132

QUADRATIC PROGRAMMING BY THE MULTIPLEX METHOD IN THE GENERAL CASE WHERE THE QUADRATIC FORM MAY BE SINGULAR

by

RAGNAR FRISCH

In previous memoranda from the Oslo University Institute of Economics some preliminary attempts were made to apply the multiplex method to quadratic programming.[1] Other preoccupations prevented the completion of these attempts. Recently I have taken this problem up again and have clarified it to a point where an algorithm for solution can be indicated.

For practical applications in the macroeconomic field it is absolutely essential to have at hand a method by which one can handle the general case where some of the variables enter *only linearly* in the preference function and where furthermore the matrix of the quadratic form for the variables that do enter quadratically, may be of a *lower rank* (sometimes of a much lower rank) than the number of degrees of freedom in the problem.[2] Most of the discussions in the sequel have, indeed, been necessary only in order to cover the perfectly general case.

1. *Definitions and symbols*

We consider $N=m+n$ variables $x_j(j=1,2,\cdots,N)$ satisfying the *m basis equations*

$$x_j=b_{j0}+\sum_{k=u\cdots w} b_{jk}\,x_k \qquad (j=1,2,\cdots)u,v,\cdots,w\,(\cdots N) \tag{1.1}$$

where $x_k(k=u,v,\cdots,w)$ are $n=N-m$ basisvariables. The inverted parenthesis to the right in (1.1) denotes "exclusion of $u,v\ldots w$". Whatever the values of the b_{j0} and b_{jk} in (1.1) the set of variables considered has n degrees of freedom.

If by convention we put

1) In particular 15 November 1956: "Kvadratisk preferansefunks jon bestemt ved intervjudata og anvendt for konveks programmering" and 21 January 1957: "The multiplex method for linear and quadratic programming".

A technique for determining the preference function in the non-linear case is discussed in 14 February 1957: "Numerical determination of a quadratic preference function for use in macroeconomic programming" (will be printed in the volume in honour of Professor Gustavo del Vecchio.)

2) This is why for instance the *Houthakker* method (*Econometrica*, January 1960) is not applicable to the type of problems I consider.

$$(1.2) \qquad b_{j0}=0 \qquad b_{jk}=e_{jk}=\begin{bmatrix}1 \text{ if } j=k\\ 0 \text{ otherwise}\end{bmatrix} \qquad \text{when } j=u, v, \cdots, w$$

the equations (1.1) will hold for all $j=1, 2, \cdots, N$, but, of course, only $N-n$ of these equations are linearly dependent.

The variables are *bounded* by the conditions

$$(1.3) \qquad \underline{x}_j \leqq x_j \leqq \bar{x}_j \qquad (j=1, 2, \cdots, N)$$

where the $\underline{x}_j$ are arbitrarily given lower bounds and the $\bar{x}_j$ arbitrarily given upper bounds. Some of the $\underline{x}_j$ may be $-\infty$ and some of the $\bar{x}_j$ may be $+\infty$. This simply means that the bounds in question do not exist. For many reasons it is much better to handle the problem in this general formulation of the bounds, than to transform it to a non-negativity form for all the variables.

The preference function is taken in the form

$$(1.4) \qquad f=p_0+\sum_{k=u\cdots w} p_k x_k-\frac{1}{2}\sum_{k=u\cdots w}\sum_{h=u\cdots w} p_{kh} x_k x_h$$

where p_0, the p_k and the p_{kh} are given real constants. The constant p_0 is unessential for the problem, but is written for the sake of formal completeness. Any of the coefficients p_k may be positive, negative or zero. The matrix p_{kh} is assumed *non negative definite but not necessarily non singular.*[1] That is to say, the quadratic form expressed by the double sum in the right member of (1.3) can never assume a strictly negative value when the x_k are real, but it may assume a zero value for one (or more) sets of values x_k even though not all the x_k in the set are zero. If— and only if— a zero value of the form can be produced by a set x_k where not all x_k are zero, the form is singular. Otherwise it is nonsingular.

2. *The desingularized preference function with or without lineo-variables*

For the reasoning in the sequel it is useful to state briefly some classical facts from the theory of quadratic forms.

If the matrix in the right member of (1.3) is singular, say of rank $r<n$, it is always possible to perform a linear and *non-singular* real transformation on the variables, i.e.

$$(2.1) \qquad y_k=\sum_{h=u\cdots w} a_{kh} x_h \qquad (k=u, \cdots, w) \qquad |a_{kh}| \neq 0$$

such that the quadratic form in (1.3) after the transformation appears as a *non-singular* form that contains only *some* of the y_k, namely precisely r of them. It is even possible to do this in an infinity of ways, i.e. there exists an infinity of non-singular real transformations a_{kh} that have the property in question.

This is only another aspect of the classical fact that it is always possible by a

1) The term "definite" is here used in the same sense as in Maxime Bocher "Introduction to Higher Algebra", 1922, § .52, that is, definiteness is not taken to imply non-singularity. If definiteness is taken to imply non-singularity, the case of Bocher-definiteness and non-singularity is usually termed semi-definiteness. Definiteness and sigularity are two rather different features and should be handled as such. A form that is (in the terminology of the text above) non-negative definite and non-singular, may be called strictly positive definite, or shorter positive definite, while a non-negative definite form may or may not be singular.

non-singular real transformation to bring the form over into a sum of squares taken with certain constant *weights* (a principle axese transformation), i.e.

$$\sum_k \sum_h p_{kh}\, x_k\, x_h = \sum_h \lambda_h\, y_h \tag{2.2}$$

where the λ_h are constant weights.

Although the *value* of the weights λ_h will depend on the specific properties of the transformation chosen, the *number* of them that are positive, the *number* of them that are negative and the *number* of them that are zero, will always remain the same. These three numbers are algebraic invariants characterizing the structure of the matrix a_{kh}. (The law of inertia of real quadratic forms).

The non-negative definiteness of the matrix is characterized by the fact that in the weighted sum of squares obtained there are *no negative terms*—i.e. no negative weights λ_h—.

The *rank* r is the same as the number of weights that are different from zero. Hence in the case of a non-negative definite form, the rank is the same as the number of weights that are strictly positive.

Furthermore, if there are no negative weights, we see immediately that the *minimum* of the quadratic form is obtained by putting all the y_k that occur with strictly positive weights, equal to zero and leave the other y_k *free*. This way of disposing of the y_k is tantamount to defining a *linear manifold* of dimentionality $(n-r)$ in the x_k space. Exactly the same reasoning applies if the original form is reduced to any non-singular form in the y_k (not necessarily a sum of squares as in (2.2)). Hence the set of points where a non-negative definite quadratic form in the x_k reaches its minimum has $(n-r)$ degrees of freedom in the x_k. If $r=n$, there is only a *point* where the minimum occurs.

A further classical fact is that there even exists (at least) one *orthogonal* transformation which carries the quadratic form over into a weighted sum of squares as in (2.2).

The orthogonal specification means

$$\sum_{g=u\cdots w} a_{kg}\, \grave{a}_{gh} = e_{kh} = 1 \text{ if } k=0, \text{ but otherwise } 0, \text{ where } \grave{a}_{gh} = a_{hg} = \text{transpose of } a_{gh}. \tag{2.3}$$

Such a transformation is necessarily non-singular since by (2.3) $|a|\cdot|\grave{a}| = |a|^2 = |e| = 1$. Furthermore its reciprocal is simply equal to its transpose which is recognized by multiplying (2.3) by a^{-1}_{sk} and summing over k.

If in the original quadratic form we insert the expressions for the x_h in terms of the y_k, we get $\sum_{st}\left[\sum_{k\,h} p_{kh}\, a_{sk}\, a_{th}\right] y_t\, y_t$. If this shall be of the form $\sum_t \lambda_t\, y_t^2$ where the λ_t form a set of so far unspecified constant weights, we must have

$$\sum_{k\,h} p_{kh}\, a_{sk}\, a_{th} = e_{st}\, \lambda_t \tag{2.4}$$

Inserting in the right member of (2.4) for e_{st} the expression taken from (2.3), namely $\sum_k a_{sk}\, a_{tk}$, and substituting $\sum_h e_{kh}\, a_{th}$ for a_{tk}, we get

$$\sum_{k\,h} (p_{kh} - \lambda_t\, e_{kh}) a_{sk}\, a_{th} = 0 \tag{2.5}$$

For any fixed t we may look upon (2.5) as a linear system $\sum_k a_{sk} U_k = 0$ to determine the unknowns $U_k = \sum_h (p_{kh} - \lambda_t e_{kh}) a_{th}$. Since a_{sk} is non-singular, the solution is $U_k = 0$ for all k, hence

$$\sum_h (p_{kh} - \lambda_t e_{kh}) a_{th} = 0 \qquad \text{for all } t \text{ and } k. \tag{2.6}$$

For any given t this is a linear system to determine the a_{th}. A solution where not all a_{th} are zero (which is obviously necessary for the non-singularity of the matrix a_{th}) is only possible if the matrix of (2.6) is zero, i.e.

$$\text{Determinant } |P_{kh} - \lambda_t e_{kh}| = 0 \tag{2.7}$$

Hence, if an orthogonal transformation to a weighted sum of squares is possible (which we know it is), the weights λ_t must be the characteristic roots of the original matrix p_{kh}.

If the roots of (2.7) are single—which is equivalent with the rank of the matrix in (2.7) being $(n-1)$—the system (2.6) determines the characteristic vectors a_{th} uniquely, apart from an arbitrary factor depending on t but not on h, and this factor is finally fixed by the condition $\sum_h a_{th}^2 = 1$ derived from (2.3). If some of the roots of (2.7) are *multiple*, the rank of the matrix in (2.6) for some t becomes *lower* than $(n-1)$ (note the difference between the rank of the matrix in (2.6) and the rank of p_{kh}). Such multiplicities produce an *indeterminateness* in the orthogonal transformation. In practice one will frequently have *extreme cases of near multiplicity*, i.e. a very high degree of indeterminateness in the principal axes. A striking example is given in (2.22)-(2.26) of my "Statistical Confluence Analysis", Oslo, 1934.

The fact that some of the characteristic roots coincide, has no effect on the rank of the quadratic form or on the possibility of reducing it to a weighted sum of squares. It only means that the orientation of the principal axes is to a certain (and sometimes large) extent arbitrary. The extreme case is that where the original form is simply the unweighted sumsquare $\sum_k x_k^2$.

The connection of the orthogonal transformation with the extremum problem for the quadratic form over the unit sphere is fully—and in very simple terms—discussed in Sections 10 and 11 of my "Maxima et Minima", Paris 1959.

In the sequel we will not be particularly interested in the reduction to a weighted sum of squares, but only in the fact that a reduction to *a non-singular form* is possible. The possibility of this reduction is expressed by the identity

$$\sum_k \sum_h p_{kh} x_k x_h = \sum_K \sum_H p_{KH} y_K y_H \qquad \text{identically in al } x_u, x_v, \cdots, x_w \tag{2.8}$$

where k and h run through the n affixes $u, v, \ldots, w$, and K and H through a subset $U, V, \cdots, W$ of $u, v, \cdots, w$. The number of affixes in the set $U, V, \cdots, W$ is equal to the rank of the matrix p_{kh}.

