# THE RELAXATION METHOD FOR LINEAR INEQUALITIES

SHMUEL AGMON

**1. Introduction.** In various numerical problems one is confronted with the task of solving a system of linear inequalities:

$$l_i(x) = \sum_{j=1}^{n} a_{ij}x_j + b_i \geqslant 0 \qquad (i = 1, \ldots, m), \tag{1.1}$$

assuming, of course, that the above system is consistent. Sometimes one has, in addition, to minimize a given linear form $l(x)$. Thus, in linear programming one obtains a problem of the latter type. To cite another example, this time from analysis, the problem of finding the polynomial of best approximation of degree less than $n$ corresponding to a discrete function defined in $N$ points is of the latter type. In this paper we shall be dealing only with the simpler problem of finding a solution of (1.1). Nevertheless, it is known (**7**, lectures IV and V) that the more difficult problem of minimization can be reduced to a system of inequalities involving no minimization by the duality (or minimax) principle. (However, this will increase considerably the number of unknowns and inequalities in the equivalent system.)

That the numerical problem of solving a system of inequalities is in general no easy task could be inferred from the fact that even in the case of equations the numerical solution is not easy, and that many ingenious methods were devised (**2**) in the hope of obtaining at least an approximate solution in a "reasonable" number of steps. The situation is much worse in the case of inequalities. Of the existing methods one could mention the double description method (**3**) and the simplex method due to Dantzig (**1**). The elimination method (proposed already by Fourier) is ruled out in general due to the huge number of elementary operations involved.

We propose to discuss here an iteration procedure of finding a solution of (1.1). The idea of the algorithm involved was communicated to the author by T. S. Motzkin.[1] This method, which uses orthogonal projection, will be seen later to be intimately connected with the so-called relaxation method in the case of equations (**4**; **5**; **6**), and it could be considered (after a suitable transformation) to be the extension of this method to inequalities. Even in the case of equations it seems to us that our results are not completely devoid of interest, for we shall get a simple geometric proof for the convergence of the

Received May 27, 1953. The preparation of this paper was sponsored (in part) by the Office of the Air Comptroller, U.S.A.F.

[1]It was through valuable conversations which the author had with T. S. Motzkin that he was led to consider the problems treated here.

relaxation procedure which will hold even if the (consistent) system of equations has a singular matrix, and we shall also establish the rate of convergence, a feature which had been absent in previous proofs (compare **6**).

**2. Preliminary remarks and lemmas.** When considering the system (1.1) it will be convenient to use a geometric language. Thus we shall look upon $x = (x_1, \ldots, x_n)$ as a point in $n$-dimensional Euclidean space, $E_n$, and each of the inequalities (1.1) as defining a half-space. The set of solutions will therefore consist of a convex polytope which we shall denote by $\Omega$. We shall also say that $l_i(x) = 0$ defines an oriented hyperplane $\pi_i$; $-l_i(x) = 0$ is oppositely oriented. A point $x$ will be said to be on the right side of $\pi_i$ if $l_i(x) \geqslant 0$, and on the wrong side of $\pi_i$ if $l_i(x) < 0$. It is clear that the set of solutions contains all those points $x$ which are on the right side of all oriented hyperplanes.

The following simple geometric lemma is basic in our discussion:

LEMMA 2.1. *Let $x$ and $y$ be two points in $E_n$ separated by the oriented hyperplane $\pi$ where $x$ is on the wrong side of $\pi$ and $y$ is on the right side of $\pi$. Let $x'$ be the orthogonal projection of $x$ on $\pi$. Then, if $0 \leqslant \lambda \leqslant 2$, we have*[2]

$$|x + \lambda(x' - x) - y| \leqslant |x - y|, \tag{2.1}$$

*where equality holds only for $\lambda = 0$ or $\lambda = 2$ and $y$ on $\pi$.*

*Proof.* Consider the two-dimensional plane $T$ through $x$, $x'$ and $y$. It cuts the hyperplane $\pi$ in a line $\tau$. Clearly $\tau$ separates $x$ and $y$ in $T$, and $x'$ is the orthogonal projection of $x$ on $\tau$. The statement (2.1) is now obvious from the geometric configuration.

