UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Probabilistic analysis of simplex algorithms

Karl-Heinz Borgwardt

Angaben zur Veröffentlichung / Publication details:

Borgwardt, Karl-Heinz. 2001. "Probabilistic analysis of simplex algorithms." In *Encyclopedia of optimization*, edited by Christodoulos A. Floudas, 368–79. Dordrecht: Kluwer. https://doi.org/10.1007/0-306-48332-7.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

Probabilistic Analysis of Simplex-Algorithms

KARL HEINZ BORGWARDT

1 Introduction

Solving Linear Programming Problems is of enormous relevance in real world applications, which contain a lot of data and of unknown variables. Hence, the computational efficiency of solution methods is a crucial criterion for their applicability.
Today, we have a competiton between the Simplex-Method (invented around 1947 by George Dantzig) and Interior-Point-Methods (starting with Karmarkar's Algorithm 1984).
This article concentrates on Simplex-Methods and on an investigation of their arithmetical effort, measured in terms of the average number of pivot steps.
Throughout the paper we discuss the following type of Linear Programming Problems

$$\begin{array}{l} \text{maximize } v^T x \\ \text{subject to } a_1^T x \leq b^1, \ldots, a_m^T x \leq b^m \\ \text{where} \qquad v, a_1, \ldots, a_m \in \mathbb{R}^n, b \in \mathbb{R}^m \text{ and } m \geq n. \end{array} \tag{1}$$

For abbreviation we use $A := \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \in \mathbb{R}^{m \times n}$ and $b := \begin{pmatrix} b^1 \\ \vdots \\ b^m \end{pmatrix}$.

The matrix A collects the m gradient vectors to the restrictions as row-vectors, and the vector b gives the m capacities. $X := \{x \mid Ax \leq b,\ x \in \mathbb{R}^n\}$ is the feasible region, resp. the feasible polyhedron to the problem (1), which can be also be written in the form

$$\text{maximize } v^T x \text{ subject to } Ax \leq b\,. \tag{2}$$

Other types of programs as

$$\text{maximize } v^T x \text{ subject to } Ax \leq b, x \geq 0 \tag{3}$$

$$\text{maximize } v^T x \text{ subject to } Ax = b, x \geq 0 \tag{4}$$

and hybrids or variations of such forms can easily be translated into (1). But for form (1) our discussion on the influence of distributions, dimensions and variants can be made much better in geometrical, verbal terms. All the stated results

hold – after adaption – for the other forms, too.
If X has vertices and if there are optimal solutions to (1), then there is a vertex in the optimal set. In each vertex of X at least n restrictions of (1) will be active or tight. And in each edge of X at least $n-1$ restrictions are active. Every nonoptimal vertex is incident to an edge improving the objective. And if an optimal vertex exists, every iterative construction of a connected path over such improving edges leads to the optimal vertex after a finite number of steps. These facts are exploited in the design of the Simplex-Method, which works in two Phases.

> **Phase I:** Find a vertex $x_0 \in X$. If there is no vertex, STOP.
> **Phase II:** Construct a sequence of vertices $x_0, \ldots, x_s \in X$, such that for $i = 0, \ldots, s-1$ the vertices x_i and x_{i+1} are adjacent and $v^T x_i < v^T x_{i+1}$.
> We stop at x_s if x_s is optimal or if at x_s the nonexistence of an optimal solution becomes obvious. (5)

Phase I works in a similar manner to Phase II. Since Phase II admits a better geometrical explanation, and is simpler to analyze, we concentrate – for the beginning – on Phase II.
Note that our definition of Phase II in (5) still gives the freedom, how we determine the successor vertex (if more then one are possible). A rule for that decision will fix a "variant" of the Simplex-Algorithm. The complexity of Phase II – the so-called "Simplex-Algorithm" – is mainly determined by the number s. Less difficult to analyze is the effort to perform a single pivot step, which costs at most $O(mn)$ arithmetic operations for updating an $(m \times n)$-tableau under all reasonable variants.
In this article we are interested in the average-case behaviour of the random number s, when our problems (1) follow a given distribution. Since nobody knows "the real world distribution", we have to introduce and to use a self-made stochastic model about the appearance of special instances of (1).

Based on that model, we will evaluate the stochastic behaviour of s. It is clear that this will massively depend on

- the variant under use
- the chosen stochastic model/resp. the distribution.

A **probabilistic analysis** of the behavior of an algorithm consists of three essential steps.

- a study of the way the algorithm is working on given, deterministic problem-instances including a characterization of the desired figures (e.g. s) for that instance
- a consensus about an underlying stochastic model on the distribution of occuring problem-instances

- a cumulation over all possible instances, weighted with their occurence probability, leading to stochastic information on the random behavior

So, we study the procedure of a deterministic algorithm, which is employed to solve random problem-instances.
This stands in contrast to the situation with **randomized algorithms**, where random parameters decide, how the algorithm shall proceed in solving a given, deterministic problem.

Throughout the paper we shall rely on a nondegeneracy assumption:

$$\text{All submatrices of } (A,b) \text{ and of } (A^T,v) \text{ are of full (maximal) rank.} \tag{6}$$

This is compatible with our models either by direct conditioning or by the fact that in such a probabilistic model the set of degenerate problems is a nullset.

In this paper, we shall briefly report on experiments and their (limited) information-value. After that we come to two different stochastic models which admitted a successful probabilistic analysis. The first is the Sign-Invariance-Model, whose analysis reached its summit in the middle of the eighties. And the second is the Rotation-Symmetry-Model, whose evaluation had started even earlier. But the refinement of that approach is still going on.

