A VIEW OF INTERIOR POINT METHODS
FOR LINEAR PROGRAMMING

Vera V. KOVAČEVIĆ-VUJČIĆ *)
Laboratory for Operations Research,
Faculty of Organizational Sciences, University of Belgrade,
Jove Ilića 154, 11000 Belgrade, Yugoslavia

Abstract: The paper gives a view of interior point methods for linear programming inspired by the work of Karmarkar. Two possible classifications of the existing methods are presented through the detailed discussion of typical representatives. The paper also discusses some implementation details and provides a selected bibliography.

Keywords: Linear programming, Interior methods, Computational complexity

1. INTRODUCTION

The introduction of polynomial-time interior methods is one of the most important developments in mathematical programming in the last decade. The first step in that direction was made in 1984 when Karmarkar in his seminal paper [35] proposed a new polynomial algorithm for linear programming. The method had not only a better complexity bound than the earlier method of Khachian [37] but it also enjoyed computational efficiency on practical problems. Karmarkar's paper stimulated an enormous research interest which has resulted in the development of numeros interior point methods both for linear and nonlinear programming problems and for certain discrete optimization problems. The ten years of research resulted in about 2000 papers dealing with interior point methods (see Kranich's bibliography [41]). Recently, Nesterov and Nemirovskii [55] have developed a general framework for polynomial-time interior point methods, based on self-concordant functions. Interior point methodology has also been used to develop a new approach to the theory of linear programming, including duality and sensitivity analysis (see [34]).

This paper reviews the most important aspects of interior point methods for linear programming inspired by work of Karmarkar. Section 2 describes the main idea

*) This research was supported by Science Fund of Serbia, grant number 0401F, through Mathematical Institute
of interior point methods and gives two possible classifications. Section 3 is devoted to projective methods with the emphasis on the original method of Karmarkar. Affine methods, including the original and modified version of Dikin’s method [18, 19] and Gonzaga’s polynomial modification [27] are analyzed in Section 4. Section 5 describes the main idea of path-following methods on the example of Gonzaga’s method [25]. Section 6 gives a survey of search directions in interior point methods with the emphasis on numerical stability problems. Finally, Section 7 discusses some details related to practical implementation of interior point methods.

2. THE BASIC PRINCIPLES OF INTERIOR POINT METHODS

Consider the linear programming problem in standard form

$$\begin{align*}
\min & \ c^T x \\
Ax &= b \\
x &\geq 0
\end{align*}$$

where $A=[a_{ij}]_{m \times n}$, $c \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $m < n$, $\text{rank} A = m$ and assume that $A$, $b$ and $c$ have integer entries (which is equivalent to the assumption that the entries are rational). Let $X$ be the feasible set of (1) and, for the sake of simplicity, assume that $X$ is bounded and $\text{ri} X = \{x \in \mathbb{R}^n \mid Ax = b, x > 0\} \neq \emptyset$. It is well known that the problem dual to (1) can be written in either of the two forms:

$$\begin{align*}
\max & \ b^T y \\
A^T y &\leq c \\
A^T y + z &\geq c, \quad z \geq 0
\end{align*}$$

The idea of interior point methods for solving (1) is quite opposite to that of the simplex method: instead of visiting different extreme points of $X$, a typical interior point method generates a sequence $(x^k) \subset ri \ X$, with the property that for $k$ large enough the approximate solution $x^k$ can be transformed into an optimal solution $x_{opt}$.

The most important issue in this approach is to estimate polynomially the number of elementary operations (additions, multiplications, comparisons, etc.) needed to get an optimal solution. In addition, it has to be shown that generated points have polynomial descriptions. This is usually done by showing that it is sufficient to do all the calculations in certain fixed precision which depends on the data of problem (1) (see e.g. [54]). We shall here shortly discuss a theoretical background for the transformation of $x^k$ into $x_{opt}$.

We shall denote by $\langle h \rangle$ the number of bits needed for binary representation of an integer $h$. Then $\langle h \rangle = 1 + \left\lceil \log_2 (|h| + 1) \right\rceil$. Let $r$ be the bit length of the input data in problem (1), i.e.

$$r = \langle a_{11} \rangle + \ldots + \langle a_{mn} \rangle + \langle c_1 \rangle + \ldots + \langle c_n \rangle + \langle b_1 \rangle + \ldots + \langle b_m \rangle$$

and let $L = r + \langle n \rangle$. The next Lemma, basically due to Khachian [37], shows that the quantity $2^{\langle L \rangle}$ can be used for the identification of the optimal solution.
Lemma 1. Suppose that the feasible set $X$ of problem (1) is nonempty and bounded and let $f_{\text{opt}} = c^T x_{\text{opt}}$ be the optimal objective function value. If $\bar{x}$ is an arbitrary extreme point of $X$ then either $c^T \bar{x} = f_{\text{opt}}$ or $c^T \bar{x} > f_{\text{opt}} + 2^{-L}$.

Interior point methods for linear programming use the results of Lemma 1 in the following way: Suppose that the optimal value $f_{\text{opt}}$ is known and that we seek a point where it is achieved. Let $(x^k)$ be a sequence generated by an interior method and suppose that for some $k$ we have

$$c^T x^k - f_{\text{opt}} \leq 2^{-L}$$

Furthermore, suppose that the point $x^k$ is transformed into an extreme point $\bar{x}$ of $X$ with the property $c^T \bar{x} \leq c^T x^k$. Then according to Lemma 1 $c^T \bar{x} = f_{\text{opt}}$, i.e. $\bar{x}$ is an optimal solution (equal to $x_{\text{opt}}$ in the case of the unique solution). An obvious difficulty with this approach lies in the fact that in general case $f_{\text{opt}}$ is not known and there are several ways to get around it. Karmakar in [35] proposes replacing problem (1) with a new problem in which the duality gap, known to have value 0 as the optimal value, is minimized. Several authors propose generating a sequence $(y^k, z^k)$ of dual feasible solutions. Since $c^T x_{\text{opt}} - b^T y_{\text{opt}} = 0$ and $b^T y^k \leq b^T y_{\text{opt}}$ we have $(z^k)^T x^k = c^T x^k - b^T y^k = c^T x^k - c^T x_{\text{opt}} + b^T y_{\text{opt}} - b^T y^k \geq c^T x^k - c^T x_{\text{opt}}$ and the test (3) can be replaced by $c^T x^k - b^T y^k \leq 2^{-L}$ or $(z^k)^T x^k \leq 2^{-L}$. There are also more efficient approaches which avoid explicit evaluation of dual variables (see e.g. [63]).

