ORF 522

Linear Programming and Convex Analysis

Interior Point Methods

Marco Cuturi

Reminder

- The ellipsoid method.
- A method that tries to exhibit a point in a polyhedron or state its emptiness.
- 2 cases:
  - some assumptions: the bounded/full-dimensional case
  - the general case
- Ellipsoid method for optimization is either
  - a direct application to the primal/dual feasible problem
  - a sliding ellipsoid method
- Poor performance in practice

Today

- A brief description of 3 interior point methods.
  - Affine scaling algorithm (goes back to '67, Dikin)
  - Potential reduction algorithm (Karmarkar, '84)
  - Central Path following algorithm, just an intro (traces back to Frisch, '56)

Then...

Breakthrough in Problem Solving

By JAMES GLEICK

A 28-year-old mathematician at A.T.&T. Bell Laboratories has made a startling theoretical breakthrough in the solving of systems of equations that often grow too vast and complex for the most powerful computers.

The discovery, which is to be formally published next month, is already circulating rapidly through the mathematical world. It has also set off a deluge of inquiries from brokerage houses, oil companies and airlines, industries with millions of dollars at stake in problems known as linear programming.

Faster Solutions Seen

These problems are fiendishly complicated systems, often with thousands of variables. They arise in a variety of commercial and government applications, ranging from allocating time on a communications satellite to routing millions of telephone calls over long distances, or whenever a limited, expensive resource must be spread most efficiently among competing users. And investment companies use them in creating portfolios with the best mix of stocks and bonds.

The Bell Labs mathematician, Dr. Narendra Karmarkar, has devised a radically new procedure that may speed the routine handling of such problems by businesses and Government agencies and also make it possible to tackle problems that are now far out of reach.

"This is a path-breaking result," said Dr. Ronald L. Graham, director of mathematical sciences for Bell Labs in Murray Hill, N.J.

"Science has its moments of great progress, and this may well be one of them."

Because problems in linear programming can have billions or more possible answers, even high-speed computers cannot check every one. So computers must use a special procedure, an algorithm, to examine as few answers as possible before finding the best one — typically the one that minimizes cost or maximizes efficiency.

A procedure devised in 1947, the simplex method, is now used for such problems,

Continued on Page A19, Column 1

THE NEW YORK TIMES, November 19, 1984

Karmarkar at Bell Labs: an equation to find a new way through the maze

Folding the Perfect Corner

A young Bell scientist makes a major math breakthrough

Every day 1,200 American Airlines jets crisscross the U.S., Mexico, Canada and the Caribbean, stopping in 110 cities and bearing over 80,000 passengers. More than 4,000 pilots, copilots, flight personnel, maintenance workers and baggage carriers are shuffled among the flights; a total of 3.6 million gal. of high-octane fuel is burned. Nuts, bolts, altimeters, landing gears and the like must be checked at each destination. And while performing these scheduling gymnastics, the company must keep a close eye on costs, projected revenue and profits.

Like American Airlines, thousands of companies must routinely untangle the myriad variables that complicate the efficient distribution of their resources. Solving such monstrous problems requires the use of an abstruse branch of mathematics known as linear programming. It is the kind of math that has frustrated theoreticians for years, and even the fastest and most powerful computers have had great difficulty juggling the bits and pieces of data. Now Narendra Karmarkar, a 28-year-old Indian-born mathematician at Bell Laboratories in Murray Hill, N.J., after only a years' work has cracked the puzzle of linear programming by devising a new algorithm, a step-by-step mathematical formula. He has translated the procedure into a program that should allow computers to track a greater combination of tasks than ever before and in a fraction of the time.

Unlike most advances in theoretical mathematics, Karmarkar's work will have an immediate and major impact on the real world. "Breakthrough is one of the most abused words in science," says Ronald Graham, director of mathematical sciences at Bell Labs. "But this is one situation where it is truly appropriate."

Before the Karmarkar method, linear equations could be solved only in a cumbersome fashion, ironically known as the simplex method, devised by Mathematician George Dantzig in 1947. Problems are conceived of as giant geodesic domes with thousands of sides. Each corner of a facet on the dome

TIME MAGAZINE, December 3, 1984

Affine Scaling

Main ideas

- We consider a primal (standard) dual (canonical, free variables) pair,

$$\begin{array}{ll}\text{minimize} & \mathbf{c}^T\mathbf{x}\\ \text{subject to} & A\mathbf{x}=\mathbf{b}\\ & \mathbf{x}\geq 0\end{array}, \qquad \begin{array}{ll}\text{maximize} & \mathbf{b}^T\mu\\ \text{subject to} & A^T\mu\leq\mathbf{c}\end{array}$$

- consider the feasible set $P=\{\mathbf{x}\in\mathbf{R}^n \mid A\mathbf{x}=\mathbf{b},\ \mathbf{x}\geq 0\}$
- its interior $\overset{\circ}{P}=\{\mathbf{x}\in\mathbf{R}^n \mid A\mathbf{x}=\mathbf{b},\ \mathbf{x}>0\}$ is called the set of interior points.
- The algorithm
  - **starts** from a point $\mathbf{x}_0$ in $\overset{\circ}{P}$ and defines $E_0=E(\mathbf{x}_0,D)\subset\overset{\circ}{P}$.
  - **loops** starting with $t=0$
    - ▷ finds the optimum $\mathbf{x}_{t+1}$ of $\mathbf{c}^T\mathbf{x}$ in E_t (analytical solution)
    - ▷ defines a new ellipsoid $E_{t+1}=E(\mathbf{x}_{t+1},D_{t+1})$ still inscribed in $\overset{\circ}{P}$.
    - ▷ $t\leftarrow t+1$
  - **controls convergence** checking dual gap.

