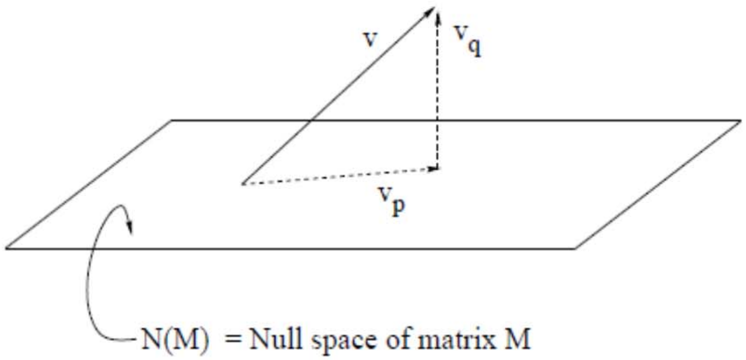
<u>**Candidate**</u>: projected negative gradient

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

Projection mapping

- A projection mapping projects the negative gradient vector **–c** into the null space of matrix A

Formula for projection: $v = v_p + v_q$



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

$$\begin{aligned} \mathcal{N}(M) &= \{ \mathbf{x} \mid M\mathbf{x} = 0 \} \\ v_p &= [\, I - M^T(MM^T)^{-1}M \,]v \\ v_q &= M^T(MM^T)^{-1}Mv \end{aligned}$$

Q3 – How to move to a new solution?

2. How far to go?

- To satisfy every linear constraint

Since

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

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

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

the step-length can be real number.

- To stay to be an interior solution, we need

$$\mathbf{x}^{k+1} > 0.$$

How to choose step-length?

- One easy approach

  - in order to keep

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{d}_{\mathbf{x}}^k > 0$$

  we may use the "minimum ratio test" to determine the step-length.

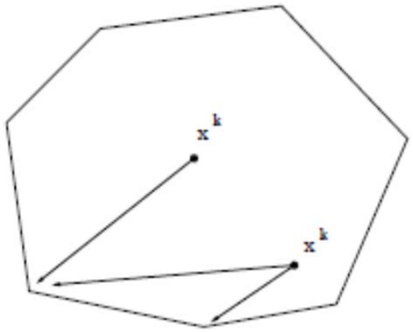
Observation:

- when $\mathbf{x}^k$ is close to the boundary, the step-length may be very small.

Question: then what?

Observations

- If a current solution is near the center of the feasible domain (polyhedral set), in average we can make a decently long move.



- If a current solution is not near the center, we need to re-scale its coordinates to transform it to become “near the center".

Question: but how?

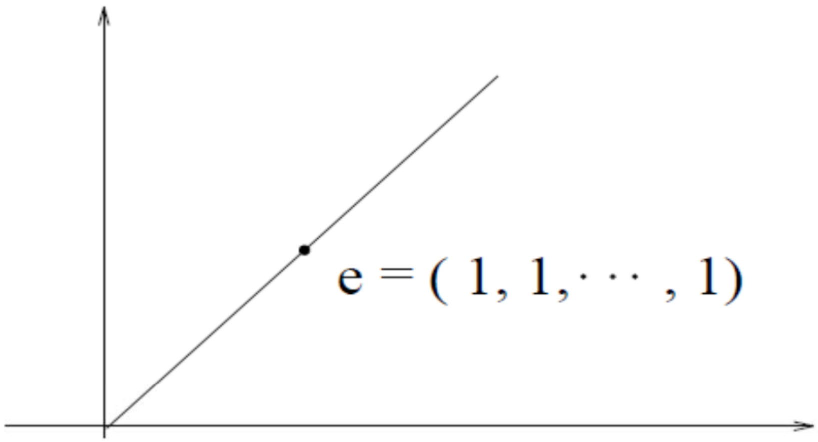
Where is the center?

- We need to know where is the "center" of the

  non-negative/first orthant $\{\mathbf{x} \in R^n \mid \mathbf{x} \geq 0\}$.

  -Concept of equal distance

  to the boundary



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

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

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

Question: If not, what to do?