It remains to be seen how a point $x^k$ satisfying (3) can be transformed into an extreme point $\bar{x}$ with the property $c^T \bar{x} \leq c^T x^k$. Such transformations are often called rounding procedures. The idea of the rounding procedure below is to start from $x^k$ and at each step generate a feasible point which has at least one more zero coordinate than the previous point and at the same time does not increase the objective function value. The Algorithm is taken from [20], see also [40].

Let $x^k > 0$ be a feasible point and let the vectors $z^1, \ldots, z^{n-m}$ constitute a basis of the vector space $N(A) = \{ x \in \mathbb{R}^n | Ax = 0 \}$. Let us note that $z^1, \ldots, z^{n-m}$ can be obtained with $0(m^3)$ operations by pivoting on the matrix $A$. Namely, if matrix $A$ is transformed into the matrix $[I_m, M]$ then the rows of the matrix $[M^T - I_{n-m}]$ are a basis of $N(A)$, where $I_m$ and $I_{n-m}$ are the unit matrices of order $m$ and $n-m$, respectively.

**Algorithm 1.**

Step 0: Set $s = n-m$, $y = x^k$

Step 1: Set $z = z^s$. If $c^T z^s < 0$ set $z = -z^s$, Compute

$$\lambda = \min \{ y_j/|z_j| | z_j > 0 \} = y_r/|z_r|.$$

Step 2: Compute $\bar{y} = y - \lambda z$. If $s = 1$ set $\bar{x} = \bar{y}$ STOP. Otherwise compute
\[
\bar{z}_j = z_j^j - (z_j^j / z_r)z_r, \quad j = 1, \ldots, s - 1.
\]

Step 3: Set \( y = \bar{y} \), \( z_j^j = \bar{z}_j \), \( j = 1, \ldots, s - 1 \), replace \( s \) by \( s - 1 \) and go to Step 1. \( \blacksquare \)

It can be shown that the point \( \bar{x} \) generated by Algorithm 1 is an extreme point of \( X \) with the property \( c^T \bar{x} \leq c^T x^k \). Since each cycle of the algorithm uses \( O(n^2) \) elementary operations and the number of cycles is \( O(n) \), the total complexity is \( O(n^3) \).

All known interior point methods use the outlined ideas for the transformation into an exact solution. The difference occurs in the rule for generating the sequence \( (x^k) \) and the number of steps needed to get a solution satisfying (3) or a related stopping criterion. A possible classification of interior methods, based on the geometric interpretation, is the following:

1. **Projective methods** (Karmarkar [35], Anstreicher [3], De Ghellinck and Vial [14], Todd and Ye [64], Ye [73], Yamashita [72], Gonzaqa [26], etc.).

2. **Affine methods** (Dikin [18, 19], Barnes [9], Gonzaqa [27], Vanderbei et al. [68], Monma and Morton [51], Monteiro et al. [53], Ye [74], etc.).

3. **Path-following methods** (Renegar [57], Gonzaqa [25, 30, 31], Roos and Vial [59, 60], Den Hertog et al. [16], Monteiro and Adler [52], Kojima et al. [38], etc.).

In Sections 2, 3, and 4 we will shortly outline typical representatives of the three classes. Interior methods can also be classified according to the type of problem being solved as:

1. **Primal methods**, working on (1), (Karmarkar [35], Anstreicher [3], De Ghellinck and Vial [14], Ye [73, 74], Gonzaqa [30, 31], Roos and Vial [59, 60], Den Hertog et al. [16], Dikin [18, 19], Barnes [9], Vanderbei et al. [68], etc.).

2. **Dual methods**, working on (2), (Yamashita [72], Gonzaqa [25], Monma and Morton [51], Renegar [57], etc.).

3. **Primal-dual methods** (Todd and Ye [64], Monteiro et al. [53], Monteiro and Adler [52], Kojima et al. [38], etc.).

3. **PROJECTIVE METHODS**

The original version of Karmarkar's method [35] is the most important representative of the class of projective methods. Karmarkar's method is formulated for the problems in the following special form:

\[
\min c^T x
\]

\[
Ax = 0, \quad e^T x = 1, \quad x \geq 0
\]
where \( A = [a_{ij}]_{\text{max}}, \ rank A = m, \ x, c \in \mathbb{R}^n, \ e = (1, \ldots, 1) \in \mathbb{R}^m \) and all the coefficients in \( A \) and \( c \) are integer. Let \( X \) be the feasible set of problem (4), i.e. \( X = N(A) \cap S \), where \( S \) is the \((n-1)\) - dimensional simplex, \( S = \{ x \in \mathbb{R}^n \ | \ e^T x = 1, \ x \geq 0 \} \). It is assumed that \( A e = 0 \), i.e. the center of \( S \) defined by \( a = (1/n)e \) is feasible, and that the optimal objective function value \( f_{\text{opt}} = 0 \).

