Randomized simplex algorithms on Klee-Minty cubes

Bernd Gärtner
Institut für Informatik
Freie Universität Berlin

Günter M. Ziegler
Konrad-Zuse-Zentrum für
Informationstechnik Berlin (ZIB)

Abstract

We investigate the behavior of randomized simplex algorithms on special linear programs.

For this, we develop combinatorial models for the Klee-Minty cubes [16] and similar linear programs with exponential decreasing paths. The analysis of two randomized pivot rules on the Klee-Minty cubes leads to (nearly) quadratic lower bounds for the complexity of linear programming with random pivots. Thus we disprove two bounds conjectured in the literature.

At the same time, we establish quadratic upper bounds for random pivots on the linear programs under investigation. This motivates the question whether some randomized pivot rules possibly have quadratic worst-case behavior on general linear programs.

1 Introduction

Linear programming is the problem of minimizing a linear objective function over a polyhedron $P \subseteq \mathbb{R}^n$ given by a system of m linear inequalities.

Without loss of generality [22] we may assume that the problem is primally and dually nondegenerate, that the feasible region is full-dimensional and bounded, and that the objective function is given by the last coordinate. In other words, we consider the problem of finding the "lowest vertex" (minimizing x_n) of a simple n-dimensional polytope $P \subseteq \mathbb{R}^n$ with at most m facets, where the last coordinate x_n is not constant on any edge, and thus the lowest vertex is unique.

In this setting, the (geometric interpretation of the) simplex algorithm proceeds from some starting vertex of P along edges in such a way that the objective function decreases, until the unique lowest vertex of P is found. The (theoretical and practical) efficiency of the simplex algorithm [23] depends on a suitable choice of decreasing edges that "quickly leads to the lowest vertex". Connected to this are two major problems of linear programming: the diameter problem "Is there a short path to the lowest vertex?", and the algorithm problem "Is there an algorithm which quickly finds a (short) path to the lowest vertex?".

The **diameter problem** is closely related to the "Hirsch conjecture" (from 1957) and its variants [5, 14, 26]. Currently there is no counterexample to the "Strong monotone Hirsch conjecture" [26] that there always has to be a decreasing path, from the vertex which maximizes x_n to the lowest vertex, of length at most $m - n$. On the other hand, the best arguments known for upper bounds establish paths whose length is roughly bounded by $m^{\log_2 2n}$ [11].

The **algorithm problem** includes the quest for a strongly polynomial algorithm for linear programming. Klee & Minty [16] showed in 1972 that linear programs with exponentially long decreasing paths exist, and that the "steepest descent" pivot rule can be tricked into selecting such a path. Using variations of the Klee-Minty constructions, it has been shown that the simplex algorithm may take an exponential number of steps for virtually every deterministic pivot rule [14]. (A notable exception is Zadeh's rule [25, 14], locally minimizing revisits, for which Zadeh's \$1,000.– prize [14, p. 730] has not been collected, yet.)

No such evidence exists for some extremely natural *randomized* pivot rules, among them the following three rules:

RANDOM-EDGE: At any nonoptimal vertex x of P, follow one of the decreasing edges leaving x with equal probability.

RANDOM-FACET: If x admits only one decreasing edge, then take it. Otherwise restrict the program to a randomly chosen facet containing x. This yields a linear program of smaller dimension in which x is nonoptimal, and which can be solved by recursive call to RANDOM-FACET.

RANDOM-SHADOW: Start at the unique vertex $y \in P$ which maximizes x_n. Choose a random unit vector $\mathbf{c}$ orthogonal to $\mathbf{e}_n$. Now take the path from y to the lowest vertex given by $\{x \in P : \mathbf{c}x \leq \mathbf{c}z$ for all $z \in P$ with $z_n = x_n\}$.

RANDOM-FACET is a randomized version, due to Kalai [11], of Bland's PROCEDURE A [2], which assumes that the facets are numbered, and always restricts to the facet with the smallest index. Interestingly enough, very elementary arguments imply a recursion

$$f(n,m) \leq f(n-1,m-1) + \frac{1}{n}\sum_{i=1}^{n} f(n,m-i)$$

for the maximal expected number of steps $f(n,m)$ on an n-dimensional linear program with m inequalities. From this one can get subexponential upper bounds of roughly $e^{O(\sqrt{n \log m})}$ for the number of steps of RANDOM-FACET — see Kalai [11], and (in a very similar dual setting) Matoušek, Sharir & Welzl [19].

The RANDOM-SHADOW rule is a randomized version of Borgwardt's SHADOW VERTEX ALGORITHM [1] (a. k. a. the GASS-SAATY RULE [15]), for which the auxiliary function $\mathbf{c}$ is deterministically obtained in such a way that it is minimized on the starting vertex. Borgwardt [1] has successfully analyzed this algorithm under the assumption that P is random in a suitable model (where the secondary objective function $\mathbf{c}$ can be fixed arbitrarily), and obtained polynomial upper bounds for the expected number of simplex steps.

None of the available evidence contradicts the possibility that the expected running time of all three randomized algorithms we consider is bounded from above by a polynomial, even a quadratic function, in n and m. In this connection, we report investigations of the performance of such algorithms on infinite families of "test problems": specific linear programs which have decreasing paths of exponential length.

It is not generally believed that polynomial upper bounds can be achieved; it is equally conceivable that subexponential bounds such as those by Kalai [11] are essentially best possible. An interesting open problem in this context is to find linear programs on which the algorithms in [11, 19] actually behave superpolynomially; Matoušek [18] has constructed abstract optimization problems — more general than linear programs — for which the subexponential analysis is tight.

In this extended abstract we concentrate on the analysis of the "Klee-Minty cubes", see Section 2. These are very interesting linear programs whose polytope is a deformed n-cube, but for which some pivot rules follow a path through all the vertices and thus need an exponential number of steps.

Our main results are quadratic, respectively nearly quadratic, lower bounds for the expected number of steps taken by the RANDOM-FACET and the RANDOM-EDGE simplex algorithms. For the RANDOM-EDGE rule this seems to be the first superlinear bound.

Specifically, our analysis of random pivots on the Klee-Minty cubes yields the following two theorems.