Alternatively, to prove the lemma analytically we may assume that $\pi$ is the hyperplane $x_1 = 0$, and that $x$ is the point $(\xi, 0, \ldots, 0)$ with $\xi < 0$, and $y$ is the point $(\eta_1, \ldots, \eta_n)$ with $\eta_1 \geqslant 0$. Then $x' = (0, \ldots, 0)$. Hence we have

$$|x + \lambda(x' - x) - y| = |(1 - \lambda)\, x - y| = \left\{[(1 - \lambda)\,\xi - \eta_1]^2 + \sum_2^n \eta_i^2\right\}^{\frac{1}{2}}, \tag{2.2}$$

and

$$|x - y| = \left\{(\xi - \eta_1)^2 + \sum_2^n \eta_i^2\right\}^{\frac{1}{2}}. \tag{2.2'}$$

The result is now obvious since $[(1 - \lambda)\xi - \eta_1]^2 \leqslant (\xi - \eta_1)^2$, and equality holds only if $\lambda = 0$ or $\lambda = 2$ and $\eta_1 = 0$.

For future reference we note that (2.2) and (2.2′) imply the following more precise form of (2.1) for $0 \leqslant \lambda \leqslant 2$:

$$|x + \lambda(x' - x) - y|^2 \leqslant |x - y|^2 - [1 - (1 - \lambda)^2]\,|x' - x|^2. \tag{2.3}$$

LEMMA 2.2. *Let $\Omega$ be a polytope defined by the inequalities* (1.1) *none of which is superfluous. Let $x$ be a point exterior to $\Omega$ and let $y$ be the nearest point to $x$ on*

[2]In what follows we do not distinguish between the point $x$ and the vector joining the origin to this point, and whose magnitude we denote by $|x|$.

https://doi.org/10.4153/CJM-1954-037-2 Published online by Cambridge University Press

$\partial\Omega$. (*The boundary of* $\Omega$, *denoted by* $\partial\Omega$, *consists of those points of* $\Omega$ *which lie on at least one of the hyperplanes* $\pi_i$). *Let* $i_k$ $(k = 1, \ldots, s)$ *be the sub-set of indices for which* $l_i(y) = 0$, *and let* $\Omega_y$ *be the polyhedral cone defined by*

$$l_{i_k}(x) \geqslant 0 \qquad (k = 1, \ldots, s). \tag{2.4}$$

*Then* $x$ *is exterior to* $\Omega_y$ *and* $y$ *is also the nearest point to* $x$ *on* $\partial\Omega_y$.

*Proof.* Let us assume, on the contrary, that $x$ is not exterior to $\Omega_y$ and consequently is on the right side of all oriented hyperplanes $\pi_{i_k}$. It follows from this and from the fact that $y$ is on the right side of all hyperplanes $\pi_i (i=1, \ldots, m)$, that any $\pi_i$ $(i \neq i_k)$ having $x$ on its wrong side intersects the open interval $xy$ at a point $y_i$. At least one such hyperplane exists since $x$ is exterior to $\Omega$. Let $y^*$ be the nearest $y_i$ to $y$. Then it is easy to see that $y^* \in \partial\Omega$. But this leads to the contradiction: $|x - y^*| < |x - y|$, which establishes the first part of our contention.

Let now $\bar{y}$ be any point on $\partial\Omega_y$ different from $y$. Obviously the whole segment $\bar{y}y$ is contained in $\partial\Omega_y$. Also, there exists a spherical neighborhood $N_\epsilon$ in $E_n$, around $y$, such that its points are on the right side of all hyperplanes $\pi_i$ with $i \neq i_k$. Thus, the segment which is the intersection of $N_\epsilon$ and the segment $\bar{y}y$ is contained in $\partial\Omega$. In particular there exists a point $y' \in \partial\Omega$ which is between $\bar{y}$ and $y$ on the segment $\bar{y}y$. We therefore have: $|x - y'| \leqslant \alpha_1|x - y| + \alpha_2|x - \bar{y}|$ with $\alpha_1 + \alpha_2 = 1$, $\alpha_1 > 0$, $\alpha_2 > 0$. But since $|x - y'| > |x - y|$ ($y$ being the nearest point to $x$ on $\partial\Omega$) we conclude that $|x - \bar{y}| > |x - y|$. This proves the second part of the lemma.

LEMMA 2.3. *Let a polyhedral cone* $C$ *be given by*:

$$l_i(x) = \sum_{j=1}^{n} a_{ij}x_j \geqslant 0 \qquad (i = 1, \ldots, m). \tag{2.5}$$

*Let* $E$ *be the set of points* $x$ *such that*:

(a) $x$ *is exterior to* $C$.