Karmarkar’s algorithm can be described as the steepest descent method with the projective transformation at each iteration. At the iteration \( k \) the transformation is defined by

\[
y = \frac{D^{-1}x}{e^T D^{-1}x}
\]

where \( D = \text{diag}(x_1^k, \ldots, x_n^k) \) and \( x^k > 0 \) is a point generated at the iteration \( k-1 \). It is easy to see that the transformation maps \( S \) into itself, that the extreme points remain fixed and that \( x^k \) is mapped into \( a \). Let \( \bar{X} \) be the image of \( X \). Then \( \bar{X} = N(A) \cap S, \ A = AD \). The inverse transformation is given by

\[
x = \frac{Dy}{e^T Dy}
\]

The next so-called potential function plays an important role in the analysis of Karmarkar’s algorithm:

\[
f: \ R^n \to R, \ f(x) = n \ln(c^T x) - \sum_{i=1}^n \ln x_i
\]

Let \( \bar{f} \) be the transformed potential function defined by \( \bar{f}(y) = f(x) \). Then

\[
\bar{f}(y) = n \ln(c^T D y) - n \sum_{i=1}^n \ln x_i^k \bar{y}_i = n \ln c^T y - \sum_{i=1}^n \ln y_i - \sum_{i=1}^n \ln x_i^k
\]

where \( \bar{c} = Dc \), i.e. functions \( \bar{f} \) and \( f \) are of the same form.

The \( k \)-th iteration of Karmarkar’s algorithm can now be shortly outlined as follows:

- With the projective transformation \( y = D^{-1}x / e^T D^{-1}x \) map the feasible set \( X \) onto \( \bar{X} \); the image of \( x^k \) is \( a = (1/n)e \).
- Take one step of the projected gradient method for decreasing \( \bar{f} \) on \( \bar{X} \), i.e. compute \( \bar{y} = a - \lambda P \nabla \bar{f}(a) / \| P \nabla \bar{f}(a) \| \).
- Get \( x^{k+1} \) as the inverse image of \( \bar{y} \).

The precise formulation of the method is:
Algorithm 2 (Karmarkar):

Step 0: Let $L>0$ and $0<\alpha<1$ be given. Set $\lambda = \alpha / \sqrt{n(n-1)}$, $x^0 = a$, $k = 0$.

Step 1: Set $D = \text{diag}(x_1^k, \ldots, x_n^k)$. Compute $\bar{A} = AD$, $\bar{c} = Dc$. Let

$$B = \begin{bmatrix} \bar{A} \\ \bar{c}^T \end{bmatrix}_{(m+1) \times n}$$

be the matrix defining $\bar{X}$.

Step 2: Let $P = I - B^T (BB^T)^{-1} B$ be the projection matrix onto $N(B)$. Compute

$$h = PVf(a)/\|PVf(a)\| = P\bar{c}/\|P\bar{c}\|$$

Step 3: Set $\bar{y} = a - \lambda h$.

Step 4: Compute $x^{k+1} = D\bar{y} / e^T D\bar{y}$.

Step 5: Replace $k$ by $k+1$. If $c^T x^k < 2^{-L}$ STOP, else go to Step 1. $\blacksquare$

Let us note that

$$BB^T = \begin{bmatrix} \bar{A} \\ \bar{c}^T \end{bmatrix} \begin{bmatrix} \bar{A} \bar{A}^T & 0 \\ 0 & n \end{bmatrix} = (BB^T)^{-1} = \begin{bmatrix} (\bar{A} \bar{A}^T)^{-1} & 0 \\ 0 & 1/n \end{bmatrix}$$

so that the main computational effort in one iteration of Algorithm 2 is the inversion of $\bar{A} \bar{A}^T = AD^2 A^T$. We shall see that this is the case with all interior point methods.

The next theorem shows that Algorithm 2 is polynomial:

**Theorem 1** ([35]). Let $r$ be the length of the input data in (4), let $L = r + <n>$ and $\alpha \leq 1/3$. Then Algorithm 2 in $O(nL)$ iterations finds a point $x^k$ satisfying $c^T x^k < 2^{-L}$. $\blacksquare$

Since the number of elementary arithmetic operations per iteration of Algorithm 2 is $O(n^3)$, it follows that the overall complexity of Algorithm 2 is $O(n^4L)$. Karmarkar also shows that Algorithm 2 can be modified using partial updating so that the complexity bound is reduced to $O(n^3.5L)$. The rate of convergence of Karmarkar’s algorithm is analysed by Asić, Kovačević-Vujčić and Radosavljević-Nikolić in [7]. Under the assumption that $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_m, 0, \ldots, 0)$, $\bar{x}_1 > 0, \ldots, \bar{x}_m > 0$ is the unique solution of (4) it is shown that

$$\lim_{k \to \infty} \frac{c^T x^{k+1}}{c^T x^k} = \frac{\sqrt{(n-1)(n-m-1)(m+1) - \alpha(m+1)}}{\sqrt{(n-1)(n-m-1)(m+1) + \alpha(n-m-1)}}$$
Furthermore, Ašić et al. show in [8] that the normalized sequence of search directions in Karmarkar’s method converges, i.e.

$$\lim_{k \to \infty} \frac{x^{k+1} - x^k}{\|x^{k+1} - x^k\|} = s,$$

where the components of $s$ are uniquely determined by the reduced cost vector at the optimal vertex. Intuitively speaking, this means that the successive iterates tend to line up along a direction which points towards the solution. In [42] Kovačević-Vujčić uses this property in order to improve the rate of convergence of Karmarkar’s method by taking long steps in promising directions. Namely, given a sequence $(x^k)$, $x^k \to \bar{x}$, generated by Karmarkar’s method, let $\bar{x}^k$, $k \in N$, be defined as a point at which the ray $x^{k-1} + t(x^k - x^{k-1})$ intersects the boundary of $X$. Then

$$\lim_{k \to \infty} \frac{\|\bar{x}^k - \bar{x}\|}{\|x^k - \bar{x}\|} = 0,$$

i.e. the auxiliary sequence $(\bar{x}^k)$, generated at a low cost ($O(n)$ per step), converges faster than $(x^k)$.

