An upper bound for the average number of iterations required in phase II of an interior-point-method

Petra Huhn, Karl-Heinz Borgwardt

Angaben zur Veröffentlichung / Publication details:

Huhn, Petra, and Karl-Heinz Borgwardt. 1998. "An upper bound for the average number of iterations required in phase II of an interior-point-method." *Operations Research Proceedings* 1997: 19–24. https://doi.org/10.1007/978-3-642-58891-4_3.

An Upper Bound for the Average Number of Iterations Required in Phase II of an Interior-Point-Method

Petra Huhn[1] *and Karl Heinz Borgwardt*[1]

[1]Department of Mathematics, University of Augsburg, 86135 Augsburg, Germany

1 Introduction

In this paper we discuss the solution process of linear programming problems (LPs) of the following type:

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

when an *interior-point-method* (IPM) is employed to solve the problem.

We call $v^T x$ the *objective function* and $X = \{x \,|\, a_1^T x \leq 1, \ldots, a_m^T x \leq 1\}$ the *feasible region.* The *interior of* X is denoted by $Int\,X$. The matrix $A^T = (a_1, \ldots, a_m)$ collects all the restriction vectors a_i. Furthermore, we agree in the following *assumption on nondegeneracy*:

Each n*-elementic subset of* $\{a_1, \ldots, a_m, v\}$ *is linearly independent and each* $(n+1)$*-elementic subset of* $\{a_1, \ldots, a_m\}$ *is in general position.*

Note, that the origin 0 is a point of $Int\,X$ and X is pointed because of nondegeneracy.

IPMs solve LPs with optimal solutions via performing the following steps.

Phase I: Determination of a point x_0 belonging to $Int\,X$ and satisfying some start-criteria (which depend on the specific IPM).

Phase IIa: Construction of a sequence $x_0, x_1, \ldots, x_k, x_{k+1}, \ldots, x_K$ in $Int\,X$, such that the duality gap (or a barrier-/potential-function) is successively reduced until a certain desired size is achieved at x_K.

Phase IIb: Determination of the optimal point x_{opt} based on the information about the last iterate x_K (usually called the termination procedure).

We denote the *duality gap* at x by $z - v^T x$, where z is an upper bound of $v^T x_{opt}$. Notice, that the reduction of the duality gap (or barrierfunction via the barrier-parameter) will result in a reduction of the gap $v^T x_{opt} - v^T x_k$ in the end, so we will focus our considerations on this gap and on the number of iterations K.

IPMs as presented in [3] and other publications are superior to the Simplex-Method in two aspects. First, the computational effort of IPMs can be bounded from above by a polynomial in the encoding length L of the specific problem instance. And for large dimensions (m, n) and sparse matrices the IPMs stand the test versus the

Simplex-Method very well. But, finding a suitable x_0 (Phase I), doing the iteration process and finding x_{opt} (Phase II) may be a great challenge for such a method.

In this paper we leave the starting problem aside, but mention:

Remark: If X provides an optimal vertex, then it is possible to find a suitable starting point (compare [8]). Although $0 \in Int\, X$, 0 is not in general a suitable starting point. Starting at 0 we can apply a "Phase I"-Algorithm. The effort for doing Phase I is not counted here, where we deal with the complexity of Phase II.

The worst-case analysis makes use of the following facts (compare [3], [6]).

Lemma 1:

a) It is possible to find a starting point x_0, whose duality gap is less than $O(2^L)$.

b) The iteration procedure assures a linear reduction of the duality gap in each step with a reduction rate depending on m.

c) There are termination procedures for Phase IIb, which will deliver the optimal vertex x_{opt}, if $v^T x_k \leq v^T x_{II}$. Here x_{II} is the second best vertex of X, i. e. $v^T x_{opt} > v^T x_{II}$, but there is no vertex $\bar{x}$ of X such that $v^T_{opt} > v^T \bar{x} > v^T x_{II}$.

d) Two vertices x_1 and x_2 with $v^T x_1 \geq v^T x_2$ differ in their objective values at least by 2^{-L}. In particular, this means that $v^T x_{opt} - v^T x_{II} > 2^{-L}$.

