On Modern Convex Optimization
The Interior-Point Revolution

André Tits

University of Maryland - College Park

May 16, 2008
My plan for today: Convey to you the **pervasiveness**, the **power**, the **elegance**, and I hope the **excitement**, of modern convex optimization.

Disclaimer: This is but a glimpse at convex optimization, with **focus on simple key ideas** rather than on the most efficient methods. Rigor is not guaranteed.

No need to take notes: Slides are available for download.

Good news, bad news.
Outline

2. Simplest Case: Linear Optimization
3. General Case: Conic Optimization
4. Recent Accomplishments at Maryland
Convex set, Convex function

- A set $S$ is **convex** if it contains every line segment who endpoints are in $S$.
- A function $f$ is **convex** if its epigraph is convex.
- When $S$ and $f$ are convex, **all stationary points** for the problem
  \[
  \text{minimize } f(x) \text{ subject to } x \in S.
  \]
  are global minimizers.
- Convex optimization problems are “easy” to solve.

In the past two decades, a **quasi-revolution** took place in the development and analysis of **effective and efficient methods** for the solution of convex problems, in particular **interior-point methods**, starting with the ground-breaking work of Nesterov and Nemirovski (following that of Karmarkar in linear optimization).
Convex problems are ubiquitous

A few areas of application:
  - signal processing
  - VLSI design
  - estimation
  - communication networks
  - circuit design
  - automatic control
  - statistics
  - finance
  - etc.
Example: VLSI gate sizing

Let $t_0 = 0$, and let $T > 0$, $s_i, \overline{s}_i > 0$, and $a_{ij} > 0$ be given.

The minimal area gate sizing problem can be written as

$$\text{minimize } \sum_{i} s_i \quad \text{s.t. } t_j + d_i(s) \leq t_i \quad \text{whenever } j \in \text{fanin}(i), \quad t_n \leq T, \quad s_i \leq s_i \leq \overline{s}_i \forall i$$

where $s_i$ is the size (area) of gate $i$, $t_i$ is an upper bound on the delay from the primary input to the output of gate $i$, and

$$d_i(s) := a_{i0} + \sum_{k \in \text{fanout}(i)} a_{ik} \frac{s_k}{s_i}.$$ 

Equivalently,

$$\text{minimize } \sum_{i} s_i \quad \text{s.t. } a_{i0} + \sum_{k \in \text{f.o.}(i)} a_{ik} s_k s_i^{-1} \leq t_i - t_j, \quad t_n \leq T, \quad s_i \leq s_i \leq \overline{s}_i$$

The cost and constraint functions being “posynomials”, this is a geometric programming problem. It is NOT convex, but it can be “convexified”, as we show next.
VLSI gate sizing: Recasting into a convex problem

The change of variables $x_i := \log(s_i)$ yields

$$\minimize_{x_i, t_i} \sum_i e^{x_i} \quad \text{s.t.} \quad a_{i0} + \sum_{k \in \text{f.o.}(i)} a_{ik} e^{x_k - x_i} \leq t_i - t_j, \quad t_n \leq T, \quad \log(\bar{s}_i) \leq x_i \leq \log(s_i)$$

which is convex. Because exponential functions tend to cause numerical difficulties, typically, the following reformulation (obtained by taking logs) is used instead:

$$\minimize \log \left( \sum_i e^{x_i} \right)$$

s.t. $\log \left( a_{i0} + \sum_{k \in \text{f.o.}(i)} a_{ik} e^{x_k - x_i} \right) - \log(t_i - t_j) \leq 0, \quad t_n \leq T, \quad \log(\bar{s}_i) \leq x_i \leq \log(s_i)$

which is still convex.
Outline


2. Simplest Case: Linear Optimization

3. General Case: Conic Optimization

4. Recent Accomplishments at Maryland
Duality in linear optimization

Primal: \( v_P := \min \{ c^T x : Ax = b, x \in \mathbb{R}_+^n \} \)

Dual: \( v_D := \max \{ b^T y : A^T y + s = c, s \in \mathbb{R}_+^n \} \)

- For feasible \((x, y, s)\), the "duality gap" is \( s^T x \):
  \[
c^T x - b^T y = (A^T y + s)^T x - y^T b = s^T x.
  \]

- **Weak duality**: \( v_P \geq v_D \). For feasible \((x, y, s)\),
  \[
v_P - v_D = \inf_{x, y \text{ feasible}} \{ c^T x - b^T y \} = \inf_{x, s \text{ feasible}} \{ s^T x \} \geq 0.
  \]

  Note: For feasible \((x, s)\), \( s^T x = 0 \) iff \( x^i s^i = 0 \) for all \( i \).