The connection between the x_k and the y_k in (2.8) is given by the non-singular

transformation (2.1). A simple substitution for the x_k using this transformation shows that

$$P_{KH}=\sum_{k}\sum_{h} p_{kh}\, a^{-1}_{kK}\, a^{-1}_{hH} \tag{2.9}$$

When this transformation is made, the total preference function (1.4) assumes the form

$$f=p_0+\sum_{k} P_k y_k-\frac{1}{2}\sum_{K}\sum_{H} P_{KH}\, y_K\, y_H \tag{2.10}$$

where k runs through all the n affixes $u, v, \cdots, w$ and K and H through the subset consisting of the r affixes $U, V, \cdots, W$.

The P_k are given by

$$P_k=\sum_{h=u\cdots w} p_h\, a^{-1}_{hk} \qquad (k=u, v, \cdots, w) \tag{2.11}$$

The above transformation only concerns the *basis* variables in the original problem. All the *dependent* variables x_j $(j=1, 2, \cdots)u, \cdots, w(\cdots, N)$ are retained, but their expression in term of the new basis variables y_k $(k=u, v, \cdots, w)$ must be determined.

A simple substitution in (1.1) gives

$$x_j=b_{j0}+\sum_{k=u\cdots w} b'_{jk}\, y_k \qquad (j=1, 2, \cdots)u, \cdots, w(\cdots, N) \tag{2.12}$$

where

$$b'_{jk}=\sum_{k=u\cdots w} b_{jh}\, a^{-1}_{hk} \qquad \begin{matrix}(j=1, 2, \cdots)u, \cdots, w(\cdots, N)\\ (k=u, v, \cdots, w)\end{matrix} \tag{2.13}$$

If the original programming formulations contained bounds also for the original basis variables (which will usually be the case), we must continue to consider them in the new formulation. That is, we must now introduce as n new *dependent* variables

$$x_j=\sum_{k=u\cdots w} a^{-1}_{jk}\, y_k \qquad (j=u, v, \cdots, w) \tag{2.14}$$

This being done, we must continue to take account of all the $2N$ bounds (1.3). But for the new dependent variables y_k, there are no bounds, unless there should for concrete reasons be necessary or desirable to introduce special bounds also for them. To cover the most general case we will reckon also with such bounds.

It will be seen that the whole problem is now in exactly the same form as originally formulated except for the following two features:

First, the quadratic form in the preference function is now positive definite and non-singular.

Second, the linear part of the preference function may contain *more* basis variables than the quadratic part. This is a fundamental aspect of the problem which must be considered in realistic programming.

The basis variables that occur in the (non-singular) quadratic part of the preference functions—and possibly also in its linear part—we will term the *quadrato* variables.

And the basis variables that occur *only* in the linear part, but not in the quadratic part, we will term the *lineo* variables.

With a slight and obvious change of notation we can now reformulate the problem as follows :

Let x_j $(j=1, 2, \cdots, N)$ be N variables, bounded by (1.3).

We assume that the variables have n $(\leqq N)$ degrees of freedom and that within this domain of variation the n variables x_k $(k=u, v, \cdots, w)$ are linearly independent. We take them as basis variables and express all the other variables in term of these basis variables. Assuming the relations to be linear, we can put them in the form (1.1).

We consider a preference function of the form

$$f=p_0+\sum_{\kappa=u\cdots)U\cdots W(\cdots w} p_\kappa x_\kappa+\sum_{K=U\cdots W} p_K x_K-\frac{1}{2}\sum_{K=U\cdots W}\sum_{H=U\cdots W} P_{KH} x_K x_H \tag{2.15}$$

where P_{KH} is a strictly positive definite—i.e. positive definite and non-singular—matrix of order $r\leqq n$. Obviously it does not restrict generality if we assume P_{KH} to be *symmetric*. If it were not, we could interchange the summation affixes K and H and take the arithmetic average between the two expressions obtained.

The r variables x_K $(K=U, V, \cdots, W)$ are the quadrato variables, and the $(n-r)$ variables x_κ $(\kappa=u, v, \cdots)U, \cdots, W(\cdots, w)$ the lineo variables. The total set of n variables x_k $(k=u, v, \cdots, w)$ are the basis variables.

As a special case there may be no quadrato variables. As another special case there may be no lineo variables. In this case (2.15) will be said to be lineo empty.

The problem is to maximize (2.15) subject to the conditions (1.3). In this formulation the problem is quadratic in the preference function but linear in the bounds. Since the bounds are linear, the admissible region is *convex*, which, of course, is a fundamental simplification of the problem.

The form (2.15) may be called the *desingularized* form of the preference function.

A simple example will show how necessary it is to desingularize the preference function and to take full account of the distinction between lineo and quadrato variables.

Suppose we start from the formulation (1.1)–(1.2), and that our preferences are only concerned with one specific of the variables, say x_J, and that the preference function depends quadratically on this variable, i.e.

$$f=p_0+px_J-\frac{1}{2}Px_J^2 \tag{2.16}$$

where p_0, p and P are given constants, P strictly positive. If x_J happens to be amongst the basis variables, the form (2.16) is a natural starting point for the analysis. But if x_J is one of the dependent variables, and the preference function is expressed in terms of the basis variables, we would get a preference function of the form (1.4) where the matrix p_{kh} is very singular. It would, indeed, only be of rank 1, and before any useful computations could begin, we would have to proceed to a more natural formulation.

In other more complicated cases there may be other special sets of variables that enter, in part linearly and in part quadratically. By introducing the desingularized form (2.15), we clear the ground for effective computation.

3. *The unconstrained maximum of the preference function*

Let us first consider the problem of maximizing the preference function (2.15) without any constraints, neither in the form of *equations* connecting the basis variables, nor in the form of the *inequations* (1.3), expressed through the basis equations (1.1).

A first remark is that if at least one lineo variable occurs, the maximization of f has no proper meaning, because in this case f can be rendered arbitrarily great. Indeed, in this case we can choose any arbitrary values for the quadratic variables, keep these values constant and either increase ad infinitum one or more of the lineo variables with positive preference coefficient p_κ, or decrease ad infinitum one or more of the lineo variables with negative p_κ. Or we can do both.

In order to give a meaning to the maximalization in this case, some sort of specification must be introduced.

We do it by saying that we attribute *some arbitrary and fixed* values to the lineo variables and then proceed to the maximization, considering all the quadrato variables as independent. This leads to a uniquely determined maximum. And the maximum value of the preference function will be a linear function of the lineo variables.

Indeed, letting partial derivatives be denoted by subscripts, the first order conditions for a maximum are

$$\sum_H P_{KH}\, x_H = p_K \qquad (K=U, \cdots, W) \tag{3.1}$$

Since P_{KH} is non-singular, this leads to the solution

$$x_H^{\text{des}} = \sum_K P_{HK}^{-1}\, p_K \qquad (H=U, \cdots, W) \tag{3.2}$$

The superscript *des* indicates "desired".

Since P_{KH} is strictly positive definite, we certainly have a maximum, neither a minimum nor an indefinite situation.

The maximum value of the preference function is

$$f^{\text{des}} = \left[p_0 - \frac{1}{2} \sum_K \sum_H P_{KH}^{-1}\, p_K\, p_H \right] + \sum_\kappa p_\kappa\, x_\kappa \tag{3.3}$$

where the x_κ are the chosen values of the lineo functions. If there are no lineo functions, the bracket in (3.3) gives the uniquely determined maximum value.

4. *The single step building up of the inverse of a symmetric matrix*

Before proceeding to the discussion of the maximization of (2.15) under a set of constraints in the form of linear equations we need to indicate two computational techniques that are particularly useful in handling linear equations and linear dependencies. The first is a technique for building up the inverse of a symmetric matrix, the second a tool for determining the rank of a matrix and reducing it to a non-singular form. The former of these techniques will be discussed in the present section and the latter in sections 5 and 8.

Suppose we start from a symmetric, non-singular matrix

$$a_{ih} \qquad \begin{pmatrix} i=\alpha, \beta, \cdots, \gamma \\ h=\alpha, \beta, \cdots, \gamma \end{pmatrix} \tag{4.1}$$

whose inverse

$$A_{ih}=a_{hi}^{-1} \qquad \begin{pmatrix} i=\alpha,\beta,\cdots,\gamma \\ h=\alpha,\beta,\cdots,\gamma \end{pmatrix} \tag{4.2}$$

is *known*.

We also assume that the determinant value of (4.1)

$$\Delta=|a_{ih}| \tag{4.3}$$

is known.

The data (4.2)–(4.3) are recorded in the upper part of a table built like (4.5). A numerical example is given in tab. (4.26).

Because of the symmetry we only need to record the elements in the upper right triangle including the principal diagonal.

The row sums

$$A_{i\cdot}=\sum_{h=\alpha\cdots\gamma} A_{ih} \qquad (i=\alpha,\beta,\cdots,\gamma) \tag{4.4}$$

are also indicated. When computing them we work along the "angle-road". That is to say, we start from the top of the *column* that corresponds to the *row* for which we want to compute the sum. We work vertically downwards to the diagonal, include the diagonal element and then continue horizontally and finish by recording the total $A_{i\cdot}$ in its proper place on row i.

The determinant value Δ is recorded in a separate cell on the row γ and below the text indicating the level $\alpha,\beta\cdots\gamma$. To the left of this text column is a column for zero checks which are not used when the inverse elements on the level $\alpha,\beta\cdots\gamma$ are *given*, but whose analogue is used in the matrix on the next higher level $\alpha,\beta\cdots\gamma,\delta$ —compare (4.16) and (4.19). For the sake of standardization of the work sheets, this zero checks column is included also in the datum table $\alpha,\beta\cdots\gamma$ from which we start.

The same applies to the bottom row "Final checks".

The original elements a_{ih} $(i=\alpha,\beta,\cdots,\gamma,\ h=\alpha,\beta,\cdots,\gamma)$ are not indicated in tab. (4.5), only the *new column* $a_{i\delta}$ that defines the *enlargement* of the matrix. The new column δ is the same as the new row. These data are recorded in the *detached column* "Original column δ" to the right in the upper part of the table. In larger matrices it is frequently adviceable to record each column in the original matrix separately. Or one may fold the table of the complete matrix and use this folded sheet instead of the detached column indicated. In electronic computation something analogous to the separate column recording is always used.

Summary

The following is a summary of the successive steps and the formulae to be used. The Arabic numbers of the formulae indicate the order in which they are to be used. These numbers are also indicated in the lower part of tab. (4.5):

I. (4.6), (4.7) with check (4.8).
II. (4.9) with two signatures.
III. (4.13), (4.14), (4.15) with check (4.16).
IV. (4.17), (4.18) with checks (4.19).
V. (4.20) with check (4.21), possibly (4.22).
VI. (4.23) with check (4.24).