We call the region $\{x \,|\, x \in X,\ v^T x > v^T x_{II}\}$ the *safety region* and $v^T x_{opt} - v^T x_{II}$ the *safety gap*. Often the reduction rate is given by a factor $\left(1 - \frac{1}{R(m)}\right)$ where $R(m)$ depends on the interior-point variant (typically, $R(m) = O(m)$ or $O(\sqrt{m})$). So, we get an upper bound on the duality gap and $v^T x_{opt} - v^T x_k$:

$$v^T x_{opt} - v^T x_k \;\leq\; \left(1 - \frac{1}{R(m)}\right)^k (z_0 - v^T x_0). \tag{1}$$

This and the results of Lemma 1 lead to the following worst-case result.

Theorem 2:

a) It is sufficient to proceed with Phase IIa until $v^T x_{opt} - v^T x_K < 2^{-L}$.

b) To reduce the duality gap from $O(2^L)$ to 2^{-L} not more than $O(R(m)L)$ iterations are necessary.

So, the IPMs are called polynomial – but not strongly polynomial, because the upper bound polynomial does not depend on the dimensions (m, n) only. Theorem 2a) shows why L enters the "finishing effort". Practical experience suggests that the mentioned bounds overestimate the usual number of iterations dramatically.

If we want to have a convincing theoretical explanation for that effect, we should try to carry out a probabilistic analysis. For that purpose one has to make assumptions on the distribution of the data of (P). We will base our investigation on the Rotation-Symmetry-Model (RSM), which had been used by Borgwardt [2] in his analysis of the Simplex-Method. So we can compare IPMs with the Simplex-Method on the basis of their average-case-behaviour. The RSM demands that

$$a_1, \ldots, a_m, v \text{ are distributed on } \mathbb{R}^n \setminus \{0\}$$
$$\text{identically, independently and symmetrically under rotations.}$$

We specialize this model here to the uniform distribution on the *unit sphere* ω_n in $\mathbb{R}^n$ and refer to it as *uni*-RSM. Note, that the RSM gives nondegeneracy the probability 1. So, the assumption of nondegeneracy does not influence our results.

Some approaches towards such a probabilistic analysis have been done already (compare [1], [5], [7] and [8]), but they apply to special variants of IPMs or often artificial stochastic assumptions are necessary for the probabilistic evaluation.

The main result of this paper is that the expected number of iterations in Phase II is strongly polynomial in m and n, (i. e. the encoding length L will not appear in upper bounds for the expected behaviour of Phase II).

We proceed in the following steps: First, we explain a termination procedure, which will work for every bounded problem, when $v^T x_K > v^T x_{II}$. Secondly, we give an algorithm (II) which enables us to recognize that the safety region is reached. In section 3 the information on the guaranteed reduction of the IPM and information on the distribution of the safety gap are composed to a result on the expected number of iterations. In section 4 we report on an analysis in [4], where the stochastic variable $v^T x_{opt} - v^T x_{II}$ and its distribution have been evaluated. At this point we can derive an explicit upper bound on the expected number of iterations in Phase II. At last, we give another algorithm (III), based on Algorithm II, which solves all complication cases, too (i. e. no optimal point, no starting point, etc.) and show that our result on the average effort applies to Algorithm III as well.

2 A Termination Procedure and an Algorithm for Phase II

Suppose that a point $x_k \in Int\, X$ is available. Let $v^T x$ be bounded on X. Then consider the following vertex-finding procedure.

Vertex-Finding Procedure: VF(A, v, x_k)

Initialization: $l := 0$, $d_0 := v$, $\xi_0 := x_k$, $I_0 = \emptyset$. Provide space for a matrix $\bar{A}$ with n columns and a specified number of rows. $\bar{A}_0$ is an "empty" matrix.