In [24] Gill, Murray, Saunders, Tomlin and Wright show that Karmarkar’s algorithm can be viewed as the Newton’s method applied to the logarithmic barrier function. Todd and Burrell [63] were the first to show that the assumption of known optimal value in Karmarkar’s method can be removed using a lower bound technique. A variant of Karmarkar’s method which works with linear programming problems in standard form was developed by Gonzaga [26] and several other authors. Further analysis of Karmarkar’s method and descriptions of some other related methods belonging to the same class can be found in [3, 12, 14, 20, 23, 28, 64, 72, 73, 76], etc.

4. AFFINE METHODS

Soon after Karmarkar’s method was published several authors have come up with the idea to simplify it using affine rather than projective transformations. (Barnes [9], Vanderbei et al. [68], etc.). It has come out later that the idea with the affine transformation was used as early as 1967 in a paper by Dikin [18]. We shall outline here Dikin’s method and its polynomial modification.

Consider the following linear programming problem:

$$\min c^T y$$

$$Ax = b, \quad x \geq 0$$

where $A = [a_{ij}]_{m \times n}$, $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $\text{rank}A = m$. We shall assume that an $x^0$ such that $Ax^0 = b$, $x^0 > 0$ is known. At the $k$-th iteration we have a feasible point $x^k$, $x^k > 0$.
which is used to define a linear transformation $y = D^{-1}x$, where, as in Karmarkar's method, $D = \text{diag}(x_1^k, \ldots, x_n^k)$. The image of $x^k$ is $e=(1,\ldots,1)$, and the inverse transformation is $x= Dy$. In the transformed space the problem is the following:

$$\min \tilde{c}^T y$$

$$\tilde{A} y = b \quad y \geq 0$$

where $\tilde{c} = Dc$, $\tilde{A} = AD$. The algorithm takes one step of the projected gradient method for decreasing the objective function of problem (6) starting from the point $e$, i.e. computes $y = e - \lambda P \tilde{c} / \|P \tilde{c}\|$, and then maps it by the inverse transformation to get $x^{k+1}$. The details are given in:

**Algorithm 3** (Dikin)

Step 0: Let $\lambda$ and $x^0$ satisfying $0<\lambda<1$ and $x^0>0$, $Ax^0=b$ be input parameters. Set $k=0$.

Step 1: Set $D = \text{diag}(x_1^k, \ldots, x_n^k)$. Compute $\tilde{A} = AD$, $\tilde{c} = Dc$.

Step 2: Let $P = (I - \tilde{A}^T \tilde{A}^{-1})^{-1}$ be the projection matrix onto $N(A)$. Compute $h = P \tilde{c} / \|P \tilde{c}\|$.

Step 3: Compute $\tilde{y} = e - \lambda h$.

Step 4: Set $x^{k+1} = Dy$

Step 5: Replace $k$ by $k+1$ and go to Step 1. $\blacksquare$

It can be shown that for a suitable value of $\lambda$ (e.g. $\lambda=1/8$) the sequence generated by Algorithm 3 converges to an optimal solution $\bar{x}$ (see e.g. [9, 18, 65]). As in the case of Karmarkar’s algorithm it can be shown that the rate of convergence is linear and that the sequence of normalized search directions converges [8, 56]. The main result in [56] is that under the appropriate nondegeneracy assumptions we have

$$\lim_{k \to \infty} \frac{c^T x^{k+1} - c^T x}{c^T x^k - c^T x} = 1 - \frac{\lambda}{\sqrt{n-m}}$$

The procedure proposed in [42] can again be used to speed up the convergence.

Recently, Dikin [19] has proposed an interesting modification of the affine method. It is easy to see that Algorithm 3 can be reformulated as follows:

$$x^0 = e, \quad x^{k+1} = x^k - \lambda_k D_k^2 z^k, \quad \lambda_k = \lambda / ((x_1^k z_1^k)^2 + \ldots + (x_n^k z_n^k)^2)^{1/2}$$
where now \( D_k = \text{diag}(x_1^k, \ldots, x_n^k) \), \( z^k = c - A^T y^k \), \( y^k = (AD_k^2 A^T)^{-1} AD_k^2 c \). The modified method differs from the original only in the choice of the step-size:

\[
x^0 = e, \quad x^{k+1} = x^k - \lambda_k D_k^2 z^k, \quad \lambda_k = \lambda / \min(x_j^k, s_j^k)
\]

and it has a nice property that without any nondegeneracy assumptions it can be shown that both sequences \((x^k)\) and \((y^k)\) converge to optimal solutions in the interior of the optimal faces of primal and dual problem, respectively.

Up to now it has remained unknown whether Algorithm 3 is polynomial or not, although the results obtained by Megiddo and Shub [48] indicate that a negative answer is to be expected. However, it is not hard to obtain different modifications that are polynomial. One of such modifications is proposed by Gonzaga [27]. Gonzaga assumes that the optimal objective function value is 0 and uses the following potential function:

\[
f: \quad R^n \to R, \quad f(x) = q \ln(c^T x) - \sum_{i=1}^n \ln x_i,
\]

where \( q > 0 \). The linear transformation \( y = D^{-1} x \), \( D = \text{diag}(x_1^k, \ldots, x_n^k) \), transforms the potential function into

\[
\tilde{f}(y) = q \ln(c^T D y) - \sum_{i=1}^n \ln x_i^k y_i = q \ln \tilde{c}^T y - \sum_{i=1}^n \ln y_i - \sum_{i=1}^n \ln x_i^k
\]

and it is easy to see that

\[
\nabla \tilde{f}(e) = \frac{q}{c^T e} \tilde{c} - e.
\]

**Algorithm 4** (Gonzaga)

Step 0: Let \( L \) and \( x^0 \) satisfying \( L > 0, x^0 > 0 \), \( Ax^0 = b \) be input parameters. Set \( k = 0 \).

Step 1: Set \( D = \text{diag}(x_1^k, \ldots, x_n^k) \). Compute \( \tilde{A} = AD, \quad \tilde{c} = Dc \).

Step 2: Let \( P = (I - A^T (\tilde{A} A^T)^{-1} \tilde{A}) \) be the projection matrix onto \( \mathcal{N}(\tilde{A}) \).

