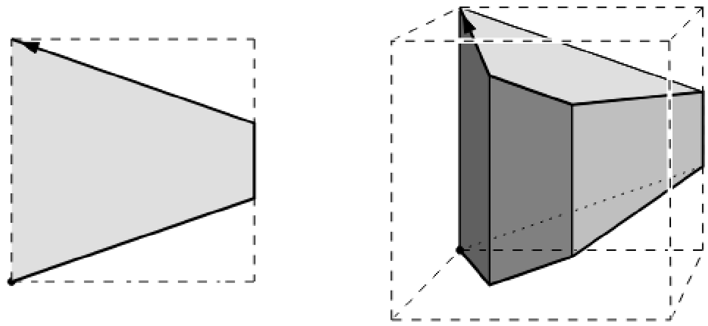
E. Bach, S. Huiberts: Optimal Smoothed Analysis of the Simplex Method
presented by Martin Koreček

Theorem (and lemma, ...) numbers are kept as they appear in the paper. Other results and definitions are labelled alphanumerically.

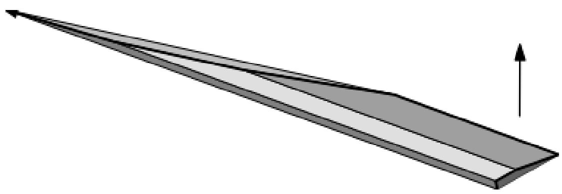
Definition: A Linear Programming problem instance is described by input $A \in \mathbb{R}^{n\times d},\ b \in \mathbb{R}^n,\ c \in \mathbb{R}^d$ and is written as

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & Ax \le b. \end{aligned} \tag{LP}$$

Fact A: Exponential lower bounds exist for simplex methods (almost every major pivot rule suffers from this). They are based on *deformed products*. To illustrate this, consider the two- and three-dimensional Klee-Minty cube:



The idea is that the simplex method needs to visit **all** 2^d **vertices** to traverse from the bottom-most vertex to the topmost one.
This is not enough however. To fool (e.g.) Bland's pivot rule, we need to deform the cube further:



These illustrations are taken from ***Matoušek: Lineární programování pro informatiky****.*

Fact B: Worst-case polynomial algorithms *do exist* for LP. While some are very inefficient in practice (such as the ellipsoid method), others (notably primal-dual interior-point methods) perform well on practical instances.

Fact C: Decades of evidence show that simplex methods take a number of pivot steps *proportional to the dimension d* on practical instances.

In practice (in fast commercial solvers such as CPLEX, Gurobi, FICO), an **exponential algorithm** (the simplex method) is used **instead of efficient and provably polynomial** algorithms!

Definition D: Smoothed analysis:
We imagine that a base LP problem is adversially constructed (i.e. arbitrarily)

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & \overline{A}x \le \overline{b} \end{aligned}$$

and scaled in such a way that the rows of the combined matrix $(\overline{A}, \overline{b})$ each have (Euclidean) norm at most 1. This input data gets randomly perturbed: for a parameter $\sigma > 0$, one samples $\hat{A}$ and $\hat{b}$ with independent entries each drawn from a Gaussian distribution with mean 0 and variance σ^2. The *smoothed complexity* of an algorithm is the expected running time to solve the perturbed problem

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & (\overline{A} + \hat{A})x \le \overline{b} + \hat{b}. \end{aligned} \tag{Input LP}$$

In other words, our LP instance corresponds to input data $A = \overline{A} + \hat{A},\ b = \overline{b} + \hat{b},\ c$.

Algorithm

Definition E: Basic feasible solution:
Given a basis $I \in \binom{[n]}{d}$ we write the corresponding solution as $x_I = A_I^{-1} b_I$.
The set $F(A, b) \subseteq \binom{[n]}{d}$ consists of all feasible bases.