Preliminary results: Inscribed ellipsoid

- an interesting prior lemma

  Lemma 1. *Let* $\beta \in (0,1)$ *be a scalar and* $\mathbf{y} \in \mathbf{R}^n$ *satisfy* $\mathbf{y} > \mathbf{0}$*. if*

$$S = \left\{ \mathbf{x} \in \mathbf{R}^n \mid \sum_{i=1}^{n} \frac{(x_i - \boldsymbol{y_i})^2}{\boldsymbol{y_i^2}} \leq \beta^2 \right\},$$

  then $\mathbf{x} > \mathbf{0}$ *for* ***every*** $\mathbf{x} \in S$*.*

  Proof. Let $\mathbf{x} \in S$.

  - then for all i, $(x_i - y_i)^2 \leq \beta^2 y_i^2 < y_i^2$,
  - hence $|x_i - y_i| < y_i$, in particular $-x_i + y_i < y_i$ thus $x_i > 0$

  ■

Preliminary results: Linear objective on ellipsoids

- For a vector $\mathbf{y} \in \mathbf{R}^n$ such that $\mathbf{y} > 0$, and $A\mathbf{y} = \mathbf{b}$, let $Y = \mathbf{diag}(y_1, y_2, \cdots, y_n)$ denote a $n \times n$ invertible diagonal matrix.
- S in lemma 1 is equivalent to $S = \{\mathbf{x} \mid \|Y^{-1}(\mathbf{x} - \mathbf{y})\|^2 \leq \beta^2\} = E(\mathbf{y}, \beta^2 Y^2)$.
- We consider the problem of minimizing $\mathbf{c}^T\mathbf{x}$ over $\{A\mathbf{x} = \mathbf{b}\} \cap \{E(\mathbf{y}, \beta Y)\}$,

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \|Y^{-1}(\mathbf{x} - \mathbf{y})\| \leq \beta \end{array}$$

which can be reformulated, using $\mathbf{d} = \mathbf{x} - \mathbf{y}$,

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{d} \\ \text{subject to} & A\mathbf{d} = 0, \\ & \|Y^{-1}\mathbf{d}\| \leq \beta \end{array} \tag{1}$$

- This can be solved **explicitly**.

Preliminary results: Linear programming on ellipsoids

Lemma 2. *Assume that the rows of A are linearly independent and that $\mathbf{c}$ is not a linear combination of the rows of A. Let $\mathbf{y}$ be a positive vector. Then an optimal solution to problem*

$$\begin{array}{ll} \textit{minimize} & \mathbf{c}^T\mathbf{d} \\ \textit{subject to} & A\mathbf{d} = 0, \\ & \|Y^{-1}\mathbf{d}\| \leq \beta, \end{array}$$

is given by

$$\mathbf{d}^\star = -\beta \frac{Y^2(\mathbf{c} - A^T\mu)}{\|Y(\mathbf{c} - A^T\mu)\|},$$

where

$$\mu = (AY^2A^T)^{-1}AY^2\mathbf{c},$$

and we observe that

$$\mathbf{c}^T\mathbf{x} = \mathbf{c}^T\mathbf{y} - \beta\|Y(\mathbf{c} - A^T\mu)\| < \mathbf{c}^T\mathbf{y}.$$

Proof

- **existence of $\mathbf{d}^\star$**:
  - AY^2A^T is invertible. If not,
    $\exists \mathbf{x} \mid \mathbf{x}^T AY^2 A\mathbf{x} = 0 \Rightarrow YA\mathbf{x} = \mathbf{0} \Rightarrow A\mathbf{x} = \mathbf{0} \Rightarrow$ rows of A l.d.
  - since $\mathbf{c}$ is not a linear combination of rows of $A, \forall \mu, \mathbf{c} - A^T\mu \neq \mathbf{0}$.
- **feasibility of $\mathbf{d}^\star$**:
  - $Y^{-1}\mathbf{d}^\star = -\beta \frac{Y(\mathbf{c}-A^T\mu)}{\|Y(\mathbf{c}-A^T\mu)\|} \Rightarrow \|Y^{-1}\mathbf{d}^\star\| = \beta$.
  - From the definition of μ, $AY^2(\mathbf{c} - A\mu) = \mathbf{0}$.
- **Optimality of $\mathbf{d}^\star$**:
  - Using the Cauchy-Schwartz inequality, for any feasible $\mathbf{d}$,

$$\begin{aligned}\mathbf{c}^T\mathbf{d} = (\mathbf{c}^T - \mu^T A)\mathbf{d} = (\mathbf{c}^T - \mu^T A)YY^{-1}\mathbf{d} \\ \geq -\|Y(\mathbf{c} - A^T\mu)\| \cdot \|Y^{-1}\mathbf{d}\| \geq -\beta\|Y(\mathbf{c} - A^T\mu)\|\end{aligned}$$

Proof

◦ on the other hand, for $\mathbf{d}^\star$:

$$\mathbf{c}^T\mathbf{d}^\star = (\mathbf{c}^T - \mu^T A)\mathbf{d}^\star = -(\mathbf{c}^T - \mu^T A)\beta\frac{Y^2(\mathbf{c} - A^T\mu)}{\|Y^2(\mathbf{c} - A^T\mu)\|}$$

$$= -\beta\frac{\big(Y(\mathbf{c}^T - \mu^T A)\big)^T Y(\mathbf{c} - A^T\mu)}{\|Y(\mathbf{c} - A^T\mu)\|} = -\beta\|Y(\mathbf{c} - A^T\mu)\|$$

hence $\mathbf{d}^\star$ is optimal.

- **Improved objective**: $\mathbf{c}^T\mathbf{x} = \mathbf{c}^T\mathbf{y} + \mathbf{c}^T\mathbf{d}^\star = \mathbf{c}^T\mathbf{y} - \beta\|Y(\mathbf{c} - A^T\mu)\| < \mathbf{c}^T\mathbf{y}$.

Taking a close look at μ

- We have just seen that μ, defined as

$$\mu = (AY^2A^T)^{-1}AY^2\mathbf{c},$$

  plays an important role in the solution. Why write it μ? confusing?

- Suppose $\mathbf{y}$ is actually a nondegenerate feasible solution with $\mathbf{I}$ corresponding basis.

