One-step behaviour of a method for finding a maximal solution of a system of linear inequalities

HELMUTH SPÄTH* AND G. ALISTAIR WATSON†

Abstract. *It has been observed empirically that a simple linearization method applied to the problem of finding maximal solutions to systems of linear inequalities frequently terminates in one step. This is normally a Kuhn-Tucker point, and depending on the starting point, can be a global solution. An explanation of this phenomenon is given, and this is illustrated by examples.*

Key words: *sequential linear programming, one-step behaviour, maximal solution of linear inequalities*

AMS subject classifications: 65K05, 65K10

Received May 1, 2001 Accepted October 23, 2001

1. Introduction

Let $A \in R^{m\times n}, \mathbf{b} \in R^m$ be given and define the set

$$S = \{\mathbf{x} \in R^n : \ A\mathbf{x} \leq \mathbf{b}\}.$$

If it is required to determine a maximal element of S in the sense of finding an element of S of the largest l_2 norm, then this can be expressed as

$$\text{find } \mathbf{x} \in S \text{ to maximize } \|\mathbf{x}\|, \tag{1}$$

where the norm is the least squares norm on R^n.

This problem normally has many local solutions, and is NP-hard. It is a special case of the global maximization of a convex function (which is equivalent to the global minimization of a concave function) over a polyhedral set. There are many papers devoted to this kind of problem and quite efficient global solution methods are available, in particular when, as here, the objective function is quadratic: see for example [3] and the references which are given there.

*Fachbereich Mathematik, Carl von Ossietzky Universität Oldenburg, Postfach 2503, D-26 111 Oldenburg, Germany, e-mail: spaeth@mathematik.uni-oldenburg.de

†Department of Mathematics, University of Dundee, Dundee DD1 4HN, Scotland, e-mail: gawatson@maths.dundee.ac.uk

A simple approach to the solution of (1) is to solve a sequence of linearized problems of the form

$$\text{find } \mathbf{x} \in S \text{ to maximize } \mathbf{x}^T\mathbf{x}^{(t)}, \tag{2}$$

where $\mathbf{x}^{(0)}$ is given, and the solution to (2) gives $\mathbf{x}^{(t+1)}$. This is just a linear programming problem, and so can be easily solved by standard methods. There is a particular feature of the iteration process which is of interest: it has been observed empirically that the iteration frequently terminates in one step. The point reached is a Kuhn-Tucker point or a stationary point of the problem, which, depending on the starting point, may be a global solution. Our intention here is to investigate and try to understand what is happening. A similar phenomenon has been observed in [6].

In the next section we examine the relationship between solutions of (2) and stationary points of (1). We briefly consider the relevance of this kind of iteration to some extensions of the main problem, including that of finding minimal solutions. Finally, in *Section 3* we illustrate the main findings by some examples.

2. One-step convergence

For any $\mathbf{x} \in S$ define the index set $I(\mathbf{x})$ by

$$I(\mathbf{x}) = \{i : \ (A\mathbf{x} - \mathbf{b})_i = 0\}.$$

We can define a stationary point (or Kuhn-Tucker point) of (1) as a point satisfying first order necessary conditions for a solution.

Theorem 1. *Necessary conditions for* $\mathbf{x}^*$ *to solve (1) are that there exist non-negative numbers* $\lambda_i, i \in I(\mathbf{x}^*)$ *such that*

$$\mathbf{x}^* = A^T\lambda,$$

where $\lambda_i = 0$ *if* $i \notin I(\mathbf{x}^*)$.

Proof. Because the constraints of (1) are linear, the Kuhn-Tucker conditions are necessary for a solution (for example, Fletcher [2], Theorem 9.1.1 and Lemma 9.2.2.). □

Theorem 2. *Let* $\mathbf{x}^*$ *be a Kuhn-Tucker point of (1), and let* $\mathbf{x}^{(t)} = \mathbf{x}^*$. *Then* $\mathbf{x}^*$ *solves (2).*

Proof. Let $\mathbf{x}^{(t)} = \mathbf{x}^*$ in (2). Then because the Kuhn-Tucker conditions are necessary and sufficient for a solution, $\mathbf{x}$ is a solution of (2) if and only if there exists λ such that

$$\mathbf{x}^* = A^T\lambda,$$

$$\lambda \geq 0, \ \lambda^T(\mathbf{b} - A\mathbf{x}) = 0. \tag{3}$$

The result follows using *Theorem 1.* □

This result shows that Kuhn-Tucker points of (1) are fixed points of the iteration process based on (2) provided that the linear programming problem (2) has a unique solution there. This ensures that if a Kuhn-Tucker point is reached, then the method will terminate there.