2 Numerical experiments and comparison of variants

The first idea to learn more about the average-case-behaviour of s is to carry out controlled numerical (Monte-Carlo-)experiments. For that purpose, one has to fix several dimension-pairs (m,n), to use a stochastic model for generating the data, and to solve the created problem-instances by application of a given variant.
These experiments can be employed for a variety of purposes, as for forecasting the number of pivot steps s (on the basis of (m,n) for a fixed variant and model) or for comparing different variants or for recognizing the different influences of stochastic models.
All of that had been done and tried in the past. Studying the huge number of reports on such experiments leads to a very confusing and frustrating impression. Since stochastic models, employed variants, dimensions and problem-types vary excessively, the results and methods can hardly be compared. In particular, it is not possible to summarize the outcome briefly, since all the test parameters would have to be mentioned exactly. So, we refer to the very informative survey

by Shamir in [19], who comes to the following overall conclusion.

> The smaller dimension n (resp. the dimension of the polyhedra) enters the mean-value function of s in a slightly superlinear way and the larger dimension m (resp. the number of inequality-restrictions, including sign-constraints) has only a significantly sublinear influence. (7)

Easier to understand and to interpret are experiments, when they are done parallel to a theoretical study, because then both results, the empirical and the theoretical one, can be checked whether they justify and confirm each other. This has been achieved in experiments for the so-called Rotation-Symmetry-Model (RSM).

> Let $b = \mathbb{1}$ and let $a_1, \ldots, a_m$, v (and an auxiliary vector u) be distributed independently, identically and symmetrically under rotations on $\mathbb{R}^n \setminus \{0\}$. (8)

Note that only $b > 0$ is essential. Choosing $b = \mathbb{1}$ means a simplifying standardization only.

The experiments could more or less confirm the theoretical results on $E_{m,n}(s)$ (the expected number of pivot steps required for (m, n)-problems). These theoretical results will be presented later. Rather informative was the comparison of the behaviour of different variants and of the influence of different stochastic distributions. In [8] we tested seven variants belonging to three categories (A, B, C), whose geometric description can be given as follows.
Note that in each vertex a decision has to be made, which one of the (exactly) n tight restrictions should be deactivated. This means that a choice among the subset of the improving edges (originating from that vertex) is made. The current basis is the set of the n gradients (a_i) corresponding to the active restrictions at the current vertex.

Category A: Variants exploiting information on the shape of X and on the objective $v^T x$.

> **Steepest Ascent:** Chooses that incident improving edge with smallest angle to the gradient of the objective function.
>
> **Greatest Improvement:** Takes that edge which leads to the maximal improvement of $v^T x$ in the next step.

Category B: Variants exploiting information on the objective only.

> **Dantzig's Rule:** Since a vertex x is optimal iff v is in the cone of the gradients a_i of the current basis, we can calculate in each step the representation of v by that basis of $\mathbb{R}^n$. So every basis-gradient is associated with its v-representation coordinate. Since optimality requires

a completely nonnegative representation, Dantzig's Rule suggests to take that edge, which deactivates the restriction whose gradient has the most negative coordinate.

Shadow-Vertex-Algorithm or Parametric Variant: This is the variant for which theoretical studies worked very well. The results will be presented in the following sections. Therefore, we explain it in detail. This variant leads us from a vertex optimizing an alternative objective $u^T x$ to the optimal solution for $v^T x$ (or an unbounded edge), by providing all optimal vertices to the family of objectives $(\lambda v + u)^T x$ with $\lambda \in [0, \infty)$.
For λ starting at 0 and increasing, the sequence of optimal vertices gives just the parametric Simplex-Path, which had also been constructed in the early parametric variant of Gass and Saaty. They had introduced the parametric concept for another type of problems and without the geometric shadow-vertex interpretation: If we project X on $\mathrm{Span}(u, v)$, then our variant constructs a path along the boundary of the two-dimensional image of X, visiting only shadow vertices. These are vertices which keep their vertex-property even after the projection. The choice of the right edge is organized by calculation of the basis representations of v and u (as in Dantzig's Rule) and by minimizing the quotient of corresponding coordinates.

Category C: Rules evaluating combinatorial principles only.

Rule of Random Choice: selects randomly one of the improving edges.

Rule of Justice: deacitivates that tight restriction which had been active most often.

Rule of Bland: deactivates that tight restriction which has the least original index.

These variants were compared under three different rotation-symmetric distributions for the vectors a_i :

Uniform distribution on ω_n (the unit sphere of $\mathbb{R}^n$),

Uniform distribution on Ω_n (the full unit ball of $\mathbb{R}^n$),

Gaussian distribution on $\mathbb{R}^n$.

In general, it turned out that results for Gaussian distribution were better (smaller) than for unit ball and the latter were better than the unit sphere results. The reason simply comes from the redundancy rate. A restriction is redundant, if its existence or nonexistence has no impact on the shape of X . Here, the i-th restriction

is redundant, iff a_i belongs to the interior of $Conv(0, a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_m)$. This will never happen when all points come from ω_n and rather seldom when all points come from Ω_n, but very often when the points are Gaussian-distributed. And, it is obvious, that under normal circumstances a problem becomes easier, if more restrictions are redundant, resp. if the redundancy-rate is high.

The quality of the different variants can be ordered consistently. The best performance shows the Rule of Steepest Ascent. It is slightly better than the Rule of Greatest Improvement. These two variants show a very good perfomance in particular when the current vertex is still far away from the optimal one.
A bit worse are the variant of Dantzig and the Shadow-Vertex-Algorithm. The reason may be that they do not exploit information on the polyhedron itself (which may make the edge-choice more ineffective, but saves computation time in the single pivot step).
Considerably worse is the performance of the combinatorial variants. The best among these is Random Choice, followed by Rule of Justice, and finally comes Bland's Rule.

The overall impression is that the differences between Category A and Category B are not dramatic, but that the differences between Category B and Category C are striking.

We have also tested the standard deviation of s and the more meaningful quotient between standard deviation and mean value (for the number of pivot steps). This quotient was less, but close to 1, when m was in the order of n. But the quotient became quite small for $m \gg n$. We understand this as a hint that in the RSM for $m \to \infty$ and fixed n (the "asymptotic case") the shape of X, the number of facets as well as their size, and the length of edges will stabilize more and more.