- Reorder for convenience the indices, $\mathbf{I} = \{1, \cdots, m\}$ and the $n-m$ remaining coefficients are nonbasic.

- $Y = \mathbf{diag}(y_1, \cdots, y_m, \underbrace{0, \cdots, 0}_{n-m})$ and $AY = [BY_0 \;\; \mathbf{0}]$ and

$$\mu = (AY^2A^T)^{-1}AY^2\mathbf{c} = B_{\mathbf{I}}^{-1}Y_0^{-2}B_{\mathbf{I}}^{-1}Y_0^2\mathbf{c}_{\mathbf{I}} = B_{\mathbf{I}}^{-1}\mathbf{c}_{\mathbf{I}}$$

- when $\mathbf{y}$ is a primal BFS, μ is a dual BS. It is thus natural to see μ as a *dual estimate* corresponding to the current primal solution.

Taking a close look at μ

- Moreover, $\mathbf{c} - A^T\mu$ also appears everywhere..
- the (transposed) **reduced cost coefficient**.
- Suppose $\mathbf{r} = \mathbf{c} - A^T\mu$ is nonnegative. Then μ is dual feasible and

$$\mathbf{r}^T\mathbf{y} = (\mathbf{c} - A^T\mu)^T\mathbf{y} = \mathbf{c}^T\mathbf{y} - \mu^T A\mathbf{y} = \mathbf{c}^T\mathbf{y} - \mu^T\mathbf{b},$$

  a quantity called the **duality gap**.
- If $\mathbf{r}^T\mathbf{y} = 0$ then by weak duality $\mathbf{y}$ and μ are the primal and dual optima.

Dual gap and closeness to optima

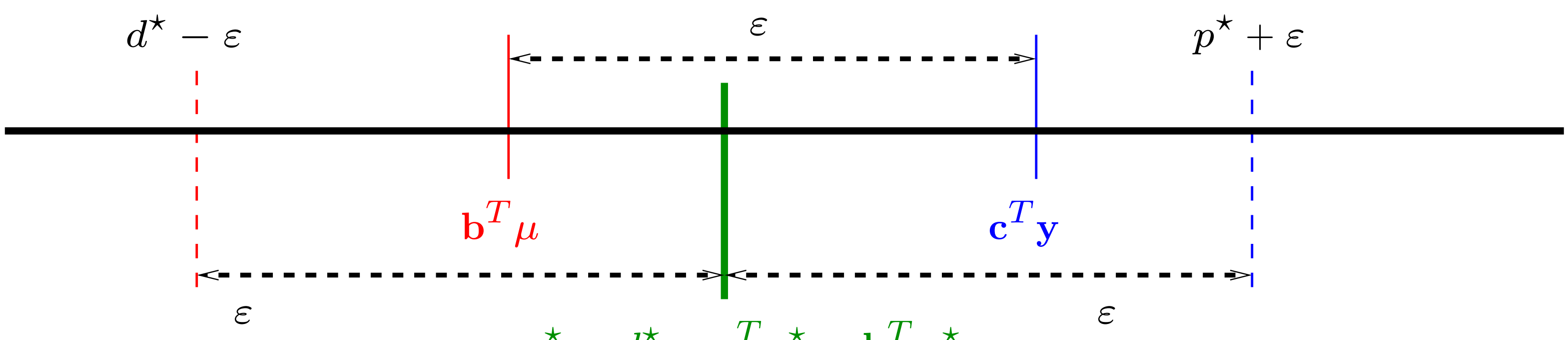
Lemma 3. *Let* $\mathbf{y}$ *and* μ *be a primal and dual feasible solution respectively such that* $\mathbf{c}^T\mathbf{y} - \mu^T\mathbf{b} < \varepsilon$. *Let* $\mathbf{y}^\star$ *and* $\mu^\star$ *be* ***optimal*** *primal and dual solutions respectively. Then,*

$$\begin{array}{ccccc} \mathbf{c}^T\mathbf{y}^\star & \leq & \mathbf{c}^T\mathbf{y} & < & \mathbf{c}^T\mathbf{y}^\star + \varepsilon \\ \mathbf{b}^T\mu^\star - \varepsilon & < & \mathbf{c}^T\mu & \leq & \mathbf{b}^T\mu^\star. \end{array}$$

or, using the notations below,

$$\begin{array}{ccccc} p^\star & \leq & \mathbf{c}^T\mathbf{y} & < & p^\star + \varepsilon \\ d^\star - \varepsilon & < & \mathbf{c}^T\mu & \leq & d^\star. \end{array}$$

- **Proof**: We use this small drawing



Dual gap: a stopping criteria

- To sum up:
  - run iteration that update the center and the **inscribed** ellipsoid
  - whenever $\mathbf{r} \geq \mathbf{0}$ we have a dual feasible solution.
  - when $\mathbf{r}^T\mathbf{y} < \varepsilon$ then we are *near-optimal*.
- convergence is thus **parameterized** by an **optimality tolerance** ε.
- a conceptual change w.r.t simplex which gives **exact solutions**
- on the other hand we have a hint of how to round off values, since the solution **must be sparse** (have a lot of zeros).

