Our nondegeneracy assumption (a)(i) fails for LCPs arising from linear programming, but we would like results in this case also.

Consider the linear program (P):

$$\begin{aligned} \min \left\{c^T x\right\} \\ \text{subject to} \\ Ax \geq b \\ x \geq 0, \end{aligned}$$

and its dual (D):

$$\begin{aligned} \max \left\{b^T y\right\} \\ \text{subject to} \\ A^T y \leq c \\ 0 \leq y, \end{aligned}$$

where A is $m \times p$. If (P) and (D) are generated from any probability distribution satisfying some nondegeneracy assumptions, and the sign invariance property that for all sign matrices S_b and S_c,

$$(S_b A S_c, S_b b, S_c c) \sim (A, b, c)$$

(this is equivalent to switching the directions of the inequalities in (P) and (D) all possible ways), the expected number of steps for the lexicographic Lemke algorithm to "solve" the corresponding LPs (show infeasible, unbounded, or find optimal solutions) is bounded by a quadratic in $\min\{m, p\}$. This was proved independently by Adler and Meggido, by Adler, Karp, and Shamir, and by Todd in 1983. Later, Adler and Megiddo gave a quadratic lower bound. These results are part of a great deal of research in expected behavior of the simplex method, notably by Smale and Borgwardt (see Borgwardt's book *The Simplex Method: A Probabilistic Analysis*). Unfortunately, all of these probabilistic models are far from realistic.

For example, look at the sign-invariant model with $m = p = 2$ and $n = 4$, and consider all $2^4 = 16$ instances arising from all choice of directions of the inequalities for a particular instance. Then, as seen in the figure below, there are only 11 problems which have feasible regions out of the 16; 5 are infeasible. Of these 11 problems, only 6 have optimal solutions. (This is not a coincidence: the numbers 11 and 6 are partial sums of binomial coefficients.)

- Suppose $p = \alpha m$, $\alpha > 1$, and $p, m \to \infty$; then almost all problems (D) are infeasible.
- Suppose $m = \alpha p$, $\alpha > 1$, and $p, m \to \infty$; then almost all problems (D) are unbounded.

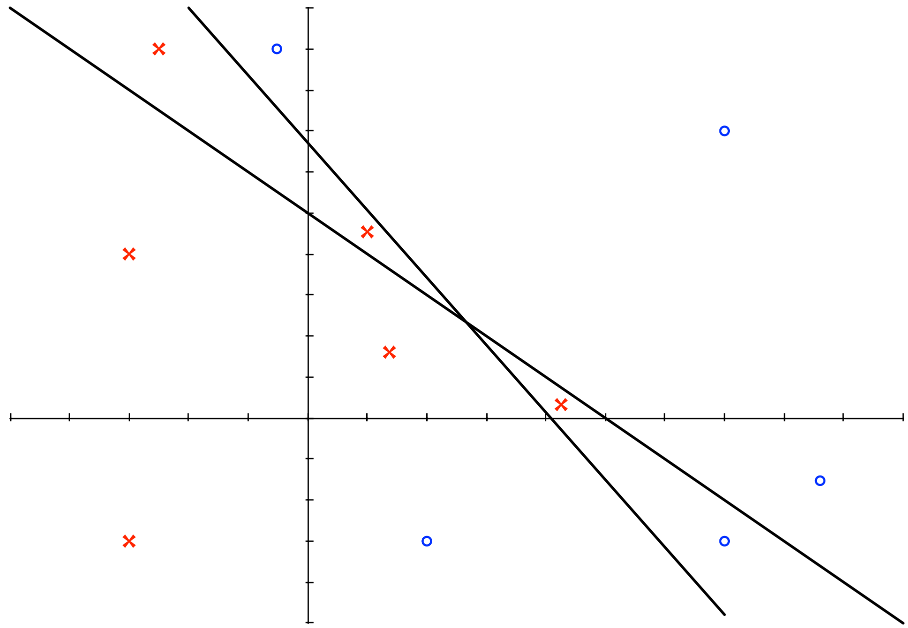

Figure 1: Consider a LP where the direction of maximization is from left to right, and slightly up. The red x's mark feasible combinations of constraints for which there is an optimal solution, and the blue o's mark feasible combinations for which the objective is unbounded above.