Algorithm 1 Shadow vertex method $\textsc{ShadowVertex}$(A, b, y, y', I)

Input: non-degenerate polyhedron $= \{x \in \mathbb{R}^d : Ax \leq b\}$
objective functions $y, y' \in \mathbb{R}^d$
feasible basis $I \subseteq [n]$, optimal for y
Output: basis $I \subseteq [n]$ optimal for y' or *unbounded*

$i \leftarrow 0$ // Iteration counter
$\lambda_i \leftarrow 0$ // Shadow progress
while $\lambda_i \neq 1$ **do**
  $i \leftarrow i + 1$
  $\lambda_i \leftarrow$ supremum λ such that $y_\lambda^\top A_I^{-1} \geq \mathbf{0}$ // Maximal λ such that I is optimal for $\lambda y' + (1 - \lambda) y$
  if $\lambda_i \geq 1$ **then**
    return I // If basis is optimal for y, return said basis
  end if
  $j \leftarrow j \in I$ such that $(y_\lambda^\top A_I^{-1})_j = 0$ // Pivot rule. Will be unique
  $x_I \leftarrow A_I^{-1} b_I$
  $s_i \leftarrow$ supremum over all s such that $A(x_I - s A_I^{-1} e_j) \leq b$ // Find simplex step length s
  if $s_i = \infty$ **then**
    return *unbounded*
  end if
  $l \leftarrow l \in [n] \setminus I$ such that $a_l^\top (x_I - s_i A_I^{-1} e_j) \leq b_l$ // Ratio test. Will be unique
  $I \leftarrow I \setminus \{l\} \cup \{j\}$
end while

Definition 15: For linearly independent $c, c' \in \mathbb{R}^d$, the subset $P(A, b, c, c') \subseteq F(A, b)$ is called *shadow path* from c to c', and consists of all bases I such that x_I maximizes an intermediate objective $\max y^T x_I$ for $y \in [c, c']$.
The vertices v_1, v_2 maximizing c or c' are called *endpoints*.

Initialization

1. Sample $Z \in \mathbb{R}^d$ with independent entries, solve auxiliary LP

$$\begin{aligned} \max \quad & Z^T x \\ & Ax \leq 1 \\ & (Rs_i)^T x \leq 1 \qquad \forall i = 1, \ldots, d. \end{aligned} \tag{Unit LP'}$$

The zero vector is strictly feasible. Let $\bar{s}_1, \ldots \bar{s}_d$ be the vertices of a regular $(d-1)$-simplex such that $e_d^T \bar{s}_i = 3$ and $\|e_d - \bar{s}_i\| = \frac{1}{10\sqrt{\ln d}}$ for each i. Each s_i is sampled randomly with mean $\bar{s}_i$ and deviation $\sigma > 0$, R is a uniformly random rotation matrix.

Lemma 22: With probability ≥ 0.3, (Unit LP') admits an optimum x^* with $(Rs_i)^T x^* \leq 0$ for all $i \in [d]$.

Lemma 23: Conditional on the rows of A having norm at most 2, with probability ≥ 0.9, independent of A, the basic solution $(RS)^{-1}1$ is feasible and satisfies $(Re_d)^T (RS)^{-1} 1 \geq 0$.

So with constant probability, we can use the shadow vertex method with starting basic solution $(RS)^{-1}1$ optimizing fixed objective Re_d, to optimize for random objective Z and optain the optimum for (Unit LP).

$$\begin{aligned} \max \quad & Z^T x \\ & Ax \leq 1 \end{aligned} \tag{Unit LP}$$

2. Sample a $(d+1)$-th coordinate $Z_{d+1} \in \mathbb{R}$ for $Z \in \mathbb{R}^d$ (such that Z_{d+1} is standard Gaussian distributed). Denote $Z' = (Z, Z_{d+1})$. Bases $I \in \binom{[n]}{d}$ to (Unit LP) index edges in the *interpolation LP*:

$$Ax + (1-b)t \leq 1. \tag{Int-LP}$$

We consider objectives $\min t, \max Z'^T x, \max t$.