Summary of the Affine Scaling Algorithm

- To sum up:
  - **Input** $(A, \mathbf{b}, \mathbf{c})$, initial point $\mathbf{x}$, tolerance ε, $\beta \in (0, 1)$.
  - **Initialization** $\mathbf{x}_0 = \mathbf{x}$, $t = 0$.
  - **Loop**
    - ▷ Given a feasible $\mathbf{x}_t$, consider the corresponding diagonal matrix X_t and

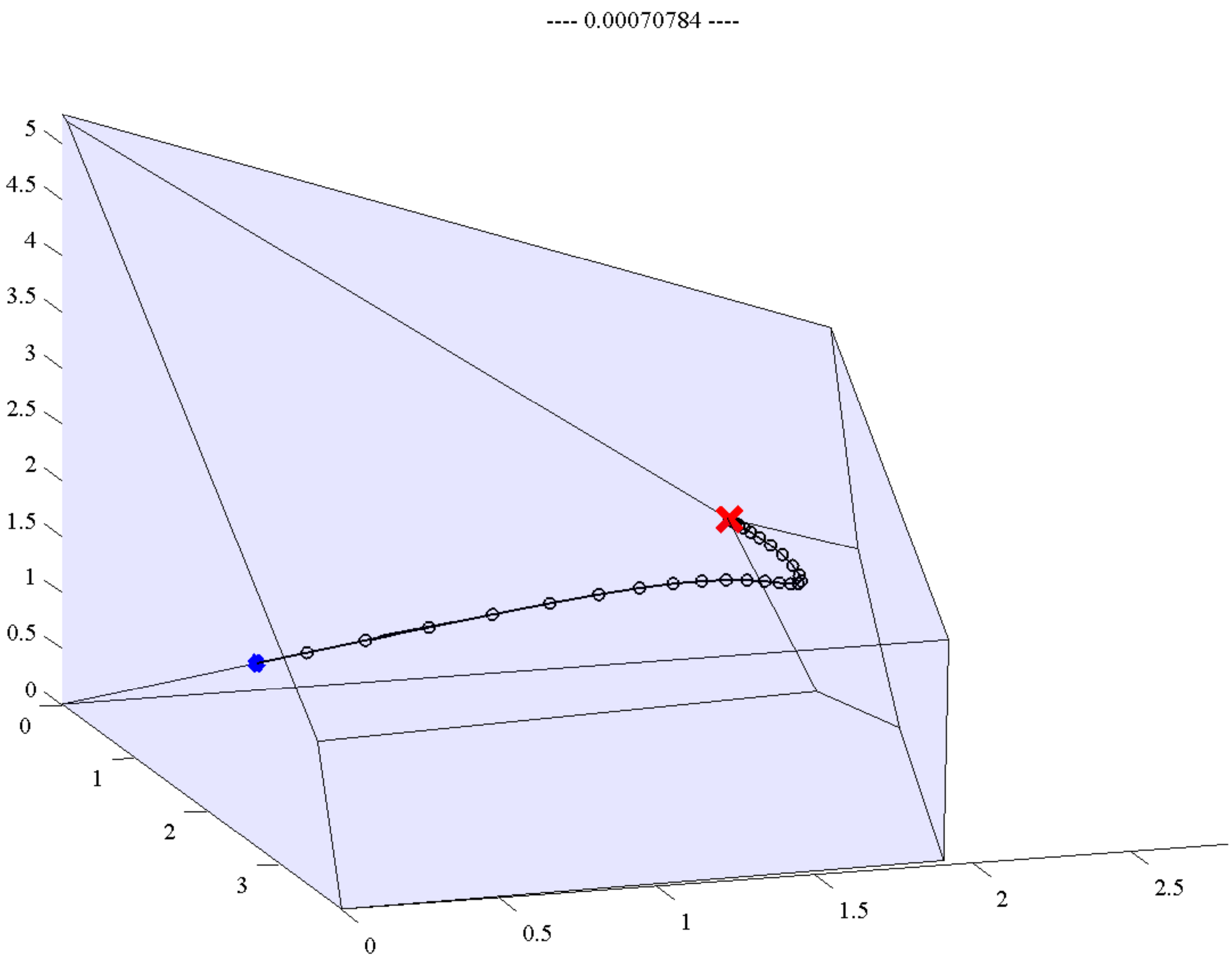
$$\begin{aligned} \mu_t &= (AX_t^2A^T)^{-1}AX_t^2\mathbf{c} \\ \mathbf{r}_t &= \mathbf{c} - A^T\mu_t \end{aligned}$$

    - ▷ **optimality check** if $\mathbf{r}_t \geq 0$ and $\mathbf{r}_t^T\mathbf{x}_t < \varepsilon$ then stop with ε-optimal $\mathbf{x}_t$.
    - ▷ **unboundedness check** if $-X_t^2\mathbf{r}_t \geq 0$ stop, optimal cost is $-\infty$.
    - ▷ **update** the new solution to

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \beta\frac{X_t^2\mathbf{r}_t}{\|X_t\mathbf{r}_t\|} \tag{2}$$

    - ▷ **next iteration** set with $t \leftarrow t + 1$

Short Matlab Demo



Variations: short and long steps

- Define the alternative norms for vectors

$$\begin{aligned} \|\mathbf{u}\|_\infty &= \max_i \|u_i\|, \\ \gamma(\mathbf{u}) &= \max\{u_i | u_i > 0\}. \end{aligned}$$