- **(Gyula, aka Julius) Farkas’s Lemma** (1902; present form due to Albert Tucker):
  \[
  \begin{align*}
  \begin{cases}
  Mu = d \\
  u \in \mathbb{R}_+^n
  \end{cases}
  \quad \text{is consistent XOR} \quad 
  \begin{cases}
  -M^T v \in \mathbb{R}_+^n \\
  d^T v > 0
  \end{cases}
  \quad \text{is consistent.}
  \end{align*}
  \]

- **Strong duality**: If the primal (or the dual) is feasible,
  \[
v_P = v_D =: v_{PD}
  \]

  Proof follows from Farkas’s lemma: see next slide (skipped).
Proof of strong duality in linear optimization

Below, “icst” means “is consistent”. Given $\theta \in \mathbb{R}$, using Farkas’s lemma, we get

$$ \begin{cases} \begin{align*} Ax &= b \\ x &\geq 0 \text{ icst} \\ c^T x &\leq \theta \end{align*} \end{cases} \iff \begin{bmatrix} A & 0 \\ c^T & 1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} = \begin{bmatrix} b \\ \theta \end{bmatrix} \quad \text{XOR} \begin{bmatrix} -A^T & c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y \\ u \end{bmatrix} \in \mathbb{R}^{m+1}_+ \quad \text{icst.} \end{cases} $$

Now suppose that the dual is feasible and $\theta < \nu_P$. Then the left-hand side is inconsistent, and hence the right-hand side must be consistent. That is, there exists $(y, u) \in \mathbb{R}^{m+1}$ such that

$$ A^T y + cu \leq 0, \quad u \leq 0, \quad b^T y + \theta u > 0. $$

If $u = 0$ then $\nu_D = +\infty$. (Indeed, take $\hat{y} = y_0 + \alpha y$, with $y_0$ feasible (we assumed dual feasibility) and $\alpha \to +\infty$.) Hence $\nu_D \geq \nu_P$. On the other hand, if $u < 0$ then, w.l.o.g., $u = -1$. We then get

$$ A^T y \leq c \quad \text{and} \quad b^T y > \theta $$

The existence of such $y$ implies that $\nu_D > \theta$. We conclude that $\nu_D > \theta$ whenever $\nu_P > \theta$, again implying that $\nu_D \geq \nu_P$. In view of weak duality, strong duality follows.

With primal feasibility assumed instead, the claim follows with a similar proof, starting by taking $\theta > \nu_D$ and expressing “$A^T y \leq c, \ b^T y \geq \theta$” as

$$ A^T(y_+ - y_-) + s = c, \ b^T(y_+ - y_-) - t = \theta, \ (s, t, y_+, y_-) \geq 0 $$
Primal-dual interior-point methods for LO: Set up

Notation: Capital letters $X$, $S$, $X_k$, $S_k$ denote diagonal matrices who entries are the components of the corresponding lower case vector. E.g, $X = \text{diag}(x^i)$.

In view of strong duality, a triple $(x, y, s)$ solves the primal-dual pair iff:

$$Ax = b, x \geq 0 \quad \text{(primal feasibility) (also $\nabla_y L_D = 0$)}$$

$$A^T y + s = c, s \geq 0 \quad \text{(dual feasibility) (also $\nabla_x L_P = 0$)}$$

$$x^T s = 0 \quad \text{(zero duality gap), i.e., } Xs = 0 \text{ (complementary slackness)}$$

Let $\beta \in (0, 1)$. Let $(x_k, y_k, s_k) \in \mathcal{N}_2(\beta)$, where

$$\mathcal{N}_2(\beta) = \{(x_k, y_k, s_k) : \text{PD feasible, } \|X_k s_k - \mu_k e\| < \beta \mu_k\},$$

with $\mu_k := \frac{x_k^T s_k}{n}$ (the “duality measure”).