We now consider the question of what constitutes an effective choice of $\mathbf{x}^{(t)}$ (or $\mathbf{x}^{(0)}$). Let the rows of A be defined by

$$A^T = [\mathbf{a}_1, \ldots, \mathbf{a}_m].$$

Theorem 3. *Let $\mathbf{x}^*$ be a Kuhn-Tucker point of (1) and let $I^* = I(\mathbf{x}^*)$. Let K^* denote the convex cone generated by the vectors*

$$\{a_i, i \in I^*\},$$

and let $\mathbf{x}^{(t)} \in K^*$. Then if (2) has a unique solution, $\mathbf{x}^{(t+1)} = \mathbf{x}^*$.

Proof. By definition

$$\mathbf{x}^{(t)} = A^T\lambda,$$

for some λ satisfying the conditions (3) with $\mathbf{x} = \mathbf{x}^*$. In other words, $\mathbf{x}^*$ solves (2). The result follows from uniqueness. □

It follows from this *Theorem* that a choice of starting point $\mathbf{x}^{(0)}$ in the union of all the cones defined by the set of Kuhn-Tucker points of (1) will result in the solution of (2) producing one of the stationary points provided only that (2) has a unique solution. In other words, there will be convergence in one step. The particular stationary point reached will be the one associated with the cone inside which $\mathbf{x}^{(0)}$ lies. Notice that if $\mathbf{x}^{(0)}$ lies on the boundary of two of the cones, then (2) cannot have a unique solution: one of the relevant stationary points may be generated, but there is no guarantee of this.

If $\mathbf{x}^{(0)}$ lies outside the union of the cones defined by the stationary points, it is of course possible that the solution to (2) at $\mathbf{x}^{(0)}$, which will give $\mathbf{x}^{(1)}$, will lie inside the union of the cones, in which case the method will (normally) terminate at the next step. There is again no guarantee of this, although numerical evidence suggests that usually there is termination at a stationary point in a few steps.

Many methods for solving (2) will terminate at a vertex of S, that is a point with $|I(\mathbf{x})| \geq n$. Usually equality will hold here, but the vertex may be degenerate. It is well known that any local maximum of a strictly convex function (for example the objective function of (1)) over a polyhedral set must occur at a vertex [5]. Therefore a solution of (1) will normally also be a solution of (2) for some $x^{(t)}$.

The above analysis applies to some extensions of the main problem (1), for example

$$\text{find } \mathbf{x} \in S \text{ to maximize } \mathbf{x}^T\mathbf{g}(\mathbf{x}), \tag{4}$$

where $\mathbf{g}$ is a nonlinear function. Then we can consider the possibility of solving this through a sequence of linear problems having the form

$$\text{find } \mathbf{x} \in S \text{ to maximize } \mathbf{x}^T\mathbf{g}(\mathbf{x}^{(t)}). \tag{5}$$

Analogous to *Theorem 3* we have

Theorem 4. *Let $\mathbf{x}^*$ be a Kuhn-Tucker point of (4) and let $I^* = I(\mathbf{x}^*)$. Let K^* denote the convex cone generated by the vectors*

$$\{\mathbf{a}_i,\ i \in I^*\},$$

and let $\mathbf{g}(\mathbf{x}^{(t)}) \in K^*$. Then if (5) has a unique solution, $\mathbf{x}^{(t+1)} = \mathbf{x}^*$.

Therefore again a Kuhn Tucker point can be obtained in one step with a suitable starting point. However, the algorithm may not terminate there. For the argument of *Theorem 2* to go through, we require $\mathbf{g}(\mathbf{x}^*) \in K^*$, and except for the case when $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, there seems to be no guarantee of this. In addition, unless $\mathbf{x}^T\mathbf{g}(\mathbf{x})$ is convex, then Kuhn-Tucker points need not be vertices of S.

Another modification of (1) is the corresponding minimization problem:

$$\text{find } \mathbf{x} \in S \text{ to minimize } \|\mathbf{x}\|, \tag{6}$$

where S is defined as before. Indeed we may consider a more general problem

$$\text{find } \mathbf{x} \in S \text{ to minimize } \|B\mathbf{x} - \mathbf{d}\|, \tag{7}$$

where $B \in R^{t\times n}$, $\mathbf{d} \in R^t$ are given: the previous problem is clearly a special case when $t = n$, $B = I$ and $\mathbf{d} = 0$. A suitable linearized problem for (7) is then

$$\text{find } \mathbf{x} \in S \text{ to mimimize } (B\mathbf{x} - \mathbf{d})^T(B\mathbf{x}^{(t)} - \mathbf{d}) \tag{8}$$

The problems (6) and (7) are convex problems and these can be treated by modifications of standard algorithms for least squares problems (for example [4], Chapter 23). However, the use of (8) is anyway of less value in the minimization context for two main reasons: (i) there are no local solutions which are not global, and so (in the case of a unique solution) only one cone inside which the initial approximation must lie for one-step termination, and (ii) the solution may not occur at a vertex. It is easy to find examples when the solution to (7) does not occur at a vertex.