Theorem 1. The RANDOM-FACET simplex algorithm on the n-dimensional Klee-Minty cube, started at the vertex $\overline{v}$ "opposite" (on the n cube) to the optimal vertex, takes a quadratic expected number of steps $F_n(\overline{v})$:

$$F_n(\overline{v}) = n + 2\sum_{k=1}^{n} \frac{(-1)^{k+1}}{k+2}\binom{n-k}{2} \approx \left(\frac{\pi}{4} - \frac{1}{2}\right) n^2.$$

Moreover, for a random starting vertex the expected number of steps is

$$F_n = \frac{n^2+3n}{8}.$$

We note that one easily gets a linear lower bound and a quadratic upper bound $F_n(x) \leq (n^2+3n)/4$ for the expected number of steps from an arbitrary starting vertex x. Furthermore, there are starting points for which the FACET RANDOM rule will take only linearly many steps. The fact that for some starting vertices the expected number of steps is quadratic follows from an explicit formula for the expectation value, given in Section 2, or from the bound for a random starting vertex.

A result very similar to Theorem 1, in the setting of dual simplex algorithms, was earlier obtained by Matoušek [18, Sect. 4], who analyzed the behavior of the Matoušek-Sharir-Welzl dual simplex algorithm on a special class of linear programs.

Similarly, for RANDOM-EDGE one gets an upper bound $E_n(x) \leq \binom{n+1}{2}$ for the expected number of steps starting at any vertex x of the n-dimensional Klee-Minty cube, see Section 2. This was first observed by Kelly [12], see also [24].

Theorem 2. The expected number E_n of steps that the RANDOM-EDGE rule will take, starting at a random vertex on the n-dimensional Klee-Minty cube, is bounded by

$$\Theta(\frac{n^2}{\log n}) \leq E_n \leq \binom{n+1}{2}.$$

This superlinear lower bound requires substantially harder work, see Section 3. It implies that there is a vertex x with $E_n(x) = \Omega(n^2/\log n)$, but compared to the case of RANDOM-FACET we are not able to show this bound for a specific starting vertex, e.g. the top vertex.

Our proof is based on a combinatorial model for the Klee-Minty cubes, which describes the RANDOM-EDGE

algorithm as a random walk on an acyclic directed graph (see Section 2).

The combinatorial model also makes it possible to do simulation experiments. Our tests in the range $n \leq 1,000$ suggest that the quadratic upper bound is close to the truth. Also, it seems that a (nearly) quadratic lower bound is valid also if the starting vertex is chosen to be the top vertex of the program, but as mentioned above, our method does not prove this.

Still, our result contradicts Exercise 8.10* in [21, p. 188], where it is claimed that $E_n(x) = O(n)$. It also disproves a conjecture of Kelly [12] that $E_n(x) = O(n \log^2 n)$ for all starting vertices x.

Another conjecture of Kelly [12], according to which the expected number of RANDOM-EDGE pivots is maximal if the starting vertex is diametrically opposite to the lowest vertex, also turned out to be false. One finds, by explicit computation of expectation values (in rational arithmetic, using REDUCE) that the smallest dimension in which this fails is $n = 18$.

The RANDOM-SHADOW algorithm has not yet been studied on special programs. Goldfarb [7, 8] has constructed a variant of the Klee-Minty cubes for which the deterministic SHADOW VERTEX ALGORITHM takes an exponential number of steps. There is hope for a successful analysis since Borgwardt's work [1] shows that methods of integral geometry can be very powerful when applied in this context.

Besides the Klee-Minty cubes and their variants, there are other natural classes of "test problems" for (randomized) linear programming algorithms. They include the deformed products of Klee & Minty [16], for which a combinatorial model is produced in Section 4. Also there is a natural model on polars of cyclic polytopes, for which the actual program has not been constructed, yet. This relates to the unsolved "upper bound problem for linear programs".

2 Combinatorial Models

The *Klee-Minty cubes* [16, 21] are the polytopes of the linear programs in $\mathbb{R}^n$ with $m = 2n$ facets given by

$$\begin{aligned} \min\, & x_n : \\ & 0 \leq x_1 \leq 1 \\ & \varepsilon x_{i-1} \leq x_i \leq 1 - \varepsilon x_{i-1} \end{aligned}$$

for $2 \leq i \leq n$ and $0 < \varepsilon < 1/2$. Our illustration shows the 3-dimensional Klee-Minty cube for $\varepsilon = 1/3$.

Considering the geometry in the limit $\varepsilon \to 0$, one sees that the feasible region is a (slightly) deformed unit cube. Thus the feasible vertices of the program are in bijection with the set $\{0,1\}^n$ of all 0/1-vectors

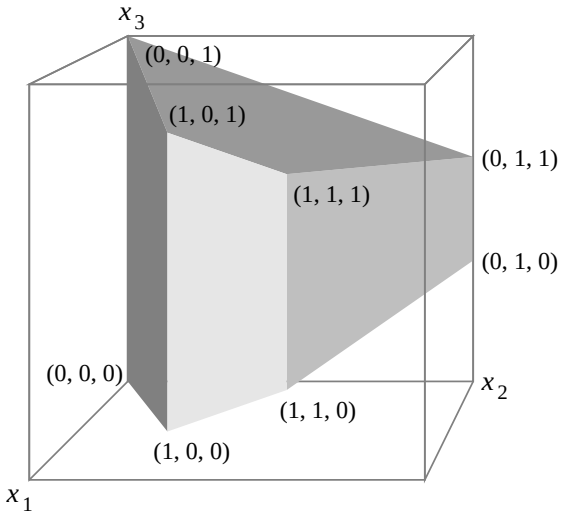

Figure 1: Klee-Minty cube for $n = 3, \varepsilon = 1/3$

of length n, where we obtain the 0/1-vector for any vertex by rounding the coordinates. Two vertices are adjacent if the corresponding 0/1-vectors differ in exactly one coordinate. (The identification of $\{0,1\}$ with $GF(2)$ will turn out useful in the next section, where linear algebra over $GF(2)$ is a key tool in our approach to lower bounds.)

In the following, we identify the vertices of the Klee-Minty cubes with the corresponding 0/1-vectors. Since the simplex algorithm proceeds along decreasing edges, we have to describe the edge orientations. It is easy to see, by induction on the dimension, that if x is a 0/1-vector with k ones, at positions $s_1 < s_2 < \ldots < s_k$, then the x_n-coordinate of the corresponding vertex of the Klee-Minty cube is

$$\varepsilon^{n-s_k} - \varepsilon^{n-s_{k-1}} + \ldots + (-1)^{k-1}\varepsilon^{n-s_1}$$

From this we obtain that if $x, x' \in \{0,1\}^n$ differ in their i-th component, then the corresponding edge is directed from x to x' if and only if the sum $x_i + x_{i+1} + \ldots + x_n$ is odd. We write $x \to x'$ in this situation.