Tab. (4.5) *Singel step building up of the inverse of a symmetric matrix*

A numerical example is given in tab. (4.26)

	$h=\alpha$	β	⋯	⋯	γ	Row sums	Zero chocks	
$i=\alpha$								
β								
⋯				A_{ih}		$A_{i\cdot}$		Level $\alpha\beta\cdots\gamma$
⋯								
γ								Δ
Final checks								

	Original column δ
α	
β	
⋯	$a_{i\delta}$
⋯	
γ	
δ	

Extension δ	C_h (4.6)					C_0 (4.7)	Check (4.8)	$\frac{\Delta^\delta}{\Delta}$ (4.9)	Signat. Signat.
	$h=\alpha$	β	⋯	⋯	γ	δ	Row sums	Zero checks	
$i=\alpha$									
β									
⋯				A^δ_{ih} (4.17)		$A^\delta_{i\delta}$ (4.14)	$A^\delta_{i\cdot}$ (4.18)	(4.19)	
⋯									Level $\alpha\beta\cdots\gamma\delta$
γ									
δ						$A^\delta_{\delta\delta}$ (4.13)	$A^\delta_{\delta\cdot}$ (4.15)	(4.16)	Δ^δ (4.23)
Final checks	(4.22)						(4.21)		Check (4.24)

Auxiliary row sums used for final checks	
α	
β	
⋯	a_i
⋯	(4.20)
γ	
δ	

In the sequel these computations I–VI are explained in detail. The lay out is very mechanical and the explanations are given in such a way that they can easily be followed by a semi-trained computor without supervision of a qualified mathematician. If the checks are carefully followed, the computations are absolutely safe. Otherwise they are not.

Only the practical computation rules are indicated, without proofs. The proofs follow most easily by using Sylvester's formula for the minors in the adjoint of a given matrix.[1)]

*

We start by adding a row—the row "Extension δ"—which may be looked upon either as an additional *bottom* row on the level $(\alpha, \beta, \cdots, \gamma)$, or, perhaps more logically, as a *top* row on the level $(\alpha, \beta, \cdots, \gamma, \delta)$ as is done in tab. (4.5). On this row we record the elements C_h $(h=\alpha, \beta, \cdots, \gamma)$ and $C.$ computed by the formulae

$$C_h = -\sum_{i=\alpha\cdots\gamma} a_{i\delta} A_{ih} \qquad (h=\alpha, \beta, \cdots, \gamma) \tag{4.6}$$

$$C. = -\sum_{i=\alpha\cdots\gamma} a_{i\delta} A_{i.} \tag{4.7}$$

In making the computations (4.6), we use the elements A_{ih} from the angle-road that starts at the top of column h in the upper part of the table. The elements $a_{i\delta}$ are taken from the detached column to the right.

This computation of all the C_h and $C.$ is checked by

$$\sum_{h=\alpha\cdots\gamma} C_h - C. = 0 \tag{4.8}$$

where the C_h and $C.$ are now taken from the recordings in the row "Extension δ".

Because of rounding errors the expression (4.8) may not be exactly zero. The value actually obtained is recorded in the cell to the right of $C.$.

When (4.8) checks within rounding errors, we compute the item

$$\frac{\Delta^\delta}{\Delta} = \sum_{h=\alpha\cdots\gamma} C_h a_{h\delta} + a_{\delta\delta} \tag{4.9}$$

At this stage the ratio $\frac{\Delta^\delta}{\Delta}$ is looked upon as a single figure defined by the right member of (4.9).

The part of the computation which is defined by (4.9) is of strategic importance, and no simple sum-check is here available. The result from a first computation according to (4.9) should be recorded in the cell indicated to the right on the row "Extension δ" and the operator who makes the computation should put his signature in the upper half cell "Signat." to the right of the item (4.9) recorded.

When this is done, *another* computor should perform the same operation (4.9) and at the end (without taking a total) enter in the machine the previously computed item $\frac{\Delta^\delta}{\Delta}$ as it is *read off* from the recording in the table, taking, however, the item

1) Sylvester's formula is stated in simple terms is (13.41) in my "Maxima et Minima", Dunod, Paris 1959.

with its *opposite sign*. Only *now* a grand total is taken. If this total is exactly zero (with no rounding error in the digits recorded), the second operator enters his signature in the lower half cell "Signat.", below the first signature.

The fiilling in of the elements in the enlarged inverse can now begin. These elements are denoted

(4.10) A^{δ}_{ih} = inverse elements in the matrix obtained by adding the row δ and the column δ to the original matrix $\begin{pmatrix} i=\alpha,\beta,\cdots,\gamma,\delta \\ h=\alpha,\beta,\cdots,\gamma,\delta \end{pmatrix}$

(4.11) $$A^{\delta}_{i.}=\sum_{h=\alpha\cdots\gamma\delta} A^{\delta}_{ih}$$

(4.12) Δ^{δ} = determinat value of the matrix obtained by adding the row δ and the column δ to the original matrix.

We first fill in the bottom element $A^{\delta}_{\delta\delta}$ in the last column in the enlarged inverse. It is computed by the formula

(4.13) $$A^{\delta}_{\delta\delta}=\frac{1}{\left(\frac{\Delta^{\delta}}{\Delta}\right)}$$

where the denominator $\left(\frac{\Delta^{\delta}}{\Delta}\right)$ is the element recorded (and double signed) to the right on the top row in the lower part of the table.

The rest of the elements in the last coulumn of the enlarged inverse are now computed by using $A^{\delta}_{\delta\delta}$ as a fixed factor in the formula

(4.14) $$A^{\delta}_{i\delta}=C_i A^{\delta}_{\delta\delta} \qquad (i=\alpha,\ \beta,\cdots,\gamma)$$

Here C_i is taken from the cell No. i in the top *row* in the lower part of the table and the results $A^{\delta}_{i\delta}$ is recorded on row i in the *column* δ in the lower part.

We also compute

(4.15) $$A^{\delta}_{\delta.}=\frac{C.+1}{\left(\frac{\Delta^{\delta}}{\Delta}\right)}$$

and record the result in the cell to the right of $A^{\delta}_{\delta\delta}$.

The sum in the last column in the lower pant is now checked by

(4.16) $$\sum_{i=\alpha\cdots\gamma\delta} A^{\delta}_{i\delta}-A^{\delta}_{\delta.}=0$$

The sum in the fiirst term of (4.16) is taken straightforward as the sum of the recorded elements in the column δ in the lower part of the table, and $A^{\delta}_{\delta.}$ in (4.16) is taken as the element now recorded to the right of $A^{\delta}_{\delta\delta}$. The value obtained for (4.16) is recorded in the zerocheck cell to the right of $A^{\delta}_{\delta.}$. This value should be zero apart from rounding errors.

The elements in the triangle outside the last column in the enlarged inverse are computed *row by row* according to the formula

(4.17) $$A^{\delta}_{ih}=a_{ih}+C_h A^{\delta}_{i\delta} \qquad \begin{pmatrix} i=a,\beta,\cdots,\gamma \\ h=\alpha,\beta,\cdots,\gamma \end{pmatrix}$$

On any such row i one only needs to compute the diagonal element A_{ii}^{δ} and the elements to the right of this diagonal element.

For each row i the element $A_{i\delta}^{\delta}$ to the right in (4.27) is a *constant factor* read off from the now recorded element on row i in the column δ ine the lower part of the table, while C_h is a *variable* factor read off as the C_h element that is situated in the top row and in the same column as that where the element A_{ih}^{δ} is to be recorded.

Another formula is obtained by interchanging i and h in (4.17). It contains a constant element C_i and a variable elerment $A_{h\delta}^{\delta}$. This formula is not so convenient when the work sheet is taken in the form (4.5).

For each row i one also computes the element $A_{i\cdot}^{\delta}$ by the formula

$$A_{i\cdot}^{\delta}=A_{i\cdot}+(C_{\cdot}+1)A_{i\delta}^{\delta} \qquad (i=\alpha,\beta,\cdots,\gamma) \tag{4.18}$$

In the right member of (4.18) the item $A_{i\cdot}$ is taken from the column "Row sums" in the upper part of the table, and $C_{\cdot}$ as well as $A_{i\delta}^{\delta}$ from the column δ in the lower part. The latter column is situated directly under that where $A_{i\cdot}$ is found.

The values obtained for the right member of (4.18) is recorded in the column "Row sums" in the lower part of the table.

This being done the computed elements A_{ih}^{δ} $(h=\alpha,\beta,\cdots,\gamma,\delta)$ on row i are checked by

$$\sum_{h=\alpha\cdots\gamma\delta} A_{ih}^{\delta}-A_{i\cdot}^{\delta}=0 \qquad (i=\alpha,\beta,\cdots,\gamma) \tag{4.19}$$

The elements in the summation over h in (4.19) are taken from the angle road that ends with the element $A_{i\delta}^{\delta}$. The value of (4.19) is recorded to the right of the element $A_{i\cdot}^{\delta}$. It should be zero within rounding errors.

All the elements in the inverse on the level $(\alpha,\beta,\cdots,\gamma,\delta)$ have now been computed, and they have been checked step by step.

A final aggregated cheek is obtatined by first computing the auxiliary sums

$$a_{i\cdot}=\sum_{h=\alpha\cdots\gamma\delta} a_{ih} \qquad (i=\alpha,\beta,\cdots,\gamma,\delta) \tag{4.20}$$

These auxiliary sums can be recorded in the column to the extreme right in the lower part of the table. The final check is performed by

$$\sum_{i=\alpha\cdots\gamma\delta} a_{i\cdot}\ A_{i\cdot}^{\delta}-(\nu+1)=0 \tag{4.21}$$

where $\nu+1=$unmber of affixes in the set $\alpha,\beta,\cdots,\gamma,\delta$.

The elements $A_{i\cdot}^{\delta}$ in (4.21) are read off from the colum "Row sums" in the lower part of the table, and the $a_{i\cdot}$ from the muxiliary column.

The value obtained for (4.21) is recorded in the cell on the bottom row and in column δ. This final check should be made in all cases.

If more detailed final checks are wanted (for instance when locating an error), one may use

$$\sum_{i=\alpha\cdots\gamma\delta} a_{ih}A_{i\cdot}^{\delta}-1=0 \qquad \text{for each } h=\alpha,\beta,\cdots,\gamma,\delta. \tag{4.22}$$

If this type of checks is used, the value of (4.22) is recorded in column h of the bottom row of the lower part of the table.

One could use a check for each separate i and h, but this would involve a

number of multiplications of the same order as that needed in an independent inversion.

The determinant value Δ^δ is computed by

$$\Delta^\delta = \Delta \cdot \left(\frac{\Delta^\delta}{\Delta}\right) \tag{4.23}$$

where Δ is read off from its place in the upper part of the table, and $\left(\frac{\Delta^\delta}{\Delta}\right)$ from the last cell in the top row of the lower part.

The value Δ^δ as computed by (4.23) is recorded in the cell to the right of the check (4.16) on row δ in the lower part.