Concept of scaling

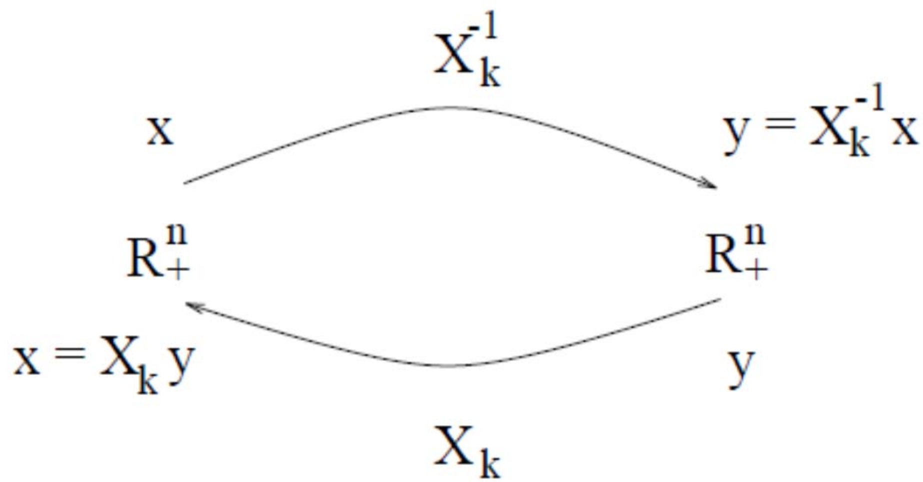
- Scale $\mathbf{x}^k$ **to be** e
- Define a diagonal scaling matrix

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

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

Transformation – affine scaling

- Affine scaling transformation



- The transformation is

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

1. one-to-one
2. onto
3. Invertible
4. boundary to boundary
5. interior to interior

Transformed LP

$$\mathbf{x} = \mathbf{X_k}\, \mathbf{y}$$

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

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

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

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

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

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

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

Step-length in the transformed space

- Minimum ratio test in the y-space

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 $\alpha \in (0, 1)$

Property 1

- Iteration in the x-space

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

Property 2

- Feasible direction in x-space

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

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

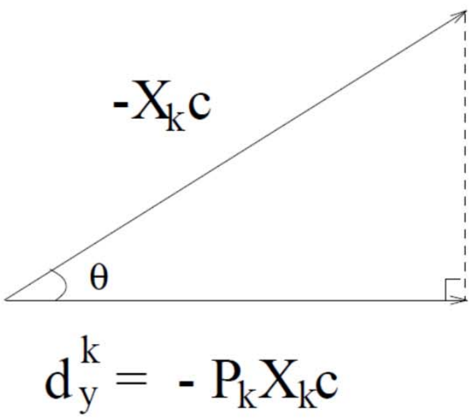
$$\therefore \quad \mathbf{A} X_k \mathbf{d}_{\mathbf{y}}^k = 0 \text{ and } \quad \mathbf{A}\mathbf{d}_{\mathbf{x}}^k = 0$$

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

Property 3

- Good direction in x-space

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



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.

Property 4

- Optimality check (Lemma 7.2)

For $\mathbf{x}^k \in P^0 = \{\mathbf{x} \in R^n \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} > 0\}$

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.*

$$X_k \mathbf{c} \in \text{row space of } (\mathbf{A}X_k)$$

$$\Rightarrow \exists\, u^k \text{ s.t. } (\mathbf{A}X_k)^T u^k = X_k \mathbf{c}$$

$$\text{or } (u^k)^T \mathbf{A} X_k = \mathbf{c}^T X_k$$

$$\Rightarrow (u^k)^T A = \mathbf{c}^T$$

For any $\mathbf{x} \in P$

$$\mathbf{c}^T \mathbf{x} = (u^k)^T \mathbf{A}\mathbf{x} = \underline{(u^k)^T \mathbf{b}} \text{ (constant)}$$

$\therefore$ Any feasible solution is optimal !!

In particular, $\mathbf{x}^k$ is optimal !

Property 5

- Well-defined iteration sequence (Lemma 7.3)

From properties 3 and 4, if the standard form LP is bounded below and $\mathbf{c}^T\mathbf{x}$ is not a constant, then the sequence $\{\mathbf{c}^T\mathbf{x}^k \mid k = 1, 2, \cdots\}$ is well-defined and strictly decreasing.