This completes the description of the combinatorial model: a directed, acyclic graph with 2^n vertices, $n2^{n-1}$ directed arcs, and a unique source and sink. It can be used as a combinatorial model for the linear program.

For instance, one can derive that the average length Φ_n of a decreasing path from the highest to the lowest vertex — taking all paths with equal probability — satisfies $\Phi_n \geq (1 + 1/\sqrt{5})^{n-1}$ [10]: it is exponential. Thus, the "average" path is exponentially long, but

the RANDOM-EDGE and RANDOM-FACET pivot rules take the long paths with low probability.

The RANDOM-EDGE algorithm moves on the digraph of the Klee-Minty cube by leaving the current vertex, using one of the outgoing edges with equal probability, until it reaches the unique sink in the digraph. For example, a legal sequence of steps for $n = 3$, starting at the highest vertex and ending at the lowest, is given by

$$\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{pmatrix} \longrightarrow \begin{pmatrix} 1 \\ \mathbf{0} \\ \mathbf{1} \end{pmatrix} \longrightarrow \begin{pmatrix} \mathbf{1} \\ 0 \\ 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Here any coordinate that can be flipped is typeset bold: from this one can read off that the first step is taken with probability $p = \frac{1}{3}$, the second one with $p = \frac{1}{2}$, and the third with probability 1. Thus this path is taken with probability $\frac{1}{6}$.

The expected number of steps $E_n(x)$ from a vertex x to the lowest vertex satisfies the recursion

$$E_n(x) = 1 + \frac{1}{\#\{x' : x \to x'\}} \sum_{x' : x \to x'} E_n(x').$$

If $i(x)$ denotes the highest index i for which $x_i = 1$, then we can easily show

$$i(x) \le E_n(x) \le \binom{i(x)+1}{2} \le \binom{n+1}{2}$$

— this implies the upper bound of Theorem 2, but only a linear lower bound. A complete analysis seems to be surprisingly difficult. In Section 3 we develop a method, based on linear algebra over $GF(2)$, that yields the nearly quadratic lower bounds "on average" of Theorem 2.

The RANDOM-FACET pivot rule can, however, be completely analyzed on the Klee-Minty cubes. For this, one first derives that

$$F_n(\mathbf{e}_i) = F_n(\mathbf{e}_i + \mathbf{e}_{i-1}) = i.$$

In particular, started at the highest vertex $\mathbf{e}_n$, the RANDOM-FACET rule only needs an expected number of $F_n(\mathbf{e}_n) = n$ steps. For an arbitrary starting vertex $x \in \{0,1\}^n$, the solution of the program restricted to a facet $x_i = 0$ delivers the lowest vertex; restricted to a facet $x_i = 1$ the algorithm yields the vector $\mathbf{e}_i + \mathbf{e}_{i-1}$, where we set $\mathbf{e}_0 = \mathbf{0}$. From this we get a recursion

$$F_n(x) = \frac{1}{n}\left(\sum_{i=1}^{n} i x_i + \sum_{i=1}^{n} F_{n-1}(x^{(i)})\right),$$

with $x^{(i)} := (x_1, \ldots, x_{i-2}, x_{i-1} + x_i, x_{i+1}, \ldots, x_n)^t \in \{0,1\}^{n-1}$ for $1 \le i \le n$. Using this recursion, it is easy to derive a linear lower bound and a quadratic upper bound for $F_n(x)$, namely

$$i(x) \le F_n(x) \le \frac{i(x)^2 + 3i(x)}{4} \le \frac{n^2 + 3n}{4}.$$

Equality in the linear lower bound holds for the infinite series of vectors $\mathbf{e}_i$ and $\mathbf{e}_i + \mathbf{e}_{i-1}$. Surprisingly, one can explicitly solve the above recursion. In particular, quadratic lower bounds as in Theorem 1 can be derived from the following result.

Proposition 3. Started at a vertex $x \in \{0,1\}^n$ of the n-dimensional Klee-Minty cube, with

$$\{t : x_t = 1\} = \{s_1, s_2, \ldots, s_k\}_<,$$

the expected number of steps of the RANDOM-FACET simplex algorithm is

$$F_n(x) = \sum_{i=1}^{k} s_i + 2 \sum_{1 \le i < j \le k} \frac{(-1)^{j-i} s_i}{s_j - s_i + 1}.$$

For a random starting vertex, the situation is substantially simpler: Let

$$G_n = \sum_{x \in \{0,1\}^n} F_n(x).$$

From the recursion we get

$$G_n = 2^{n-2}(n+1) + 2G_{n-1}$$

with $G_1 = 1$. This gives

$$G_n = 2^{n-2}\left(\binom{n+2}{2} - 1\right),$$

and the second part of Theorem 1 follows.

3 A Lower Bound

Our analysis of the RANDOM-EDGE rule on the Klee-Minty cubes starts with a coordinate transformation in $GF(2)^n$. Namely, we associate with every vertex $x \in GF(2)^n$ the label

$$(x_n, x_n + x_{n-1}, \ldots, x_n + x_{n-1} + \ldots + x_1)^t \in GF(2)^n.$$

With these new labels, the vertex set of the digraph is again given by $GF(2)^n$. An arc of the digraph now corresponds to vertices $x, x' \in GF(2)^n$ such that $x_i = 1$, and x' arises from x by replacing x_j by $x_j + 1 \pmod 2$ for every $j \ge i$. (In particular, this yields $x'_i = 0$.)

Thus, for any vector $x \in GF(2)^n$, we consider the *game* KM(x):

choose a random coordinate r for which $x_r = 1$, and flip this coordinate together with all coordinates of higher index. This operation is repeated until the zero vector is reached.

For example, the flipping sequence considered in Section 2 corresponds, after this coordinate transformation, to the sequence

$$\begin{pmatrix}1\\1\\1\end{pmatrix} \longrightarrow \begin{pmatrix}1\\1\\0\end{pmatrix} \longrightarrow \begin{pmatrix}0\\0\\1\end{pmatrix} \longrightarrow \begin{pmatrix}0\\0\\0\end{pmatrix}.$$

The version in which we prove the lower bound of Theorem 2 in this section is the following: starting with a random vector $x \in GF(2)^n$, the expected number of *rounds played* is at least $cn^2/\log n$ for some $c > 0$.