However, all these experiments and their outcome are not at all satisfactory for a final judgement. One reason is that the computation time for a sufficient number of repetitions of the experiments is irresponsibly high. Hence we cannot advance to reasonably high dimensions. Also complexity-theory investigations cannot be settled by limited experiments. A third argument concerns potential regression analysis-attempts based on the data of the results. It is almost impossible to modelize the qualitative structure of $E_{m,n}(s)$ as a function of m and n with parameters to be specified by the regression, as long as we do not understand (theoretically) the interaction between m, n and the stochastic model. Many such attempts failed as the model structures did always fit only in a bounded range of m and n.

Much more meaningful is the outcome of theoretical (arbitrary-dimension) considerations. In the following, we present two successful approaches.

3 Results under the Sign-Invariance-Model

The first investigation under this kind of model was done by Smale [20]. He analyzed problems of type (3)

$$\text{maximize } v^T x \text{ subject to } Ax \leq b,\ x \geq 0$$

and treated this problem as a special case of the Linear Complementarity Problem

$$\begin{array}{l}\text{Find } w, z \in \mathbb{R}^{m+n} \text{ such that for given } q \in \mathbb{R}^{m+n}, M \in \\ \mathbb{R}^{(m+n)\times(m+n)} \\ w - Mz = q \text{ and } w^T z = 0, w \geq 0, z \geq 0.\end{array} \tag{9}$$

When M is specialized to $\begin{pmatrix} 0 & -A \\ A^T & 0 \end{pmatrix}$ and q to $\begin{pmatrix} b \\ -v \end{pmatrix}$, then a solution of (9) yields a solution of (3) and its dual.

As solution procedure, Smale employs "Lemke's algorithm". One starts with a solution for $\mathbb{1}$ replacing q. Then one moves on $[\mathbb{1}, q]$ forward and performs pivot steps whenever one of the coordinates of the vector w reaches the value 0, in order to keep all components nonnegative.
The analysis amounted to the question, how many cones of a special type will be intersected by a line segment. This is a typical question for a parametric algorithm. The expected number of pivot steps $E_{m,n}(s^L)$ was analyzed under the following stochastic model:

1. (A, b, v) is distributed absolutely continuous.
2. The columns of (A, b) and v are independent.
3. The measure of (A, b) is invariant under coordinate permutations in columns of (A, b).

Smale proved for problems distributed under that model

Theorem 1 *[20]*

$$E_{m,n}(s^L) \leq C(n)\ (1 + \ln(m+1))^{n(n+1)}\,. \tag{10}$$

This shows polynomiality in m (for fixed n), but not in n. $C(n)$ is an (exponential) function of n.

Smale's studies gave a motivation for the analysis of the so-called Sign-Invariance-Model (SIM). It is extremely simple and only relies on a finite number of reflections and symmetries.

Let A, b and v define a nondegenerate data-set for problem (3). And let the occurence of $\boxed{\begin{array}{cc} A & b \\ v^T & 0 \end{array}}$ be equi-probable to the occurence of $\boxed{\begin{array}{cc} S_1AS_2 & S_1b \\ v^TS_2 & 0 \end{array}}$ for every sign-matrix
$S_1 \in \mathbb{R}^{m\times m}$ and $S_2 \in \mathbb{R}^{n\times n}$. (A sign-matrix is a diagonal matrix with $+1$ or -1 in the diagonal entries). To explain the impact of that model, it suffices to consider a somehow relaxed version of sign-invariance, the so-called Flipping-Model, where we consider only the sign-matrix S_1 and deal with problem instances of form (1).

$$\text{maximize } v^Tx \text{ subject to } a_1^Tx \leq \geq b^1, \ldots, a_m^Tx \leq \geq b^m. \tag{11}$$

Here $\leq \geq$ indicates, that one of the relations $\leq$ or $\geq$ shall be valid in the formulation of the instance. We get $\leq$ if $s_{ii} = 1$ and $\geq$ if $s_{ii} = -1$ in (11). Since all sign-matrices S_1 shall be equi-probable, this means that we independently determine the m directions of the relations, each one with probability $\frac{1}{2}$ for $\leq$ and with probability $\frac{1}{2}$ for $\geq$.

By the way, we generate exactly 2^m problem-instances out of one data-set. The idea of averaging is to solve all 2^m instances, to sum up the required pivot steps and to divide by the number of instances.
This set of problem-instances can be solved (as far as Phase II is concerned) by application of the shadow-vertex-algorithm explained in the section before. There we realize a Simplex-Path over all (temporarily) optimal vertices when we traverse the set of objectives $[u + \lambda v]^Tx$ for $\lambda \geq 0$.
If we add the corresponding set of (optimal) vertices for negative values of λ, then the total set will be called the set of cooptimal vertices.
With s for the number of pivot steps, S_{coop} for the number of cooptimal vertices and S for the number of shadow vertices, the following relation is obvious: $s \leq S_{coop} \leq S$.

Using simple combinatorial enumeration arguments, Adler and Haimovich [13] and showed

Theorem 2 *[13]*

$$E_{m,n}(S_{coop}|\textit{a cooptimal path exists}) \leq n\,\frac{m-n+2}{m+1} \leq n \textit{ for type (1) under SIM.}$$

So far, the analysis considers only the procedure of moving from a u^Tx-optimum to a v^Tx-optimum. But this does not fit exactly into a probabilistic analysis of a complete solution method (as Smale's method), because the u-optimum is not

given beforehand and calculating it would be as troublesome as calculating the $v^T x$-optimum.

In 1983/84, the combination of this result with a design of a complete algorithm was done in three papers by Todd [21], Adler & Megiddo [3], Adler, Karp, Shamir [2]. They all came to the same result for $(E_{m,n}(s^t))$, the expected number of pivot steps required to solve the LP completely (including Phases I and II).