Property 6

- Dual estimate, reduced cost and stopping rule

We may define

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

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

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.*,

$$\text{Therefore, if } \underbrace{\mathbf{r}^k \geq 0 \ \text{ and } \ e^T X_k \mathbf{r}^k = 0}_{\text{(Stopping rule)} \nearrow}$$

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

Property 7

- Moving direction and reduced cost

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

Primal affine scaling algorithm

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

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

<u>Step2</u> 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,

<u>Step3</u> 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,

<u>Step4</u> Find

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

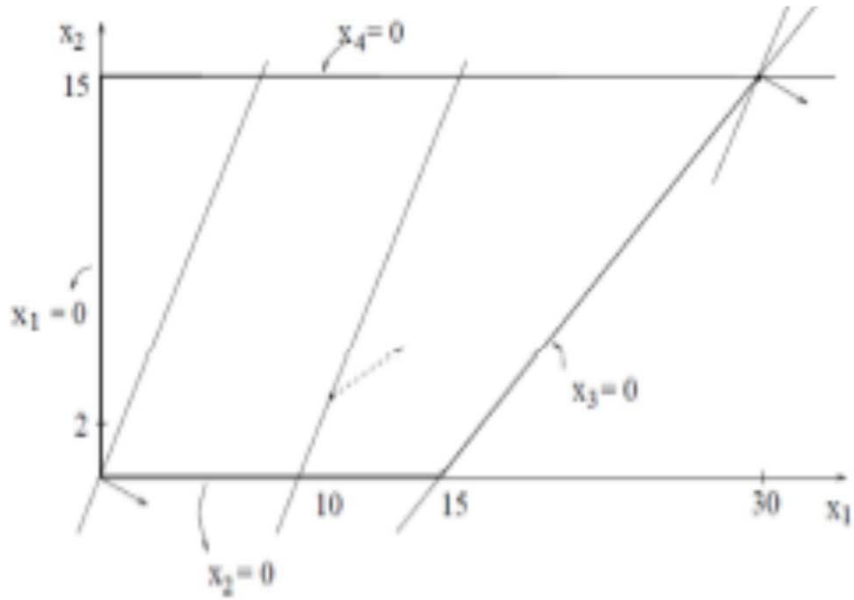
$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k X_k \mathbf{d}_\mathbf{y}^k$$

$$k \leftarrow k + 1$$

Go to Step 2.

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

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

$$\mathbf{x}^0 = \begin{pmatrix} 10 \\ 2 \\ 7 \\ 13 \end{pmatrix} \text{ is feasible} \qquad \mathbf{X}_0 = \begin{bmatrix} 10 & & & 0 \\ & 2 & & \\ 0 & & 7 & \\ & & & 13 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 15 \\ 15 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

Example

$$\mathbf{X}_0 = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 13 \end{bmatrix} \quad \text{and} \quad \mathbf{w}^0 = (\mathbf{A}\mathbf{X}_0^2\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{X}_0^2\mathbf{c} = [-1.33353 \quad -0.00771]^T$$

Moreover,

$$\mathbf{r}^0 = \mathbf{c} - \mathbf{A}^T\mathbf{w}^0 = [-0.66647 \quad -0.32582 \quad 1.33535 \quad -0.00771]^T$$

Since some components of $\mathbf{r}^0$ are negative and $\mathbf{e}^T\mathbf{X}_0\mathbf{r}^0 = 2.1187$, we know that the current solution is nonoptimal. Therefore we proceed to synthesize the direction of translation with

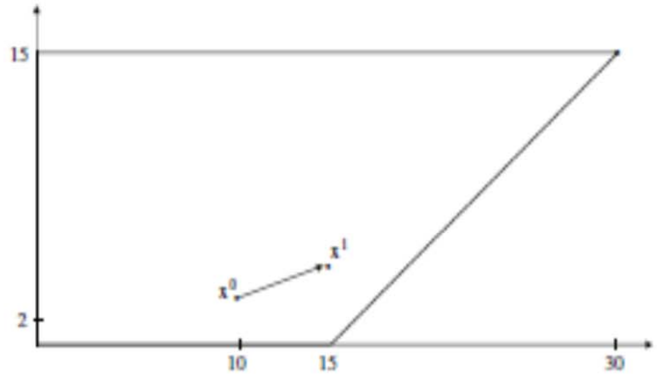
$$\mathbf{d}_y^0 = -\mathbf{X}_0\mathbf{r}^0 = [6.6647 \quad 0.6516 \quad -9.3475 \quad 0.1002]^T$$