Our optimal basis for (Unit LP) indexes a set of constraints that is tight for some edge of (Int-LP) that passes through the $t = 0$ slice. Both endpoints of this edge are part of the *combined shadow path*

$$P\left((A,(1-b)), 1, -e_{d+1}, Z'\right) \cup P\left((A,(1-b)), 1, Z', e_{d+1}\right).$$

We can therefore use the shadow vertex method for (Int-LP) to find a basic feasible solution to (Input LP) which is optimal for the random objective $\max Z$ (*or certify that* (Input LP) *is infeasible*).

Now we can use the shadow vertex method to optimize from Z to c.
Goal: show that the shadow vertex auxiliary initialization runs, and the one from Z to c, all require

$$\mathcal{O}\left(\sqrt{\sigma^{-1}\sqrt{d^{11}\log^7 n}}\right)$$

pivot steps in expectation.

Upper bound

Definition 28: Given $A \in \mathbb{R}^{n\times d}$ and c, c', and a threshold $m > 0$, the set of bases with *good multipliers* is

$$M(A, c, c', m) = \left\{ I \in \binom{[n]}{d} \mid \exists y \in [c, c'] \text{ s.t. } yA_I^{-1} \geq m \right\}.$$

Definition 33: For $A \in \mathbb{R}^{n\times d}, b \in \mathbb{R}^n$, define the set of *feasible bases with relative gap* $g > 0$ as

$$G(A, b, g) = \left\{ I \in F(A, b) \mid A_{[n]\setminus I} x_I \leq b_{[n]\setminus I} - g \cdot \|x_I\| \right\}.$$

Definition 35: For a graph $G = (V, E)$ and $S \subseteq V$, write $T(S)$ for the vertices $v \in S$ that have at least 2 neighbors in S.

Theorem 37: *"Most vertices of the shadow path are in $T(G(A, b, g) \cap M(A, c+Z, c'+Z, m))$"*
Let $A \in \mathbb{R}^{n\times d}$ have independent Gaussian distributed entries, each with standard deviation $\sigma \leq \frac{1}{4\sqrt{d \ln n}}$. Assume that the rows of $\mathbb{E}[A]$ have norm ≤ 1. Take $b \in \mathbb{R}^n$ and $c, c' \in \mathbb{R}^d$ to be fixed. If $Z \in \mathbb{R}^d$ has a 1-log-Lipschitz probability density function then

$$\mathbb{E}\left[|P(A, b, c+Z, c'+Z)|\right] \leq 500 + 2\mathbb{E}\left[\left|T(G(A, b, \frac{\sigma}{5000 d^{3/2}\ln(n)^{3}/2}) \cap M(A, c+Z, c'+Z, \frac{\ln(1/0.99)}{2d}))\right|\right].$$

Theorems 51, 52: Let the constraint matrix $A \in \mathbb{R}^{n\times d}$ have independent Gaussian distributed entries, each with standard deviation $\sigma > 0$. Let $c \in \mathbb{R}^d$ be arbitrary and fixed. If $Z \in \mathbb{R}^d$ has a 1-log-Lipschitz probability density function that satisfies $\Pr[\|Z\| \geq 2ed\ln(n)] \leq n^{-d}$ and is independent of A,

- if the rows of $\mathbb{E}[A]$ each have norm at most 1, then the semi-random shadow path on $\{x : Ax \leq 1\}$ has length bounded as

$$|P(A, 1, Z, c)| \leq \mathcal{O}\left(\sqrt{\frac{1}{\sigma}\sqrt{d^{11}\log^7 n}} + d^3\log(n)^2\right),$$

- if also the vector $b \in \mathbb{R}^n$ has independent Gaussian distributed entries, each with standard deviation σ, and A, b such that the rows of $\mathbb{E}[(A, b)]$ each have norm at most 1, then the semi-random shadow path on $\{x : Ax \leq b\}$ has length bounded as

$$|P(A, b, Z, c)| \leq \mathcal{O}\left(\sqrt{\frac{1}{\sigma}\sqrt{d^{11}\log^7 n}} + d^3\log(n)^2\right).$$

Lower bound

We construct a hard instance $\max c^T x;\ \ \overline{A}x \leq 1$.