Example 1. *In (6), let* $m = 1$, $n = 2$,

$$A = \begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -6 \\ 0 \\ 0 \end{pmatrix}.$$

Then the solution occurs at the point $\mathbf{x}^* = \frac{6}{5}(1,2)^T$ *and* K^* *is the cone generated by the single vector* $(1,2)^T$. *With* $\mathbf{x}^{(t)} \in K^*$, *it is clear that* x^* *is a solution to (8). However, the simplex method will return one of the two vertices* $(0,3)$ *or* $(6,0)$, *so using (8) is unlikely to be useful for this problem.*

3. Some examples

We now illustrate the results of the previous section by considering some problems of the form (1). In fact, all the examples have a nonnegativity constraint on $\mathbf{x}$, so we will consider the feasible region S to be defined by

$$S = \{\mathbf{x} \in R^n : \ A\mathbf{x} \le \mathbf{b}, \ \mathbf{x} \ge 0\}.$$

This is as before if $\mathbf{x} \ge 0$ is incorporated into the other constraints $\mathbf{Ax} \le \mathbf{b}$ by adding $-\mathbf{Ix} \le \mathbf{0}$. To motivate our computational strategy we will first illustrate *Theorem 3*.

Example 2. *Let*

$$A = \begin{pmatrix} 1 & 1 \\ 5 & 8 \\ 1 & 0 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 6 \\ 40 \\ 4 \\ 3 \end{pmatrix}$$

The feasible region $\mathbf{Ax} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ *is shown in* Figure 1.

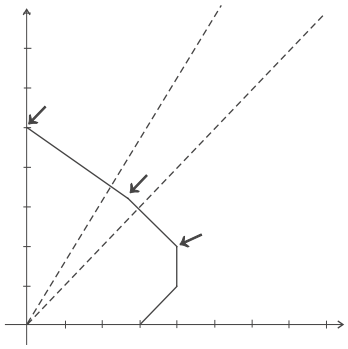

Figure 1.

There are three Kuhn-Tucker points for the maximization problem in this case, namely $(0, 5), (8/3, 10/3)$, *and* $(4, 2)$, *which are indicated by arrows on the diagram. The corresponding three cones are given by* $r(-1, 0)^T + s(5, 8)^T, r(5, 8)^T + s(1, 1)^T$, *and* $r(1, 1)^T + s(1, 0)^T$, *where* $r, s \geq 0$, *and these are also shown in* Figure 1. *If* $\mathbf{x}^{(0)}$ *lies in one of those cones and if the corresponding linear program (2) has a unique solution, then the method will converge to the corresponding Kuhn-Tucker point in one step. Clearly this requirement is satisfied if* $\mathbf{x}^{(0)}$ *lies in the interior of one of the cones.*

For actual problems of course we neither know the Kuhn-Tucker points nor the corresponding cones. If the union of the cones contains $\mathbf{x} \geq \mathbf{0}$ as in the above example, we can take M randomly generated starting values $\mathbf{x}^{(0)}$ satisfying $\mathbf{x} \geq 0$ and solve (2). Clearly $\mathbf{x}^{(0)}$ and $\alpha\mathbf{x}^{(0)}, \alpha > 0$, are in the same cone and so different values of α will change the objective function but make no difference to the solution vector. Now there may be several termination points, and we can select the one for which $\mathbf{x}^T\mathbf{x}$ is maximal. Of course, we are not certain to get the absolute maximum in this way. It is also clear that for larger m, M has also to be increased.