Suppose that $\alpha = 0.99$ is chosen, then the step-length

$$\alpha_0 = \frac{0.99}{9.3475} = 0.1059$$

Therefore, the new solution is

$$\mathbf{x}^1 = \mathbf{x}^0 + \alpha_0\mathbf{X}_0\mathbf{d}_y^0 = [17.06822 \quad 2.13822 \quad 0.07000 \quad 12.86178]^T$$



Notice that the objective function value has been improved from -18 to -31.99822. The reader may continue the iterations further and verify that the iterative process converges to the optimal solution $\mathbf{x}^* = [30 \quad 15 \quad 0 \quad 0]^T$ with optimal value -45.

How to find an initial interior feasible solution?

- Big-M method

Idea: add an artificial variable with a big penalty

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

$$\text{(big-M)}\begin{cases} \min & \mathbf{c}^T\mathbf{x} + Mx^a \\ \text{s.t.} & \mathbf{A}\mathbf{x} + (\mathbf{b} - \mathbf{A}e)x^a = \mathbf{b} \\ & \mathbf{x} \geq 0,\ x^a \geq 0 \end{cases}$$

- Objective

to make $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ be feasible, i.e., $\mathbf{A}e = \mathbf{b}$?

Properties of (big-M) problem

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

(2) *e* is an interior feasible solution of (big-M).

(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).

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 of (Phase-I) problem

(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.

Facts of the primal affine scaling algorithm

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

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

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.

More facts

(5) In practice, VMF reported

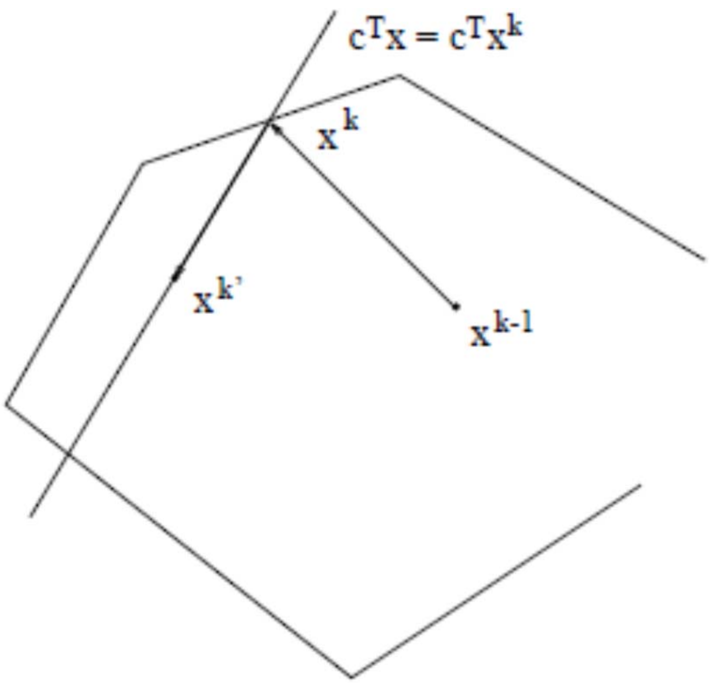
	# iterations
Simplex	$0.7159\ m^{0.9522}\ n^{0.3109}$
Affine Scaling	$7.3385\ m^{-0.0187}\ n^{0.1694}$

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

(7) May be sensitive to primal degeneracy.

Improving performance – potential push

- Idea: (Potential push method)
  - Stay away from the boundary by adding a potential push.