Definition 52: For $\eta > 0$, a set $S \subset \mathbb{S}^{d-1}$ is called η-dense if for any $x \in \mathbb{S}^{d-1}$ there exists $s \in S$ such that $\|x - s\| \leq \eta$.

Fact F: There exists an η-dense set $S \subset \mathbb{S}^{d-1}$ with cardinality $|S| \leq (4/\eta)^d$.

Lemma 54: Let $\{s_1, \ldots, s_n\} \subset \mathbb{S}^{d-1}$ be η-dense for $\eta \leq \frac{1}{8}$ and let the rows of $A \in \mathbb{R}^{n \times d}$ satisfy $\|a_i - s_i\| \leq \eta$ for every $i \in [n]$. Given a vector $b \in [1-\eta, 1+\eta]^n$, the polyhedron $\{x \in \mathbb{R}^d :\ Ax \leq b\}$ satisfies

$$(1-2\eta)\mathbb{B}^d \subseteq \{x \in \mathbb{R}^d :\ Ax \leq b\} \subseteq (1+4\eta)\mathbb{B}^d.$$

Definition G: The polar set of $P \subseteq \mathbb{R}^d$ is $P^\circ = \{y \in \mathbb{R}^d :\ \langle x, y\rangle \leq 1\ \ \forall x \in P\}$.

Lemma 55: For $d \geq 2, R > 0$, let $P \subseteq R \cdot \mathbb{B}^d$ be a simple bounded polytope containing the origin in its interior.
If every facet of P° has geometric diameter at most $\gamma > 0$, then *for any unit-length objective vector* $c \in \mathbb{S}^{d-1}$, any maximizing vertex $v^+ \in P$ and any minimizing vertex $v^- \in P$, there is no simplex path from v^- to v^+ of length less than $(d-1)(\frac{2}{R\gamma} - 2)$.

Theorem 56: Given $d \geq 2$ and $\sigma > 0$ such that $\sigma\sqrt{\ln(4/\sigma)} \leq \frac{1}{32d}$, we form adversial bounds

$$\overline{A}x \leq 1$$

by having the rows $\overline{a}_1, \ldots, \overline{a}_n$ for $n = \lceil (4/\sigma)^d \rceil$ form a σ-dense set as guaranteed by **Fact F**.
If A, b have their entries independently Gaussian distributed with variance σ^2 and expectation $\mathbb{E}[A] = \overline{A},\ \mathbb{E}[b] = 1$, then the combinatorial diameter of the polytope $\{x :\ Ax \leq b\}$ satisfies

$$\Pr\left[\operatorname{diam}(\{x :\ Ax \leq b\}) \geq \frac{\sqrt{d-1}}{24\sqrt{\sigma\sqrt{\ln(4/\sigma)}}}\right] \geq 1 - n^{-d}.$$

Moreover, with probability at least $1 - n^{-d}$, for any fixed nonzero objective vector $c \in \mathbb{R}^d$, its minimizing and maximizing vertices attain this lower bound.

Conclusion: At least one of the paths $P(A, b, -c, Z), P(A, b, Z, c)$ of any combined semi-random shadow path $P(A, b, -c, Z) \cup P(A, b, Z, c)$ must have length at least $\frac{\sqrt{d-1}}{48\sqrt{\sigma\sqrt{\ln(4/\sigma)}}}$. In particular there must exist a nonzero objective c such that

$$\mathbb{E}\left[|P(A, b, c, Z)|\right] \geq \frac{\sqrt{d-1}}{96\sqrt{\sigma\sqrt{\ln(4/\sigma)}}}.$$

The upper bound shown therefore has optimal noise dependence (*up to a polylog factor*) in the sense that any upper bound on the shadow path length of the form $\operatorname{poly}(d, \sigma^{-1}, \log n)$ must have a monomial term with dependence at least $\sigma^{-1/2}$.