The flipping operation. The flip at index r (in the new coordinate system) can conveniently be expressed as a linear transformation over $GF(2)^n$, i.e., there is a matrix A^r such that

$$x^r := (x_1, ..., x_{r-1}, 0, x_{r+1} + x_r, ..., x_n + x_r)^t = A^r x$$

for all vectors $x = (x_1, ..., x_n)$.

The columns of A^r are the images of the unit vectors under the flip at r, i.e.

$$A^r = \begin{pmatrix} 1 & & & \downarrow \text{column } r & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \leftarrow \text{row } r \\ & & & 1 & 1 & \\ & & & \vdots & & \ddots \\ & & & 1 & & & 1 \end{pmatrix},$$

and all other entries are zero. Note that for $j \neq r$, ${\mathbf{e}_j}^r = \mathbf{e}_j$; in general, a flip with $x^r = x$ is called *void*, and although KM(x) does not perform void flips, this more general flipping concept is a key ingredient in our approach.

Flip sequences. Let $\mathcal{S}$ be the set of all finite sequences $(s_1, ..., s_{|s|})$ with elements in $\{1, ..., n\}$. We refer to the members of $\mathcal{S}$ as *flip sequences*. For a flip sequence s we let $x^{(s)}$ be the result of 'applying' s to x, i.e.

$$x^{(s)} := A^{(s)}x, \text{ with } A^{(s)} := A^{s_{|s|}}...A^{s_2}A^{s_1}.$$

For a sequence s let s^k denote the prefix of length k (or s itself if $k > |s|$). Finally, let s^*_x denote the subsequence of indices inducing nonvoid flips when s is applied to x.

The *length* L_x of game KM(x) is the number of rounds played, i.e., the number of flips performed until termination. L_x is a random variable over $\mathcal{S}^*_x$, the set of all flip sequences that may be traversed during KM(x), and we are interested in its expectation $E[L_x]$.

Fixed length sequences

For any integer l, let $\mathcal{S}^l \subseteq \mathcal{S}$ be the probability space consisting of all flip sequences of length exactly l, equipped with the uniform distribution

$$\text{prob}_{\mathcal{S}^l}(s) = n^{-l}.$$

It is clear that we can simulate game KM by flipping with a random $r \in \{1, ..., n\}$ in each step and ignoring the void flips. This means that the expected length of game KM(x) is just the expected number of nonvoid flips encountered during the simulation, and this number can only decrease if we stop playing after the l-th flip, for any integer l (or keep on playing until the l-th flip if the game ends early – that only adds void flips). This boils down to the following formula:

Fact 4. Let l be a fixed integer. Then

$$E[L_x] \geq \frac{1}{n^l} \sum_{s \in \mathcal{S}^l} |s^*_x|.$$

Note that in the limit ($l \to \infty$) we have equality, so our objective will be to estimate the right hand side for large values of l, where the appropriate choice will be determined only at the end of the proof.

The relation to unit vector probabilities

Fact 4 leaves us with the problem of estimating the expected number of nonvoid flips in a random fixed-length sequence $s \in \mathcal{S}^l$. For this, by linearity of expectation, it is sufficient to estimate the probability that the k-th flip in such a sequence is nonvoid, for all $k \leq l$.

Fact 5.

$$\begin{aligned} \frac{1}{n^l} \sum_{s \in \mathcal{S}^l} |s^*_x| &= \sum_{k=1}^{l} \text{prob}_{\mathcal{S}^l}(x^{(s^k)} \neq x^{(s^{k-1})}) \\ &= \sum_{k=1}^{l} \text{prob}_{\mathcal{S}^k}(x^{(s^k)} \neq x^{(s^{k-1})}). \end{aligned}$$

Recall that $x^{(s^k)} \neq x^{(s^{k-1})}$ if and only if the k-th flip hits a one-entry of the current vector $x^{(s^{k-1})}$, which

implies that the probability for a nonvoid k-th flip is just the expected number of one-entries in $x^{(s^{k-1})}$, divided by n, i.e.

Fact 6.

$$\mathrm{prob}_{\mathcal{S}^k}(x^{(s^k)} \neq x^{(s^{k-1})}) = \frac{1}{n^{k-1}} \sum_{s \in \mathcal{S}^{k-1}} \frac{|x^{(s)}|}{n},$$

where $|x|$ is the support size of x, i.e. its number of one-entries. By linearity of expectation this can be rewritten as

Fact 7.

$$\mathrm{prob}_{\mathcal{S}^k}(x^{(s^k)} \neq x^{(s^{k-1})}) = \frac{1}{n} \sum_{r=1}^{n} \mathrm{prob}_{\mathcal{S}^{k-1}}({x^{(s)}}_r = 1).$$

What is the probability for ${x^{(s)}}_r = (A^{(s)}x)_r = 1$ if s runs through all sequences of some fixed length? For an individual vector x we can give no estimate, but averaging over all $x \in GF(2)^n$ (i.e. considering a random vector) will allow us to come up with a bound in terms of "unit vector probabilities".

Definition 8. For $r = 1...n$ and arbitrary k, the numbers

$$p_{r,k} := \mathrm{prob}_{\mathcal{S}^k}({\mathbf{e}_r}^{(s)} \neq \mathbf{0})$$

are called the *unit vector probabilities*.

Thus $p_{r,k}$ tells us how likely it is that the r-th unit vector 'survives' a random flip sequence of length k.

Now suppose for a moment, we were not interested in ${x^{(s)}}_r = (A^{(s)}x)_r$ but in $(x^tA^{(s)})_r$, which is the product of x and the r-th column of $A^{(s)}$, the latter being nothing else than ${\mathbf{e}_r}^{(s)}$. This product evaluates to 0 if ${\mathbf{e}_r}^{(s)} = \mathbf{0}$; otherwise — and this is the crucial observation — it evaluates to 1 with probability exactly $1/2$, if x is a random vector! This is where the unit vector probabilities come in.

Observation 9.

$$\frac{1}{2^n} \sum_{x \in GF(2)^n} \mathrm{prob}_{\mathcal{S}^k}((x^tA^{(s)})_r = 1) = \frac{1}{2}p_{r,k}.$$

In order to be able to apply this observation to our original situation, we need a lemma which allows us to relate $x^tA^{(s)}$ and $A^{(s)}x$.