Consider

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

Use

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

Improving performance – logarithmic barrier

- Idea: (Logarithmic barrier function method)

Consider

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

Properties:

(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 algorithm

- Affine scaling method applied to the dual LP

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

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

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

<u>**Objective**</u> 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

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

Key knowledge

- Dual scaling (centering)
- Dual feasible direction
- Dual good direction – increase the dual objective value
- Dual step-length
- Primal estimate for stopping rule

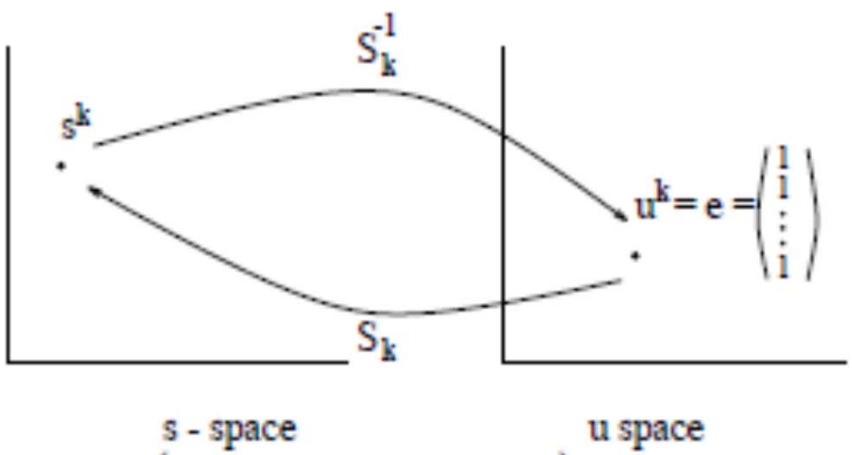
Observation 1

- Dual scaling (centering)

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

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

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



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

Observation 2

- Dual feasibility (feasible direction)

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

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

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

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

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

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

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

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

Observation 3

- Increase dual objective function (good direction)

$$\mathbf{b}^T\mathbf{w}^{k+1} = \mathbf{b}^T\mathbf{w}^k + \beta_k \mathbf{b}^T\mathbf{d}_\mathbf{w}^k \geq \mathbf{b}^T\mathbf{w}^k$$

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

We can choose

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

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

Thus

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

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

Observation 4

- Dual step-length

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

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

Observation 5

- Primal estimate

We define

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

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 feasibl

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

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

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 $S_k = \text{diag}\,(\mathbf{s}^k)$

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

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

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 !

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

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.

Find an initial interior feasible solution

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)

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

Properties of (big-M) problem

(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.

Performance of dual affine scaling

- No polynomial-time proof !
- Computational bottleneck

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

- Less sensitive to primal degeneracy and numerical errors, but sensitive to dual degeneracy.
- Improves dual objective value very fast, but attains primal feasibility slowly.

Improving performance

1. Logarithmic barrier function method

$(\mu > 0)$

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

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

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

Polynomial-time proof

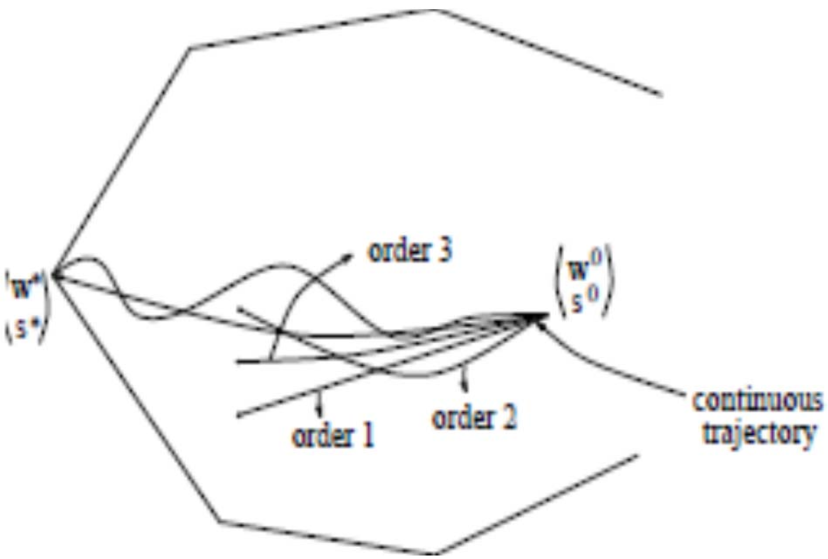
J. Renegar $O(n^{3.5}L)$

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

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

Improving performance

- Power series method
  - Basic idea: following a higher order trajectory



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

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.