Typical step:

for $l = 0, 1, \ldots, n-1$ *do begin*

1) Determine $\bar{\lambda} := \text{Min}\left\{\frac{1-a_i^T\xi_l}{a_i^T d_l} \,|\, i \notin I_l, a_i^T d_l > 0\right\}$ and the arg-min index $i_l \notin I_l$ such that $a_{i_l}^T(\xi_l + \bar{\lambda} d_l) = 1$.

2) Set $\xi_{l+1} := \xi_l + \bar{\lambda} d_l$, $I_{l+1} := I_l \cup \{i_l\}$ and $\bar{A}_{l+1} := \begin{pmatrix} \bar{A}_l \\ a_{i_l}^T \end{pmatrix}$.

3) Calculate $d_{l+1} := (E - \bar{A}_{l+1}^T(\bar{A}_{l+1}\bar{A}_{l+1}^T)^{-1}\bar{A}_{l+1})v$.

end.

Output: ξ_n

Lemma 3:
Let $v^T x$ be bounded on X. Under the assumption of nondegeneracy we know:
a) The procedure VF delivers a vertex ξ_n of X.

b) If $v^T x_k > v^T x_{II}$, then the procedure VF delivers x_{opt}.

c) VF requires a computational effort of $O(n^3)$ arithmetical operations.

If $v^T x_k \leq v^T x_{II}$, then VF leads to a vertex, but not necessarily to x_{opt}.

But still, we do not know $v^T x_{II}$ and therefore we do not recognize being better than $v^T x_{II}$ or not. Our expedient lies in applying the VF-procedure as a test in each iteration.

Algorithm II "Test-and-Iterate" for Phase II

Input: A point $x_0 \in Int\, X$, which is a suitable starting point for the interior point variant under consideration, A and v. Set $k = 0$.

Typical Step:

1) Given x_k, ξ_n (the result of VF(A, v, x_k)) decide whether ξ_n is optimal or not (by checking the dual variables at ξ_n).
2) If ξ_n is optimal, then STOP with **Output:** $\xi_n = x_{opt}$ and $v^T \xi_n = v^T x_{opt}$.
3) Apply an iteration step of the interior-point-variant under consideration to calculate x_{k+1} at the base of x_k.
4) Set $k := k + 1$ and goto 1).

Remark: The work of step 1) can be done in $O(n^3)$ operations, this does not exceed the effort of an iteration of an IPM.

3 The random size of the safety gap and its impact on the number of iterations

Assume that Phase I has produced a suitable starting point x_0 with a duality gap of at most C, where C is a given constant, and let $v^T x$ be bounded on X. Moreover, suppose that the IPM used in step 3) of Algorithm II provides a reduction guarantee as described in (1). It follows that $v^T x_{opt} - v^T x_k \leq \left(1 - \frac{1}{R(m)}\right)^k C$. Now we can derive an upper bound for the number of iterations for reaching the safety gap.

Lemma 4:

If $v^T x$ is bounded on X and if U denotes the safety gap $v^T x_{opt} - v^T x_{II}$, then Algorithm II requires at most

$$K = \begin{cases} 0 & \text{for } U \geq C, \\ \ln |\frac{U}{C}| \cdot R(m) & \text{for } U < C. \end{cases} \qquad \text{iteration steps.}$$

In a probabilistic analysis it will be desirable to know the distribution function of U, which is a function $F_U(\varepsilon) : [0, \infty] \to [0, 1]$ (the knowledge of F_U is desirable at least for $U < C$). We conclude

$$E_{F_U}[K] \leq R(m) \int_0^C |\ln \frac{\varepsilon}{C}|\, dF_U(\varepsilon) = R(m) \left(F_U(C) \ln C - \int_0^C \ln \varepsilon \, dF_U(\varepsilon) \right) \qquad (2)$$

Remark: We will increase the expectation $E_{F_U}[K]$, if we replace F_U in (2) by another distribution function $\widetilde{F}_U$ such that $F_U(\varepsilon) \leq \widetilde{F}_U(\varepsilon) \quad \forall \varepsilon$ with $0 < \varepsilon < C$.