- Suppose $m = p$ and $m, p \to \infty$; then the probability that (P) and (D) have optimal solutions is $\Omega(\frac{1}{\sqrt{m}})$, and there is a bound on the number of pivots, *conditional* on the problem having an optimal solution, of $O(m^{2.5})$.

Is it possible to obtain more practical results?

1. Smoothed Analysis (Spielman and Teng):

Theorem 1 *(S&T, 2004) There is a two-phase simplex method such that, for any $A \in \mathbb{R}^{m \times p}$, b, and c, if we generate all entries of G and h independently from $\mathcal{N}(0, \max \|(A_j, c_j)\|^2 \sigma^2)$, then the expected number of steps to solve*

$$\max\{b^T y\}$$
$$\text{subject to,}$$
$$(A+G)^T y \leq c + h$$

is polynomial in m, p, and $\frac{1}{\sigma}$.

This is a beautiful interpolation between the worst case ($\sigma \to 0$, bound $\to \infty$), and the average case ($\sigma \to \infty$, bound $\to$ poly(m, p)). But, the polynomial is huge: $p^{86} m^{55} \sigma^{-30}$. This bound was improved to $O(\log^7(p)(m^9 + m^3\sigma^{-4}))$ by Vershynin (2009).

2. Randomized Simplex Methods, Worst-Case Instance

Theorem 2 *(Kalai, Matousek-Sharir-Welzl, 1997) There is a randomized simplex method algorithm that solves*

$$\max \left\{ b^T y \right\}$$
subject to,
$$A^T y \leq c,$$

where A is $m \times p$, in an expected number of steps $O(\exp(K\sqrt{m \log p}))$ for some constant K.

In contrast, the bound on the diameter we obtained before is $O(\exp(K \log m \log(p)))$. See Kalai's paper *Linear programming, the simplex method, and simple polytopes* on the homepage.

1 Informational Complexity of Nonlinear Programming

How hard is it to find $\varepsilon-$approximate solution to a nonlinear programming problem, where access to "items" in the problem (e.g., sets, functions) is by calls to an *oracle*? We measure the complexity by the number of oracle calls. The types of bounds are:

1. Polynomial in dimension and in $\frac{1}{\varepsilon}$.
   (a) For example, von Neumann had an algorithm for approximating the value of a two person zero-sum $m \times n$ game in $O(\frac{m+n}{\varepsilon^2})$ work.
2. Polynomial in dimension and $\ln \frac{1}{\varepsilon}$.
   (a) The ellipsoid method with $O(n^4 \ln \frac{1}{\varepsilon})$ work is an example, or interior point methods (but the latter are very different from oracle methods). Here linear in $\ln \frac{1}{\varepsilon}$ corresponds to (global) linear convergence.
3. A function of the instance plus polynomial in dimension and $\ln \ln \frac{1}{\varepsilon}$.
   (a) Newton's method is a good example. The "$\ln \ln$" corresponds to quadratic convergence.

We'll see that the situation is hopeless unless the objective function to be minimized is convex. Recall, $f : D \to \mathbb{R}$, with $D \subseteq \mathbb{R}^n$ convex, is convex if

$$f(\lambda x + (1-\lambda)y) \leq (1-\lambda)f(x) + \lambda f(y), \quad \forall x, y \in D, \lambda \in (0,1),$$

and also that $g \in \mathbb{R}^n$ is a subgradient of f at $\bar{x} \in D$ if

$$f(x) \geq f(\bar{x}) + g^T(x - \bar{x}), \quad \forall x \in D.$$

For an example of a convex function and its subgradient provided by an oracle, consider Lagrangian relaxation. Suppose we wanted to solve