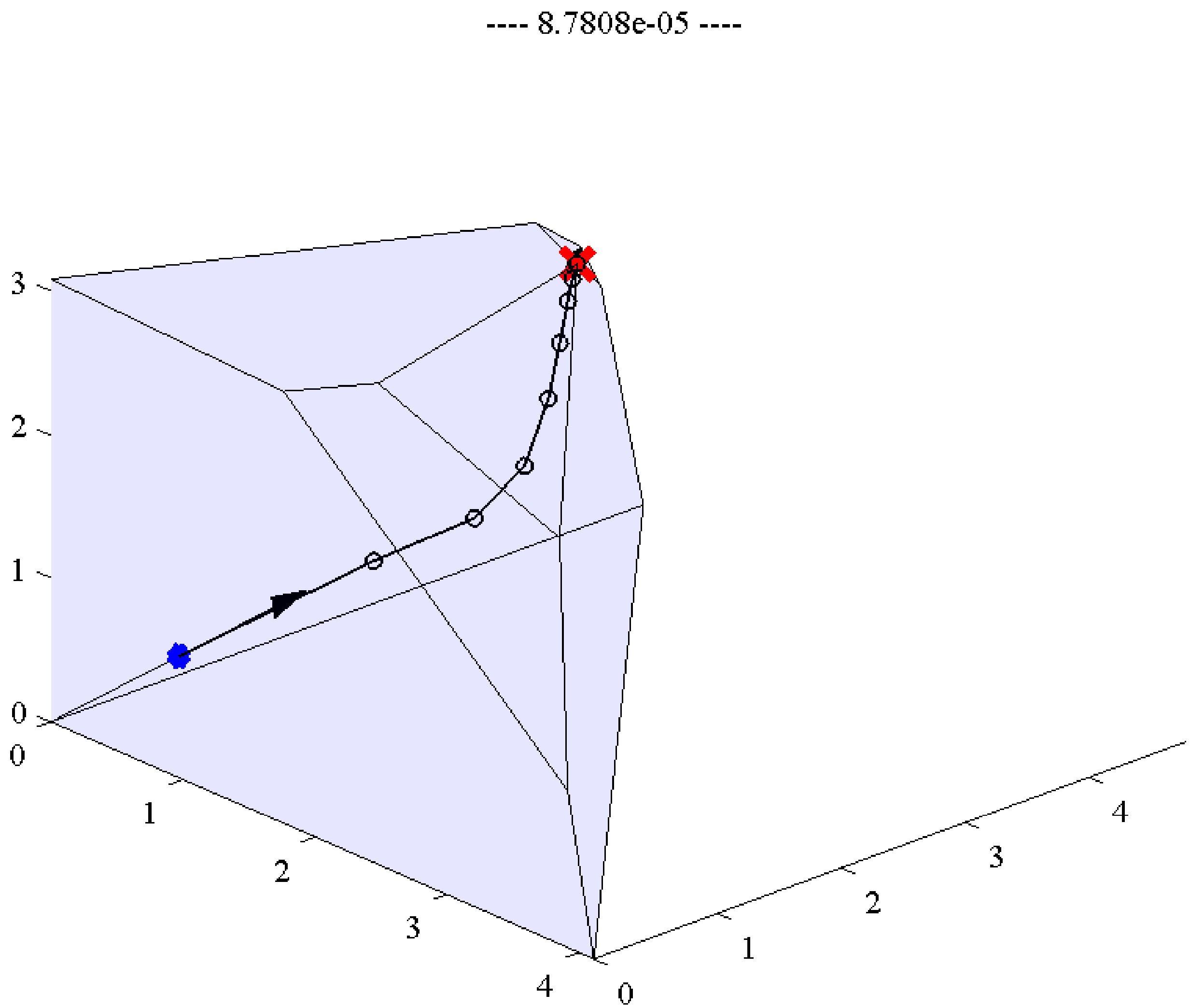
  then $\gamma(\mathbf{u}) \leq \|\mathbf{u}\|_\infty \leq \|\mathbf{u}\|$.

- The previous version we proposed is usually called a short step method.
- long step variants update $\mathbf{x}_t$ in the same direction $-X_t^2\mathbf{r}_t$ but with longer steps.
- Instead of a step $-\beta\frac{X_t^2\mathbf{r}_t}{\|X_t\mathbf{r}_t\|}$ as in Equation (2),

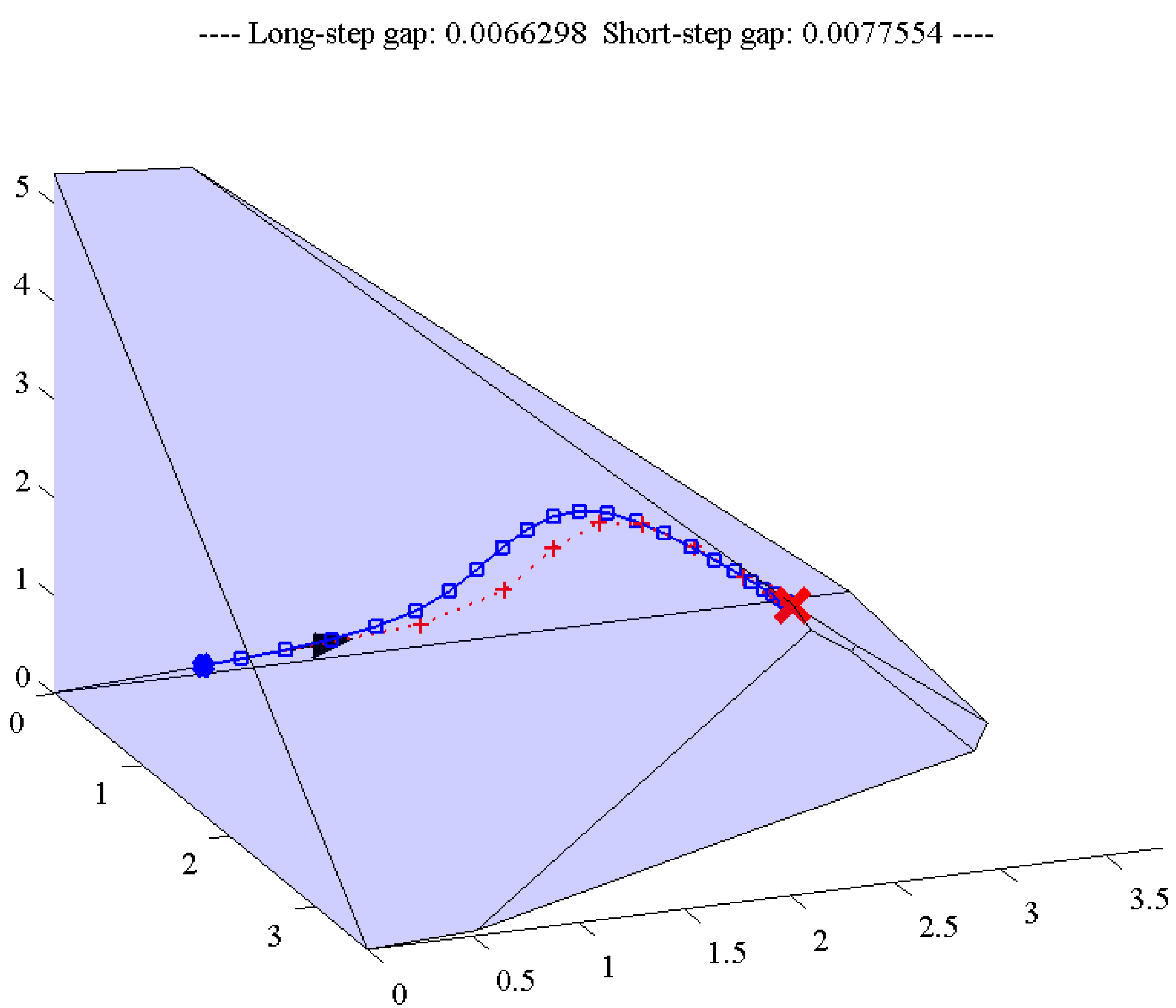
$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{x}_t - \beta\frac{X_t^2\mathbf{r}_t}{\|X_t\mathbf{r}_t\|_\infty}, \text{ or} \\ \mathbf{x}_{t+1} &= \mathbf{x}_t - \beta\frac{X_t^2\mathbf{r}_t}{\gamma(X_t\mathbf{r}_t)}. \end{aligned}$$