Compute \( h = P \nabla \tilde{f}(\tilde{c}) / \|P \nabla \tilde{f}(\tilde{c})\| \).

Step 3: Compute \( \tilde{y} = e - 0.3h \).

Step 4: Set \( x^{k+1} = D \tilde{y} \).

Step 5: Replace \( k \) by \( k + 1 \). If \( c^T x^k < 2^{-L} \) STOP. Otherwise go to Step 1.
Theorem 2 ([27]). Suppose that problem (5) has integer coefficients and let \( r \) be the length of the input data and \( L=r+n \).

(i) If \( q=n \) then Algorithm 4 stops after \( O(n^2L) \) iterations.

(ii) If \( q \geq n + \sqrt{n} \) then Algorithm 4 stops after \( O(nL) \) iterations. \( \blacksquare \)

The number of elementary arithmetic operations per step of Algorithm 4 is \( O(n^3) \), so that the overall complexity of finding \( \bar{x} \) by Algorithm 4 is \( O(n^5L) \) for \( q=n \), i.e. \( O(n^4L) \) for \( q \geq n + \sqrt{n} \). As in the case Karmarkar's method, using partial updating, the later bound can be reduced to \( O(n^{3.5}L) \). It should be noted that in Karmarkar's method we have \( \bar{P}\bar{c} / \| \bar{P}\bar{c} \| = P\nabla f(a) / \| P\nabla f(a) \| \), which does not hold in the case of affine methods. Namely, Dikin's method uses the direction \( \bar{P}\bar{c} \), while Gonzaga's polynomial modification uses \( P\nabla f(e) \). Using a different potential function Ye [74] proposes an affine algorithm which finds a satisfactory approximate solution in \( O(n^{0.5}L) \) iterations. The overall complexity of Ye's algorithm is \( O(n^{3.5}L) \), i.e. \( O(n^3L) \) if the partial updating is included. The bound \( O(n^{3}L) \) is the best complexity bound obtained by now and it also holds for most of the path-following methods, which are the subject of the next section.

Affine methods have been studied by many other authors, e.g. Anstreicher [4], Freund [21], Anstreicher and Bosch [5], Monteiro et al. [53], Kojima et al. [39], etc.

5. PATH-FOLLOWING METHODS

The idea of the path following methods will be illustrated on the example of a simple dual method proposed by Gonzaga [25]. Let us consider the dual problem:

\[
\begin{align*}
\max & \quad b^T y \\
& \quad A^T y \leq c
\end{align*}
\]

and denote by \( Y \) its feasible set. We shall assume that

\[ \text{int } Y = \{ y \in \mathbb{R}^n \mid A^T y < c \} \neq \emptyset \]

and for the sake of simplicity that \( Y \) (instead of \( X \)) is compact. Let

\[ f(y, t) = b^T y + t \sum_{i=1}^n \ln(c_i - A_i y) \]

be a barrier function associated with (7), where \( t > 0 \) is the penalty parameter and \( A_i, i=1, \ldots, n \) are the rows of \( A^T \). For a fixed \( t \) the function \( f \) is strictly concave and the problem

\[
\begin{align*}
\max & \quad f(y, t) \\
& \quad A^T y < c
\end{align*}
\]
has the unique solution \( y(t) \). The trajectory \( y(t), t > 0 \) is called the central path associated with (7). The trajectory has the important property that it converges to an optimal solution of (7) when \( t \to 0 \).

The classical barrier method generates a monotone decreasing sequence \((t_k)\), \( t_k \to 0 \) and the corresponding sequence \((y(t_k))\). In that way the constrained optimization problem is replaced by an infinite sequence of uncostrained problems. Although the computation of \( y(t_k) \) is usually replaced by the computation of a suitable approximation \( y^k \), since the number of steps needed to compute \( y^k \) is not known in advance we can roughly say that classical methods generate infinitely many infinite sequences. Path-following methods for linear programming are characterized not only by the fact that for each \( k \) a fixed (independent of \( k \)) number of steps is taken in order to obtain \( y^k \), but it is also possible to estimate polynomially the overall complexity needed to produce a solution of (7). The simplest algorithm of this type is Gonzaga’s method [25], where \( p = 1 \), i.e. the penalty parameter is changed at each iteration.

**Algorithm 5** (Gonzaga)

Step 0: Let \( y^0 \in Y \) and \( t_0 > 0 \) be input parameters. Set \( k = 0 \).

Step 2: Compute \( t_{k+1} \):

\[
t_{k+1} = (1 - \frac{1}{41\sqrt{n}})t_k
\]

Step 3: Compute \( y^{k+1} \):

\[
y^{k+1} = y^k + \nabla^2_{yy} f(y^k, t_{k+1})^{-1} \nabla_y f(y^k, t_{k+1})
\]

Step 3: Replace \( k \) by \( k + 1 \) and go to Step 1.

Let us note that Gonzaga’s method uses Newton steps for approximate maximization, which is the case with most of the path-following methods. It is easy to see that

\[
\nabla f_y(y, t) = b - tAD^{-1}c \quad \nabla^2_{yy} f(y, t) = -tAD^{-2}A^T
\]

where \( D = \text{diag}(c_1 - A_1y, \ldots, c_n - A_ny) \), so that the main computational effort per iteration of Algorithm 5 is the inversion of the positively definite matrix \( AD^{-2}A^T \). The proof of polynomiality of Algorithm 5 can be obtained directly [25] or as a consequence of Kantorovich’s theorem [58]. Namely, if \( f_{opt} \) is the optimal value of the objective function in (7) the following result holds:

**Theorem 3.** Let all the data in (7) be integer with the bit length \( r \) and let \( L = r + \langle n \rangle \). If \( y^0 \) is such that

\[
\| DA^T(y^0 - y(t_0)) \| \leq 1/20
\]
where $D = \text{diag}(c_1 - A_1 y^0, \ldots, c_n - A_n y^0)$, then Algorithm 5 in $0(n^{0.5}L)$ iterations generates a point $y^k$ such that

$$ f_{\text{opt}} - b^T y^k \leq 2^{-L}. \quad (10) $$

It can be shown that a starting point $y^0$ satisfying (9) can be obtained either directly after a suitable reformulation of problem (7) (e.g. [47, 25]) or using so called centralization procedures. It also should be noted that $y^k$ satisfying (10) can be transformed into an exact solution $\tilde{y}$ with $O(n^3)$ operations using algorithms related to Algorithm 1 (see e.g. [57]). If Algorithm 5 is modified to include a stopping rule and if $y^0$ satisfies the conditions of Theorem 3, it is easy to see that the overall complexity is $O(n^{3.5}L)$. The algorithm can be modified using the idea of partial updating so that the complexity bound is reduced to $O(n^2L)$ (see [25, 58]).