The value thus recorded is checked by computing

$$\frac{\Delta}{\Delta^\delta_{\delta\delta}} - \Delta^\delta = 0 \tag{4.24}$$

In this formula Δ is taken from its cell in the upper part of the table, $A^\delta_{\delta\delta}$ is taken from its cell in the lower part and Δ^δ from the item just recorded on row δ in the lower part. The value of (4.24)—which should be zero, apart from rounding errors—is recorded immediately below the recorded value of Δ^δ.

The computation and checking of the inverse on the level $(\alpha, \beta, \cdots, \gamma, \delta)$ and of the determinant value Δ^δ is now completed.

The following is a numerical example of the technique described above.

A numerical example

Tab. (4.25) gives a 5-rowed symmetric matrix of original elements a_{ij}, and the upper part of tab. (4.26) gives the inverse of the matrix consisting of the first 4 rows and columns of tab. (4.25) as well as the determinant value of this 4-rowed. The lower part of tab. (4.26) gives the inverse of the complete datum table (4.25) as well as the determinant value of this 5-rowed.

Tab. (4.25). *Datum matrix*

	$h=1$	2	3	4	5	Row sums
$i=1$	-5.2090	-5.4480	-5.0550	-3.8690	-3.6620	-23.2430
2		-7.0000	-6.5000	-6.1000	-5.7000	-30.7480
3			-6.2000	-5.7300	-5.3000	-28.7850
4				-1.8600	-1.8520	-19.4110
5					-1.2520	-17.7660
Column sums	-23.2430	-30.7480	-28.7850	-19.4110	-17.7660	-119.9530

Tab. (4.26) *Numerical example of tab.* (4.5)

	$h=1$	2	3	4	Row sums	Zero checks	
$i=1$	-0.84055086	0.56666049	-0.10441753	0.21170681	-0.16660109	0	
2		-5.70152102	5.78000377	-0.28631850	0.35882475	0	Level
3			-6.05397527	-0.08857727	0.46696630	0	1234
4				0.23386936	0.07068040	0	-4.9276
Final checks							

	Original column 5
1	-3.6620
2	-5.7000
3	-5.3000
4	-1.8520
5	-1.2520

Extension 5	-0.00946435	-0.31980092	-0.31353046	-0.89307859	-0.90881341	0.00000001	0.59779380	EKK / SV
	$h=1$	2	3	4	5	Row sums	Zero checks	
$i=1$	-0.84040102	0.57172362	-0.10938139	0.22584615	-0.01583213	-0.16804477	0	
2		-5.53043756	5.61227481	0.19145051	-0.53496861	0.31004279	-0.00000002	Level
3			-5.88953504	-0.55697848	0.52447928	-0.41914082	0	12345
4				1.56809091	-1.49395760	-0.06554865	0.00000014	
5					1.67281762	0.15253853	0.00000003	-2.94568873
Final checks	0.00000056	0.00000085	0.00000082	0.00000023	0.00000025	0.00000271		-0.00000001

	Auxiliary row sums used for final checks
1	-23.2430
2	-30.7480
3	-28.7850
4	-19.4110
5	-17.7660

5. *Determining the rank of a nonnegative definite matrix and finding a large principal minor in it*

The technique of Section 4 can also be used to great advantage in the case where an inverse is known and we only want to compute the *determinant value* of the enlarged matrix. In this case we can disregard the lower part of tab. (4.5) and only compute *the bottom row in the upper part*, i. e. the row with the C_h. This leads to the item $\frac{\Delta^\delta}{\Delta}$, and from this we can immediately deduce Δ^δ since Δ is recorded on the row immediately above.

In this way we can easily compute all the superdeterminants obtained by adding one row and column to the matrix whose inverse is known. Any such computation would only involve *one row* added in a list which forms an elongation of the given inverse.

This technique is particularly useful if we want to determine a *large* principal minor in a non-negative definite matrix.

In this case we first note that if a non-negative definite (but not necessarily non-singular) matrix is given—which is the same that a moment matrix is given—and we *normalize* it by dividing each element, $m_{\alpha\beta}$, by $|\sqrt{m_{\alpha\alpha}\, m_{\beta\beta}}|$, i. e. if we introduce the correlation coefficients

$$r_{\alpha\beta}=\frac{m_{\alpha\beta}}{|\sqrt{m_{\alpha\alpha}\, m_{\beta\beta}}|}, \tag{5.1}$$

we get a matrix where the following lemma holds:

Lemm (5.2). In a matrix of correlation coefficients each minor is *at most* as large as any of the principal subminors contained in it.

This lemma is proved by complete induction using the general formula[1]

$$\Delta^{(\alpha\cdots\gamma i)}\ \Delta^{(\alpha\cdots\gamma j)}-\Delta^{(\alpha\cdots\gamma)}\ \Delta^{(\alpha\cdots\gamma ij)}=D^{(\alpha\cdots\beta)\,(ij)}\ D^{(\alpha\cdots\gamma)\,(ji)} \tag{5.2}$$

where $\Delta^{(\alpha\cdots\gamma)}$ is the principal minor formed by $(\alpha\cdots\gamma)$ and $D^{(\alpha\cdots\gamma)\,(ij)}$ the (possibly non-principal) minor obtained from $\Delta^{(\alpha\cdots\gamma)}$ by adding the row i and the column j.

In the case of a real symmetric total matrix the determinant $D^{(\alpha\cdots\gamma)(ij)}$ is symmetric in i and j, and hence

$$\Delta^{(\alpha\cdots\gamma i)}\ \Delta^{(\alpha\cdots\gamma j)}\gtreqless\Delta^{(\alpha\cdots\gamma)}\ \Delta^{(\alpha\cdots\gamma ij)} \tag{5.3}$$

If n (the number of affixes $\alpha,\cdots,\beta$) is equal to 1, the lemma (5.2) is obviously true because it simply states $1\gtreqless 1-r_{ij}^2$. And if the lemma holds for n, it must hold for $(n+1)$. Indeed, if it holds for n, we have

$$\Delta^{(\alpha\cdots\gamma)}\gtreqless\Delta^{(\alpha\cdots\gamma j)} \tag{5.4}$$

Hence by inserting $\Delta^{(\alpha\cdots\gamma j)}$ for $\Delta^{(\alpha\cdots\beta)}$ in the right member of (5.3), and remembering that none of the principal minors cansidered can be negative, we see that we must certainly have

$$\Delta^{(\alpha\cdots\gamma i)}\ \Delta^{(\alpha\cdots\gamma j)}\gtreqless\Delta^{(\alpha\cdots\gamma j)}\ \Delta^{(\alpha\cdots\gamma ij)} \tag{5.5}$$

If $\Delta^{(\alpha\cdots\gamma j)}$ *is* $\neq 0$, which in the case of a non-negative definite total matrix must mean $\Delta^{(\alpha\cdots\gamma j)}>0$, we can divide by $\Delta^{(\alpha\cdots\gamma j)}$ and ohtain

$$\Delta^{(\alpha\cdots\gamma i)}\gtreqless\Delta^{(\alpha\cdots\gamma ij)} \tag{5.6}$$

1) This formula is only a special case of Sylvester's fundamental formula on the minors in an adjoint.

If $\Delta^{(\alpha\cdots ij)}=0$ but $\Delta^{(\alpha\cdots i)}>0$, we again deduce (5.6) by interchanging i and j.

The above reasoning only shows that a principal minor in a correlation determinant must be not larger than any of its principal subminors that are non-singular. But by a passage to the limit, remembering that all the minors are continuous in their various elements, we see that the lemma must hold in any case.

We can express it by saying that if in a correlation matrix any principal minor tends towards zero, it will chase all its principal superminors of all orders to zero, because they must be even smaller (or at least not larger) than the one that tends towards zero.

This lemma being established, a natural technique for finding a non-singular and "large" minor of the highest possible order in a non-negative definite matrix is the following.

First reduce the matrix to a correlation form by normalizing all elements using (5.1)

Next compute all the two rowed principal minors—in all combinations $(\alpha\beta)$—directly by the formula

$$\Delta^{(\alpha\beta)}=1-r_{\alpha\beta}^2 \tag{5.7}$$

Consider the *largest* value of (5.7) obtained.

If this maximum value is zero, we conclude that all rows in the total matrix are proportional and hence the total matrix of rank 1. We can disregard the trivial case when all the elements of the total matrix are zero, and hence the rank is 0.

If the maximum value of (5.7) is strictly positive, we pick the combination $(\alpha\beta)$ that yields the maximum. If more than one combination yields the maximum, we make a random pick amongst the combinations that yield the maximum.

In the $(\alpha\beta)$ combination chosen we compute directly the inverse

$$\begin{pmatrix} \dfrac{1}{1-r_{\alpha\beta}^2} & \dfrac{-r_{\alpha\beta}}{1-r_{\alpha\beta}^2} \\ \dfrac{-r_{\alpha\beta}}{1-r_{\alpha\beta}^2} & \dfrac{1}{1-r_{\alpha\beta}^2} \end{pmatrix} \tag{5.8}$$

From now on we use the technique described in Section 4.

We compute the values of all the three rowed principal superdeterminants $\Delta^{(\alpha\beta\delta)}$ obtained by adding one row and the same column—namely No. δ—to the two rowed matrix in the set $(\alpha\beta)$.

Again we look for the highest value obtained. If this highest value is zero, we conclude that the matrix is of rank 2. And we now know a two rowed principal determinant that is large, and we also know its inverse and can proceed accordingly for whatever purpose we may have in mind.

If the maximum obtained by adding a row and the same column to $(\alpha\beta)$ is strictly positive, we pick the three-set $(\alpha\beta\gamma)$ that yields the maximum (or, more generally: one such three-set chosen at random amongst those that yield the maximum) and we compute its inverse by the method of Section 4.

In this way we can continue until we have determined the rank of the total matrix. As a by-product we have obtained the inverse of one of the principal minors that are "large".

6. *The equationally quadro constrained maximum of the preference function*

Suppose that we impose the condition that the basis variables x_k $(k=u, v, \cdots, w)$ shall satisfy the equations obtained by putting the ν $(\leqq r)$ variables x_i $(i=\alpha, \beta, \cdots, \gamma)$ equal to prescribed values x_i^*, for instance one of the bounds $\underline{x}_i$ or $\bar{x}_i$, or any other prescribed values. The variables x_i may be basis variables or dependent variables. Compare the formulation (1.1).

The set of variables x_i $(i=\alpha, \beta, \cdots, \gamma)$ shall be *linearly independent over the quadro variables.* That is to say, we assume the $\nu \times r$ matrix

$$\left\| \begin{matrix} b_{\alpha U} & b_{\alpha V} & \cdots & b_{\alpha W} \\ b_{\beta U} & b_{\beta V} & \cdots & b_{\beta W} \\ \cdots & \cdots & \cdots & \cdots \\ b_{\gamma U} & b_{\gamma V} & \cdots & b_{\gamma W} \end{matrix} \right\| \tag{6.1}$$

is of rank ν. In other words we assume that there exists at least one $\nu \times \nu$ determinant in (6.1) that is different from zero. If any of the variables in the set $\alpha, \beta, \cdots, \gamma$ is a basis variable, the elements in the corresponding row of (6.1) are given by (1.2).