4 Results of Stochastic Geometry

The methodology of stochastic geometry, which had been successfully used in the probabilistic analysis of the Simplex-Method in [2] was exploited in [4] for getting information on the required figures. Here we only report on the results.

The first result concerns a distribution function $\widetilde{F}_U$, which dominates F_U, the conditional distribution function of U when x_{opt} exists and $\|x_{opt}\| \leq \frac{1}{q}$, $q \in (0,1)$.

Theorem 5: *Let $a_1, \ldots, a_m, v$ be distributed according to uni-RSM, let $n \geq 3$, $m \geq 5n^{\frac{3}{2}}$, $q \in (0, \tilde{q}(\frac{2n}{m}))$ with $\tilde{q}(\eta) := \sqrt{1 - (\frac{|\omega_n|(n-1)}{|\omega_{n-1}|}\eta)^{\frac{2}{n-1}}}$. Then defining the distribution function*

$$\widetilde{F}_U(\varepsilon) := \begin{cases} s_1(n,q)\varepsilon^{\frac{1}{3}} + s_2(m,n,q)\varepsilon + s_3(m,n,q)\varepsilon^{\frac{1}{3}} & \text{for } \varepsilon \in [0, \bar{\varepsilon}], \\ 1 & \text{for } \varepsilon > \bar{\varepsilon}, \end{cases}$$

it is guaranteed that $F_U(\varepsilon) \leq \widetilde{F}_U(\varepsilon)$. Here: $\bar{\varepsilon} = \inf\{\varepsilon \mid \widetilde{F}_U(\varepsilon) \leq 1\} \approx \frac{1}{5}\, q^{6n} n^{-9} m^{-\frac{3}{n-1}}$, $s_1(n,q) = n^2 q^{-n}$, $s_2(m,n,q) = 4.2 n^{\frac{5}{2}} q^{-2n} m^{\frac{3}{n-1}}$ and $s_3(m,n,q) = 3.3 n^3 q^{-2n} m^{\frac{1}{n-1}}$.

We set $C = \frac{1}{q}$, $(q \in (0,1))$. Now we can bound the probability of not satisfying the condition that an optimal vertex exists with $\|x_{opt}\| \leq \frac{1}{q}$.

1) $P(X \text{ is unbounded}) \leq \frac{C(n)}{(m-1)2^{m-1}}$ with $C(n) = \pi^{\frac{n-2}{2}}(2n)^{\frac{n}{2}}(n-2)^{\frac{1}{2}}(n(n-2)-1)^{-\frac{1}{2}}$.
2) $P(\text{there is an optimal vertex, but } \|x_{opt}\| \geq \frac{1}{q}) \leq 0.8G(q)^{m-n}\binom{m}{n}$, where G is the marginal distribution function of the last (any) component of a random vector distributed according to the RSM.

All unfavourable cases (where our assumptions are not valid and therefore Algorithm II may not work) are covered by these two events.

5 Results for the Average Complexity of IPMs

Since the sum of the quantities in 1) and 2) tends to 0 drastically in the *asymptotic case* ($m \to \infty$, n fixed) for $q \in (0, \tilde{q}(\frac{2n}{m}))$, the probability of satisfying our conditions tends to 1. That means that the function F_U asymptotically converges to the total distribution function and we can use Theorem 4 in a calculation of $E[K]$. Insertion of Theorem 4 in formula (2) for $E_{\widetilde{F}_U}[K]$ delivers one of the main results:

Theorem 6: *Let $a_1, \ldots, a_m, v$ be distributed according to uni-RSM. Let $n \geq 3$, $m \geq 5n^{\frac{3}{2}}$, $q \in (0, \tilde{q}(\frac{2n}{m}))$. Let a suitable starting point x_0 be given and $\|x_{opt}\| \leq \frac{1}{q}$. Then an IPM with a reduction guarantee as in (1) (i. e. Algorithm II) does – on the average – reach the safety region in less than $R(m) \cdot O(\ln m + \ln n + n|\ln q|)$ iterations.*

But still the unfavourable cases may not be solved by Algorithm II.