(b) *The nearest point to* $x$ *on* $\partial C$ *is the origin.*

*Let us denote by* $i(x)$ *the subset of indices for which* $l_i(x) < 0$, *and by* $d_i(x)$ *the distance of* $x$ *from the hyperplane* $\pi_i$: $l_i(x) = 0$. *Then*

$$\operatorname*{Inf}_{x \in E} \max_{i(x)} \frac{d_i(x)}{|x|} = \lambda(C) > 0. \tag{2.6}$$

*Proof.* From the homogeneity of (2.5) it follows that there is no loss of generality in replacing $E$ by the subset $E^*$ consisting of those points of $E$ which are also on the unit hypersphere: $|x| = 1$. The set $E^*$ is clearly compact. But, for any $x \in E^*$, we have

$$\max_{i(x)} \frac{d_i(x)}{|x|} = \max_{i(x)} d_i(x) > 0,$$

since $x$ is on the wrong side of at least one hyperplane. This and the compactness give (2.6).

**3. The method of orthogonal projection.** We shall discuss here a special iteration procedure to solve (1.1) which we shall call the method of orthogonal projection, and where the sequence of iterates is defined in the following way:

$x^{(0)}$ is arbitrary;

$x^{(\nu+1)} = x^{(\nu)}$ if $x^{(\nu)}$ is a solution of (1.1);

$x^{(\nu+1)}$ is the orthogonal projection of $x^{(\nu)}$ on the farthest hyperplane $\pi_i$ with respect to which it is on the wrong side, if $x^{(\nu)}$ is not a solution of (1.1). (If this hyperplane is not unique one chooses one of the hyperplanes with respect to which $x^{(\nu)}$ is on the wrong side and whose distance from $x^{(\nu)}$ is maximum.)

Numerically, if $x^{(\nu)}$ is not a solution we consider all indices $i'$ for which $l_{i'}(x^{(\nu)}) < 0$, and among them pick an $i_0$ for which $-l_{i'}(x^{(\nu)})/|a_{i'}|$ has its greatest value; $a_i$ being the vector: $(a_{1i}, \ldots, a_{ni})$. Then, $x^{(\nu+1)} = x^{(\nu)} + ta_{i_0}$ where $t = -l_{i_0}(x^{(\nu)})/|a_{i_0}|^2$.

We shall establish now:

THEOREM 3. *Let* (1.1) *be a consistent system of linear inequalities and let* $\{x^{(\nu)}\}$ *be the sequence of iterates defined above. Then* $x^{(\nu)} \to x$ *where* $x$ *is a solution of* (1.1). *Furthermore, if* $R$ *is the distance of* $x^{(0)}$ *from the nearest solution, we have*

$$|x^{(\nu)} - x| \leqslant 2\, R\theta^{\nu}, \qquad \nu = 0, 1, \ldots, \tag{3.1}$$

*where* $0 < \theta < 1$ *depends only on the matrix* $(a_{ij})$.

*Proof.* We first claim that if $y$ is a solution of (1.1) then $x^{(\nu)}$ approaches $y$ steadily, or, more precisely

$$|x^{(\nu+1)} - y|^2 \leqslant |x^{(\nu)} - y|^2 - |x^{(\nu+1)} - x^{(\nu)}|^2. \tag{3.2}$$

Indeed, (3.2) is trivial if $x^{(\nu)}$ is a solution. If $x^{(\nu)}$ is not a solution then (3.2) follows from the refinement (2.3) of Lemma 2.1 with $x = x^{(\nu)}$, $x' = x^{(\nu+1)}$ and $\lambda = 1$. (We use also the fact that $y$ is a solution and hence is on the right side of $\pi_i$.)

Let us consider now the polyhedral cone $C_{i_1,\ldots,i_s}$ defined by

$$\sum_{j=1}^{n} a_{i_k j} x_j \geqslant 0 \qquad (k = 1, \ldots, s), \tag{3.3}$$

where $i_k$ is a subset of the set $i = 1, \ldots, m$, $(a_{i_k j})$ being a submatrix of the matrix $(a_{ij})$ of (1.1). Let $\lambda_{i_1,\ldots,i_s}$ be the "norm" associated with $C_{i_1,\ldots,i_s}$ which was introduced in (2.6). Then, by Lemma 2.3, $\lambda_{i_1,\ldots,i_s} > 0$. Therefore:

$$\mu = \min_{i_1,\ldots,i_s} \lambda_{i_1,\ldots,i_s} > 0, \tag{3.4}$$

where the minimum is taken over all possible choices of $i_1, \ldots, i_s$ from $i = 1, \ldots, m$. Let

$$\theta = \sqrt{1 - \mu^2},$$

and let us denote by $y^{(\nu)}$ the nearest point to $x^{(\nu)}$ on $\partial\Omega$. We assert that

$$|x^{(\nu+1)} - y^{(\nu)}| \leqslant \theta |x^{(\nu)} - y^{(\nu)}| \qquad (\nu = 0, 1, \ldots). \tag{3.5}$$