This is the same as to say that there does *not* exist any set of ν numbers ω_i $(i=\alpha, \beta, \cdots, \gamma)$—independent of H—not all zero, and such that

$$\sum_{i=\alpha \cdots \gamma} \omega_i b_{iH} = 0 \qquad \text{for all } H=U, V, \cdots, W \tag{6.2}$$

If there exists at least one $\nu \times \nu$ determinant in (6.1) which is different from zero, such a set of ν numbers cannot exist. Indeed the ω_i would then have to satisfy the ν homogeneous equations (6.2) obtained by letting H run hrough the ν columns in (6.1) that form the non-singular determinant in question. And this non-singular system the ω_i could not satisfy without being all equal to zero.

Note that we require that the ν vectors b_{ih} to be not only linearly independent over all $h=u, v, \cdots, w$, but even linearly independent over the subset $h=U, V, \cdots, W$. This will in general be a stronger requirement if lineo variables occur.

A set of ν variables x_i $(i=\alpha, \beta, \cdots, \gamma)$ of this sort will be said to form a *quadro operational set.* This terminology is derived from the fact that in (6.1) only columns pertaining to quadro variables occur.

Instead of saying that the set is linearly independent over the quadro variables, we may sometimes say for brevity that the variables in the operational set considered are *quadro independent.*

Suppose that the values x_i^* $(i=\alpha, \beta, \cdots, \gamma)$ in a quadro operational set are *given.* This is the same as to impose on the n basis variables x_k $(k=u, v, \cdots, w)$ the following ν equations

$$x_i^* = b_{i0} + \sum_{k=u \cdots w} b_{ik} x_k \qquad (i=\alpha, \beta, \cdots, \gamma) \tag{6.3}$$

We may speak of these equations as the equational quadro constraints or shorter the quadro conditions.

Since the vectors b_{ik} are quadro independent, the ν equations (6.3) are certainly linearly independent over the quadro basis variables.

To be more explicit we will write these equations in the form

$$x_i^* = b_{i0} + \sum_{\kappa} b_{i\kappa}\, x_\kappa + \sum_{K} b_{iK}\, x_K \qquad (i=\alpha, \beta, \cdots, \gamma) \tag{6.4}$$

where κ runs through the lineo affixes and K through the quadro affixes. As a special case all the terms in (6.4) containing x_κ may be absent. If so, the conditions may be termed *lineo empty*.

We consider the problem of maximizing (2.15) under the constraints (6.4).

We will first approach the problem by assuming given values k_κ for all the lineo variables and then maximize (2.15) over the quadro variables under the ν constraints (6.4). The solution obtained will be a function of the x_κ. Subsequently we will consider the maximization of the new preference function, now a function only of the x_κ. This two-stage approach to the problem is very effective. The solution of the first stage may be called the quadro optimum.

The first stage maximization leads to the equations

$$\sum_{H} P_{KH}\, x_H = P_K + \sum_{i} \mu_i\, b_{iK} \qquad (K=U, V, \cdots, W) \tag{6.5}$$

where the ν numbers μ_i $(i=\alpha, \beta, \cdots, \gamma)$ are Lagrange multipliers. Solving (6.5) for the x_H we get (remembering that the non-singular matrix P_{KH} and hence P^{-1}_{HK} are symmetric[1])

$$x_H = x_H^{\mathrm{des}} + \sum_{i} \mu_i \left[\sum_{K} b_{iK}\, P^{-1}_{KH} \right] \qquad (H=U, V, \cdots, W) \tag{6.6}$$

where x_H^{des} is given by (3.2).

Inserting the values (6.6) of x_H into (6.4), we get for the μ_i the ν equations

$$\sum_{j} Q_{ij}\, \mu_j = x_i^* - b_{i0} - \sum_{\kappa} b_{i\kappa}\, x_\kappa - \sum_{K} b_{iK}\, x_K^{\mathrm{des}} \qquad (i=\alpha, \beta, \cdots, \gamma) \tag{6.7}$$

where

$$Q_{ij} = \sum_{K} \sum_{H} P^{-1}_{KH}\, b_{iK}\, b_{jH} \qquad \begin{pmatrix} i=\alpha, \beta, \cdots, \gamma \\ j\alpha=, \beta, \cdots, \gamma \end{pmatrix} \tag{6.8}$$

The $\nu \times \nu$ matrix Q_{ij} is obviously symmetric since P^{-1}_{KH} is so. Furthermore, Q_{ij} is positive definite and *non-singular*. Indeed, P_{KH} is positive definite and non-singular, its inverse P^{-1}_{HK} or, if we like, P^{-1}_{KH}, is also positive definite and non-singular. Hence it possesses a real non-singular (but not necessarily positive definite) square root $P^{-\frac{1}{2}}$, i. e. a matrix such that

$$P^{-1}_{KH} = \sum_{G} P^{-\frac{1}{2}}_{KG}\, P^{-\frac{1}{2}}_{GH} \tag{6.9}$$

---

1) To verify the symmetry of P^{-1}_{HK} we miltiply $\sum_{G} P_{KG}\, P^{-1}_{GH} = e_{KH}$ by P^{-1}_{KL} and perform a summation over K. We get in the left member, since P_{KG} is symmetric,

$$\sum_{G}^{KL} \Big[\sum_{K} P_{GK}\, P^{-1}_{KL} \Big] P^{-1}_{GH} = \sum_{G} e_{GL}\, P^{-1}_{GH} = P^{-1}_{LH}$$

In the right member we get $\sum_{K} e_{KH}\, P^{-1}_{KL} = P^{-1}_{HL}$.

The square root $P_{KG}^{-\frac{1}{2}}$ must be symmetric[1].

Introducing (6.9) into (6.8) we see that Q_{ij} can be written in the form

$$Q_{ij}=\sum_{G} B_{iG}\, B_{jG} \qquad \begin{pmatrix} j=\alpha,\beta,\cdots,\gamma \\ j=\alpha,\beta,\cdots,\gamma \end{pmatrix} \tag{6.10}$$

where

$$B_{iG}=\sum_{K} b_{iK}\, P_{KG}^{-\frac{1}{2}} \qquad \begin{pmatrix} i=\alpha,\beta,\cdots,\gamma \\ G=U,V,\cdots,W \end{pmatrix} \tag{6.11}$$

The vectors B_{iG} cannot be linearly independent over G. Indeed, if numbers ω_i exist with the property that

$$\sum_{i} \omega_i B_{iG}=0 \qquad \text{for all } G \tag{6.12}$$

we get by inserting (6.11) into (6.12)

$$\sum_{K}\left[\sum_{i} \omega_i\, b_{iK}\right] P_{KG}^{-\frac{1}{2}}=0 \qquad \text{for all } G \tag{6.13}$$

The bracket in (6.13) depends on K, and hence (6.13) can be looked upon as a system of equations to determine this bracket. Since $P_{KG}^{-\frac{1}{2}}$ is non-singular, the bracket must vanish for all K, i. e.

$$\sum_{i} \omega_i B_{iK}=0 \qquad \text{for all } K \tag{6.14}$$

Since the vectors b_{iK} are assumed to be linearly independent over K—i. e. (6.1) of rank ν—the homogeneous equations (6.14) are of the same rank as the number of unknowns ω_i, and hance all the ω_i must be zero. Consequently the B_{iG} are linearly independent over G.

Since Q_{ij} by (6.10) is the moment matrix of vectors that are linearly independent over G, we conclude that Q_{ij} is positive definite and non-singular. The equations (6.7) can therefore be solved in the form

$$\mu_i=R_{i0}+\sum_{\kappa} R_{i\kappa}\, x_\kappa \qquad \begin{pmatrix} i=\alpha,\beta,\cdots,\gamma \\ \kappa=\text{lineo affixes in the preference function} \end{pmatrix} \tag{6.15}$$

---

1) Indeed, if we premultiply (6.9) by $P_{LK}^{\frac{1}{2}}$ and perform a summation over K (and afterwards write K insted of H), we get $P_{LK}^{-\frac{1}{2}}=\sum_{G} P_{LG}^{\frac{1}{2}} P_{GK}^{-1}$. Similarly, if we postmultiply (6.9) by $P_{HL}^{\frac{1}{2}}$ and perform a summation over H we get $P_{KL}^{-\frac{1}{2}}=\sum_{G} P_{KG}^{-1} P_{GL}^{\frac{1}{2}}$. Adding these two equations, and taking afterwards the difference of the expression far L, K and for K, L, we get $\sum_{G}\left(P_{KG}^{\frac{1}{2}}-P_{GK}^{\frac{1}{2}}\right)P_{LG}^{-1}+\left(P_{LG}^{\frac{1}{2}}-P_{LG}^{\frac{1}{2}}\right)P_{KG}^{-1}=0$ for all K and L. In particular for $K=L$ this gives $\sum_{G}\left(P_{KG}^{\frac{1}{2}}-P_{GK}^{\frac{1}{2}}\right)P_{KG}^{-1}=0$ for all K. Since P_{KG}^{-1} is non-singular, this gives $P_{KG}^{\frac{1}{2}}=P_{GK}^{\frac{1}{2}}$ for all G and K. From the symmetry of $P_{LG}^{\frac{1}{2}}$ (and the symmetry of P_{KG}^{-1}) we conclude (by considering the difference between the first two equations in this footnote) that also $P_{LK}^{-\frac{1}{2}}$ is symmetric.

where

$$R_{i0}=\sum_j Q_{ij}^{-1}\left[x_j^x-b_{j0}-\sum_K b_{jK}\,x_K^{\mathrm{des}}\right] \tag{6.16}$$

$$R_{i\kappa}=-\sum_j Q_{ij}^{-1}\,b_{j\kappa} \tag{6.17}$$

Inserting (6.15) in (6.6), we get in the quadro optimum

$$x_K=S_{K0}+\sum_\kappa S_{K\kappa}\,x_\kappa \qquad (K=U, V, \cdots, W) \tag{6.18}$$

where

$$S_{K0}=x_K^{\mathrm{des}}+\sum_i\sum_G R_{i0}\,b_{iG}\,P_{GK}^{-1} \qquad (K=U, V, \cdots, W) \tag{6.19}$$

$$S_{K\kappa}=\sum_i\sum_G R_{i\kappa}\,b_{iG}\,P_{GK}^{-1} \qquad \begin{pmatrix} G=U, V, \cdots, W \\ \kappa=u, v, \cdots)U, \cdots, W(\cdots, w \end{pmatrix} \tag{6.20}$$

By (6.18)–(6.20) the optimum values of the quadro variables are expressed explicitly as linear functions of the given values of the lineo variable.