Lemma 10. Let T be the $n \times n$ matrix with 1-entries on the co-diagonal and below, i.e.,

$$T_{ij} = \begin{cases} 1 & \text{if } i+j \geq n+1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

(i) $|Tx| \geq |x|/2$ for all $x \in GF(2)^n$.

(ii) $A^rT = (A^{n+1-r}T)^t$.

Note that T is exactly the matrix corresponding to the coordinate transformation introduced at the beginning of this section.

Proof. (i) We have $(Tx)_r = \sum_{j \geq n+1-r} x_j$, so $(Tx)_r$ just gives the parity of the last r positions of x, and consequently $(Tx)_{r-1} \neq (Tx)_r$ if and only if $x_{r-1} = 1$. This means that by scanning x from bottom to top the parity (starting from 0) changes whenever a one-entry is encountered, so at least $\lceil |x|/2 \rceil$ of the parities must be one.

(ii) Let E be the unit matrix. Then we can write the i-th row ${A^r}_i$ of A^r as

$${A^r}_i = E_i + E_r[i \geq r].$$

Therefore

$$A^rT = T + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ E_r \\ \vdots \\ E_r \end{pmatrix} T.$$

The product evaluates to a matrix with a block of one entries in rows r through n and columns $n+1-r$ through n, and exactly the transpose is obtained when we start with $A^{n+1-r}T$. By observing that T is symmetric, the claim follows. □

Recall that in order to estimate the probability of a nonvoid k-th flip, we would like to bound the expected support size after $k-1$ flips, i.e. the expected number of one-entries in $A^{(s)}x$ for a random flip sequence s of length $k-1$ (Fact 6). First of all, it is an immediate consequence of part (ii) of Lemma 10 that

$$A^{(s)} = T(A^{(s')})^tT^{-1},$$

where

$$s' = (n+1-s_{|s|}, ..., n+1-s_1),$$

i.e., s' arises from s by reversion and replacement of every entry r with $n+1-r$. Obviously, if s is random, this also holds for s'. Furthermore, if x is random, so is $T^{-1}x$. Consequently, the expected support size of

$$A^{(s)}x$$

equals the expected support size of

$$T(A^{(s)})^tx$$

when averageing over all sequences *and* all vectors.

Furthermore, by Lemma 10(i) we have

$$|T(A^{(s)})^t x| \geq \frac{|(A^{(s)})^t x|}{2} = \frac{|x^t A^{(s)}|}{2}.$$

This line of argumentation together with Observation 9 gives

$$\begin{aligned} & \sum_{x \in GF(2)^n} \sum_{r=1}^{n} \mathrm{prob}_{\mathcal{S}^{k-1}}({x^{(s)}}_r = 1) \\ = & \sum_{x \in GF(2)^n} \sum_{r=1}^{n} \mathrm{prob}_{\mathcal{S}^{k-1}}((T(A^{(s)})^t x)_r = 1) \\ \geq & \frac{1}{2} \sum_{x \in GF(2)^n} \sum_{r=1}^{n} \mathrm{prob}_{\mathcal{S}^{k-1}}((x^t A^{(s)})_r = 1) \\ = & 2^{n-2} \sum_{r=1}^{n} p_{r,k-1}. \end{aligned}$$

Via Fact 7 we get the main result of this subsection.

Lemma 11.

$$\frac{1}{2^n} \sum_{x \in GF(2)^n} \mathrm{prob}_{\mathcal{S}^k}(x^{(s^k)} \neq x^{(s^{k-1})}) \geq \frac{1}{4n} \sum_{r=1}^{n} p_{r,k-1}.$$

This shows in particular that as long as the expected number of unit vectors that are 'alive' is linear, the expected probability (over all vectors) for a nonvoid flip will be at least constant; this again means that if linearly many unit vectors 'survive' a random flip sequence of length l, we can expect $\Theta(l)$ nonvoid flips on a random vector. To prepare the choice of a suitable l, the next subsection will investigate the unit vector probabilities; actually, the analysis will be more general and establish bounds for arbitrary vectors.

General vector probabilities

Let $x \in GF(2)^n, x \neq \mathbf{0}$ be fixed. The objective of this subsection will be to find the values of k, for which the general vector probabilities

$$\mathrm{prob}_{\mathcal{S}^k}(x^{(s)} \neq \mathbf{0})$$

are at least some constant. Obviously, for $k = 0$ the probability is 1, while for $k \to \infty$ it tends to zero, and our goal will be to determine k asymptotically as large as possible (depending on n and x) such that we still can find some constant bounding the above probability from below.

To this end we will analyze how x evolves when applying a random flip sequence $s \in \mathcal{S}^k$ to it. Actually, the analysis will only trace the *leading dimension*, which records the one-entry with lowest index. Therefore, the considerations for x are valid as well for any other vector y with the same lowest one-entry.

Definition 12. For a nonzero vector $y \in GF(2)^n$, the leading dimension is the number

$$d(y) := n + 1 - \min\{r \mid y_r = 1\}.$$

Furthermore, $d(\mathbf{0}) := 0$.

During the traversal of a flip sequence, the leading dimension of the vector under consideration can only decrease, and we will show that it does not decrease too fast on the average.

With respect to x, any flip sequence s subdivides into *phases*, where a flip terminates a phase if it decreases the leading dimension of the current vector, or if it is the last flip in s. Obviously there are at most n phases, and every single phase makes some progress, which is the decrease in leading dimension during the phase.

Definition 13. For $p \leq n$ and $s \in \mathcal{S}^k$ let s_j be the flip terminating phase p (if it exists). We define $d_p(s)$ as the leading dimension during phase p, i.e.,

$$d_p(s) := d(x^{(s^{j-1})}).$$

The *progress* in phase p is defined by

$$\Delta_p(s) := d_p(s) - d(x^{(s^j)}).$$

Δ_p is a random variable over $\mathcal{S}^k$. The progress in the last phase may be 0, and to a nonexisting phase we assign progress 0 in any case. The next lemma shows that the expected progress in a single phase is no more than logarithmic in the leading dimension; this will be the basis of the final bound derived in this subsection.

Lemma 14.

$$E[\Delta_p] \leq H_{d(x)},$$

where $H_i := 1 + \frac{1}{2} + \ldots + \frac{1}{i}$ is the i-th harmonic number.