Theorem 3 *[21], [3], [2]*
For problems of type (1) resp. of type (3) distributed under SIM the expected number of pivot steps for the complete solution by a lexicographic version of Lemke's algorithm (s^t) is

$$E_{m,n}(s^t) \;\leq\; 2(n+1)^2, \quad \textit{resp.} \quad \leq\; 2Min(m^2, n^2). \tag{12}$$

In the first two papers,analyzing type (3)-problems, the proof was based – as in Smale's analysis – on the evaluation of a probability that a typical cone is intersected by a line. But this time, this is the line

$$[d, \mathbb{1}] \;\text{ with }\; 0 < d = (\delta, \delta^2, \ldots, \delta^{m+n})^T \quad \text{with} \quad \delta \;\text{ as small as desired.} \tag{13}$$

Closer to our geometrical interpretation and easier to explain is the third approach in [2].
To explain the solution process of a type (1)-problem, we use

$$X^{(n+k)} \;=\; \{x \in \mathbb{R}^n | a_1^T x \leq b^1, \ldots, a_{n+k}^T x \leq b^k\} \;\text{ for }\; 0 \leq k \leq m-n, \;\text{ and }\; X^{(m)} = X.$$

The following complete algorithm works directly in the space $\mathbb{R}^n$ and is called

Lexicographic Constraint-By-Constraint-Method
Initialization:
Determine the unique vertex $\bar{x} \in X^{(n)}$ and choose u as $u = \delta^1 a_1 + \delta^2 a_2 + \ldots + \delta^n a_n$ with $\delta > 0$ sufficiently small.
Stage k $(1 \leq k \leq m-n)$
Start at $\bar{x}$, the maximal vertex for $u^T x$ on $X^{(n+k-1)}$. If $\bar{x} \in X^{(n+k)}$, then go to stage $k+1$.
Else use the shadow-vertex-algorithm to improve the value of $a_{n+k}^T x$ (note that so far $a_{n+k}^T \bar{x} > b^{n+k}$), start at $\bar{x}$ and minimize $a_{n+k}^T x$ on $X^{(n+k-1)}$. Stop as soon as $a_{n+k}^T x \leq b^{n+k}$.
On the last traversed edge find a point $\tilde{x}$ with $a_{n+k}^T \tilde{x} = b^{n+k}$.
Enter stage $k+1$ with $\tilde{x}$ and replace $\bar{x}$. This is possible, because we have moved on a cooptimal path, hence $\tilde{x}$ maximizes $u^T x$ on $X^{(n+k)}$.
STOP if it is impossible to achieve $a_{n+k}^T x \leq b^{n+k}$, because then the original

problem is infeasible.

Stage $m-n+1$**:**
Start at $\bar{x}$, which maximizes $u^T x$ on $X^{(m)} = X$. Apply the shadow-vertex-algorithm and find the optimal point for $v^T x$ or discover that $v^T x$ is unbounded on X.

In principle, this amounts to solving $(m-n+1)$ problems, for which the average number of steps is less n each (Theorem 2). But now, due to the lexicographical choice of u, it can be exploited that – when we enter stage $k+1$ – most of the work to optimize the current objective has already been done in earlier stages. Thus, the effort of stage $k+1$ becomes much smaller than n.
Finally, the order of the total average number of steps is $O(n^2)$ instead of $O(mn)$.

With slight additional conditions on the distributions of the A-entries, Adler & Megiddo [3] could establish also a lower bound of type $C \cdot n^2$.
And they argued that for $m \leq 2n$, since the share of feasible problems is at least $n^{-\frac{1}{2}}$, the conditional expected number of pivot steps for solving LP's of that model, under the condition that the problem-instance is feasible, is $O(n^{2.5})$.

As for every probabilistic model, one should ask about the direct impact of the model on the results.
An important feature of SIM is the fact that many instances will be infeasible, precisely

$$\frac{\text{number of feasible instances}}{\text{number of generated problems}} = \frac{\binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{n}}{2^m} \to 0 \quad (m \to \infty,\ n \text{ fixed}). \tag{14}$$

Only conditioning on feasible problem-instances avoids averaging over a lot of easy problems. But even if we do so, we meet a remarkably small expected number of vertices.

$$E_{m,n}\ (\text{vertices per feasible instance}) = \frac{2^n \binom{m}{n}}{\binom{m}{0} + \cdots + \binom{m}{n}}, \tag{15}$$

which is less 2^n and converges to that value for $m \to \infty$, n fixed.
Now it is not astonishing, that for a large class of variants the average number of pivot steps for the complete solution will be bounded from above by a function of n only (compare [1]).
But the most important reason for simplification of problem-instances with $m \gg n$ comes from the average redundancy rate (the share of the restrictions without impact on X). This expected number (conditioned on feasible problems) is

$$\frac{\binom{m-1}{n}}{\binom{m}{0} + \cdots + \binom{m}{n}} \to 1 \quad \text{for } m \to \infty,\ n \text{ fixed}. \tag{16}$$

Simultaneously, the absolute number of nonredundant constraints (conditioned on feasibility) tends to $2n$ for $m \to \infty$ and fixed n.

So, for $m \gg n$, even for the feasible instances, the nonredundancy rate will be very small. This will of course make these problems easy. And it says that the Sign-Invariance-Model gives reasonable and meaningful information only for $m = O(n)$.
SIM relies on symmetries and reflections only. The combinatorial methods for evaluation make it unlikely that a calculation of higher moments of the s-distribution can easily be done. Besides that, the model is somehow inflexible. For every set of data, the reflection procedure leads to exactly the same characteristics in the 2^m instances. There is no way to choose a desired redundancy share or a size of the expected number of vertices and to parametrize certain figures in order to study their impact.
Apparently (in particular for $m \gg n$) the small upper bounds in (Theorem 2) and (Theorem 3) do rather reflect the special properties of the model than confirm the efficiency of the Simplex-Method, which had been pointed out in [1]).