The quadro optimum value of the preference function under the constraints (6.4) with the given values x_κ of the lineo variables is obtained by inserting (6.18) in (2.15). This gives

$$f=M_0+\sum_\kappa M_\kappa\,x_\kappa-\frac{1}{2}\sum_\kappa\sum_\delta M_{\kappa\delta}\,x_\kappa\,x_\delta \tag{6.21}$$

where

$$M_0=p_0-\sum_K p_K\,S_{K0}-\frac{1}{2}\sum_K\sum_H P_{KH}\,S_{K0}\,S_{H0} \tag{6.22}$$

$$M_\kappa=p_\kappa+\sum_K p_K\,S_{K\kappa}-\sum_K\sum_H P_{KH}\,S_{K0}\,S_{H\kappa} \quad (\gamma=u, v, \cdots)U, \cdots, W(\cdots, w) \tag{6.23}$$

$$M_{\kappa\delta}=\sum_K\sum_H P_{KH}\,S_{K\kappa}\,S_{K\delta} \qquad \begin{pmatrix} \kappa=u, v, \cdots)U, \cdots, W(\cdots, w \\ \delta=u, v, \cdots)U, \cdots, W(\cdots, w \end{pmatrix} \tag{6.24}$$

The expression (6.21) considered as a function of the original lineo variables x_κ, may be called the *quadrato optimal shape* of the preference function.

As a special case the original preference function (2.15) may have been lineo-empty, and if so, the quadrato optimal shape is simply a constant. That is, the quadrato optimal value of the preference function is then uniquely determined. Its value is simply given by the first term M_0 in (6.21). In this case no further problem will arise.

If the original preference function (2.15) actually contained one or more lineo variables, the further problem arises of what will happen if we now change the x_κ so as to obtain a still wider maximum of the preference function. This is the second stage of the problem.

We now have to seek the maximum of (6.21) over the x_κ $(\kappa=u, v, \cdots)U, \cdots, W(\cdots, w)$. This new maximization may be attempted with or without equational constraints. The equations (6.4) are certainly fulfilled for any values of the x_κ we may

choose (this we have assured), but there may perhaps be additional equations now to be imposed.

Whether in the second stage we seek the constrained or unconstrained maximum of (6.21), the nature of the $(n-r)$ rowed matrix $M_{\kappa\delta}$ in (6.2) is essential. We can show that this matrix is a moment matrix and hence is positive definite, *but we cannot take for granted that it is non-singular.* The solution of the new maximum problem will, therefore, depend essentally on the rank of $M_{\kappa\delta}$. If $M_{\kappa\delta}$ is non-singular, i. e. of rank equal to its dimensionality, the new function (6.21) will contain no variables that are lineo within this function. A unique maximum will therefore always exist, and it can be determined by the method of Section 3. But if $M_{\kappa\delta}$ is of lower rank, it may (and as a rule will) happen that the new preference function (6.21) contains certain variables that are lineo in the new preference function when it is reduced to a shape where its quadratic form is non-singular.

To see that this is so, and to give a necessary and sufficient criterion for the singularity of $M_{\kappa\delta}$, and indicate a method of desingularizing it, we will first reduce the formulae (6.24) by inserting for $S_{K\kappa}$ and $S_{H\delta}$ from (6.20). This gives

$$M_{\kappa\delta}=\sum_{KHiGjL} P_{KH}\, R_{i\kappa}\, b_{iG}\, P^{-1}_{GK}\, R_{j\delta}\, b_{jL}\, P^{-1}_{LH} \tag{6.25}$$

In the six fold summation in (6.25) the four affixes $KHGL$ run independently over the r values $U, V, \cdots, w$ while the two affixes i, j run independently over the ν values $\alpha, \beta, \cdots, \gamma$.

In (6.25) we first perform $\sum_K P_{KH}\, P^{-1}_{GK}=e_{HG}$. Next we perform $\sum_H e_{HG}\, P^{-1}_{LH}=P^{-1}_{LG}$, and finally $\sum_{GL} P^{-1}_{LG}\, b_{iG}\, b_{jL}=Q_{ij}$ (by (6.8)). Hence

$$M_{\kappa\delta}=\sum_{ij} Q_{ij}\, R_{i\kappa}\, R_{j\delta}. \tag{6.26}$$

Inserting here for $R_{i\kappa}$ and $R_{j\delta}$ from (6.17), we get

$$M_{\kappa\delta}=\sum_{ijrs} Q_{ij}\, Q^{-1}_{ir}\, b_{r\kappa}\, Q^{-1}_{js}\, b_{s\delta}. \tag{6.27}$$

In the four fold summation in (6.27) all the affixes i, j, r, s run independently through the ν values $\alpha, \beta, \cdots, \gamma$.

In (6.27) we first perform $\sum_i Q_{ij}\, Q^{-1}_{ir}=e_{jr}$ and next $\sum_j e_{jr}\, Q^{-1}_{jr}=Q^{-1}_{rs}$. This gives finally

$$M_{\kappa\delta}=\sum_{rs} Q^{-1}_{rs}\, b_{r\kappa}\, b_{s\delta} \tag{6.28}$$

Here the two affixes r, s run independently through the ν values $\alpha, \beta, \cdots, \gamma$, while κ and δ may each of them be any of the $(n-r)$ affixes $(u, v, \cdots)U, \cdots, W(\cdots, w)$.

Since Q^{-1}_{rs} is positive definite and hence possesses a symmeric square root, we see by an argument similar to the one in (6.8)—(6.10) that $M_{\kappa\delta}$ is a moment matrix and hence non-negative definite. In (6.11)—(6.14) we could build on the linear independency of the ν row vectors b_{iK} $(i=\alpha, \beta, \cdots, \gamma)$ over the field $K=U, V, \cdots, W$—compare

(6.1)—and we could therefore conclude that Q_{ij} is nonsingular, i. e. is strictly positive definite. But in (6.28) we cannot take for granted that the $(n-r)$ column vectors $b_{i\kappa}$ $(\kappa=u, v, \cdots)U, \cdots, W(\cdots, w)$ are linearly independent over the field $i=\alpha, \beta, \cdots, \gamma$. Whether they are so or not depend on the rank of the $\nu\times(n-r)$ matrix

$$(6.29) \qquad \left\| \begin{matrix} b_{\alpha u}\, b_{\alpha v} \cdots) \, b_{\alpha U} \cdots b_{\alpha W} (\cdots b_{\alpha w} \\ b_{\beta u}\, b_{\beta v} \cdots) \, b_{\beta U} \cdots b_{\beta W} (\cdots b_{\beta W} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ b_{\gamma U}\, b_{\gamma v} \cdots) \, b_{\gamma U} \cdots b_{\gamma W} (\cdots b_{\gamma w} \end{matrix} \right\|$$

The $(n-r)$ column vectors $b_{i\kappa}$ are linearly independent over the field $i=\alpha, \beta, \cdots, \gamma$ if and only if the matrix (6.29) is exactly of rank $(n-r)$. Indeed, $(n-r)$ weights ω_κ —independent of i and not all zero—and such that

$$(6.30) \qquad \sum_\kappa b_{i\kappa}\, \omega_\kappa = 0 \qquad \text{for all } i=\alpha, \beta, \cdots, \gamma$$

exists, if and only if the rank of the matrix $b_{i\kappa}$ is *less than* the number of the weights[1] ω_κ, i. e. less than $(n-r)$. This means that such a system of weights does *not* exist if and only if the matrix (6.29) is of rank equal to or larger than $(n-r)$. This is the same as the case the rank is exactly $(n-r)$, because the rank can obviously never be larger than $(n-r)$. In other words we have the following

Proposition (6.31). The $(n-r)$ column vectors $b_{i\kappa}$ $(\kappa=u, v, \cdots)U, \cdots, W(\cdots, w)$ are linearly independent if and only if the rank of (6.29) is exactly $(n-r)$.

From this we conclude in particular:

Lemma (6.32). If the number ν of the quadro independent conditions (6.4) which are to be fulfilled when we seek the quadro optimum, is less than the number $(n-r)$ of the variables x_κ—the lineo variables in the original preference function (2.15) —the matrix $M_{\kappa\delta}$ in (6.29) is always singular. This applies for instance if there are no conditions and at least one lineo variable in the original preference function.

We can sharpen (6.31) to saying that the $(n-r)$ column vectors $b_{i\kappa}$ are π fold linearly dependent if and only if the rank of the rectangular matrix (6.29) is $(n-r)-\pi$.

Applying this to (6.28) we conclude—because of the non-singularity of the positive definite matrix Q_{ij}—that we have:

Proposition (6.33). The rank of the square matrix $M_{\kappa\delta}$ defined by any of the formulae (6.24)—(6.28) is exactly equal to the rank of the rectangular matrix (6.29).

7. *A simple example*

To illustrate the various cases that may occur, let us take the example where there are only two variables x and y in the preference function (2.15), and that this function is of the form

$$(7.1) \qquad f=p_0+ax+by-\frac{1}{2}Py^2$$

where P_0, a, b are any three real constants and P a strictly positive constant. In other words, x is a lineo variable in the preference function (2.15), and y is a quadro variable, both x and y being basis variables.

1) See for instance theorem (8.40) in my "Maxima et Minima". Dunod, Paris 1959.

We consider only one side condition which we take in the form

$$Ax+By=C \tag{7.2}$$

here A, B, C are any three constants. In terms of (6.4) we can say that C represents $(x_i^*-b_{i0})$ where x_i^* is the given value of some variable, basis variable or dependent variable.

The problem is to maximize (7.1) under the condition (7.2).

In approaching this problem the values of P_0 and b are *unessential* in the sense that whatever they are, we can only insert them in the final formulae and read off the solution which this gives. We therefore need not distinguishing these values in the discussion of how the solution is obtained.

On the other hand the two cofficients A, B are *essential* in the sense that we must distinguish between zero and non zero values when we work out the solution. This gives $2^2=4$ cases. In addition some of these 4 cases are such that we have to split them in two, accordingly as C, respectively a, is equal to zero or unequal to zero. This gives the 7 principal cases listed below as I-VII.

Another way to count is to say that we consider all the four parameters A, B, a, C as being equal to zero or unequal to zero. This gives $2^4=16$ subcases out of which any two may be joined if they are such that one of the parameters in the specification is arbitrary, and any four subcases may be joined if they are such that two of the parameters are arbitrary. In the 7 principal cases below this gives the following number of subcases

$$2+1+1+4+2+2+4=16. \tag{7.3}$$

This is a check that all subcases are concerned, and each of them only once. We will discuss the 7 principal cases one by one.

Case I. $A=B=0$, $C\neq 0$, *a arbitrary.*

In this case the side-condition is infeasible. It is not possible to choose any (finite) values of the variables x and y which will satisfy the side-condition. In this case the problem is clearly impossible.

Case II. $A=B=C=0$, $a=0$.

In this case no side-condition exists (or, if we like, there is one side-condition which is "feasible" but has no effect for any of the two variables x and y), and the original preference function is lineo empty. The final solution in this case can be written down at sight, namely

$$\text{Final: } x=\text{arbitrary}, \quad y=\frac{b}{P}, \quad f=p_0+\frac{1}{2}\frac{b^2}{P} \tag{7.4}$$

We can interpret this solution by saying that the optimum point has one degree of freedom, but the optimum value of the preference function is unique and remains the same regardless of the way in which we choose the maximum point within the degree of freedom which it has. This character of the solution is due to the fact that x is in this case essentially an *extraneous* variable that has nothing to do with the maximum problem, and further such that no condition is imposed on it.