Indeed, let $\Omega_{y(\nu)}$ be the polyhedral cone of Lemma 2.2 generated by the oriented hyperplanes $\pi_i$ containing $y^{(\nu)}$. From the lemma it follows that $x^{(\nu)}$ is exterior to $\Omega_{y(\nu)}$, and that $y^{(\nu)}$ is also the nearest point to $x^{(\nu)}$ on $\partial\Omega_{y(\nu)}$. Translating $y^{(\nu)}$ to the origin, and applying Lemma 2.3, taking into account (3.4) and the fact that $|x^{(\nu+1)} - x^{(\nu)}|$ is the distance of $x^{(\nu)}$ from the farthest hyperplane with respect to which it is on the wrong side, we find that

$$|x^{(\nu+1)} - x^{(\nu)}| \geqslant \max_{i'(x^{(\nu)})} \text{dist}\,(x^{(\nu)}, \pi_i) \geqslant \lambda' |x^{(\nu)} - y^{(\nu)}| \geqslant \mu |x^{(\nu)} - y^{(\nu)}|, \tag{3.6}$$

where $i'(x)$ has the meaning of Lemma 2.3 and $\lambda'$ is the associated "norm" of (2.6), the "primes" indicating that we are dealing with the polyhedral cone $\Omega_{y(\nu)}$. Combining now (3.6) and (3.2) we get

$$|x^{(\nu+1)} - y^{(\nu)}|^2 \leqslant |x^{(\nu)} - y^{(\nu)}|^2 - \mu^2 |x^{(\nu)} - y^{(\nu)}|^2, \tag{3.7}$$

which establishes (3.5).

The remainder of the proof now follows easily. We first note that the iterates $x^{(0)}, x^{(1)}, \ldots, x^{(\nu)}, \ldots$ are all included in the hypersphere $S^{(0)} : |x - y^{(0)}| \leqslant |x^{(0)} - y^{(0)}| = R$. This follows from (3.2). For the same reason $x^{(1)}, \ldots, x^{(\nu)}, \ldots$ lie in the hypersphere $S^{(1)} : |x - y^{(1)}| \leqslant |x^{(1)} - y^{(1)}|$. But since $y^{(1)}$ is the nearest point to $x^{(1)}$ on $\partial\Omega$, and on account of (3.5), we have

$$|x^{(1)} - y^{(1)}| \leqslant |x^{(1)} - y^{(0)}| \leqslant \theta R.$$

In the same way we get that $x^{(\nu)}, x^{(\nu+1)}, \ldots$ are contained in the hypersphere $S^{(\nu)} : |x - y^{(\nu)}| \leqslant |x^{(\nu)} - y^{(\nu)}| \leqslant \theta^\nu R$. It is now evident, since we have a sequence of hyperspheres with non-zero intersection and whose radii tend to zero, that

$$\bigcap_{\nu=0}^{\infty} S^{(\nu)} = x.$$

Thus we get $\lim x^{(\nu)} = x = \lim y^{(\nu)}$ which proves the convergence of $x^{(\nu)}$ to a solution of (1.1). Moreover, since both $x^{(\nu)}$ and $x$ belong to the hypersphere $S^{(\nu)}$ whose diameter does not exceed $2R\theta^\nu$, (3.1) follows, and the proof is complete.

In the above theorem we have established that the rapidity of convergence of the iterates to the solution is at least linear. However, the positive constant $\mu$ appearing in the definition of $\theta$ was obtained from Lemma 2.3 where we had only an existence statement. More elaborate considerations can give a lower bound to $\mu$ in terms of the matrix $(a_{ij})$. Let $C = (c_{ij})$ $(i, j = 1, \ldots, r)$ be a rectangular matrix. We shall denote by $|C|$ the determinant of $C$, and by $\Gamma(C)$ the expression

$$\Gamma(C) = \left[\sum_{j=1}^{r} \left(\sum_{i=1}^{r} |C_{ij}|\right)^2\right]^{\frac{1}{2}} \tag{3.8}$$

where the $C_{ij}$'s are the cofactors of the elements $c_{ij}$. With this notation the following result may be established:

THEOREM. *If in Theorem 3 the $l_i(x)$ are normalized so that*

$$\sum_{j=1}^{n} a_{ij}^2 = 1, \tag{3.9}$$

*and if $r$ is the rank of the matrix $(a_{ij})$, then*

$$\mu \geqslant \min_i \Big[ \sum |A_{i_1,\ldots,i_r}^{j_1,\ldots,j_r}|^2 \Big]^{\frac{1}{2}} \Big/ \Big[ \sum (\Gamma_{i_1,\ldots,i_r}^{j_1,\ldots,j_r})^2 \Big]^{\frac{1}{2}}, \tag{3.10}$$

*where the summations are taken over the range $1 \leqslant j_1 < \ldots < j_r \leqslant n$, and $i_1, \ldots, i_r$ are $r$ linearly independent rows of $(a_{ij})$ (the rows are held fixed in the brackets);*

$$A_{i_1,\ldots,i_r}^{j_1,\ldots,j_r}$$

*is the $r \times r$ sub-matrix formed by the indicated rows and columns and*

$$\Gamma_{i_1,\ldots,i_r}^{j_1,\ldots,j_r}$$

*is the associated quantity* (3.8), *while the minimum is to be taken over all different combinations of the $i$'s which correspond to linearly independent rows.*

We omit the somewhat lengthy proof of this theorem. We remark only that in its proof we make use of the invariance under orthogonal transformations of the numerator and denominator in (3.10), and use induction with respect to $n$.

**4. More general procedures.** The method discussed above admits different variants which all yield (when the system is consistent) a sequence $\{x^{(\nu)}\}$ of iterates converging to a solution of the system. The following are few examples which may prove useful in computations. In all these cases the convergence is proved easily and the rate of convergence to the solution $y$ is found to be: $|x^{(\nu)} - y| = O(\theta^\nu)$ for some $0 < \theta < 1$.

(i) *The maximal residual method.* This method differs only slightly from the method of orthogonal projection of §3. Instead of choosing $x^{(\nu+1)}$ as the orthogonal projection of $x^{(\nu)}$ on the farthest hyperplane $\pi_i$ with respect to which it is on the wrong side, we choose $x^{(\nu+1)}$ as the orthogonal projection of $x^{(\nu)}$ on this hyperplane $\pi_i$ for which the negative residual $l_i(x^{(\nu)})$ is the greatest in absolute value. The two methods coincide if the system (1.1) is normalized so that (3.9) holds.

(ii) *Over and under projection with a fixed ratio.* Here the iterates are defined in the following way: $x^{(\nu+1)} = x^{(\nu)}$ if $x^{(\nu)}$ is a solution of (1.1), or, if $x^{(\nu)}$ is not a solution we let

$$x^{(\nu+1)} = x^{(\nu)} + (1 + \beta)(\xi^{(\nu)} - x^{(\nu)}),$$

where $\xi^{(\nu)}$ is the projection of $x^{(\nu)}$ on the farthest hyperplane with respect to which it is on the wrong side, and where $0 < |\beta| < 1$. We say that we overproject, or underproject, according as $\beta > 0$ or $\beta < 0$.

https://doi.org/10.4153/CJM-1954-037-2 Published online by Cambridge University Press

In connection with the last procedure we remark that it is very plausible that overprojecting with a small positive constant $\beta$ will accelerate the convergence of $\{x^{(\nu)}\}$. Indeed, slowness in the convergence of the method of orthogonal projection as described in §3 may arise if some of the "solid angles" of the polytope $\Omega$ are very small. Overprojecting has the effect of opening the angles and this may have the effect of accelerating the convergence.

(iii) *Systematic projection.* In this procedure the $m$ hyperplanes $\pi_i : l_i(x) = 0$, are arranged in a periodic infinite sequence $\pi_\nu$ ($\nu = 1, 2, \ldots$) where $\pi_\nu = \pi_i$ and $l_\nu(x) = l_i(x)$ if $\nu \equiv i \pmod{m}$. The sequence of iterates $x^{(\nu)}$ ($\nu = 0, 1, \ldots$) is defined then in the following way: $x^{(0)}$ is arbitrary; $x^{(\nu+1)} = x^{(\nu)}$ if $x^{(\nu)}$ is on the right side of $\pi_{\nu+1}$ (i.e., if $l_{\nu+1}(x^{(\nu)}) \geqslant 0$) while if $x^{(\nu)}$ is on the wrong side of $\pi_{\nu+1}$, then $x^{(\nu+1)}$ is the orthogonal projection of $x^{(\nu)}$ on $\pi_{\nu+1}$.