- Steps are **larger** and one can show that the updates are still **feasible**.

Short Matlab Demo



Short Matlab Demo



Complete Implementation: looking for an initial point

- The algorithm requires an initial feasible point $\mathbf{x}_0$.
- As with the simplex, we may not have it.
- As with the simplex, we
  - **augment** the problem,
  - use a **trivial first feasible point**,
  - **make sure artificial variables disappear** on convergence.
- Suppose $M \gg 0$ and consider the problem

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} + Mx_{n+1} \\ \text{subject to} & A\mathbf{x} + (\mathbf{b} - A\mathbf{1})x_{n+1} = \mathbf{b}, \ , \\ & \left[\begin{smallmatrix} \mathbf{x} \\ x_{n+1} \end{smallmatrix}\right] \geq 0 \end{array}$$

- $\left[\begin{smallmatrix} \mathbf{x} \\ x_{n+1} \end{smallmatrix}\right] = \left[\begin{smallmatrix} \mathbf{1} \\ 1 \end{smallmatrix}\right]$ is a positive feasible solution,
- Convergence yields $x_{n+1} = 0$ if the problem is feasible.

Convergence

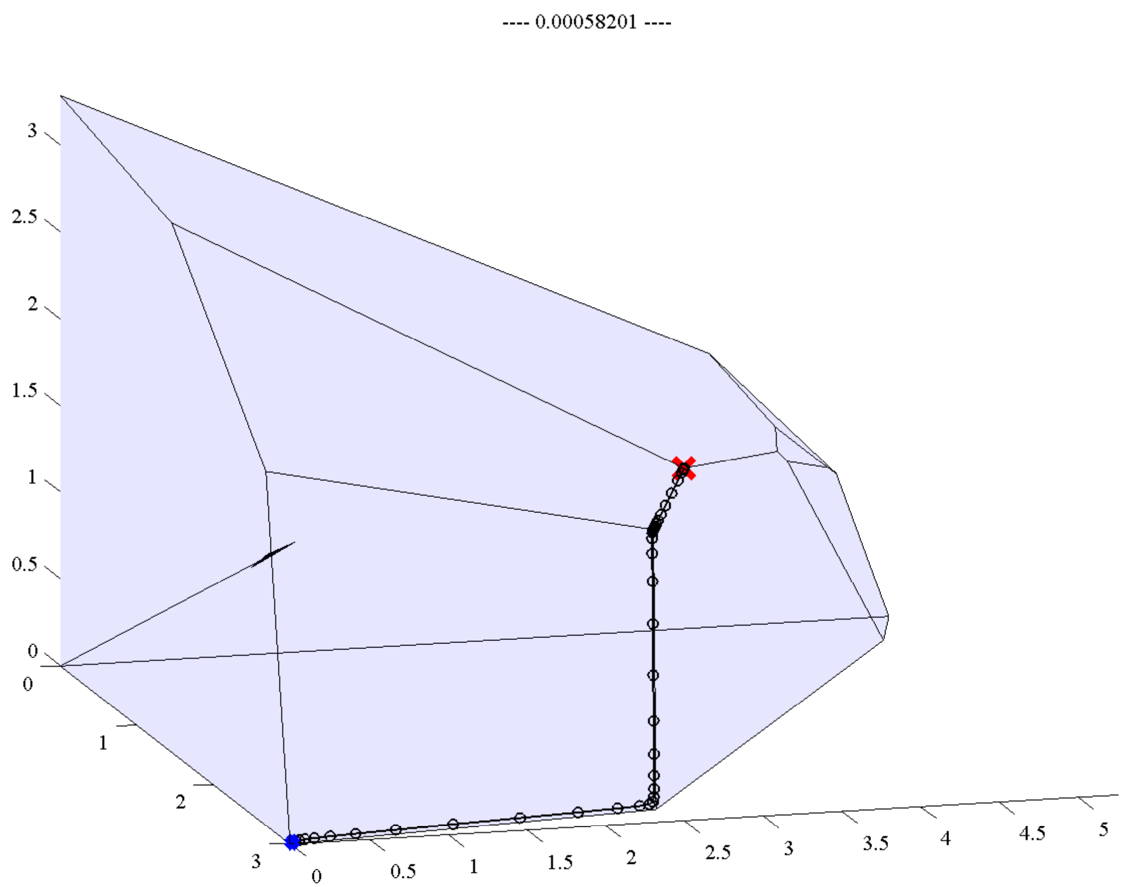
- Convergence can be proved for both primal and dual under certain assumptions.
- In particular, if
  - rows of A are l.i. ,
  - $\mathbf{c}$ is not a linear combination of the rows of A,
  - there exists an optimal solution,
  - there exists a **positive** feasible solution,

  then **everything works with** $\beta < 2/3$.
- if in addition
  - All BFS of the primal problem are nondegenerate,
  - The reduced cost of nonbasic coefficients corresponding to a BFS are nonzero,

  everything works with $0 < \beta < 1$.