If we follow the reasoning of (6.4)-(6.21), i. e., first determine the maximum under a given value of the original lineo variable x, we reach a quadro optimal solution which is identical with (7.4). We can interpret it by saying that the

quadratic form in the quadrato optimal shape (6.21) is now of rank zero, i. e. it is singular ($n-r=1$ being in our example the rank that would indicate non-singularity). This is in conformity with lemma (6.32) since we have no side-condition and one lineo variable. Nevertheless, it so happens that when the quadratic form in (6.21) is "reduced to a non-singular form", the function (6.21) will contain no variable that is lineo within this function. This is because of a peculiarity of the numerical values of the coefficients in this case, namely $a=0$.

Case III. $A=B=C=0. \quad a\neq 0.$

In this case no side-condition exists, and the original preference function actually contains a lineo variable. We recognize at sight that in this case the value of the preference function can be rendered arbitrarily great (by choosing y arbitrarily and afterwards increasing x an infinitum when $a>0$, but decreasing it when $a<0$).

From the viewpoint of the reasoning (6.4)—(6.21) we now get the quadro optimal solution

$$\text{(7.5)} \qquad \text{Quadro}: \quad x=\text{arbitrary}, \quad y=\frac{b}{P}, \quad f=\left[P_0+\frac{1}{2}\frac{b^2}{P}\right]+ax$$

The quadro optimum point has one degree of freedom. The quadratic form of the quadro optimal shape is singular—also now in conformity with lemma (6.32)—and when the quadratic form is "reduced to a non-singular form", there remains a variable which is lineo within the new function. Taking the quadro optimal solution (7.5) as a starting point, we see that we can produce an arbitrarily large value of the preference function if we proceed to seeking the enlarged maximum.

The essential difference between this case and the foregoing is that now the quadro shape of the preference function exists (as it always will), but no finite value of the preference function exists in the final stage.

Case IV. $A\neq 0$, $B=0$, *C arbitrary, a arbitrary.*

In this case a side-condition exists, the final solution determined by taking account of this side-condition can be written down at sight as

$$\text{(7.6)} \qquad \text{Final}: \quad x=\frac{C}{A}, \quad y=\frac{b}{P}, \quad f=P_0+\frac{1}{2}\frac{b^2}{P}+\frac{aC}{A}$$

Following the reasoning of (6.4)—(6.21) we see that while it is quite true that a side-condition exists, this condition *does not contain the quadro variable y.* Expressed in general language, the side-condition is not "quadro independent", hence we can not take account of it when seeking the quadro optimum. This means that the quadro maximization has to take place *under no side-condition.* The result is

$$\text{(7.7)} \qquad \text{Quadro}: \quad x \text{ arbitrary}, \quad y=\frac{b}{P}, \quad f=\left[P_0+\frac{1}{2}\frac{b^2}{P}\right]+ax$$

The quadro solution is the same as in case III, but if we start from this quadro solution and search for the enlarged optimum, *we must now take account of an additional condition*, namely

$$\text{(7.8)} \qquad \text{Additional condition}: \quad Ax=C$$

Using this in addition to (7. 7), we verify that we get the same final solution as we wrote down at sight in (7.6).

The essential difference between this case and Case III is that the *final* maximum is now uniquely determined, it cannot be rendered arbitrarily great. This result emerges even though the quadro optimal shape contains a variable that is lineo within this function.

In the present case there is no need to distinguish between a zero and a non-zero value of a. This difference will appear automatically when we insert the value of a in (7.6) or, if we like, in (7.7).

Case V. $A=0$, $B\neq 0$, C *arbitrary*, $a=0$.

In this case there also exists a side-condition. The final solution, written down at sight, is

$$\text{(7.9)} \qquad \text{Final: } x \text{ arbitrary, } y=\frac{C}{B},\ f=P_0+\frac{C(2Bb-PC)}{2B^2}$$

To determine the quadro solution we must note that the existing condition is now quadro independent and must therefore be taken account of when working out the quadro solution. This condition is

$$\text{(7.10)} \qquad \text{Quadro independent condition: } By=C$$

Using this we get a quadro solution which is identical with (7.9). Interpreting now this solution from the quadro viewpoint, we can say the following: The quadratic form in the quadro optimal shape is singular. It is *not* so by virtue of lemma (6.32), but because the matrix (6.29) is now

$$\text{(7.11)} \qquad \|0\|$$

hence, by proposition (6.33), $M_{\varepsilon\delta}$ of rank 0. It is therefore possible that in the quadro shape we may now have a variable which is lineo within this function, Because of the particular value of a in this case, this variable drops, however, out, so that the quadro shape of the preference function becomes the same as that written to the right in (7.9).

The case now considered is similar to Case II in the respect that the quadro solution is identical with the final solution, but the *reasons* for the identity of the two solutions are different. And the actual numerical forms of the solution are also different (except for the fact that the same numerical values may be produced by coincidence).

Case VI. $A=0$, $B\neq 0$, C *arbitrary*, $a\neq 0$.

In this case we recognize at sight that the final value of the preference function may be rendered arbitrarily great.

The quadro solution is obtained by using one quadro independent condition. We get

$$\text{(7.12)} \qquad \text{Quadro: } x \text{ arbitrary, } y=\frac{C}{B},\ f=P_0+\frac{C(2Bb-PC)}{2B^2}+ax$$

The quadro shape is singular as in the Case V, and for the same reason, namely (7.11). But in distinction to Case V the singularity is now followed by the presence of a variable that is lineo in the new function. Hence we conclude from the quadro shape (and from the fact that we now have no *additional* constraints) that the preference function can be rendered arbitrarily great.

Case VII. $A \neq 0$, $B \neq 0$, C *arbitrary*, a *arbitrary.*

In this case we can solve directly by inserting

$$x = \frac{C - B_y}{A} \tag{7.13}$$

in the preference function, which gives

$$f = \left[P_0 + \frac{aC}{A}\right] + \frac{Ab - aB}{A} y - \frac{1}{2} P y^2. \tag{7.14}$$

Maximizing this over y, we get

$$\text{Final:} \quad x = \frac{ACP - B(Ab - aB)}{A^2 P}, \; y = \frac{Ab - aB}{AP}, \; f = P_0 + \frac{1}{2} \frac{(Ab - aB)^2}{A^2 P} + \frac{aC}{A} \tag{7.15}$$

This is a final solution which in character is only a generalization of (7.6) in Case IV. We can, indeed, derive (7.6) from (7.15) simply by putting $B = 0$. But the character of the quadro maximum is so different that the two cases should be kept distinct.

In determining the quadro maximum we note the matrix (6.1) is now

$$\| B \| \tag{7.16}$$

which is of rank 1 (which it was not in the Case IV). We therefore have to consider the condition (7.2) when working out the quadro solution. We can do so by means of a Lagrange multiplier or simply by expressing y in terms of the given x as

$$y = \frac{C - Ax}{B} \tag{7.17}$$

The expression for the quadro preference function will in any case come out as the same. We get

$$\text{Quadro:} \quad \begin{cases} x \text{ arbitrary}, \; y = \dfrac{C - Ax}{B}, \\ f = \left[P_0 + \dfrac{C(2Bb - PC)}{2B^2}\right] + \left[a - \dfrac{A(Bb - PC)}{B^2}\right] x - \dfrac{1}{2} \dfrac{PA^2}{B^2} x^2 \end{cases} \tag{7.18}$$

In this case the quadratic form in the quadro shape is non-singular. This checks with the fact that the matrix (6.29) is now

$$\| A \| \tag{7.19}$$

and hence now of rank 1.

We therefore know that if we now maximize f in (7.18) over x, we get a unique solution. If we actually carry out the computation of the maximum over x in (7.18), we reach the same final solution as in (7.15).

*

The above gives a survey of the various cases and illustrates the fact that in all cases where a unique maximum exists, we can find it as surely by proceeding through the first stage represented by the quadro maximum, as by the direct approach. And furthermore the quadro maximum has a sense in all cases, even in those where the final maximum of the preference function has no finite value.

8. *A method of determining the rank of any rectangular matrix and of expressing it in a basis form*

Sometimes it may be easy to determine by sight the rank of the rectangular

matrix (6.29) and to perform the subsequent computations in a way appropriate to the situation. But in the case of many variables—and in particular if the computations are done automatically—we need a general and mechanical procedure.

The method now to be considered is applicable to any rectangular matrix, but in order to keep the notation conform to the particular problem here envisaged, we consider the matrix (6.29).

The method will also be applicable to the description of the structure of a statistical cluster of points in several dimensions (describing its "rank" and "basis form" etc.), but here we shall not discuss the problem from this angle.

In (6.29) we can leave out any column that consists exclusively of zeros, because the rank of the remaining matrix will be equal to that of (6.29). Suppose that (6.29) is the result obtained after having left out these zero columns. In other words we assume that none of the columns in (6.29) consists exclusively of zeros.

We start from the remark that the rank of the rectangular matrix (6.29) is by proposition (6.33) equal to that of the square matrix (6.28). This is essentially due to the fact that Q^{-1}_{rs} is non-singular. The same conclusion would be obtained if Q^{-1}_{rs} is replaced by any other non-singular matrix. For convenience we replace Q^{-1}_{rs} by the unit matrix e_{rs}. This leads to the moments

$$m_{\kappa\delta} = \sum_{i=\alpha,\cdots\gamma} b_{i\kappa}\, b_{i\delta} \tag{8.1}$$

Since the rank of the rectangular matrix (6.29) is exactly equal to that of the square moment matrix $m_{\kappa\delta}$ defined by (8.1), we can approach the problem by studying $m_{\kappa\delta}$.

We will for convenience normalize the elements by transforming $m_{\kappa\delta}$ to the corresponding correlation coefficient

$$r_{\kappa\delta} = \frac{m_{\kappa\delta}}{|\sqrt{m_{\kappa\kappa}\, m_{\delta\delta}}|} \qquad \begin{pmatrix}\kappa = u,\ v,\cdots)\ U,\cdots,W\ (\cdots,w\\ b = u,\ v,\cdots)\ U,\cdots,W\ (\cdots,w\end{pmatrix} \tag{8.2}$$

The matrix (8.2) can be treated by the very mechanical and streamlined method of Section 5 (using the single step building up method of Section 4) and thus we can find a "large" principal minor chosen amongst the principal minors of highest order, that do not vanish.

Let this "large" principal minor be the one formed by the rows $(\rho\varepsilon\cdots\theta)$ and the same columns from the matrix (8.2). The number of affixes in the set $(\rho\varepsilon\cdots\theta)$ is at most equal to $(n-r)$ which is the number of columns in (6.29), that is to say the number of lineo variables in the preference function.