4 Results under the Rotation-Symmetry-Model

The theoretical analysis based on the Rotation-Symmetry-Model (8) started in 1977 [4] and is still developing. The main result– a polynomial upper bound for the expected number of shadow vertices – was derived in 1996/97 [11] and it had predecessors with slightly cruder bounds in 1987 [7] and 1982 [6].

Theorem 4 *[11]*
For every rotation-symmetry distribution as in (8) and for every pair (m, n) with $m \geq n$ the expected number S of shadow vertices (and of pivot steps s in Phase II) satisfies

$$4E_{m,n}(s) \approx E_{m,n}(S) \leq Const. \cdot m^{\frac{1}{n-1}} \cdot n^2. \tag{17}$$

This result and its predecessors have been derived by translating the question about S into the dual space of the vectors a_i. Candidates for being a vertex are only the $\binom{m}{n}$ basic solutions x_Δ solving a system of n equations

$$a_{\Delta^1}{}^T x = 1, \ldots, a_{\Delta^n}{}^T x = 1 \text{ with } \Delta = \{\Delta^1, \ldots, \Delta^n\} \subset \{1, \ldots, m\}. \tag{18}$$

x_Δ is actually a vertex if all other restrictions are satisfied, i.e. $a_i^T x_\Delta \leq 1 \forall i \notin \Delta$. It becomes a shadow vertex, if the projection on $Span(u, v)$ preserves its vertex-property.

Now there is a one-to-one correspondence

$$x_\Delta \;\leftrightarrow\; \Delta = \{\Delta^1, \ldots, \Delta^n\} \;\leftrightarrow\; Conv(a_{\Delta^1}, \ldots, a_{\Delta^n}) \tag{19}$$

Besides $X = \{x | Ax \leq b\}$ we consider its "polar polyhedron" $Y = Conv(0, a_1, \ldots, a_m)$.

The following equivalencies enable us to derive the average number of shadow vertices directly from the input data

Lemma 1

1. *x_Δ is a vertex of X iff $Conv(a_{\Delta^1}, \ldots, a_{\Delta^n})$ is a facet of Y.*
2. *A vertex x_Δ is a shadow vertex of X iff $Conv(a_{\Delta^1}, \ldots, a_{\Delta^n}) \cap Span(u, v) \neq \emptyset$.*

The addition theorem for expectation values and the symmetry of index-choices yield

$$E_{m,n}(S) \;=\; \binom{m}{n} \cdot P\left(Conv(a_1, \ldots, a_n) \text{ is a facet and intersected by } Span(u, v)\right) \tag{20}$$

Here, one integrates over all possible configurations of $a_1, \ldots, a_m, u, v$ and weights with regard to the underlying distribution. The resulting multiple integral was very hard to evaluate. For the case of moderate dimensions (m, n arbitrary), we could only compare our integral with known results about a closely related integral. Much more efficient are the tools for evaluating the so-called asymptotic case ($m \to \infty$, n fixed), because there the integrals behave like Laplace-integrals and can conveniently be evaluated. So it was much easier to derive asymptotic results for specific RSM-distributions.

In the following we write $E_{m,n}(S) \sim f(m,n)$ for $m \to \infty$, n fixed, when we mean that

$$C_1 \leq \liminf_{\substack{m\to\infty \\ n\, fixed}} \frac{E_{m,n}(S)}{f(m,n)} \leq \limsup_{\substack{m\to\infty \\ n\, fixed}} \frac{E_{m,n}(S)}{f(m,n)} \leq C_2 \quad \text{for some constants} \quad C_1, C_2 > 0.$$

Besides that, we speak of a distribution on Ω_n with algebraically decreasing tail , if
$P(\|x\| \geq r) \sim (1-r)^\gamma$ for $r \to 1$ for a $\gamma > 0$.

Theorem 5 *[4],[7],[16],[18],[14]*
For fixed n and $m \to \infty$, the following distributions lead to the following behaviour of $E_{m,n}(S)$.

1. *Gaussian distribution on $\mathbb{R}^n$: $E_{m,n}(S) \sim \sqrt{\ln m}\; n^{\frac{3}{2}}$*

2. *Uniform distribution on* Ω_n : $E_{m,n}(S) \sim m^{\frac{1}{n+1}}\, n^2$

3. *Uniform distribution on* ω_n : $E_{m,n}(S) \sim m^{\frac{1}{n-1}}\, n^2$

4. *There are distributions such that* $E_{m,n}(S) \sim C(n)$

5. *For distributions on* Ω_n *with algebraically decreasing tail:* $E_{m,n}(S) \sim m^{\frac{1}{n-1+2\gamma}} n^2$.

These results should be compared with corresponding results on the average number of vertices of X, resp. facets of Y, denoted by $E_{m,n}(V)$ in our model.

Theorem 6 *[7],[16],[10]*
For fixed n *and* $m \to \infty$ *the following distributions lead to the following behaviour of* $E_{m,n}(V)$

1. *Gaussian distribution on* $\mathbb{R}^n$: $E_{m,n}(V) \sim [\ln m]^{\frac{n-1}{2}}\, 2^n \cdot \pi^{\frac{n-1}{2}}\, \frac{1}{\sqrt{n}}$

2. *Uniform distribution on* Ω_n : $E_{m,n}(V) \sim m^{\frac{n-1}{n+1}}\, 2^{\frac{n}{2}}\, \pi^{\frac{n}{2}}\, n^{-\frac{3}{4}}\, (n+1)^{\frac{n-1}{2}} \cdot \left(1+\frac{n}{2}\right)^{-\frac{1}{2}+\frac{n-1}{n+1}}$

3. *Uniform distribution on* ω_n : $E_{m,n}(V) \sim m^{\frac{n-1}{n-1}}\, 2^{\frac{n}{2}}\, \pi^{\frac{n}{2}}\, n^{-\frac{3}{4}}\, (n-1)^{\frac{n-1}{2}} \cdot \left(\frac{n}{2}\right)^{-\frac{1}{2}}$

4. *For distributions on* Ω_n *with algebraically decreasing tail:*
$E_{m,n}(V) \sim m^{\frac{n-1}{n-1+2\gamma}}\, 2^{\frac{n}{2}}\, \pi^{\frac{n}{2}}\, n^{-\frac{3}{4}}\, (n-1+2\gamma)^{\frac{n-1}{2}} \cdot \left(\frac{n}{2}\right)^{-\frac{1}{2}\cdot\frac{n-1}{n-1+2\gamma}}$.