Speed and efficiency

- Still a research topic.
- Although the following example gives a hint of why things may not always be great:



Potential Reduction Algorithm

Using two (almost) standard forms for primal and dual problems

- Closest to Karmarkar's algorithm of 1984.
- We consider the **primal-dual pair** in *standardized* forms

$$\begin{array}{lr} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \quad , \quad \begin{array}{lr} \text{maximize} & \mathbf{b}^T\mu \\ \text{subject to} & A^T\mu + \mathbf{s} = \mathbf{c} \\ & \mathbf{s} \geq 0 \end{array}$$

- Note that primal variable $\mathbf{x} \in \mathbf{R}^n$. dual variable $\mu \in \mathbf{R}^m$ but $\mathbf{s} \in \mathbf{R}^n$
- We assume that
  - A has linearly independent rows
  - There exist $\mathbf{x} > 0$ and$(\mu, \mathbf{s}), \mathbf{s} > 0$ which are feasible for the primal and dual problem respectively.
- We define the **potential function**

$$G(\mathbf{x}, \mathbf{s}) = q \log s'\mathbf{x} - \sum_{j=1}^{n} \log x_j - \sum_{j=1}^{n} \log s_j$$

with $q \geq n$.

Reducing the potential function

- Note that the dual gap

$$\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mu = (\mathbf{s}^T + \mu^T A)\mathbf{x} - \mathbf{x}^T A^T \mu = \mathbf{s}^T\mathbf{x}$$

  corresponds to $\mathbf{s}^T\mathbf{x}$ which the potential function aims to minimize.

- Furthermore, the two sums penalize the **proximity** to the **boundary** of the **feasible sets** of the primal and dual respectively.

  Theorem 1. *Let $\mathbf{x}_0 > 0$ and $(\mu_0, \mathbf{s}_0)$ with $\mathbf{s}_0 > 0$ be* ***feasible primal and dual solutions*** *and $\varepsilon > 0$ the* ***optimality tolerance****.* ***Any algorithm*** *that maintains primal and dual feasibility and* ***reduces*** *$G(\mathbf{x}, \mathbf{s})$ by at least $\delta > 0$ at each iteration* ***finds a solution to the primal and dual problems*** *with duality gap $\mathbf{s}_K^T\mathbf{x}_K \leq \varepsilon$ with*

$$K = \left\lceil \frac{G(\mathbf{x}_0, \mathbf{s}_0) - (q-n)\log\varepsilon - n\log n}{\delta} \right\rceil$$

  iterations

The potential function

Proof. • We show $G(\mathbf{x},\mathbf{s}) \geq n\log n + (q-n)\log \mathbf{s}^T\mathbf{x}$.

  ◦ $G(\mathbf{x},\mathbf{s}) = n\log \mathbf{s}^T\mathbf{x} - \sum_{j=1}^n x_j - \sum_{j=1}^n \log s_j + (q-n)\log \mathbf{s}^T\mathbf{x}$.
  ◦ let's do some minimization on $n\log\sum s_j x_j - \sum_{j=1}^n \log x_j s_j$
  ◦ alternatively study $n\log\sum_{j=1}^n u_j - \sum_{j=1}^n \log u_j$ with $\sum u_j = \mathbf{s}^T\mathbf{x}$, which is minimized for $u_j = \mathbf{s}^T\mathbf{x}/n$ (think entropy)
  ◦ Replace and get the lower bound.

- Fix δ and suppose $G(\mathbf{x}_{k+1},\mathbf{s}_{k+1}) - G(\mathbf{x}_k,\mathbf{s}_k) \leq -\delta$
- Thus $G(\mathbf{x}_k,\mathbf{s}_k) - G(\mathbf{x}_0,\mathbf{s}_0) \leq -k\delta$.
- In particular $G(\mathbf{x}_K,\mathbf{s}_K) \leq (q-n)\log\varepsilon + n\log n$.
- Using the inequality above, $G(\mathbf{x}_K,\mathbf{s}_K) \geq n\log n + (q-n)\log \mathbf{s}_K^T\mathbf{x}_K$
- Combining the two, $\mathbf{s}_K^T\mathbf{x}_K \leq \varepsilon$

■

Potential Reduction Algorithm

- Advantage: **stay far from boundary** at each iteration while improving gap.
- What we need: algorithm that reduces **steadily** the potential while maintaining feasibility.
- Hence the name of **Potential Reduction Algorithms**.

A Proposal for a Potential Reduction Algorithm

- We start with a primal solution $\mathbf{x} > 0$ and a dual feasible solution with $\mathbf{s} > 0$.
- We look for a direction $\mathbf{d}$ such that $G(\mathbf{x}+\mathbf{d}, \mathbf{s}) < G(\mathbf{x}, \mathbf{s})$.
- Similarly to affine scaling, we can proceed by having $\mathbf{d}$ satisfy $A\mathbf{d} = \mathbf{0}$, $\|X^{-1}\mathbf{d}\| \leq \beta < 1$ so that $\mathbf{x}+\mathbf{d}$ is still feasible.
- **Important difference**: we minimize a **nonlinear** function, not $\mathbf{c}^T\mathbf{x}$.
- Using a local Taylor approximation,

$$\begin{array}{ll} \text{minimize} & \nabla_{\mathbf{x}} G(\mathbf{x}, \mathbf{s})^T \mathbf{d} \\ \text{subject to} & A\mathbf{d} = \mathbf{0} \\ & \|X^{-1}\mathbf{d}\| \leq \beta \end{array}$$

- Difference with affine scaling, objective "$\hat{\mathbf{c}}$" is $\nabla_{\mathbf{x}} G(\mathbf{x}, \mathbf{s})$.
- Namely $\hat{\mathbf{c}}$ is such that $\hat{c}_i = \frac{\partial G(\mathbf{x},\mathbf{s})}{\partial x_i} = \frac{q s_i}{\mathbf{s}^T\mathbf{x}} - \frac{1}{x_i}$.

