Exponential Gaps in Mathematical Programming

Michael J. Todd

May 3, 2012

School of Operations Research and Information Engineering, Cornell University

http://people.orie.cornell.edu/∼miketodd/todd.html

Conference in honor of Mike Shub, Fields Institute, May 2012

1. Gaps ...

Duality gaps: convex problems, combinatorial optimization problems

Optimality gaps: how close to optimality can we solve problems in various classes?

Algorithmic gaps in our understanding and in bounds for iteration complexity of convex programming.

2. Algorithmic Gaps

Gaps between the worst-case and typical-case behavior of an algorithm (or in the theory supporting an algorithm).

Prime example: the simplex method for linear programming:

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

where A is $m \times n$.

For (almost) all (local) pivoting rules, there is a family of instances requiring an exponential (in the dimension) number or iterations.

In practice, for (almost) all instances, the number of iterations grows (almost) linearly in the smaller dimension of the problem.

An exponential gap!

3. Related Theory

There is an (almost) exponential gap between the upper and lower bounds known on the diameter of a polytope:

a d-polytope with n-facets has diameter at most

$$n^{2+\log(d)}$$

(Kalai and Kleitman);

there are pairs (d, n) and d-polytopes with n facets and diameter at least

$$\frac{21}{20}(n-d)$$

(Santos, improved by Matschke-Santos-Weibel).

How can we explain the good behavior of the simplex method in practice?

4. Probabilistic Analysis

There is a family of distributions on triples (A, b, c) (A an $m \times n$ matrix, b an m-vector, c an n-vector) so that

Theorem 1 *(Adler-Karp-Shamir, Adler-Megiddo, T., 1983) If the data for a linear programming problem is drawn from a distribution in this family, the expected number of iterations for a particular simplex variant to "solve" the instance is at most*

$$\min\{\frac{m^2 + 5m + 11}{2}, \frac{2d^2 + 5d + 5}{2}\},$$

where $d := n - m$.

There is related work by Smale, Borgwardt, Haimovich, and others.

5. Smoothed Analysis

Theorem 2 *(Spielman and Teng, 2004) For any* (A, b, c)*, if the data of a linear programming problem are drawn independently from Gaussian distributions centered at* (A, b, c) *with variances* σ^2*, then the expected number of iterations of a particular simplex variant to solve the problem is polynomial in* m*,* n*, and* $1/\sigma$*.*

This is a beautiful interpolation between worst-case and average-case analyses.

But now we have polynomial algorithms for LP! What's the big deal?

6. Polynomial Algorithms, I

The ellipsoid method of Yudin-Nemirovskii (1976) and Shor (1977), as applied to linear programming by Khachiyan (1979), obtains an ϵ-approximate solution to a linear programming problem in $O(d^2 \ln(1/\epsilon))$ iterations and $O(n\, d^3 \ln(1/\epsilon))$ arithmetic operations. (d is the dimension, n the number of inequalities.)

Polynomial, but it seems to need this many iterations, which is not competitive with the simplex method in practice. No exponential gap, but not the practical answer!

7. Polynomial Algorithms, II

Primal-dual interior-point methods obtain an ϵ-approximate solution to a linear programming problem in $O(\sqrt{n}\ln(1/\epsilon))$ or $O(n\ln(1/\epsilon))$ iterations and $O(n^{3.5}\ln(1/\epsilon)$ or $O(n^4\ln(1/\epsilon))$ arithmetic operations.

Polynomial, but in practice these algorithms seem to need a number of iterations which is either constant, or maybe grows logarithmically with n. This is why they are successful in practice.

Another exponential gap to be explained!

8. Polynomial Algorithms, III

Is this a real gap, or is the analysis too loose? We want lower bounds, as given by the exponential instances for the simplex method.

Megiddo-Shub (1989) showed that the affine-scaling algorithm gave rise to trajectories which could visit small neighborhoods of every vertex of the Klee-Minty cube. But this method had no polynomial bound.

T. (1993) and T.-Ye (1996) show that, for a large class of long-step primal-dual interior-point methods, the number of iterations required to decrease the duality gap by a constant is $\Omega(n^{1/3})$.