Proof. Observing that $\Delta_p \leq d(x)$, we can write the expectation as

$$\begin{aligned} E[\Delta_p] &= \sum_{q \geq 1} q \ \mathrm{prob}_{\mathcal{S}^k}(\Delta_p = q) \\ &= \sum_{q=1}^{d(x)} \mathrm{prob}_{\mathcal{S}^k}(\Delta_p \geq q). \end{aligned}$$

Now we claim that

$$\mathrm{prob}_{\mathcal{S}^k}(\Delta_p \geq q) \leq \frac{1}{q},$$

which implies the lemma.

To see this inequality, consider the leading one-entry during phase p (which has index $n+1-d_p(x)$) and the next higher $q-1$ indices (if this is not defined, the probability is zero anyway). It is easy to see that in order to make progress at least q, phase p of the flip sequence must necessarily hit the leading one-entry before it hits any of the $q-1$ higher indices (and this may not even be sufficient). However, if any of the q positions are flipped at all, then the probability that the leading one comes first is only $1/q$. □

Let

$$\Delta_{\leq p} := \sum_{p' \leq p} \Delta_{p'}$$

be the progress in the first p phases. By linearity of expectation we immediately get

Corollary 15.

$$E[\Delta_{\leq p}] \leq pH_{d(x)}.$$

By setting

$$p = p(x) := \frac{d(x)}{eH_{d(x)}}$$

we see that the expected progress in the first $p(x)$ phases of a random flip sequence is at most $d(x)/e$ when starting with x. Of course, the overall progress is at most $d(x)$, and Markov's inequality shows that it actually is $d(x)$ with probability at most $1/e$; this, on the other hand implies that with probability at least $1-1/e$ the progress is smaller, which exactly means that x has not been reduced to $\mathbf{0}$ within $p(x)$ phases.

For a flip sequence s let us denote the prefix of s that consists of the first (at most) p phases of s by

$$s|_p.$$

What we have shown so far is that

$$\text{prob}_{\mathcal{S}^k}(x^{(s|_{p(x)})} \neq \mathbf{0}) \geq 1 - \frac{1}{e}, \tag{1}$$

i.e., with probability larger than $1/2$, x 'survives' the first $p(x)$ phases of a random flip sequence. What we want, however, is that it survives the whole sequence with constant probability. To achieve this, it will be sufficient to ensure that with probability larger than $1/2$ the whole sequence actually consists of at most $p(x)$ phases, and this can be done by choosing the parameter k appropriately. To get an intuition what the value for k should be, observe that we can expect a phase to terminate after n flips, so for $k = np(x)$ a random sequence should have $p(x)$ phases on the average, and if k is chosen slightly smaller, with high probability there will be no more than $p(x)$ phases.

Formally, we go about as follows: $s \in \mathcal{S}^k$ determines a sequence $B(s)$ of k Bernoulli trials, where success in a trial either means that the leading dimension decreases, or — in the case that x has already been reduced to $\mathbf{0}$ — position 1 is hit. When we traverse a random sequence s, success occurs independently in each trial with probability $1/n$; moreover, each phase (except possibly the last one) contributes a success. This means that the number of phases in s is bounded by the number of successes in $B(s)$ plus one. Let Z be a random variable counting the total number of successes in a sequence of k Bernoulli trials, with probability of success equal to $1/n$. We have

$$E[Z] = \frac{k}{n}$$

and

$$\text{prob}(Z \geq \frac{k}{n} + t) \leq e^{-t^2 n/4k(1-1/n)}$$

for any positive integer t, by the well-known Chernoff bound. Setting

$$k := n(p(x) - 2\sqrt{p(x)}) \text{ and } t := 2\sqrt{p(x)}$$

gives

$$\text{prob}(Z \geq p(x)) \leq e^{-p(x)/(p(x)-2\sqrt{p(x)})(1-1/n)} \leq \frac{1}{e}.$$

This means, the number of phases in a random $s \in \mathcal{S}^k$ is larger than $p(x)$ with probability at most $1/e$. Hence

$$\text{prob}_{\mathcal{S}^k}(s = s|_{p(x)}) \geq 1 - \frac{1}{e}, \tag{2}$$

for $k \leq n(p(x) - 2\sqrt{p(x)})$.

We finally obtain the lower bounds on the general vector probabilities.

Lemma 16. For

$$k \leq n(\frac{d(x)}{eH_{d(x)}} - 2\sqrt{\frac{d(x)}{eH_{d(x)}}})$$

we have

$$\text{prob}_{\mathcal{S}^k}(x^{(s)} \neq \mathbf{0}) \geq 1 - \frac{2}{e} \approx 0.26.$$

Proof. Consider over $\mathcal{S}^k$ the events

$$\begin{aligned} A &= (x^{(s|_{p(x)})} \neq \mathbf{0}), \\ B &= (s = s|_{p(x)}). \end{aligned}$$

$A \cap B$ implies $x^{(s)} \neq \mathbf{0}$, and by invoking inequalities (1) and (2) we see that $\text{prob}_{\mathcal{S}^k}(x^{(s)} \neq \mathbf{0})$ is bounded from below by

$$\begin{aligned} & \text{prob}_{\mathcal{S}^k}(A \cap B) \\ = \; & \text{prob}_{\mathcal{S}^k}(A) + \text{prob}_{\mathcal{S}^k}(B) - \text{prob}_{\mathcal{S}^k}(A \cup B) \\ \geq \; & (1 - \frac{1}{e}) + (1 - \frac{1}{e}) - 1 = 1 - \frac{2}{e}. \end{aligned}$$

□

Putting everything together

The previous subsections have developed all the machinery we need to prove Theorem 2, which — adapted to the terminology of this section — reads as follows.