**5. The equivalence of the (generalized) relaxation method, and the method of orthogonal projection.** The method of projections described in the last two sections can of course be applied to equations by replacing each equality by a pair of inequalities. An equivalent procedure would be to change slightly the algorithm defining the points $x^{(\nu)}$ by considering the absolute value of all residuals and not only the negative ones. We shall now describe a procedure which will be the generalization of the relaxation method to inequalities, a procedure which we assume the reader to be familiar with in the case of equations. Let

$$L_i(y) = \sum_{j=1}^{m} g_{ij} y_j + b_i \geqslant 0 \qquad (i = 1, \ldots, m) \tag{5.1}$$

be a set of $m$ linear inequalities in $m$ unknowns having a symmetric and positive semi-definite matrix $G = (g_{ij})$. Clearly, one may assume that no row of $G$ is identically zero. This, together with the previous assumptions, will imply that $g_{ii} > 0$. Let $e_i$ ($i = 1, \ldots, m$) be the $m$ unit vectors directed along the axes in the $E_m$ space, and let us define the sequence $\{y^{(\nu)}\}$ by the following iteration scheme:

$$\begin{aligned} &y^{(0)} \text{ is arbitrary;} \\ &y^{(\nu+1)} = y^{(\nu)} \text{ if } y^{(\nu)} \text{ is a solution of (5.1);} \\ &y^{(\nu+1)} = y^{(\nu)} + t_\nu e_{i_\nu} \text{ if } y^{(\nu)} \text{ is not a solution of (5.1),} \end{aligned} \tag{5.2}$$

where $i_\nu$ is such that

$$L_{i_\nu}(y^{(\nu)}) < 0, \qquad -L_{i_\nu}(y^{(\nu)}) = \max_i \left(-L_i(y^{(\nu)})\right),$$

and $t$ is a scalar chosen so that

$$L_{i_\nu}(y^{(\nu+1)}) = 0,$$

or, more explicitly,

$$t_\nu = -L_{i_\nu}(y^{(\nu)})/g_{i_\nu i_\nu} \tag{5.3}$$

The above procedure can be considered as the extension of the relaxation method to inequalities. We shall now establish the following theorem:

https://doi.org/10.4153/CJM-1954-037-2 Published online by Cambridge University Press

THEOREM 5. *If the previously discussed system* (5.1) *is consistent, then the sequence* $\{y^{(\nu)}\}$ *tends to a solution of the system. Moreover, we have*

$$|y^{(\nu)} - y| \leqslant K|y^{(0)} - y|\theta^{\nu}, \qquad \nu = 0, 1, \ldots, \tag{5.4}$$

*where* $0 < \theta < 1$, *and where the constants* $K$ *and* $\theta$ *depend only on the matrix* $G$.

*Proof.* We shall show that the relaxation procedure can be interpreted as an orthogonal projection procedure,which will enable us to use our previous results.

Since $G$ is symmetric and positive semi-definite, there exists a real matrix $A = (a_{ij})$ $(i, j = 1, \ldots, m)$ such that

$$G = AA^*,$$

where we denote by $A^*$ the transposed matrix. Let us introduce the new variable $x = (x_1, \ldots, x_m)$ connected with $y = (y_1, \ldots, y_m)$ by

$$x = yA. \tag{5.5}$$

We shall associate with the system (5.1) (which in matrix notation can be written as $yAA^* + b \geqslant 0$) the system

$$l_i(x) = \sum_{j=1}^{m} a_{ij}x_j + b_i \geqslant 0, \qquad i = 1, \ldots, m, \tag{5.6}$$

or in matrix notation

$$xA^* + b \geqslant 0.$$

Obviously if (5.1) is consistent the same will be true for the system (5.6). Let us define the sequence $\{x^{(\nu)}\}$ by:

$$x^{(\nu)} = y^{(\nu)}A, \qquad \nu = 0, 1, \ldots. \tag{5.7}$$

We claim that $\{x^{(\nu)}\}$ is also a sequence obtained from (5.6) by the method of orthogonal projection (i) of §4. Indeed, we note that $L_i(y^{(\nu)}) = l_i(x^{(\nu)})$ so that the residuals are the same for the two systems. Now, if $y^{(\nu)}$ is a solution of (5.1) then $x^{(\nu)}$ is a solution of (5.7), and $x^{(j)} = x^{(\nu)}, j \geqslant \nu$. If $y^{(\nu)}$ is not a solution, then:

$$y^{(\nu+1)} = y^{(\nu)} + t_\nu e_{i_\nu}$$

where

$$-L_{i_\nu}(y^{(\nu)}) = \max_i (-L_i(y^{(\nu)})), \; L_{i_\nu}(y^{(\nu+1)}) = 0.$$