The central path can also be defined for the primal problem:

$$ \min c^T x \quad (11) $$

$$ Ax = b, \quad x \geq 0 $$

Consider the logarithmic barrier function associated with (11):

$$ g(x, t) = c^T x - t \sum_{i=1}^{n} \ln x_i, \quad x \in \mathbb{R}^n $$

The function $g(x, t)$ is strictly convex and it is easy to see that

$$ \nabla_x g(x, t) = c - tD_x^{-1}c \quad \nabla^2_{xx} g(x, t) = tD_x^{-2} $$

where $D_x = \text{diag}(x_1, \ldots, x_n)$. Under the assumption $\text{int} \ Y \neq \emptyset$ it follows that the problem

$$ \min g(x, t) \quad (12) $$

$$ Ax = b, \quad x > 0 $$

has the unique solution $x(t)$. The central path associated with (11) is defined as the trajectory $x(t)$, $t > 0$. There is an interesting relationship between central paths associated with primal and dual problems which can be obtained from the optimality conditions for (12). The necessary and sufficient optimality conditions have the form

$$ c - tD_x^{-1}c = A^T y \quad Ax = b, \quad x > 0 $$

If $z = c - A^T y$ is the slack variable and $D_z = \text{diag}(z_1, \ldots, z_n)$ it follows that $D_z e = z$ and (13) reduces to

$$ D_x D_z e = te \quad (14) $$

$$ Ax = b, \quad x > 0 $$

$$ A^T y + z = c, \quad z > 0 $$
A point $x(t)$ is an optimal solution to (12) if and only if there exist $y(t)$ and $z(t)$ such that $x(t)$, $y(t)$, $z(t)$ is a solution to (14). It is easy to show that $y(t)$ is precisely the optimal solution to (8) (see [47]). Hence, we can say that the system (14) characterizes central paths of both primal and dual problems and for that reason the central path is often defined as the triple $x(t), y(t), z(t)$ satisfying (14) (see e.g. [71]). Many path-following methods operate directly on the system (14) or on some of its equivalent formulations. We shall illustrate it on the example of a primal-dual method which uses Newton's method for finding approximate solutions of (14).

Let $(x^k, y^k, z^k)$ be an approximate solution of the system (14) for $t=t_k$ and let $(x^{k+1}, y^{k+1}, z^{k+1})$ be an approximate solution of the system:

$$\Phi(x, y, z) = \begin{bmatrix} D_x D_z e - l_{k+1} e \\ A x - b \\ A^T y + z - c \end{bmatrix} = 0$$

obtained from $(x^k, y^k, z^k)$ using one step of Newton's method. Let $p = (p_x, p_y, p_z)$, $p_x = x^{k+1} - x^k$, $p_y = y^{k+1} - y^k$, $p_z = z^{k+1} - z^k$ be the Newton's direction. As it is well known, $p$ is the solution of the system

$$J(x^k, y^k, z^k) p = -\Phi(x^k, y^k, z^k)$$

where $J$ is the Jacobian of $\Phi$, i.e. $p_x, p_y$ and $p_z$ satisfy the system

$$J(x^k, y^k, z^k) = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} A \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} l_{k+1} e - D_x D_z e \\ b - A x \\ c - A^T y - z \end{bmatrix}$$

where $D_x = \text{diag}(x^1, \ldots, x^n)$, $D_z = \text{diag}(z^1, \ldots, z^n)$.

There are obviously three sets of equations induced by the blocks of $J$. From the third set we obtain

$$p_z = -A^T p_y + c - A^T y^k - z^k$$

and the first two sets of equations can be equivalently transformed into

$$\begin{bmatrix} D_x \quad A^T \\ A \quad 0 \end{bmatrix} \begin{bmatrix} p_x \\ -p_y \end{bmatrix} = \begin{bmatrix} l_{k+1} D_x e - c + A^T y^k \\ b - A x^k \end{bmatrix}$$

From the first subsystem we have

$$p_y = D_x^{-1} D_z^{-1} A^T p_y + D_x^{-1} D_z^{-1} (l_{k+1} D_x^{-1} e - c + A^T y^k).$$
and $p_z$ can be eliminated from the second subsystem, yielding
\[ AD_x^{-1} D_z k A^T p_y = AD_x^{-1} D_z k (c - t_{k+1} D_y^{-1} c - A^T y^k) + b - Ax^k. \]

Computational effort for computing $p_y$ is $O(n^3)$, while $p_x$ and $p_z$ can be obtained by a direct substitution.

Path-following methods are subject of investigation of many authors, e.g. Renegar [57], Kojima et al. [38], Monteiro and Adler [52], Roos and Vial [59, 60], Den Hertog et al. [16]. A unified approach to path-following methods is given e.g. in [58, 71], while the recent book by Den Hertog [17] gives a systematic analysis of short-, medium- and long-step path-following methods for linear, quadratic and convex programming. Analytical properties of the central path as well as the definition and properties of trajectories associated with projective and affine methods were investigated by Bayer and Lagarias [10,11], Witzgall et al. [70], Megiddo and Shub [48] and others.