Theorem 2. There exists a constant $c > 0$ (independent of n) such that

$$\frac{1}{2^n} \sum_{x \in GF(2)^n} E[L_x] \geq c \frac{n^2}{\log n}.$$

Proof. After setting

$$l := n(\frac{n/2+1}{eH_{n/2+1}} - 2\sqrt{\frac{n/2+1}{eH_{n/2+1}}}) = \Theta(\frac{n^2}{\log n}),$$

any $k \leq l$ satisfies the condition of Lemma 16 for vectors $x = \mathbf{e}_1, ..., \mathbf{e}_{n/2}$. We thus get

$$\begin{aligned} \sum_{x \in GF(2)^n} E[L_x] & \overset{(4)}{\geq} \sum_{x \in GF(2)^n} \frac{1}{n^l} \sum_{s \in \mathcal{S}^l} |s^*_x| \\ & \overset{(5)}{=} \sum_{x \in GF(2)^n} \sum_{k=1}^{l} \text{prob}_{\mathcal{S}^k}(x^{(s^k)} \neq x^{(s^{k-1})}) \\ & \overset{(11)}{\geq} \frac{2^n}{4n} \sum_{k=1}^{l} \sum_{r=1}^{n} \text{prob}_{\mathcal{S}^{k-1}}(\mathbf{e}_r{}^{(s)} \neq \mathbf{0}) \\ & \geq \frac{2^n}{4n} \sum_{k=1}^{l} \sum_{r=1}^{n/2} \text{prob}_{\mathcal{S}^{k-1}}(\mathbf{e}_r{}^{(s)} \neq \mathbf{0}) \\ & \overset{(16)}{\geq} \frac{2^n}{4n} \sum_{k=1}^{l} \sum_{r=1}^{n/2} (1 - \frac{2}{e}) \\ & = \frac{2^n}{8}(1 - \frac{2}{e})l = 2^n \Theta(\frac{n^2}{\log n}), \end{aligned}$$

and by dividing with 2^n we see that the average expectation (over all vectors) is at least of the order

$$\Theta(\frac{n^2}{\log n}).$$

(The constant c is rather small, about 1/165). □

4 Related Models

In this final section, we provide two more combinatorial models for classes of linear programs with exponentially long decreasing paths. A main feature of these two classes — as compared to the Klee-Minty cubes — is that they include polytopes with arbitrarily large number of facets in any fixed dimension. In both classes, we can prove quadratic upper bounds for the running time of RANDOM-EDGE with arbitrary starting vertex.

Deformed products. This class of linear programs was also constructed by Klee & Minty [16]. Its polytopes are combinatorially equivalent to products of 1- and 2-dimensional polytopes. For the following, we restrict to the special case where the dimension n is even, and $P := (C_k)^{n/2}$ is a product of k-gons: an n-dimensional polytope with $m = \frac{kn}{2}$ facets. Such polytopes are now realized in $\mathbb{R}^n$ ("deformed") in such a way that they have an x_n-decreasing path through all the vertices. The geometric construction of these programs is tricky [16], but the combinatorial model is very simple, as follows.

The vertex set of P can naturally be identified with the set of vectors $\{1, \ldots, k\}^{n/2}$. Two vertices are adjacent if their vectors $x, x' \in \{1, \ldots, k\}^{n/2}$ differ in a single coordinate, and in this coordinate the difference is either 1 or $k-1$. Furthermore, the directions for these edges are given as follows: if x and x' differ in their i-th coordinate, then we get a directed edge $x \to x'$ if either $x_i < x'_i$ and $(x_1, \ldots, x_{i-1})$ contains an even number of even entries, or if $x_i > x'_i$ and $(x_1, \ldots, x_{i-1})$ contains an odd number of even entries. For example, for $n = 4$ and $k = 3$ ($m = 6$) we get
$11 \to 12 \to 13 \to 23 \to 22 \to 21 \to 31 \to 32 \to 33$
as the directed path through all the vertices.

This explicitly describes a digraph, on which algorithms such as RANDOM-EDGE take a random walk.

Proposition 17. For an arbitrary starting vertex x on a deformed product program, the expected number of steps taken by the RANDOM-EDGE algorithm is bounded by a quadratic function, namely,

$$E_{n,m}(x) \leq n \cdot m.$$

The function $E_{n,m}(x)$ is, however, not even completely analyzed for the case $n = 4$.

For the deformed products, the shortest path from the highest to the lowest vertex visits only these two vertices, while the longest decreasing path visits all the $k^{n/2} = (\frac{2m}{n})^{n/2}$ vertices. In constant dimension this yields a longest decreasing path of length $O(m^{n/2})$,

which is asymptotically sharp. However, for other interesting parameter settings, like $m = 2n$, there might be substantially longer paths — see the following construction.

Cyclic programs. Here the construction starts with the polars $C_n(m)^\Delta$ of cyclic polytopes [9, 26]. These simple polytopes have the maximal number of vertices for given m and n, namely

$$V(n,m) = \binom{m - \lceil \frac{n}{2} \rceil}{\lfloor \frac{n}{2} \rfloor} + \binom{m - 1 - \lceil \frac{n-1}{2} \rceil}{\lfloor \frac{n-1}{2} \rfloor},$$

according to McMullen's upper bound theorem [20, 26]. The facets of $C_n(m)^\Delta$ are identified with $[m] := \{1, 2, \ldots, m\}$; the vertices correspond to those n-subsets $F \subseteq [m]$ which satisfy "Gale's evenness condition": if $i, k \in [m] \backslash F$, then the set $\{j \in F : i < j < k\}$ has even cardinality.

Now any two sets $F = \{i_1, i_2, ..., i_n\}_<$ and $G = \{j_1, j_2, ..., j_n\}_<$ satisfying Gale's evenness condition are compared by the following *twisted lexicographic order*: $F < G$ if and only if $i_1 < j_1$,
or $i_1 = j_1, \ldots, i_k = j_k, i_{k+1} < j_{k+1}$, and i_k is even,
or $i_1 = j_1, \ldots, i_k = j_k, i_{k+1} > j_{k+1}$, and i_k is odd.

Thus one compares the first element in which the (sorted) sets F and G differ, and takes the natural order if the element before that is even (or doesn't exist), and the reversed order if the element before is odd. For example, for $C_4(8)^\Delta$ we get the ordering
$1678 < 1568 < 1458 < 1348 < 1238 < 1234 < 1245 < 1256 < 1267 < 1278 < 2378 < 2367 < 2356 < 2345 < 3456 < 3467 < 3478 < 4578 < 4567 < 5678$.

Now we use this ordering to construct the digraph model. Its vertices are the sets satisfying Gale's evenness condition. There is a directed edge $F \to F'$ if and only if $F' < F$ and F, F' differ in exactly one element, that is, the corresponding vertices of $C_n(m)^\Delta$ are adjacent.