When we employ the Simplex-Method to solve all these cases, then the average effort for the Simplex-Method can be bounded from above by

$$\binom{m}{n}\left(\frac{C(n)}{(m-1)2^{m-1}} + \binom{m}{n}0.8G(q)^{m-n}\right) \to 0 \ \text{ for } m \to \infty, n \text{ fixed}, q \in (0, \tilde{q}(\tfrac{(2n+1)\ln m}{m}).$$

But – when we want to solve all problems of type (P) – we have to manage the problem that we do not recognize the critical problems a priori. To get rid of that obstacle, we suggest to use the following hybrid algorithm.

Algorithm III

Phase 1: Let the work of Phase I be done for the IPM and the Simplex-Method in parallel. Then the Simplex-Method provides a vertex of X (X is pointed) and the Phase I for the IPM delivers (if successful) a suitable starting point x_0.

Phase 2: In case that the step 1) has been successful, start a parallel iteration-procedure consisting on

– a Simplex-Iteration (pivot-step)

– an iteration of the Iterate-and-Check-Algorithm (Algorithm II)

STOP as soon as Algorithm II reaches the optimal vertex or unboundedness becomes obvious.

Note, that the effort of an iteration of Algorithm II (resp. an IPM) is again greater than the effort of an pivot-step.

For this algorithm we can show that for $m \to \infty$, n fixed, the following holds:

Theorem 7: *Let $a_1, \ldots, a_m, v$ be distributed according to uni-RSM and $q \in ((mn)^{-\frac{1}{n}}, \tilde{q}(\frac{(2n+1)\ln m}{m}))$. Let the IPM, that is involved in Algorithm III, guarantee that the duality gap is reduced by a factor $\left(1 - \frac{1}{R(m)}\right)$ in each step, and let x_0 be a suitable starting point.*
Then for $m \to \infty$, n fixed, Algorithm III does not need more than $R(m)O(\ln m)$ iterations on the average to solve a linear programming problem of type (P).

References

[1] Anstreicher, K.M., Ji, J., Potra, F.A., Ye, Y. (1993): *Average performance of a self-dual interior-point algorithm for linear programming*, Complexity in Numerical Optimization (ed. P.Pardalos) p. 1-15, World Scientific, New Jersey, 1993.

[2] Borgwardt, K.H. (1987): *The Simplex Method – A Probabilistic Analysis*, Springer-Verlag, Heidelberg, 1987.

[3] den Hertog, D. (1994): *Interior Point Approach to Linear, Quadratic and Convex Programming, (Mathematics and Its Applications*, Kluwer Academic Publishers, Dordrecht, 1994.

[4] Huhn, P. (1997): *Schranken für die durchschnittliche Laufzeit des Simplexverfahrens und von Innere-Punkte-Verfahren*, (Dissertation, Universität Augsburg) Augsburger Mathematisch-Naturwissenschaftliche Schriften, Band 16, Wissner Verlag, Augsburg, 1997.

[5] Mizuno, S., Todd, M.J., Ye, Y. (1993): *On adaptive-step primal-dual interior-point algorithms for linear programming*, Mathematics of Operations Research 18, p. 964-981, 1993.

[6] Papadimitriou, C.H., Steiglitz, K. (1982): *Combinatorial Optimization, Algorithms and Complexity*, Prentice-Hall, New Jersey, 1982.

[7] Ye, Y. (1994): *Toward probabilistic analysis of interior-point algorithms for linear programming*, Mathematics of Operations Research 19, p. 38-52, 1994.

[8] Ye, Y. (1995): *Interior-Point Algorithms: Theory and Practice*, Script, 1995.