The many facets of linear programming

Michael J. Todd
School of Operations Research
and Industrial Engineering
Cornell University
Ithaca, NY

miketodd@cs.cornell.edu
http://www.orie.cornell.edu/~miketodd/

ISMP Atlanta, August 10, 2000

Anniversaries

- (At 1997 ISMP) 50th of the simplex method
- 60th of Kantorovich's 1939 paper: "Mathematical Methods in the Organization and Planning of Production"
- 50th of 0th Mathematical Programming Symposium, Chicago 1949
- 45th of Frisch's 1955 suggestion of the logarithmic barrier function.
- 25th of the awarding of the 1975 Nobel Prize in Economics
- 20th of Khachiyan's 1979 and 1980 papers
- 15th of Karmarkar's 1984 paper.

Quotations

- Kantorovich: "I want to emphasize again that the greater part of the problems ... are connected specifically with the Soviet system of economy and ... do not arise in the economy of a capitalist society."

- Koopmans: "It has been found so far that, for any computation method which seems useful in relation to some set of data, another set of data can be constructed for which that method is obviously unsatisfactory."

- Dantzig: "This column geometry gave me the insight which led me to believe that the *simplex method* would be an efficient solution technique. I earlier had rejected the method when I viewed it in the row geometry because running around the outside edges seemed so unpromising."

Decades and Scope

- '50s: theory, industrial app'ns, combinatorial app'ns;

- '60s: large-scale, structure, quadratic programming and complementarity;

- '70s: computational complexity, Klee-Minty example.

We'll concentrate on developments since then: hope restored by new polynomial-time algorithms, by results on expected number of pivots, by amazing computational studies.

Two paradigms

- linear optimization over a simplex
  ($\rightarrow$ edge-following,
  combinatorial geometry).

- linear optimization over a ball
  ($\rightarrow$ solution by calculus,
  approximation of polyhedra by ellipsoids).

What are high-dimensional polyhedra like?
Quartz crystals or disco balls?
(See the next two slides.)

- the simplex method

- the ellipsoid method

- interior-point methods

- other methods

(Thanks to Jay Schomer: image from http://www.halcyon.com/nemain /gallery/gallery.html)

(Copyright 2000, 1999, eHow, Inc. From http://www.ehow.com/eHow/eHow/0,1053,4570,00.html)

An interesting time warp

Paper by Hoffman, Mannos, Sokolowsky, and Wiegmann (1953): comparison of three methods (fictitious play, relaxation, and primal simplex) on LPs from symmetric games (of size 5×5 up to 10×10). Simplex won, and could solve large-scale problems of size about 50×100!

Talk by Bixby on solving large MIPs (and LPs): comparison of three methods (primal and dual simplex, and barrier) on LPs, with results for an LP of size $49,944 \times 177,628$. The dual simplex method was the winner.

Problems and Notation

Primal problem:

$$\text{min}\{c^T x : Ax = b, x \geq 0\},$$

with A an $m \times n$ matrix of rank m.

Dual problem:

$$\text{max}\{b^T y : A^T y \leq c\}.$$

Also $s := s(y) := c - A^T y, \qquad d := n - m.$

The Simplex Method

Remarkable fact:
the (primal) simplex method typically requires at most $2m$ to $3m$ pivots to attain optimality.

Mentioned by Hoffman et al. (1953); numerical evidence through '50s and '60s.

More recently: Bixby (1991) on 90 Netlib problems gives a ratio of pivots to row size of < 3 on 72, between 3 and 7 for 16, and 10 to 470 on 3. Bixby (1994) on 8 large problems has < 2 on 3, 4 to 9 on 4, and 18 on the last.

Chosen to be nasty!

Why?

Diameter

Let $\Delta(d,n)$ denote the largest diameter of a d-polyhedron with n facets.

Hirsch conjecture: $\Delta(d,n) \leq n - d$.

Best bounds: $\Delta(d,n) \leq \min\{2^{d-3}n, n^{1+\log d}\}$.

Conjecture holds for 0-1 polytopes, dual transportation polyhedra, ... Fails for unbounded polyhedra in general, still open for polytopes.

Survey by Klee and Kleinschmidt (1987).

Subexponential pivot rules

Found by Kalai and Matousek-Sharir-Welzl. All (thus far) randomized, with a best bound of $\exp(K\sqrt{d \log n})$ for some constant K.

One version:

- Given a vertex v, choose a facet F containing v at random;
- Apply the algorithm recursively to find the optimizing vertex w in F;
- Repeat the algorithm from w.

Survey by Kalai (1997).

Probabilistic analysis

Hot topic in late '70s and '80s — much less activity recently.

Main results:

Borgwardt: expected number of pivots $O(m^3 n^{1/(m-1)})$ for a dimension-by-dimension simplex method for the dual problem. Vectors b and a_j's generated from rotationally symmetric distribution, all c_j's equal to 1.