The special property of the ordering is that every vertex is adjacent to the previous one. Thus the digraph is acyclic with unique source and sink, and with a directed path through all the vertices. (The construction is derived from Klee [13], where the order is constructed and described recursively.)

It is not clear if one can realize the polytope $C_n(m)^\Delta$ such that the x_n-coordinate orders the vertices according to twisted lexicographic order. (Equivalently, we cannot show that this order corresponds to a Bruggesser-Mani shelling [3, 26] of some realization of the cyclic polytope.) If such a realization is possible, then this solves the "upper bound problem for linear programs ":

> "What is the largest possible number $P(n, m)$ of vertices on a decreasing path in a linear program of dimension n with m facets?"

(a very natural unsolved problem!) by showing that the bound $P(n, m) \leq V(m, n)$, from the upper bound theorem for polytopes, holds with equality.

Even without such a realization, the twisted lexicographic ordering yields an abstract objective function in the sense of [24], and thus a shelling of the cyclic polytope $C_n(m)$. Thus our digraph model is a very reasonable "worst case" (?) szenario for the performance of randomized simplex algorithms. Both the RANDOM-EDGE and the RANDOM-FACET variants can, indeed, be analyzed in terms of the digraph model, without use of a metric realization.

Proposition 18. For the RANDOM-EDGE rule, started at an arbitrary vertex F of the cyclic program, there is a linear lower bound and a quadratic upper bound for the expected number of steps. For this, we set $\ell(F) := m + 1 - \min(F)$, with $n \leq \ell(F) \leq m$, and obtain

$$\ell(F) - n \leq E_{n,m}(x) \leq \binom{\ell(F)+1}{2} - \binom{n+1}{2}.$$

Since both the diameter problem [17, 14] and the algorithm problem [4, 11] have upper bounds that are linear in m, it would be interesting to know that $E_{n,m}(x)$ indeed grows at most linearly in m for such problems. On the other hand, it is certainly challenging to strive for a nonlinear lower bound for these models.

More details for the analysis of the models in this section will appear in [6].

Thanks

We wish to thank M. Reiss, E. Welzl and M. Henk for helpful discussions, comments and contributions, and D. Kühl, W. Schlickenrieder, C. Betz-Haubold, T. Takkula and W. Neun for extensive computations.

References

[1] K.H. BORGWARDT: *The Simplex Method. A Probabilistic Analysis,* Algorithms and Combinatorics **1**, Springer 1987.

[2] R.G. BLAND: *New finite pivoting rules for the simplex method,* Math. Operations Research **2** (1977), 103-107.

[3] H. Bruggesser & P. Mani: *Shellable decompositions of cells and spheres,* Math. Scand. **29** (1971), 197-205.

[4] K. Clarkson: *A Las Vegas algorithm for linear programming when the dimension is small,* in: "Proc. 29. IEEE Symposium on Foundations of Computer Science (FOCS)" 1988, 452-457.

[5] G.B. Dantzig: *Linear Programming and Extensions,* Princeton University Press, Princeton 1963.

[6] B. Gärtner, G.M. Ziegler et al.: *Randomized simplex algorithms on special linear programs,* in preparation 1994.

[7] D. Goldfarb: *Worst case complexity of the shadow vertex simplex algorithm,* preprint, Columbia University 1983, 11 pages.

[8] D. Goldfarb: *On the complexity of the simplex algorithm,* preprint, Columbia University 1993, 21 pages; in: "IV. Workshop on Numerical Analysis and Optimization" in Oaxaca, Mexico (ed. S. Gomez), to appear.

[9] B. Grünbaum: *Convex Polytopes,* Interscience, London 1967.

[10] M. Henk: *Note on the Klee-Minty cube,* manuscript, 1993.

[11] G. Kalai: *A subexponential randomized simplex algorithm,* in: "Proc. 24th ACM Symposium on the Theory of Computing (STOC)", ACM Press 1992, 475-482.

[12] D.G. Kelly: *Some results on random linear programs,* Methods of Operations Research **40** (1981), 351-355.

[13] V. Klee: *Paths on polyhedra II,* Pacific Journal of Mathematics **17** (1966), 249-262.

[14] V. Klee & P. Kleinschmidt: *The d-step conjecture and its relatives,* Math. Operations Research **12** (1987), 718-755.

[15] V. Klee & P. Kleinschmidt: *Geometry of the Gass-Saaty parametric cost LP algorithm,* Discrete Comput. Geometry **5** (1990), 13-26.

[16] V. Klee & G.J. Minty, *How good is the simplex algorithm?,* in: "Inequalities III" (O. Sisha, ed.), Academic Press, New York 1972, 159-175.

[17] D. G. Larman: *Paths on polytopes,* Proc. London Math. Soc. **20** (1970), 161-178.

[18] J. Matoušek: *Lower bounds for a subexponential optimization algorithm,* preprint B 92-15, FU Berlin 1992, 18 pages.

[19] J. Matoušek, M. Sharir & E. Welzl: *A subexponential bound for linear programming,* in: "Proc. 8th Annual ACM Symp. Computational Geometry" (Berlin 1992), ACM Press 1992, 1-8.

[20] P. McMullen: *The maximum numbers of faces of a convex polytope,* Mathematika **17** (1970), 179-184.

[21] C.H. Papadimitriou & K. Steiglitz: *Combinatorial Optimization: Algorithms and Complexity,* Prentice-Hall, Inc., Englewood Cliffs, NJ 1982.

[22] A. Schrijver: *Theory of Linear and Integer Programming,* Wiley-Interscience, Chichester 1986.

[23] R. Shamir: *The efficiency of the simplex method: a survey,* Management Sci. **33** (1987), 301-334.

[24] K. Williamson Hoke: *Completely unimodal numberings of a simple polytope,* Discrete Applied Math. **20** (1988), 69-81.

[25] N. Zadeh: *What is the worst case behavior of the simplex algorithm?,* Technical Report No. 27, Dept. Operations Research, Stanford 1980, 26 pages.

[26] G.M. Ziegler: *Lectures on Polytopes,* Springer-Verlag New York 1994, in preparation.

Bernd Gärtner
Institut für Informatik
Freie Universität Berlin
Takustr. 9
14195 Berlin, Germany
gaertner@inf.fu-berlin.de

Günter M. Ziegler
Konrad-Zuse-Zentrum für
Informationstechnik Berlin (ZIB)
Heilbronner Str. 10
10711 Berlin, Germany
ziegler@zib-berlin.de