6. SEARCH DIRECTIONS IN INTERIOR METHODS

Consider the primal-dual pair of problems (1) and (2) and suppose that $rx\in\mathbb{R}$, $inty\in\mathbb{R}$. As we have indicated, interior methods can be classified into primal, dual and primal-dual methods. Typically, a primal method generates $(x^k)\subset rx$, with the property $x^k \to x_{opt}$, a dual method generates $(y^k)\subset inty$ with the property $y^k \to y_{opt}$ while primal-dual methods generate simultaneously $(x^k)$, $(y^k)$ and $(z^k)=(c-A^T y^k)$. All these methods can be viewed as iterative procedures which at each iteration make positive steps in certain search directions. Yamashita [72] was the first who noticed that search directions in different methods are linear combinations of the same two vectors, the so called affine scaling direction and centering direction. The investigation in that direction was continued by other authors (Gonzaga [28], Den Hertog and Roos [15], etc.). We shall shortly outline the results from [15].

Let $p_{aff}$ and $p_{cent}$ respectively denote the affine and the centering direction associated with the problem (1). For a given $x$ and $D_x = diag(x_1^k, \ldots, x_n^k)$ we define
\[ p_{aff} = -D_x (I - D_x A^T (AD_x^{-2} A^T)^{-1} AD_x) c. \]
The primal centering direction is defined as
\[ p_{cent} = D_x (I - D_x A^T (AD_x^{-2} A^T)^{-1} AD_x) c. \]

It can be proved that $p_{cent}$ is in fact Newton's direction for the problem of finding the analytical center of $X$. Roos and Den Hertog show that the search directions for a variety of primal methods and primal versions of dual methods (Gonzaga [25, 27, 28, 30, 31], De Ghellinek and Vial [14], Gay [23], Ye and Kojima [76], Ye [73, 74], Barnes [9], Dikin [18], Vanderbei et al. [68], Gill et al. [24], Roos and Vial [59, 60], Renegar [57], Den Hertog et al. [16], Freund [21], Anstreicher and Bosch [5], Iri and Imai [33] are given by
\[ P_x = P_{\text{aff}} + \mu P_{\text{cent}}. \]

where \( \mu \) varies with the method.

The affine and the centering direction associated with the dual problem (2) at \((y, z)\) are defined by

\[ d_{\text{aff}} = (AD_z^{-2}A^T)^{-1}b, \quad d_{\text{cent}} = -(AD_z^{-2}A^T)^{-1}AD_z^{-1}c \]

where \( D_z = \text{diag}(c_1 - A_1y, \ldots, c_n - A_ny) = \text{diag}(z_1, \ldots, z_n) \). It is shown in [15] that search directions in many dual methods as well as in dual versions of primal methods are of the form

\[ p_y = d_{\text{aff}} + \mu d_{\text{cent}}. \]

The values for \( \mu \) are obtained in the following cases: Gonzaga [25, 27, 29, 30, 31], Yamashita [72], Barnes [9], Dikin [18], Vanderbei et al. [68], Gill et al. [24], Roos and Vial [59, 60], Renegar [57], Den Hertog et al. [16], Freund [21], Ye [74], Anstreicher and Bosch [5], Iri and Imai [33].

In primal-dual case let \( D_xz = (D_x^{-1}D_z)^{1/2} = \text{diag}(\sqrt{x_1/z_1}, \ldots, \sqrt{x_n/z_n}) \) and let the directions at \( x \) and \((y, z)\) be defined by

\[ p_{\text{aff}}^* = -D_xz(I - D_xzA^T(AD_xz^2A^T)^{-1}AD_xz)D_xz, \quad d_{\text{aff}}^* = (AD_xz^2A^T)^{-1}b \]

\[ p_{\text{cent}}^* = D_xz(I - D_xzA^T(AD_xz^2A^T)^{-1}AD_xz)D_xzD_xz^{-1}c, \quad d_{\text{cent}}^* = -(AD_xz^2A^T)^{-1}AD_xz^{-1}c. \]

It is shown that primal-dual methods move from \( x \) in the direction

\[ p_x = p_{\text{aff}}^* + \mu p_{\text{cent}}^* \]

and from \( y \) in the direction

\[ p_y = d_{\text{aff}}^* + \mu d_{\text{cent}}^* \]

(which uniquely defines the direction at \( z \)). This result is obtained for the methods: Monteiro et al. [53], Kojima et al. [38, 39], Monteiro and Adler [52], Todd and Ye [64].

From the results in [15] it follows that interior point methods at the \( k \)-th iteration use a direction defined by the matrix \((AD_x^kA^T)^{-1}\), where \( D_k \) is equal to \( D_x^k \), \( D_z^k \) or \((D_x^{-1/2}D_x^k)^{1/2}\) and \( D_x^k \), \( D_z^k \) are the diagonal matrices with \( x^k \) and \( z^k \) at the diagonal. Another interesting property of interior point methods is that the majority of them generate a solution sequence in which every limit point satisfies the strict complementarity condition, i.e. belongs to the relative interior of the optimal face (see [32]). It is easy to see now that if the optimal face \( \mathcal{X}^* = \{ x \mid A_Bx_B = b, x_B \geq 0, x_N = 0 \} \)
is degenerate (\( rank A_B < m \)) then \( cond(AD_B^2 A^T) \to \infty \), \( k \to \infty \) and all three classes of interior point methods are faced with numerical stability problems, no matter which computational technique is used to find the direction. Stabilization procedures are analyzed by Ašić et al. \([7]\) and Kovačević-Vujčić \([43]\).

7. IMPLEMENTATION OF INTERIOR POINT METHODS

Practical implementation of interior point methods imposes many problems ranging from ill-conditioning to efficient data structures for handling large problems and it is a subject of many papers, e.g. \([1, 2, 13, 36, 44]\).