A Proposal for a Potential Reduction Algorithm

- Using Lemma 2, the optimal direction $\mathbf{d}^\star$ is

$$\mathbf{d}^\star = -\beta X \frac{\mathbf{u}}{\|\mathbf{u}\|} \text{ with } \mathbf{u} = X\left(\hat{\mathbf{c}} - A^T(AX^2X^T)^{-1}AX^2\hat{\mathbf{c}}\right)$$

- replacing the values of $\hat{\mathbf{c}}$ (with $\hat{c}_i = \frac{qs_i}{\mathbf{s}^T\mathbf{x}} - \frac{1}{x_i}$) we have

$$X\hat{\mathbf{c}} = \frac{q}{\mathbf{s}^T\mathbf{x}} X\mathbf{s} - \mathbf{1}$$

which yields,

$$\mathbf{u} = \left(I - XA^T(AX^2A^T)^{-1}AX\right)\left(\frac{q}{\mathbf{s}^T\mathbf{x}} X\mathbf{s} - \mathbf{1}\right).$$

- Lemma 2 also gives the decrease in objective: $\beta\|\mathbf{u}\| + O(\beta^2)$
- Depending on the size of $\|\mathbf{u}\|$ can we have a minimal decrease?

Potential Reduction Algorithm

- Parameters $(A, \mathbf{b}, \mathbf{c})$, β, γ, q and ε.
- **Initialization** $\mathbf{x}_0 > 0, \mathbf{s}_0 > 0$ and $\mu_0, k = 0$.
- **Optimality test** if $\mathbf{s}_k^T \mathbf{x}_k < \varepsilon$ stop. otherwise go to next step.
- **Update correction** compute considering X_k corresponding to $\mathbf{x}_k$,

$$A_k = (AX_k)^T \left(AX_k^2 A^T)^{-1}\right) AX_k$$

$$\mathbf{u}_k = (I - A_k)\left(\frac{q}{\mathbf{s}_k^T \mathbf{x}_k} X_k \mathbf{s}_k - \mathbf{1}\right)$$

$$\mathbf{d}_k = -\beta X_k \frac{\mathbf{u}}{\|\mathbf{u}\|}$$

Potential Reduction Algorithm

Check the decrease

- if $\|\mathbf{u}_k\| \geq \gamma$, **primal update**

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k,$$
$$\mathbf{s}_{k+1} = \mathbf{s}_k,$$
$$\mu_{k+1} = \mu_k.$$

- if $\|\mathbf{u}_k\| < \gamma$, **dual update**

$$\mathbf{x}_{k+1} = \mathbf{x}_k,$$
$$\mathbf{s}_{k+1} = \frac{\mathbf{s}_k^T \mathbf{x}_k}{q} X_k^{-1}(\mathbf{u}_k + \mathbf{1}),$$
$$\mu_{k+1} = \mu_k + \left(A X_k^2 A^T\right)^{-1} A X_k \left(X_k \mathbf{s}_k - \frac{\mathbf{s}_k^T \mathbf{x}_k}{q} \mathbf{1} \right).$$

Why does it work?

- One can prove that if $\|\mathbf{u}_k\| \geq \gamma$, **primal update**

$$G(\mathbf{x}_{k+1}, \mathbf{x}_{k+1}) - G(\mathbf{x}_k, \mathbf{s}_k) \leq -\beta\gamma + \frac{\beta^2}{2(1-\beta)}$$

- One can prove that if $\|\mathbf{u}_k\| < \gamma$, **dual update**

$$G(\mathbf{x}_{k+1}, \mathbf{x}_{k+1}) - G(\mathbf{x}_k, \mathbf{s}_k) \leq -(q-n) + n \log \frac{q}{n} + \frac{\gamma^2}{2(1-\gamma)}$$

- **bottom line**: if $q = n + \sqrt{n}, \beta \approx 0.285$ and $\gamma \approx 0.479$ then the potential reduction algorithm reduces $G(\mathbf{x}, \mathbf{s})$ by at least $\delta = 0.079$ at each iteration.
- **iterations**: $K = O\left(\sqrt{n} \log 1\varepsilon + n^2 \log(nU)\right)$.
- **overall complexity**: $O(n^{3.5} \log \frac{1}{\varepsilon} + n^5 \log(nU))$.

A short outlook of primal path following algorithm

The barrier function

- As usual, consider the **primal-dual pair**

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{array} \quad , \quad \begin{array}{ll} \text{maximize} & \mathbf{b}^T\mu \\ \text{subject to} & A^T\mu + \mathbf{s} = \mathbf{c} \\ & \mathbf{s} \geq 0 \end{array}$$

- Get around the issues given nonnegativity constraints.
- Encode this through a *barrier function*:

$$B_\tau(\mathbf{x}) = \mathbf{c}^T\mathbf{x} - \tau \sum_{j=1}^{n} \log x_j$$

- Consider the family of *barrier problems* indexed by $\tau > 0$,

$$\begin{array}{ll} \text{minimize} & B_\tau(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

Brief Analysis

- For all $\mu > 0$, the barrier problem has a unique optimal solution $\mathbf{x}(\mu)$
- In particular, for $\tau = \infty$, the barrier problem becomes

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log x_i \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

- A barrier problem originating from the dual problem is

$$\begin{array}{ll} \text{maximize} & \mathbf{b}^T\mu + \tau \sum_{i=1}^{n} \log s_j \\ \text{subject to} & \mu^T A + \mathbf{s} = \mathbf{c} \end{array}$$

- Similar to the potential reduction method, we consider a Taylor expansion of B_μ and update $\mathbf{x}$ by a given direction $\mathbf{d}$ while still staying in the feasible set.
- Studied in more depths in ORF 523.

Next time

- Network flows,
  - network simplex
  - transportation problems
  - maximum flow
  - assignment problem