Adler-Megiddo, Adler-Karp-Shamir, Todd: expected number of pivots $O(\min\{d^2, m^2\})$ for a lexicographic parametric simplex method. Problem generated from a sign-invariant model, possibly not feasible or unbounded.

Survey by Borgwardt (1986).

Big faces, long edges

Results of Kuhn (1953, 1991) and Goemans (1995) suggest that some interesting polytopes for combinatorial optimization can have relatively few big facets, and many small ones. By polarity, there may be polytopes with a very large number of vertices, but many of these may only be optimal for a very small set of objective functions, and so "relevant" for a very small set of simplex methods.

For $d \geq 4$, there are neighborly polytopes, with every pair of vertices adjacent.

For d from 3 to 11 and $n = 2^d$ and 2^{d+1}, I generated 100 random d-polytopes with n vertices. In many cases, the two maximally distant vertices were joined by an edge of the polytope.

These results suggest that the "quartz crystal" model may be reasonable.

The Ellipsoid Method

Originally for convex programming by Yudin-Nemirovski (1976) and Shor (1977) but adapted by Khachiyan for LP in 1979-80.

Created a whirlwind of publicity:
"Soviet Answer to 'Traveling Salesmen' "
"A Russian's Solution in Math Questioned."

Result: an LP with n inequalities and integral data with bit size L can be solved in $O(n^2L)$ iterations, requiring $O(n^4L)$ arithmetical operations on numbers with $O(L)$ digits (1980).

$LP \in P$!

Not a practical method, but highly useful theoretically: see Bland-Goldfarb-Todd (1981) and Grötschel-Lovasz-Schrijver (1988).

Traditional view and formulae

Method: start with ellipsoid known to contain solution, then cut in half and enclose in the smallest ellipsoid: repeat!

Assume we want a point in $Y := \{y : A^T y \le c\}$, with constraints $a_j^T y \le c_j,\ j = 1, ..., n$.

Assume $Y \subseteq E_0 := \{y : \|y\| \le R\}$.

Any ellipsoid can be written as

$$E(\bar{y}, B) := \{y : (y - \bar{y})^T B^{-1} (y - \bar{y}) \le 1\}.$$

Start with $E_0 = E(y_0, B_0)$ with $y_0 = 0$, $B_0 = R^2 I$.

Iteration k:
Given $E_k = E(y_k, B_k) \supseteq Y$,

Find j with $a_j^T y_k > c_j$ (if none, STOP: $y_k \in Y$);

Set $y_{k+1} := y_k - \frac{\tau B_k a_j}{(a_j^T B_k a_j)^{\frac{1}{2}}}$;

Set $B_{k+1} = \delta \left(B_k - \sigma \frac{B_k a_j a_j^T B_k}{a_j^T B_k a_j} \right)$.

Here $\tau = 1/(m+1)$, $\delta = m^2/(m^2-1)$, and $\sigma = 2/(m+1)$.

This gives $E_{k+1} = E(y_{k+1}, B_{k+1})$ as the minimum volume ellipsoid containing $\{y \in E_k : a_j^T y \le a_j^T y_k\}$.

The systematic volume reduction gives the complexity bound.

Improved formulae for double-sided or deep cuts: same as above but with different parameters.

Alternate representation

Since Y is assumed bounded, we can find lower bounds on each $a_j^T y$ for $y \in Y$.
So suppose

$$Y = \{y : \ell \leq A^T y \leq c\}.$$

Let D be a nonnegative diagonal matrix. Then

$$Y \subseteq \bar{E}(D, \ell) := \{y : (A^T y - \ell)^T D(A^T y - c) \leq 0\},$$

and this set is an ellipsoid if ADA^T is nonsingular. Note that D gives a short certificate that $\bar{E}(D, \ell)$ contains Y. From this viewpoint, the ellipsoid method generates a sequence $E_k = \bar{E}(D_k, \ell_k)$ of ellipsoids containing Y. The center of E_k is y_k, the solution of

$$AD_k A^T y = AD_k(\ell_k + c)/2.$$

Iteration k:
Given $E_k = \bar{E}(D_k, \ell_k) \supseteq Y$,

Find j with $a_j^T y_k > c_j$ (if none, STOP: $y_k \in Y$);

Possibly update the jth component of the vector ℓ_k to get ℓ_{k+1};

Increase the jth diagonal entry of the matrix D_k to get D_{k+1}.

Details in Burrell and Todd (1985).

Shows that the quadratic inequality defining each ellipsoid can be viewed as a weighted sum of quadratic constraints ensuring that each $a_j^T y$ lies in an appropriate range.

Indicates why convergence may be slow: adjusting one entry at a time of the diagonal of D to minimize the volume (like coordinate descent).

Interior-Point Methods

History and Pre-history: Frisch (1955), Fiacco and McCormick (1968).