Obviously the Simplex-Method is able to select a rather short path through the huge set of vertices. Hereby it visits (on the average) only the n-th root of the total number of available vertices.

Another very important point is the variance of the number of shadow-vertices resp. of the number of required pivot steps. Due to the technical difficulties mentioned above, so far only the asymptotic case has been analyzed. Küfer [17] showed

Theorem 7 *[17]*
For distributions with algebraically decreasing tail on Ω_n*, the quotient of variance and square of expected value behaves asymptotically as follows*

$$\frac{Var_{m,n}(s)}{E^2_{m,n}(s)} \sim \frac{1}{n} \quad \textit{and} \quad \frac{Var_{m,n}(S)}{E^2_{m,n}(S)} \sim m^{\frac{-1}{n-1+2\gamma}}. \tag{21}$$

Here s *is the number of pivot steps and* S *is the number of shadow-vertices,*

So far, we have dealt only with a fictive Phase II-algorithm, starting at an optimal vertex for an auxiliary objective. But this vertex is impracticable to find. Now let us talk about a safe Phase I.

A special feature of our problems is the feasibility of the origin, which makes – in contrast to the Sign-Invariance-Model – every instance feasible. Based on that information, we can employ a method (cf. [6] and[7]), which applies the shadow-vertex-algorithm $n-1$ times, and each time the dimension of the problem is increased. In each of these stages all the stochastic requirements of our model are satisfied.
Here we introduce $X^{(k)}$ as $\{x|Ax \leq \mathbb{1}, x^{k+1} = \ldots = x^n = 0\}$, and formulate the

Dimension-By-Dimension-Algorithm

Initialization (Stage $k = 1$)
Starting from the origin, find a vertex of $X^{(1)} = \{x|Ax \leq \mathbb{1}, x^2 = \ldots = x^n = 0\}$ maximizing $v^T x = v^1 \cdot x^1$. If this maximal vertex does not exist, STOP.

Stage k $(2 \leq k \leq n)$
Use the optimal point $(\bar{x}^1, \ldots, \bar{x}^{k-1}, 0, \ldots, 0)^T$ for $v^T x$ on$X^{(k-1)}$, which is located on an edge of $X^{(k)}$.

1. Find an adjacent vertex in $X^{(k)}$ to that edge.
2. Apply the shadow-vertex-algorithm using $e_k^T x$ and $v^T x$ as pair of objectives for maximizing $v^T x$ on $X^{(k)}$. If $v^T x$ turns out to be unbounded on $X^{(k)}$, STOP.
3. If $k < n$, set $k = k + 1$ and enter the next stage. If $k = n$, PRINT the optimal vertex for X.

One can derive an upper bound for this cumulation of $n-1$ shadow-vertex-applications by summing up all the expected numbers of shadow-vertices. But this would significantly overestimate the actual number of pivot steps in this algorithm, since we would ignore that the original distribution comes from $\mathbb{R}^n$ and that only projection-distributions (from $\mathbb{R}^n$ to $\mathbb{R}^k$) can determine the behaviour in stage k. Since the set of projection distributions is only a small subset of the RSM-distributions in dimension k, the corresponding bound for the expected number of steps in stage k is much better. Consequently, we obtain the following result, which also holds for problems of type (3), including sign-constraints.

Theorem 8 *[12]*
For every pair (m, n) with $m \geq n$ and every RSM-distribution on $\mathbb{R}^n$, the expected total number of pivot steps for the dimension-by-dimension-algorithm

$(E_{m,n}(s^t))$ *satisfies*

$$E_{m,n}(s^t) \;\leq\; m^{\frac{1}{n-1}} \cdot n^3 \cdot C \tag{22}$$

as well for problems of type (1) as for problems of type (3).

But as observed in the analysis of the Constraint-By-Constraint-Method (compare Theorem 3) it is plausible that most work of optimizing in stage $k+1$ has already been prepared in prior stages, such that the actual number of steps in stage $k+1$ is much smaller. This was precisely clarified by Gabriele Höfner [14] for the asymptotic case.

Theorem 9 *[14]*
For every RSM-distribution the expected total number of pivot steps in the dimension-by-dimension-algorithm satisfies

$$E_{m,n}(s^t) \;\sim\; m^{\frac{1}{n-1}} \cdot n^{\frac{5}{2}} \tag{23}$$

when $m \to \infty$ *and* n *is fixed for problems of type (1) and (3).*

It must be clear that this algorithm is crude and lengthy and has been introduced only for meeting the conditions of RSM and for making the probabilistic analysis possible.

In order to confirm the "folklore" observation, that Phase I can be done with an effort not exceeding that of Phase II, Höfner analyzed another complete algorithm. But unfortunately, only in the asymptotic case this method is assured to work.

1. Solve the problem
   $\max \mathbb{1}^T x$ subject to $Ax \leq \mathbb{1},\ x \geq 0$
   by use of the shadow-vertex-algorithm starting at the vertex 0. The optimal vertex $\bar{x}$ will – in the asymptotic case – with extremely high probability be a vertex of $X = \{x | Ax \leq \mathbb{1}\}$.
2. Start the shadow-vertex-algorithm at $\bar{x}$, forget about the sign-constraints and optimize $v^T x$ on X.