One of the most important issues in the implementation of interior point methods is the strategy for computing the direction. Instead of using the projection formula indicated e.g. in Algorithms 2 and 3, a suitable linear system is solved. As it is well known, the projection of a vector \( h \) onto the subspace defined by \( ADx = 0 \) is the solution to the problem

\[
\begin{align*}
\min & \|h - x\|^2 \\
\text{subject to} & \quad ADx = 0
\end{align*}
\]

In this case the Lagrange optimality conditions are given with

\[
x + DA^T r = h \\
ADx = 0
\]

or, equivalently, with

\[
AD_B^2 A^T r = ADh \\
x = h - DA^T r
\]

The first set of equations is usually solved by a stable Cholesky factorization, although some authors recommend preconditioned conjugate gradients method, and \( x \) is then obtained by a direct substitution. On the other hand, some authors (see e.g. \([6]\)) propose using the matrix structure of (15):

\[
\begin{bmatrix}
1 & DA^T \\
AD & 0
\end{bmatrix}
\begin{bmatrix}
x \\
r
\end{bmatrix}
= 
\begin{bmatrix}
h \\
0
\end{bmatrix}
\]

and choose the pivots so that the sparsity of the coefficient matrix is preserved, which is an important issue for large problems. Minimum fill in strategies were analyzed by Vanderbei \([66]\), Vanderbei and Carpenter \([67]\) and others.

Another important issue is the so-called early rounding. The idea is to apply Algorithm 1 or a similar procedure at early stages of an interior point algorithm, when a condition of the type (3) is not yet satisfied. Very often the obtained point turns out to be optimal, which can save a lot of computational effort (see e.g. \([49, 61]\)).
An interesting point is also the comparison with the simplex method. We shall quote here a part of the results reported by Lustig, Marsten and Shanno [45] who compared the OSL program for the simplex method developed by IBM and the OB1 program for the primal-dual method developed by the authors. The comparison was performed on the IBM RISC System 1600, model 530. A series of standard problems (Table 1) was tested and the results are presented in Table 2. It is evident that with growing dimension the interior methods become more efficient than the simplex method, but that the simplex method stays concurrent. The general conclusion is that both approaches to linear programming will remain in use, with the recommendation to use interior methods at early stages of the iteration process and then switch to the simplex method.

Table 1: Test problems

<table>
<thead>
<tr>
<th>NAME</th>
<th>#ROWS</th>
<th>#COLUMNS</th>
<th>#NONZEROS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARKAL</td>
<td>5,831</td>
<td>7,689</td>
<td>59,435</td>
</tr>
<tr>
<td>PILOT87</td>
<td>2,030</td>
<td>4,883</td>
<td>73,152</td>
</tr>
<tr>
<td>MFMPTE16</td>
<td>6,507</td>
<td>13,610</td>
<td>100,523</td>
</tr>
<tr>
<td>NCM</td>
<td>3,352</td>
<td>16,298</td>
<td>128,057</td>
</tr>
<tr>
<td>TRIP</td>
<td>8,335</td>
<td>21,200</td>
<td>161,160</td>
</tr>
<tr>
<td>FORD</td>
<td>43,387</td>
<td>106,908</td>
<td>189,864</td>
</tr>
<tr>
<td>OILCO</td>
<td>18,800</td>
<td>38,540</td>
<td>219,880</td>
</tr>
</tbody>
</table>

Table 2: OSL(simplex) vs. OB1(interior)

<table>
<thead>
<tr>
<th>NAME</th>
<th>OSL</th>
<th>OB1</th>
<th>OSL/OB1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#ITER</td>
<td>TIME(sec)</td>
<td>#ITER</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>MARKAL</td>
<td>16,041</td>
<td>664.22</td>
<td>41</td>
</tr>
<tr>
<td>PILOT87</td>
<td>11,578</td>
<td>1,459.42</td>
<td>42</td>
</tr>
<tr>
<td>MFMPTE16</td>
<td>27,257</td>
<td>2,017.97</td>
<td>30</td>
</tr>
<tr>
<td>NCM</td>
<td>8,938</td>
<td>309.06</td>
<td>44</td>
</tr>
<tr>
<td>TRIP</td>
<td>104,687</td>
<td>12,417.22</td>
<td>41</td>
</tr>
<tr>
<td>FORD</td>
<td>19,935</td>
<td>7,430.95</td>
<td>56</td>
</tr>
<tr>
<td>OILCO</td>
<td>95,857</td>
<td>19,889.51</td>
<td>64</td>
</tr>
</tbody>
</table>

A great scientific activity indicates that further progress in the area of interior point methods is to be expected. One of the open questions comes from the fact that interior methods, similarly as the simplex method, have much better practical performance than the theoretical bounds indicate (20-60 iterations, almost independent of the problem dimensions). Some attempts to explain this phenomenon are given in [62, 75], etc. The next important issue is how to use the information on the solution of a given program in order to solve a closely related program more efficiently.
ently. Namely, large programs are often being resolved after slight modification of
the input data, which can be done efficiently with a few iterations of the simplex
method. It is still an open question how to make interior point methods more efficient
if a possibly infeasible "warm start" is given. Attempts in that direction were made e.g.
by Freund [22]. An interesting issue is also the application of interior point methods to
linear programming problems of the special structure and discrete optimization
problems (see e.g. [46, 50, 69]). Finally we make a brief reference to infeasible interior-
point algorithms for linear programming and linear complementarity problems which
work with strictly positive but infeasible points (e.g. [77, 78]).

REFERENCES

techniques for the implementation of Karmarkar's algorithm", ORSA Journal on
of Karmarkar's method for linear programming", Mathematical Programming 46 (1990)
173-190.
behaviour of the projective and the affine rescaling algorithms', Contemporary Mathematics
Legendre transform coordinates and central trajectories', Trans. Amer. Math. Soc. 314
481-510.
[16] Den Hertog, D., Ross, C. and Terlaky, T., 'A potential reduction variant of Renegar's short-
step path-following method for linear programming', Linear Algebra and its Applications
152 (1991) 43-68.


Kovacević-Vujcić, V.V., "Improving the rate of convergence of interior point methods for linear programming", Mathematical Programming 52, (1991) 467-480.