Modern era: Karmarkar (1984),
2000 odd papers since.
"Breakthrough in Problem Solving,"
NYT (J. Gleick).
Closely related to (non-polynomial) affine-scaling method, first considered by I. I. Dikin, student of Kantorovich (1967)!

Polynomial convergence + practical importance

Projective method: $O(n^{3.5}L)$ arithmetic operations, much better in practice.

Path-following, potential-reduction methods.
Primal, dual, primal-dual.

Dual path-following

The projective method used a projective transformation at each iteration to center the current iterate. Modern methods use a local norm to make the iterate "look central."

Suppose $\bar{y}$ is a strictly feasible solution to (D): $\bar{s} := s(\bar{y}) > 0$. The largest ellipsoid centered at $\bar{s}$ in the nonnegative orthant is

$$\{s : (s - \bar{s})^T \bar{S}^{-2} (s - \bar{s}) \leq 1\},$$

and the corresponding set of y's

$$E := \{y : (y - \bar{y})^T A\bar{S}^{-2} A^T (y - \bar{y}) \leq 1\} \subseteq Y.$$

Now we have $\bar{y} \in Y$, and E is inscribed in rather than circumscribing Y, but otherwise remarkably similar to $\bar{E}(D, \ell)$ in the ellipsoid method.

The matrix $A\bar{S}^{-2}A^T$ appearing in E defines a local norm at $\bar{y}$. Unit ball is feasible! Note that the matrix is the Hessian at $\bar{y}$ of the logarithmic barrier function

$$f(y) := -\sum_j \ln(c - A^T y)_j.$$

Barrier and search directions

This is a special case of a

self-concordant barrier function,

introduced and studied by Nesterov and Nemirovski, and for these the local norm has some remarkable properties showing that the unit ball approximates the feasible region well.

The steepest ascent direction for the objective function is

$$d_{\mathsf{AFF}} := (A\bar{S}^{-2}A^T)^{-1}b,$$

while the steepest descent direction for the barrier function is

$$d_{\mathsf{CEN}} := -(A\bar{S}^{-2}A^T)^{-1}A\bar{s}^{-1},$$

with $\bar{s}^{-1}$ the vector whose components are the reciprocals of those of $\bar{s}$. Search directions in the dual-barrier method are linear combinations of these two.

Iterates are maintained in some neighborhood of the dual central path, the set of strictly feasible points where these two directions are diametrically opposed.

Note that computing the search direction requires the solution of a linear system with coefficient matrix ADA^T for some diagonal matrix D, exactly as in the ellipsoid method, but here all components of D vary from one iteration to the next.

Primal-dual methods are the methods of choice, with the primal "helping" the dual and vice versa. Similar linear systems arise.

The complexity of these methods is $O(\sqrt{n}L)$ iterations, first found by Renegar. In practice, $10 - 50$ iterations usually suffice.

Potential-reduction methods

Based on reducing a suitable potential function. Karmarkar used the primal function

$$\phi_P(x) := n \ln(c^T x - z_*) - \sum_j \ln x_j,$$

where z_* is the known optimal value of (P). Constant decrease at each iteration $\rightarrow$ the complexity bound of $O(nL)$ iterations.

The primal-dual potential function

$$\begin{aligned} \Phi_{PD}(x, y) \quad := \quad & (n + \sqrt{n}) \ln(c^T x - b^T y) \\ & - \sum_j \ln x_j - \sum_j \ln(c - A^T y)_j \end{aligned}$$

of Tanabe and Todd-Ye can also be decreased by a constant $\rightarrow O(\sqrt{n}L)$ iterations.

Note that potential-reduction methods achieve these complexities without requiring the iterates to remain close to the central path. The primal-dual search directions are similar to those for primal-dual path-following methods.

Other Methods

- Gradient-like methods: Brown and Koopmans (1951), Zoutendijk (1960), Lemke (1961), Rosen (1961), Chang and Murty (1989).

- Fictitious play: Brown (1951), Brown and von Neumann (1950), von Neumann (1947-1963, 1954).
  First complexity result: $O([m+n]/\epsilon^2)$.

- Relaxation: Agmon (1954), Motzkin-Schoenberg (1954); ellipsoid method is related (Goffin). SOR method of De Leone and Mangasarian (1988) has promising results.

- Ideas of comp. geom.: Megiddo (1984), Clarkson (1995), Gärtner-Welzl (1996).

The future

What will the next 50 (or 5?) years bring?

At present there is a rough computational parity between simplex and interior-point approaches.

Will our complacency in the status quo be shattered by another computationally effective class of methods?

I wouldn't bet on it in the next five years, but over the next ten, I'd take even odds.

Big questions: does the bounded Hirsch conjecture hold? Is there a polynomial pivot rule for the simplex method? For interior-point methods, can we give a theoretical explanation for the difference between worst-case bounds and observed practical performance?

Let us hope that the next fifty years brings as much excitement as the last!