It can be shown that both applications of the shadow-vertex-algorithm require (on the average) at most $m^{\frac{1}{n-1}} \cdot n^2 \cdot$ Const. pivot steps.
So, this is an algorithm with a Phase I-effort not exceeding that of Phase II.
So far, the plausible and natural very good behaviour of Phase I can only be guaranteed in the asymptotic case. In the moderate cases, the situation is similar to

that of the Sign-Invariance-Results, where the Constraint-By-Constraint-Method needs a factor n more pivot steps than Phase II does.

As we have discussed the advantages and drawbacks of SIM, we now consider similar questions for RSM. Seemingly it is a tremendous advantage of RSM that it generates only (the hard) feasible problems. But simultaneously it turns out to be a drawback that the given Phase I-algorithms are designed in a way such that they exploit this fact and are dependent on the guarantee of "0 being feasible". One way to overcome that drawback lies in the following idea. Remember that we want to solve all problem instances of the type

$$\text{maximize } v^T x \text{ subject to } a_1^T x \leq b^1, \ldots, a_m^T x \leq b^m \tag{24}$$

with arbitrary values of b^i (not necessarily positive).

For integrating all these problems in our analysis, we use a "homogenization method". We introduce the notation $P_n := \{x | Ax \leq b\}$ and reformulate our restrictions

$$a_i^T x \;\leq\; b^i \text{ corresponds to } a_i^T x \;\leq\; 1 - \tilde{b^i} \text{ when } b^i = 1 - \tilde{b^i} \tag{25}$$

So we can demand that

$$a_1^T x + \tilde{b^1} \cdot 1 \;\leq\; 1, \ldots, a_m^T x + \tilde{b^m} \cdot 1 \;\leq\; 1 \tag{26}$$

and define a polyhedron in $\mathbb{R}^{n+1}$ by

$$\left(a_i^1, a_i^2, a_i^3, \ldots, a_i^n, \tilde{b^i}\right) \begin{pmatrix} x^1 \\ \vdots \\ x^n \\ x^{n+1} \end{pmatrix} \;\leq\; 1 \quad (x \in \mathbb{R}^{n+1}), \tag{27}$$

which means that $a_i^T \bar{x} + \tilde{b^i} x^{n+1} \;\leq\; 1 \quad (\bar{x} \in \mathbb{R}^n)$ for $i = 1, \ldots, m$.

This system defines a new polyhedron $P_{n+1} \subset \mathbb{R}^{n+1}$. The set of feasible points with $x^{n+1} = 0$ is a one-to-one copy of $\{x | Ax \leq \mathbb{1}\}$. And the set of feasible points with $x^{n+1} = 1$ corresponds one-to-one to the set of points in P_n .

It is now clear that in level $\{x^{n+1} = 0\}$ the problem satisfies all RSM-requirements. So we can solve the optimization problem for $v^T x$ on that artificial polyhedron. But then we can use one further stage $(n+1)$ of the dimension-by-dimension-algorithm to reach level $\{\, x^{n+1} = 1 \,\}$ (by maximizing $x^{n+1} = e_{n+1}^T x$ on P_{n+1}). If we use the shadow-vertex-algorithm starting at the level $\{\, x^{n+1} = 0\}$-optimum, then we walk on a cooptimal path all the time. And there will be two possible outcomes:

1. Max $\{x^{n+1} | x \in P_{n+1}\} < 1$.
   Then level $\{x^{n+1} = 1\}$ has no feasible points and P_n is proven to be empty, resp. infeasible.
2. Max $\{x^{n+1} | x \in P_{n+1}\} \geq 1$.
   Then the shadow-vertex-path in stage $n+1$ will traverse the desired level. We calculate the intersection point, drop the last coordinate 1 and have the optimal point for max $v^T x$ subject to $Ax \leq b$. This results from the cooptimality of our path.

Now the following probabilistic result is obvious.

Theorem 10 *[9]*
If $\begin{pmatrix} a_1 \\ \tilde{b^1} \end{pmatrix}, \ldots, \begin{pmatrix} a_m \\ \tilde{b^m} \end{pmatrix}$ *are distributed on* $\mathbb{R}^{n+1}$ *according to the RSM, then general problems of type (1) can be solved for every* (m, n) *with an expected total number of pivot steps as*

$$E_{m,n}(s^t) \leq m^{\frac{1}{n}}(n+1)^3 \cdot C \tag{28}$$

But this condition has a quite artificial flavour, because the RSM-distribution of the augmented vectors may lead to dependencies between the gradients a_i and the capacities b^i.
We know one special distribution, where both wishes (RSM-distribution and independency) can be combined, namely the Gaussian distribution on $\mathbb{R}^{n+1}$. This is the only RSM-distribution, where the components of the generated vectors are independent. We obtain

Theorem 11 *[9]*
If the vectors $\begin{pmatrix} a_1 \\ \tilde{b^i} \end{pmatrix}, \ldots, \begin{pmatrix} a_m \\ \tilde{b^m} \end{pmatrix}$ *are independent and Gaussian-distributed, then*

$$E_{m,n}(s^t) \leq m^{\frac{1}{n}}(n+1)^3 \cdot Const. \tag{29}$$

For more general independent distributions of the right sides (the capacities), as for uniform distribution, we could not derive satisfactory bounds so far. However, this seems to be caused by technical difficulties. The special results in (Theorems 10 and 11) indicate that general problems with arbitrary independent capacity-distribution may be solvable on the average with the same effort.