Key idea: Focusing on the equalities in the optimality condition, perform a Newton iteration towards solving the square system, with $\sigma \in (0, 1)$,

$$Ax = b$$

$$A^T y + s = c$$

$$Xs = \sigma \mu_k e$$
**Primal-dual interior-point methods for LO: “Short Step” method**

**Why** $Xs = \sigma \mu_k e$ **rather than** $Xs = 0$? **With** $\sigma$ **close enough to 1**, this will make the full Newton step acceptable **globally**.

Thus, solve the Newton system

\begin{align*}
A \Delta x_k &= 0 \\
A^T \Delta y_k + \Delta s_k &= 0 \\
x_k s_k + X_k \Delta s_k + S_k \Delta x_k &= \sigma \mu_k e.
\end{align*}

and set

\begin{align*}
x_{k+1} &= x_k + \Delta x_k \\
y_{k+1} &= y_k + \Delta y_k \\
s_{k+1} &= s_k + \Delta s_k
\end{align*}

It turns out that, with an appropriate selection of $\sigma$ (**e.g.**, $\sigma = 1 - \frac{1}{\sqrt{n}}$), this exceedingly simple algorithm has **rather strong convergence properties**.
Primal-dual interior-point methods for LO: Polynomial complexity

The following two results follow:

1. If \((x_k, s_k)\) is strictly feasible, then 
   \[ \mu_{k+1} := \frac{x_{k+1}^T s_{k+1}}{n} = \sigma \mu_k \] 
   (immediate).

2. If \((x_k, y_k, s_k) \in \mathcal{N}_2(\beta)\), then 
   \((x_{k+1}, y_{k+1}, s_{k+1}) \in \mathcal{N}_2(\beta)\) (takes some work). In 
   particular, \((x_{k+1}, s_{k+1})\) is strictly feasible.

It follows that, if \((x_0, y_0, s_0) \in \mathcal{N}_2(\beta)\), then for all \(k\), 
\((x_k, y_k, s_k) \in \mathcal{N}_2(\beta)\) and 
\[
\begin{align*}
    c^T x_k - v_{PD} &\leq x_k^T s_k = n \mu_k = \sigma^k n \mu_0 = \sigma^k x_0^T s_0, \\
    v_{PD} - b^T y_k &\leq x_k^T s_k = n \mu_k = \sigma^k n \mu_0 = \sigma^k x_0^T s_0,
\end{align*}
\]

a global linear convergence rate.

Hence, given \(\epsilon > 0\), 
\[ c^T x_k < v_{PD} + \epsilon \] and 
\[ b^T y_k > v_{PD} - \epsilon \] are achieved after 

at most \(\sqrt{n} \log(x_0^T s_0/\epsilon)\) iterations.

Since every iteration (solution of an \((2n + m) \times (2n + m)\) linear system) takes at most 
\(O(n^3)\) flops, the total number of flops to \(\epsilon\)-solution is at most 
\(O(n^{7/2})\), which is polynomial in \(n\).
Outline

2. Simplest Case: Linear Optimization
3. General Case: Conic Optimization
4. Recent Accomplishments at Maryland
Why conic optimization?

Every convex optimization problem can be recast as a conic optimization problem. (And many convex problems arise naturally as conic optimization problems.)

Consider

\[
\begin{align*}
\text{minimize} & \quad f(x) \text{ subject to } x \in S, \\
\text{with } f \text{ and } S \text{ convex. This problem is equivalently written as} & \\
\text{minimize} & \quad t \text{ subject to } f(x) \leq t, x \in S.
\end{align*}
\]

where \( t \in \mathbb{R} \) is an additional optimization variable. Defining \( z = (x, t) \), we can recast this problem as

\[
\begin{align*}
\text{minimize} & \quad c^T z \text{ subject to } z \in G,
\end{align*}
\]

where \( c = (0 \ldots 01)^T \) and \( G \) is a convex set. This can be further recast as

\[
\begin{align*}
\text{minimize} & \quad d^T w \text{ subject to } [0 \ldots 0 1]w = 1, w \in K,
\end{align*}
\]

where \( d^T = (c^T 0) \), \( w = (z, u) \) \((u \in \mathbb{R})\), and

\[
K = \text{cl} \left( \left\{ (z, u) : u > 0, \frac{Z}{u} \in G \right\} \right),
\]

a closed convex cone. Advantages of conic optimization are shown next.
Duality in conic optimization