Obviously, we have also:

$$-l_{i_\nu}(x^{(\nu)}) = \max_i (-l_i(x^{(\nu)})), \; l_{i_\nu}(x^{(\nu+1)}) = 0,$$

so that $x^{(\nu)}$ is replaced by the point $x^{(\nu+1)}$ situated on the hyperplane $\pi_{i_\nu}$ corresponding to the negative residual with the largest absolute value. Finally, we have

$$x^{(\nu+1)} - x^{(\nu)} = t_\nu e_{i_\nu}A = t_\nu(a_{1i_\nu}, \ldots, a_{mi_\nu}),$$

which shows that $x^{(\nu+1)}$ is indeed the orthogonal projection of $x^{(\nu)}$ on $\pi_{i_\nu}$.

But, we have pointed out in §4 that the sequence $\{x^{(\nu)}\}$ converges to a solution $x$ of (5.6). More precisely, in the same manner as (3.1) was established one shows that:

$$|x^{(\nu)} - x| \leqslant 2R\theta^{\nu}, \tag{5.8}$$

where $R$ is the distance of $x^{(0)}$ from the set of solutions of (5.6), and where $0 < \theta < 1$ depends only on $A$. Assuming the non-trivial case where $y^{(\nu)}$ is not a solution, we may write

$$0 \geqslant L_{i_\nu}(y^{(\nu)}) = l_{i_\nu}(x^{(\nu)}) = l_{i_\nu}(x) + \sum_{j=1}^{m} a_{i_\nu j}(x_j^{(\nu)} - x_j). \tag{5.9}$$

But, $l_{i_\nu}(x) \geqslant 0$ and

$$\left| \sum_{j=1}^{m} a_{i_\nu j}(x_j^{(\nu)} - x_j) \right| \leqslant \left( \sum_j a_{i_\nu j}^2 \right)^{\frac{1}{2}} \left| x^{(\nu)} - x \right| = g_{i_\nu i_\nu}^{\frac{1}{2}} \left| x^{(\nu)} - x \right|, \tag{5.10}$$

so that if we define

$$Q = \max_i g_{ii}^{\frac{1}{2}} \text{ and } q = \min_i g_{ii}^{\frac{1}{2}} \tag{5.11}$$

we get from (5.8)-(5.11):

$$|L_{i_\nu}(y^{(\nu)})| \leqslant 2RQ\theta^{\nu}. \tag{5.12}$$

Combining (5.12), (5.3) and (5.2), we find that

$$|y^{(\nu+1)} - y^{(\nu)}| \leqslant |L_{i_\nu}(y^{(\nu)})|/g_{i_\nu i_\nu} \leqslant \frac{2Q}{q^2} R\theta^{\nu} \qquad (\nu = 0, 1, \ldots), \tag{5.13}$$

from which follows the convergence of $y^{(\nu)}$ to a point $y$. Since the solution $x$ of (5.6) is related to $y$ by (5.5), $y$ is also a solution of (5.1). Finally, we may write:

$$\left| y - y^{(\nu)} \right| \leqslant \sum_{j=\nu}^{\infty} \left| y^{(j+1)} - y^{(j)} \right| \leqslant \frac{2Q}{q^2} R \sum_{j=\nu}^{\infty} \theta^j = \frac{2Q}{q^2} R \frac{1}{1-\theta} \theta^{\nu}.$$

In order to get the exact statement (5.4) we note that

$$\begin{aligned} R \leqslant |x - x^{(0)}| &= \left( \sum_{j=1}^{m} \left[ \sum_{i=1}^{m} a_{ij}(y_i - y_i^{(0)}) \right]^2 \right)^{\frac{1}{2}} \\ &\leqslant \left\{ \sum_{j=1}^{m} \left[ \sum_{i=1}^{m} (y_i - y_i^{(0)})^2 \sum_{i=1}^{m} a_{ij}^2 \right] \right\}^{\frac{1}{2}} = |y - y^{(0)}| \left( \sum_{i,j} a_{ij}^2 \right)^{\frac{1}{2}} \\ &= |y - y^{(0)}| \left( \sum_i g_{ii} \right)^{\frac{1}{2}} \leqslant |y - y^{(0)}| m^{\frac{1}{2}} Q, \end{aligned}$$

which, when combined with (5.13), gives (5.4) where

$$K = \frac{2m^{\frac{1}{2}}Q^2}{q^2(1-\theta)}.$$

https://doi.org/10.4153/CJM-1954-037-2 Published online by Cambridge University Press