We conclude our report with a look on general variants. In [15] Petra Huhn proved a lower bound on the average number of pivot steps, which is valid for all

variants. Assume that Phase I has provided us with a vertex x_0 of X, and that the objective $v^T x$ had no impact on Phase I. Then we start at x_0 with Phase II and try to reach the optimal vertex x_{opt}. To bridge the distance, every variant has to use edges of the polyhedron X. Now stochastic geometry can provide information on the distribution of the length of these edges. If one can show that there are extremely few "long" edges, then a large number of "small" edges has to be used for our walk. This has been done in [15] and it gave a guarantee that no variant can – on the average – do its job with less than a certain (computable) number of steps. Quantitatively, this reads as follows. We present only the result for a special distribution, the uniform distribution on ω_n (corresponding results have been derived for a large class of distributions).

Theorem 12 *[15]*
In a typical RSM problem with uniform distribution on ω_n, every variant of the Simplex-Algorithm will – on the average – require a certain number $E^{av}_{m,n}(s)$ of pivot steps, and

$$E^{av}_{m,n}(s) \;\geq\; \textit{Const.} \cdot m^{\frac{1}{n-1}-\frac{1}{(n-1)^2}} \cdot n^{-\frac{1}{2(n-2)}} \cdot (2n-2)^{-\frac{1}{n-1}+\frac{2}{(n-1)^2}} \;\approx\; C \cdot m^{\frac{1}{n-1}} \cdot n^0 . \quad (30)$$

Despite the fact that here the n-order is n^0 (compare with n^2 for the shadow-vertex-algorithm), this shows that no variant can perform substantially better. This means that there is no algorithm (variant) running essentially faster than the shadow-vertex-algorithm, which can exploit the increasing number of options with n, and which avoids the typical order $m^{\frac{1}{n-1}}$ in the RSM-model.

Thus – a posteriori – the results on the shadow-vertex-algorithm have proved to be quite representative. It is not the very best variant, but not much worse than the very best.

Note that the lower bound in Theorem 12 is meaningful only when $m \gg n$, because only then it becomes significantly greater than 1, although the inequality is valid for all (m, n). This is different from the results about the variance (Theorem 7) and the speed-up for Phase I (Theorem 9), where it is uncertain, whether these results will be valid in moderate dimensions, too. (Perhaps not the technical difficulties are to blame). It may as well be possible that these results essentially rely on a regularization effect of the polyhedra for large number of points, as we know it from the approximation of a ball from inside by the convex hull of a huge number of random points.
To clarify these questions, remains an important challenge for future research.

References

[1] Adler, I. and Karp, R. and Shamir, R. : A family of variants solving an $m \times d$ linear program in expected number of pivot steps depending on d only. Math. Oper. Res. 11,1986, 570–590

[2] Adler, I. and Karp, R. M. and Shamir, R.: A Variant Solving an $m \times d$ Linear Program in $O(min(m^2, d^2))$ Expected Number of Pivot Steps, Journal of Complexity 3, 1987, 372–387

[3] Adler, I. and Megiddo, N.: A Algorithm where the Average Number of Steps is Bounded Between two Quadratic Functions of the Smaller Dimension, Journal of the ACM 32, 1985, 871–895

[4] Borgwardt, K.H.: Untersuchungen zur Asymptotik der mittleren Schrittzahl von verfahren in der Linearen Optimierung, Dissertation Kaiserslautern, 1977

[5] Borgwardt, K.H.: Some Distribution-Independent Results About the Asymptotic Order of the Average Number of Pivot Steps of the Method, Mathematics of Operations Research 7, 1982, 441–462

[6] Borgwardt, K.H.: The Average Number of Pivot Steps Required by the -Method is Polynomial, Zeitschrift für Operations Research 26, 1982, 157–177

[7] Borgwardt, K.H.: The Method, A Probabilistic Analysis, Springer Verlag, Heidelberg, 1987

[8] Borgwardt, K.H. and Damm, R. and Donig, R. and Joas, G.: Empirical Studies on the Average Efficiency of Variants under Rotation Symmetry, ORSA Journal on Computing 5, 1993, 249–260

[9] Borgwardt, K. H.: Probabilistic Analysis of the Method, Contemporary Mathematics 114, 1990, 21–34

[10] Borgwardt, K.H.: Average Complexity of a Gift-Wrapping Algorithm for Determining the Convex Hull of Randomly Given Points, Discrete and Computational Geometry 17, 1997, 79–109

[11] Borgwardt, K.H.: A Sharp Upper Bound for the Expected Number of Shadow-Vertices in the Rotation-Symmetry-Model, submitted to Mathematics of Operations Research, 1997, 47 pages

[12] Borgwardt, K.H.: Ideas Leading to a Better Bound on the Average Number of Pivot Steps for Solving an LP, submitted to Operations Research Proceedings, 1997, 12 pages

[13] Haimovich, M.: The Algorithm is Very Good! - On the Expected Number of Pivot Steps and Related Properties of Random Linear Programs, Columbia University, New York, Report, 1983

[14] Höfner, G.: Lineare Optimierung mit dem Schatteneckenalgorithmus – Untersuchungen zum mittleren Rechenaufwand und Entartungsverhalten, Dissertation Augsburg, 1995

[15] Huhn, P.: Schranken für die durchschnittliche Laufzeit von verfahren und Innere-Punkt-Verfahren, Dissertation Augsburg, 1997

[16] Küfer, K.: Asymptotische Varianzanalysen in der stochastischen Polyedertheorie, Dissertation Kaiserslautern, 1992

[17] Küfer, K.: On the variance of the number of pivot steps required by the algorithm, Zeitschrift für Operations Research 42, 1995, 1–24

[18] Küfer, K.: An improved asymptotic analysis of the number of pivot steps required by the algorithm, Zeitschrift für Operations Research 44, 1996, 147–170

[19] Shamir, R.: The Efficiency of the Method, Management Science 33, 1987, 301–334

[20] Smale, S.: On the Average Speed of the Method of Linear Programming, Mathematical Programming 27, 1983, 241–262

[21] Todd, M.J.: Polynomial Expected Behavior of a Pivoting Algorithm for Linear Complimentarity and Linear Programming Problems, Mathematical Programming 35, 1986, 173–192