Let $K$ be a closed, convex, solid and pointed cone, $K^*$ its dual. (Note: $(\mathbb{R}_+^n)^* = \mathbb{R}_+^n$)

Primal: $v_P := \min \{ c^T x : Ax = b, x \in K \}$

Dual: $v_D := \max \{ b^T y : A^T y + s = c, s \in K^* \}$

- For feasible $(x, y, s)$,
  \[ c^T x - b^T y = (A^T y + s)^T x - y^T b = s^T x. \]

- Weak duality: $v_P \geq v_D$. For feasible $(x, y, s)$,
  \[ v_P - v_D = \inf_{x, y \text{ feasible}} \{ c^T x - b^T y \} = \inf_{x, s \text{ feasible}} \{ s^T x \} \geq 0. \]

  Note: $s^T x = 0$ no longer equivalent to $x^i s^i = 0$ for all $i$.

- Farkas's Lemma: Always holds only if $MK$ is closed! (Quiz time!!)

  \[ \begin{cases} Mu = d \\ u \in K \end{cases} \text{ is consistent XOR } \begin{cases} -M^T v \in K^* \\ d^T v > 0 \end{cases} \text{ is consistent.} \]

  [Equivalently: $d \in MK \iff d \in (MK)^{**}$]

- Strong duality: If the primal (or the dual) is feasible and $AK$ is closed (or some weaker condition), then
  \[ v_P = v_D =: v_{PD} \]

  Same proof (under the assumption that $AK$ is closed) as in the case of linear optimization!
Key to polynomial complexity: Self-concordance

Given a proper (i.e., closed, convex, solid and pointed) cone $K \subset \mathbb{R}^n$,

$$F : \text{int}(K) \rightarrow \mathbb{R}, \text{ three times differentiable, } F''(x) \text{ invertible } \forall x$$

is a $\nu$-logarithmically homogeneous self-concordant barrier if

- $F(tx) = F(x) - \nu \log(t) \quad \forall x \in \text{int}(K), \ t > 0 \quad (\nu$-logarithmically homogeneous)
- $|F'''(x)[h, h, h]| \leq 2(F''(x)[h, h])^{3/2} \quad \forall x \in \text{int}(K), \ h \in \mathbb{R}^n \quad (\text{self-concordant})$
- $F(x_k) \rightarrow \infty \quad \text{whenever } \{x_k\} \subset \text{int}(K) \text{ and } x_k \rightarrow x \in \text{bd}(K) \quad (\text{barrier})$

Examples:

- $K = \mathbb{R}_+, \quad F(x) = -\log(x), \quad \nu = 1$
- $K = \mathbb{R}^n_+, \quad F(x) = -\sum \log(x^i), \quad \nu = n \quad (\text{linear optimization})$
- $K = S^n_+, \quad F(x) = -\log \det x, \quad \nu = n \quad (\text{semidefinite programming})$

This property is key to polynomial complexity. $\nu$ is the “complexity parameter”.

Define the “dual local norm”

$$\|s\|_{x, F}^\ast = (F''(x)^{-1}[s, s])^{1/2}$$
Primal-dual interior-point methods for CO
Let $\beta \in (0, 1)$. Let $(x_k, y_k, s_k) \in \mathcal{N}_2(\beta)$, where (with $\mu_k = \frac{x_k^T s_k}{\nu}$)

$$\mathcal{N}_2(\beta) = \{(x_k, y_k, s_k) : \text{PD feasible, } \|s_k + \mu_k F'(x_k)\|^*_{x_k,F} < \beta \mu_k\}.$$

Perform a Newton iteration towards solving the square system, with $\sigma \in (0, 1)$,

$$Ax = b \quad \text{(primal feasibility)}$$
$$A^Ty + s = c \quad \text{(dual feasibility)}$$

$$s + \sigma \mu_k F'(x) = 0 \quad \text{(generalizes } Xs = \sigma \mu_k e)$$

Thus, solve the Newton system

$$A\Delta x_k = 0$$
$$A^T \Delta y_k + \Delta s_k = 0$$
$$\mu_k F''(x_k) \Delta x_k + \Delta s_k = -\sigma \mu_k F'(x_k) - s_k$$

and set

$$x_{k+1} = x_k + \Delta x_k$$
$$y_{k+1} = y_k + \Delta y_k$$
$$s_{k+1} = s_k + \Delta s_k$$

It turns out that with, e.g., $\sigma = 1 - \frac{1}{\sqrt{\nu}}$, this simple algorithm again converges, with the same complexity bound as in the LO case, with $\nu$ replacing $n$. 