We have discussed above one type of relaxation. Similar results hold for other types, such as relaxation with maximum change of $|y^{(\nu+1)} - y^{(\nu)}|$ (this corresponds to the method of orthogonal projections of §3), over and under relaxation and the method of systematic relaxation. It is also obvious that the results hold true for the case of equations after we make the necessary change in the algorithm defining $\{y^{(\nu)}\}$.

We shall now proceed to show that conversely the orthogonal projection method can be interpreted as a relaxation procedure, at least when the initial point $x^{(0)}$ is suitably chosen. We shall suppose that a consistent system of inequalities (1.1) is given. Or, in matrix notation,

$$xA^* + b \geqslant 0,$$

where $A$ is an $m \times n$ matrix, and $x = (x_1, \ldots, x_n)$ a row vector. Let us introduce the new variable $y = (y_1, \ldots, y_m)$ which will again be connected with $x$ by

$$x = yA.$$

Let us also consider the associated system

$$yAA^* + b \geqslant 0, \tag{5.14}$$

which, when expanded, has the form of (5.1), where $G = AA^*$ is an $m \times m$ symmetric and positive semi-definite matrix. Let now $x^{(0)}$ be a starting point of the form

$$x^{(0)} = y^{(0)}A, \tag{5.15}$$

and let us define $x^{(1)}, \ldots, x^{(\nu)}, \ldots$ by the orthogonal projection method (i) of §4. That is, $x^{(\nu+1)}$ is the orthogonal projection of $x^{(\nu)}$ on the hyperplane $\pi_{i_\nu}$: $l_{i_\nu}(x) = 0$ corresponding to the negative residual $l_i(x^{(\nu)})$ with the largest absolute value. We shall also define a corresponding sequence $y^{(\nu)}$ in $E_m$ as follows: $y^{(0)}$ is the chosen starting point in (5.15), $y^{(\nu+1)}$ is the projection of $y^{(\nu)}$ on the hyperplane $L_{i_\nu}(y) = 0$ in the direction parallel to the $y_{i_\nu}$ axis ($i_\nu$ being the sequence of indices associated with the $x^{(\nu)}$'s). It is now easy to see (using (5.2)) that

$$x^{(\nu)} = y^{(\nu)}A,$$

and that, moreover, the sequence $\{y^{(\nu)}\}$ may also be obtained by the relaxation scheme (5.2). Now, since in the proof of Theorem 5 we did not use the consistency of the system (5.1), but only that of (5.6), we may use the same proof to obtain again that $y^{(\nu)}$ converges to a solution $y$, and that (5.4) holds. Thus, we have established the equivalence of the two methods, and have also obtained, as a by-product, that the two systems (1.1) and (5.14) are either both consistent or inconsistent.

https://doi.org/10.4153/CJM-1954-037-2 Published online by Cambridge University Press

## References

1. G. B. Dantzig, *Maximization of a linear form whose variables are subject to a system of linear inequalities* (U.S.A.F., 1949), 16 pp.
2. G. E. Forsythe, *Solving linear algebraic equations can be interesting*, Bull. Amer. Math. Soc., *59* (1953), 299–329.
3. T. S. Motzkin and H. Raiffa, G. L. Thompson, R. M. Thrall, *The double description method*, in *Contributions to the Theory of Games*, Annals of Mathematics Series, *2* (1953), 51–74.
4. R. V. Southwell, *Relaxation methods in engineering science* (Oxford, 1940).
5. ———, *Relaxation methods in theoretical physics* (Oxford, 1946).
6. G. Temple, *The general theory of relaxations applied to linear systems*, Proc. Roy. Soc. London, *169* (1939), 476–500.
7. *Linear programming seminar notes*, Institute for Numerical Analysis (Los Angeles, 1950).

*The Rice Institute, Houston, Texas*
*University of California at Los Angeles*
*National Bureau of Standards at Los Angeles*
*The Hebrew University, Jerusalem*