Example 3. *Consider the linear system of inequalities*

$$\begin{array}{rcrcrcl} x & - & 2y & + & z & \leq & 2, \\ x & + & y & - & z & \leq & 4, \\ -x & + & 2y & + & 2z & \leq & 8, \\ -x & - & y & & & \leq & 4. \end{array}$$

Using the Fourier-Motzkin elimination method [1] we get at first

$$\left.\begin{array}{rcrcr} -8 & + & 2y & + & 2z \\ -4 & - & y & & \end{array}\right\} \leq x \leq \left\{\begin{array}{rcrcr} 2 & + & 2y & - & z \\ 4 & - & y & + & z \end{array}\right. . \tag{9}$$

Ignoring x in these four double-sided inqualities we get

$$3z \le 10, \quad 3y + z \le 12, \quad -3y + z \le 6, \quad -z \le 8.$$

Thus we have

$$-8 \le z \le 10/3$$

and

$$z - 6 \le 3y \le 12 - z.$$

Inserting values z and y with these properties into (9) we will get all possible values of x. Looking for $x, y, z \ge 0$ with $x^2 + y^2 + z^2$ maximized we get at first $z = 10/3$, then $y = (12 - z)/3 = 26/9$ and finally $x = 2 + 2y - z = 4 - y + z = 40/9$. Thus $\mathbf{x}^* = \frac{1}{9}(40, 26, 30)^T$ is the desired global maximum. With $M = 100$ starting values our method reached $\mathbf{x}^*$ 97 times and $\mathbf{x}^{**} = (0, 4, 0)^T$ (another Kuhn-Tucker point) 3 times, always in one step.

Example 4. *Let $m = 3$, $n = 5$, and*

$$A = \begin{pmatrix} 0 & -1 & 2 & -3 & 4 \\ 5 & 6 & -7 & 8 & -9 \\ -10 & 11 & 12 & 13 & -14 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 15 \\ 16 \\ 17 \end{pmatrix}.$$

Again with $M = 100$ the method obtained two solutions

$$\mathbf{x}^* = (1.50, 0, 0, 33.80, 29.10)^T, \quad \mathbf{x}^{**} = (14.93, 0, 11.19, 2.46, 0)^T.$$

The first one (giving the likely global maximum) was obtained 82 times, the second one 18 times. Only in one case did the method take two iterations, reflecting a choice of $\mathbf{x}^{(0)}$ not inside an appropriate cone. The total computation time on a PC was only a few milliseconds.

Example 5. *For given m and n we generated the matrix A such that each element a_{ij} was (equally distributed) randomly chosen from $[0, 1]$. To guarantee feasibility we set $b_i = n$ $(i = 1, \ldots, n)$. Then we took M randomly generated starting values $x^{(0)}$ where $x_i^{(0)}$ $(i = 1, \ldots, n)$ were also randomly selected from $[0, 1]$. Note that those $x^{(0)}$ might not be within the union of the cones.*

We run several examples with similar results. In the case of $m = 20$, $n = 5$, $M = 1000$ we get 11 different Kuhn-Tucker points. The corresponding objective function values OF, the number k of nonzero elements within x^*, and the percentage p of appearance are given in Table 1.

It is most likely that the objective function value $OF = 28.03$ corresponds to the absolute maximum. The number of iterations normally was one, but according to the statements in Section 2 *it also was up to three. The computing time on a PC was about one second here and thus is neglectable.*

OF	*k*	*p*
17.00	*2*	*6.2*
19.63	*2*	*9.9*
20.36	*2*	*10.6*
25.52	*1*	*11.1*
26.01	*1*	*9.1*
26.14	*1*	*6.6*
26.26	*2*	*1.8*
26.29	*1*	*15.2*
26.39	*2*	*6.5*
28.01	*1*	*11.6*
28.03	*1*	*11.4*

Table 1.

4. Conclusions

The purpose of this note is to explain the empirically observed phenomenon that the sequence (2) can frequently give one-step convergence to a stationary point of (1). The circumstances when convergence will be obtained to the global solution of (1) have also been identified. Of course these are theoretical rather than practical results, and do not imply that using (2) will generally be an efficient way of finding the global solution: more sophisticated methods such as those to be found in [3] will clearly be necessary. On the other hand, the use of (2) with different starting points gives a simple way of identifying a number of stationary points, from which a good (though not necessarily optimal) solution may be obtained.

References

[1] V. CHVATAL, *Linear Programming*, Freeman, N.Y., 1983.

[2] R. FLETCHER, *Practical Methods of Optimization, 2nd Edition*, Wiley, Chichester, 1987.

[3] R. HORST, P. PARDALOS, N. V. THOAI, *Introduction to Global Optimization, 2nd Edition*, Kluwer, 2000.

[4] C. L. LAWSON, R. I. HANSON, *Solving Least Squares Problems*, Prentice Hall, Englewood Cliffs, 1974.

[5] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[6] H. SPÄTH, G. A. WATSON, *On a sequential linear programming approach to finding the smallest circumscribed, largest inscribed, and minimum zone circle or sphere*, Math. Comm. **6**(2001) 29–38.