Deza, Nematollahi, Terlaky, and Zinchenko (2008-2009) show that the d-dimensional Klee-Minty cube can be defined using $n = O(d^3 2^{2d})$ constraints so that the central path visits small neighborhoods of every vertex, so that closely-path-following methods require

$$2^d = \Omega\left(\sqrt{\frac{n}{\ln n}}\right)$$

iterations.

Thus it seems the upper bound is (close to) tight in the worst case!

9. Polynomial Algorithms, IV

There have been attempts to mirror the probabilistic or smoothed analysis of the simplex method.

Nemirovskii (1987) for the projective algorithm, and Gonzaga and T. (1992) and Mizuno, T., and Ye (1993) for primal-dual algorithms, gave "plausibility" arguments that for "most" problems, the number of iterations required would be $O(\ln n \ln(1/\epsilon))$.

There have also been smoothed analyses of the termination criteria or of condition numbers arising in the complexity of interior-point methods (Spielman, Teng, and others).

Dedieu, Malajovich, and Shub (2005) showed that the average curvature of the dual central paths in all the bounded feasible regions corresponding to sign switches is at most

$$2\pi m$$

(improved slightly by De Loera, Sturmfels, and Vinzant).

10. First-order methods

There are also exponential gaps in the dependence of the iteration complexity of certain algorithms on the accuracy ϵ:

The minimum-volume ellipsoid problem asks for the smallest d-dimensional ellipsoid centered at the origin that contains a set of n points, and arises in computational geometry and, via its dual, in optimal experiment design in statistics.

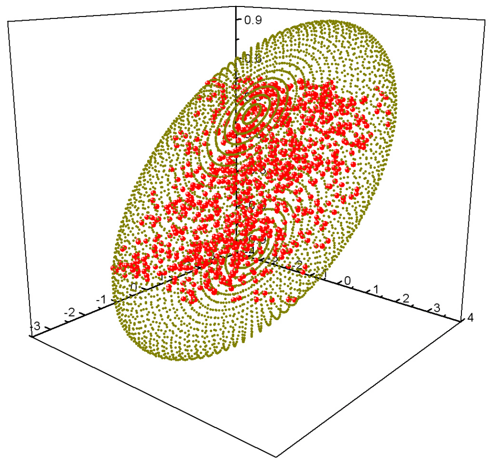

Figure 1: The minimum-volume ellipsoid problem.

11. Complexity Results

Khachiyan (1996) showed that a variant of the Frank-Wolfe algorithm (also developed by the statisticians Fedorov and Wynn) could obtain a $d(1+\epsilon)$-rounding of the points in

$$O(nd^2(\frac{1}{\epsilon} + \ln d + \ln\ln n))$$

arithmetic operations.

Ahipasaoglu, Sun, and T. (2008) showed that a variant of this method had linear convergence, so that it ultimately required only

$$O\left(\ln \frac{1}{\epsilon}\right)$$

iterations.

Can a rigorous global bound with such a dependence on ϵ be proved?

12. Convergence like $\ln(1/\epsilon)$

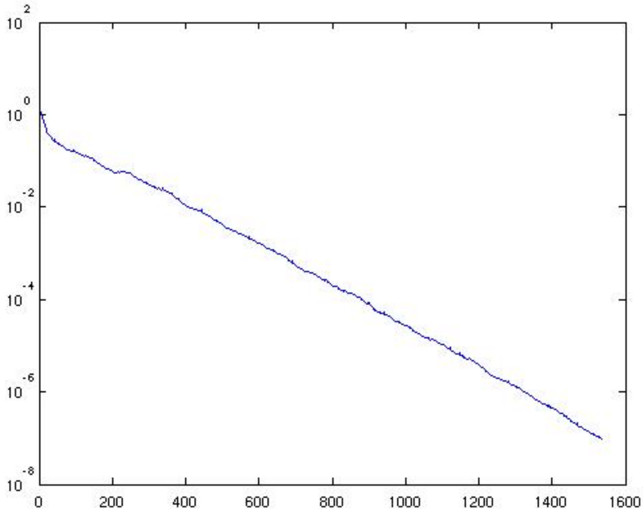

Figure 2: Linear convergence of the error.

13. Conclusion

There are several intriguing challenges in optimization to explain the excellent behavior of certain algorithms in practice by removing the exponential gaps in our understanding!