André Tits (UMD)

On Modern Convex Optimization

May 16, 2008 18 / 24
Outline

2. Simplest Case: Linear Optimization
3. General Case: Conic Optimization
4. Recent Accomishments at Maryland
Inexact function evaluation in conic optimization


In practice, it is often rather expensive, or even impossible, to evaluate \( F'(x) \) and \( F''(x) \) exactly.

Instead, consider solving the approximate Newton system

\[
A \Delta x_k = 0 \\
A^T \Delta y_k + \Delta s_k = 0 \\
\mu_k F_2(x_k) \Delta x_k + \Delta s_k = -\sigma \mu_k F_1(x_k) - s_k
\]

where the “errors” \( E_1 := F_1 - F' \) and \( E_2 := F_2 - F'' \) are small, in the sense that, for given \( \epsilon_1, \epsilon_2 > 0 \),

\[
\frac{\| E_1(x) \|_{x,F}^*}{\| F'(x) \|_{x,F}^*} \leq \epsilon_1 \quad \text{for all } x \in \text{int}(K), \\
\| F''(x)^{-1/2} E_2(x) F''(x)^{-1/2} \|_2 \leq \epsilon_2 \quad \text{for all } x \in \text{int}(K).
\]

Conditions were obtained, relating \( \beta, \sigma, \epsilon_1, \epsilon_2, \) and \( \nu \) (allowing for arbitrarily large \( \nu \)), under which the same polynomial complexity order is preserved.
Problems with many inequality constraints


Main results so far are for linear optimization. Consider the dual form

$$\max \{ b^T y : A^T y \geq c \}$$

with $A \in \mathbb{R}^{m \times n}$, with $n \gg m$, i.e., the dual problem has many more (inequality) constraints than variables. Such situation is common in practice (e.g., finely discretized semi-infinite problems).

Idea: At each iteration, select a small subset $Q$ of seemingly critical constraints, and construct a search direction based on those only. The “constraint-reduction” scheme should be implantable into popular variants of primal-dual interior-point algorithms.

Theoretical results: A simple rule for selecting $Q$ was identified (which leaves room for heuristics), under which global convergence, and local quadratic convergence, are guaranteed, under nondegeneracy assumptions, for algorithms including Mehrotra’s Predictor Corrector algorithm, the current “champion” algorithm.
Problems with many inequality constraints: Numerical tests

- $m = 200$, $n = 40000$
- Generate: $A$, $b$, $y_0 \sim \mathcal{N}(0, 1)$, $s_0 \sim \mathcal{U}(0, 1)$. Normalize columns of $A$.
- Set $c = A^T y_0 + s_0$ and $x_0 = e$. 

![Graphs showing CPU time and iterations vs. fraction of constraints kept]
Summary

- Simple primal-dual interior-points methods solve linear optimization problems in polynomial time.

- Convex optimization problems arise everywhere you look.

- Convex optimization problems are readily recast into conic optimization problems.

- Conic optimization problems are (formally) rather similar to linear optimization problems.

- “Minor” generalization of PD-IPMs for LO solve CO problems in polynomial time.
A solution to the quiz. The following is an example with $n = 3$ and $A \in \mathbb{R}^{2 \times 3}$:

$$K = \{(x, y, z) : (x - z)^2 + y^2 \leq z^2\}, \quad A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

i.e., $K$ is an (infinitely high) “ice cream cone” tangent to the vertical axis, and $A$ is the orthogonal projection onto the $(x, y)$ plane. It is readily checked that

$$AK = \{(0, 0)\} \cup \{(x, y) : x > 0\},$$

which is a convex, solid cone (as it should) but, clearly, it is not closed. It is also not pointed, so its closure is not a proper cone.

A take-home quiz: Come up with an instance where $AK$ is not closed (but $K$ is closed), and in addition its closure is a proper cone (closed, convex, solid and pointed).