$$(P) \qquad \max \left\{ b^T y : A^T y = c, y \in Y \right\},$$

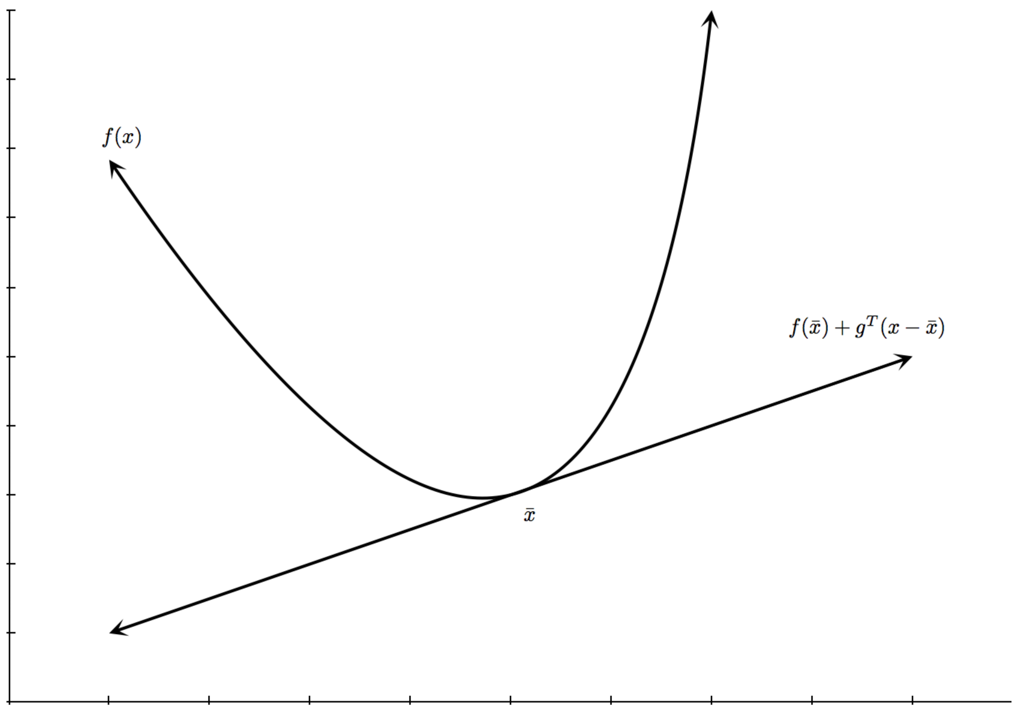

Figure 2: The subgradient of a function f at the point $\bar{x} \in D$.

where it is easy to solve, for any $v \in \mathbb{R}^m$, $\max\left\{v^T y : y \in Y\right\}$. Here $Y \in \mathbb{R}^m$ and A is $m \times n$. Choose $x \in \mathbb{R}^n$ and consider

$$(P(x)) \qquad f(x) := \max\left\{b^T y - x^T(A^T y - c) : y \in Y\right\}.$$

This is easy to solve by assumption, and is a relaxation of (P): any y feasible for (P) is feasible for $(P(x))$ with the same objective value. So the value of (P) is at most $f(x)$ for all x. To get the best possible bound, solve $\min\{f(x)\}$. This can be a very powerful technique in combinatorial optimization. Then f is convex (it is a pointwise max of linear functions). Solving $(P(x))$ gives us $f(x)$, as well as:

- $g(x) := c - A^T y$, a subgradient of f at x, where y solves $(P(x))$.

Indeed, suppose $\bar{y}$ solves $(P(\bar{x}))$. Then for all x,

$$\begin{aligned} f(x) &\geq b^T\bar{y} + (c - A^T\bar{y})^T x \\ &= b^T\bar{y} + (c - A^T\bar{y})^T\bar{x} + (c - A^T\bar{y})^T(x - \bar{x}) \\ &= f(\bar{x}) + \bar{g}^T(x - \bar{x}), \qquad \text{where } \bar{g} = c - A^T\bar{y}. \end{aligned}$$