By the method of Section 5 we have not only been able to locate such a principal minor of (8.2), but we have as a by-product obtained its inverse $r^{-1}_{\kappa\delta(\rho\varepsilon\cdots\theta)}$ $(\kappa=\rho,\varepsilon,\cdots,\theta;\ \delta=\rho,\varepsilon,\cdots,\theta)$. From this inverse we immediately derive the inverse of the corresponding minor in $m_{\kappa\delta}$, namely

$$m^{-1}_{\kappa\delta(\rho\varepsilon\cdots\theta)} = \frac{r^{-1}_{\kappa\delta(\rho\varepsilon\cdots\theta)}}{|\sqrt{m_{\kappa\kappa}\, m_{\delta\delta}}|} \qquad \begin{pmatrix}\kappa=\rho,\ \varepsilon,\cdots,\theta\\ \delta=\rho,\ \varepsilon,\cdots,\theta\end{pmatrix} \tag{8.3}$$

Since the determinant in the set $(\rho\varepsilon\cdots\theta)$ is different from zero (it is even "big"), while all the higher rowed principal minors in the matrix (8.2) are zero, we know that the vectors $b_{i\rho}, b_{i\varepsilon},\cdots,b_{i\theta}$ are *not* linearly dependent over the field $i=\alpha,\ \beta,\cdots,\gamma$, while any higher set—chosen amongst $b_{i\kappa}$ with $\kappa=u,\ v,\cdots)\ U,\cdots,W(\cdots,w-is$ linearly dependent over this field. This means that any vector $b_{i\kappa}$ can be expressed as a linear form in the

vectors $b_{i\rho}$, $b_{i\iota},\cdots,b_{i\theta}$. In other words thers exists linear expressions of the form

$$b_{i\kappa}=\sum_{\phi=\rho\cdots\theta} b_{i\phi}\,E_{\phi n} \qquad (\kappa=u,\ v,\cdots)\ U,\cdots,W\,(\cdots,w) \tag{8.4}$$

where the $E_{\phi n}$ are constants independent of i.

These constants are easily determined if we multiply (8.4) by $b_{i\Psi}$ $(\Psi=\rho,\ \varepsilon,\cdots,\theta)$ and perform a summation over i. This gives

$$\sum_{\phi=\rho\cdots\theta} m_{\Psi\phi}\,E_{\phi\kappa}=m_{\Psi\kappa} \qquad \text{(for any } \kappa \text{ and } \Psi=\rho,\varepsilon,\cdots,\theta) \tag{8.5}$$

We know that the mataix of (8.5) is non-singular, and we even know its inverse from (8.3), so that we can write down immediately

$$E_{\phi\kappa}=\sum_{\Psi=\rho\cdots\theta} m^{-1}_{\phi\Psi(\rho\iota\ldots\theta)}\,m_{\Psi\kappa} \qquad \text{(for any } \kappa \text{ and } \phi=\rho,\ \varepsilon,\cdots,\theta) \tag{8:6}$$

where $m^{-1}_{\phi\Psi(\rho\iota\ldots\theta)}$ is given by (8.3).

The formulae (8.4) and (8.6) give a reduction of the singular matrix (6.27) to a non-singular basis form.

9. *The concavity propositions*

Consider any point that is given by the values x^0_k $(k=u,\ v,\cdots,w)$ of the basis variables. The first and second order partial derivatives of the preference function (1.4) in such a point are

$$f^0_k=p_k-\sum_{h=u\cdots w} p_{kh}\,x_h \qquad (k=u,\ v,\cdots,w) \tag{9.1}$$

$$f^0_{k\,h}=p_{kh} \qquad \begin{pmatrix}k=u,\ v,\cdots,w\\ h=u,\ v,\cdots,w\end{pmatrix} \tag{9.2}$$

If the preference function is written in the desingularized form (2.15), the partial derivatives are

$$f^0_k=\begin{cases} p_k & \text{if } k=u,\cdots)\ U,\cdots,W\,(\cdots,w\\ p_k-\sum_{H=U\cdots W} P_{kH}\,x^0_H & \text{if } k=U,\cdots,W\end{cases} \tag{9.3}$$

$$f^0_{k\,h}=\begin{cases} 0 & \text{if either } k \text{ or } h \text{ (or both) is any of the numbers } u,\cdots)\ U,\cdots,W\,(\cdots,w\\ P_{kh} & \text{if } k \text{ is one of the numbers } U,\cdots,W \text{ and also } h \text{ one of the numbers } U,\cdots,W\end{cases} \tag{9.4}$$

It should be noted that if the (non-negative definite) matrix p_{kh} is singular, the variables in (2.15)—and hence in (9.3) and (9.4)—heve different meanings from those in (1.4)—and hence in (9.1) and (9.2)—.

A simple Taylor development of the preference function around the given point x^0_k gives

$$f-f^0=\sum_{k=u\cdots w} f^0_k(x_k-x^0_k)-\frac{1}{2}\sum_{k=u\cdots w}\sum_{h=u\cdots w} p_{kh}(x_k-x^0_k)(x_h-x^0_h)\ \text{[in the case of (1.4)]} \tag{9.5}$$

$$f-f^0=\sum_{k=u\cdots w} f^0_k(x_k-x^0_k)-\frac{1}{2}\sum_{K=U\cdots W}\sum_{H=U\cdots W} P_{KH}(x_K-x^0_K)(x_H-x^0_H)\ \text{[in the case of (2.15)]} \tag{9.6}$$

Again it should be noted that variables in (9.6) have different meanings from those in (9.5).

If we want to express the fact that the first order derivatives in (9.6) for $k=u,\cdots)U,\cdots,W(\cdots,w$ are independent of the point, we can write the first term in the right member of (9.6) in the form

$$\sum_{k=u\cdots w} f_k^0 (x_k - x_k^0) = \sum_{\kappa} p_\kappa (x_\kappa - x_\kappa^0) + \sum_{K} f_K^0 (x_K - x_K^0) \tag{9.7}$$

where κ runs through $u,\cdots)U,\cdots,W(\cdots,w$ and K through $U,\cdots,W$.

Let x_k^0 be any point in (9.6) and consider the *tangency plane* in this point. This plane is defined as the linear manifold in x_k that satisfy the single linear equation

$$\sum_{k} f_k^0 (x_k - x_k^0) = 0 \tag{9.8}$$

The positive side of this plane is defined as that where the linear form (9.8) is strictly positive, i. e. where

$$\sum_{k} f_k^0 (x_k - x_k^0) > 0 \tag{9.9}$$

And the negative side of the plane is defined as that where the linear form is strictly negative, i. e. where

$$\sum_{k} f_k^0 (x_k - x_k^0) < 0 \tag{9.10}$$

Since the first order partial derivatives f_k^0 defined by (9.3) are uniquely determined in any point x_k^0, the tangency plane (9.8) is also uniquely determined except in a point where all the first order partial derivatives vanish. Such a point, obviously, does not exist if the preference function actually contains at least one lineo variable —i. e. where at least one of the P_κ is different from zero—. If no lineo variables occur in the preference function, one and only one point will exist where all the partial derivatives vanish. This follows from the non-singularity of the matrix P_{KH}. The point is given by (3.2). If such a point exists, it has—by virtue of the positive definiteness of P_{KH}—the property that in this point the value of the preference func tion is larger than in any other point. If such a point exists, we may call it the glo bal unconstrained maximum point.

Now consider any point x_k^0 that does not have this global maximum property. Its tangency plane is given by (9.8) and the positive and negative sides of the plane are defined by (9.9) and (9.10) respectively.

Because of the positive definiteness and non-singularity of P_{KH}, we deduce from (9.6) the following three *concavity propositions*:

(9.11) In any point x_k on the *negative* side of the tangency plane through x_k^0— i. e. where (9.10) holds—the value of the preference function is *strictly less than* the value which it assumes in x_k^0.

(9.12) In any point x_k that is situated *in* the tangency plane through x_k^0—i. e. where (9. 8) holds—but *outside* of the point x_k^0 itself the value of the preference function is *strictly less than* the value which it assumes in x_k^0.

(9.13) On the *positive* side of the tangency plane through x_k^0—i. e. where (9.9) holds—there are some points where the preference function is strictly less than the

value which it assumes in x_k^0, and other points where the value is strictly larger than and still other points where it is equal to the value which it assumes in x_k^0.

In the case of (9.5) where the matrix P_{kh} may be singular, the above three propositions (9.11)—(9.13) still hold, except that in (9.12) the words "strictly less than" must be replaced by "not larger than". And we must now consider the possibility that there may be a whole set of points, namely those for which

$$\sum_h p_{kh}\, x_h = p_k \qquad \text{for all } k=u,\ v,\cdots,w \tag{9.14}$$

—with p_{kh} singular—where a well determined tangency plane does not exist.

The three properties expressed by (9.11)—(9.13) are essentially connected with the fact that the quadratic preference function which we consider, is a *concave* function. We can generalize the analysis by *defining* a concave and non-singular function as one for which (9.11)—(9.13) holds for any point x_k^0, except possibly for a unique point. And we can define a concave but not necessarily non-singular function as one where (9.11)—(9.13) hold with the words "strictly less than" in (9.12) replaced by "not larger than", and exception is made for a certain well defined set of points.

The properties (9.11)—(9.13) are the essential foundations of the multiplex algorithm for quadratic programming. This being so, it is possible to adopt the general definition of a concave function suggested above and develop the whole multiplex technique for this general case. At present we will, however, for simplicity confine the analysis to the quadratic case.

10. A sufficient criterion for optimum

By the concavity propositions (9.11)—(9.13) we can derive an extremely useful sufficient criterion for an optimum under quadratic programming. By considering the expression $\sum_k f_k^0 (x_k - x_k^0)$ we can, indeed, reduce the optimality argument to one that is essentially similar to the one applicable to the case of a *linear* preference function. We have the following proposition.

If the point x_k^0 satisfies the following four conditions:

(10.1) The point x_k^0 $(k=u,\ v,\cdots,w)$ is *admissible*, i. e. it satisfies all the inequalities (1.3), i. e. we have $\underline{x}_k \leqq x_k^0 \leqq \bar{x}_k$ $(k=u,v,\cdots,w)$ and $\underline{x}_j \leqq b_{j0} + \sum_{k=u\cdots w} b_{jk}\, x_k^0 \leqq \bar{x}_j$ $(j=1,\ 2,\cdots)\ u,\cdots,w\ (\cdots,N)$.

(10.2) There exists a set of ν $(\nu=0,1,2,\cdots)$ linearly independent variables $x_\alpha, x_\beta,\cdots, x_\gamma$ (where $\alpha,\beta,\cdots,\gamma$ form a set of ν affixes belonging to the set $j=1,2,\cdots,N$) such that for each $i=\alpha,\beta,\cdots,\gamma$ the value $x_i^0 = b_{i0} + \sum_{k=u\cdots w} b_{ik}\, x_k^0$ is either at its upper or lower bound, i. e.

$$x_i^0 = \begin{cases} \text{either } \bar{x}_i \\ \text{or } \quad \underline{x}_i \end{cases} \tag{10.2.1}$$

This is the same as to say that all the ν linearly independent variables $x_\alpha,\cdots,x_\gamma$ are *bound-attained.*

(10.3) The preference vector f_k^0 in the point x_k^0 is *linearly expressible* in terms of the ν vectors $b_{\alpha k},\cdots,b_{\gamma k}$, i. e. there exists ν constants B_i^0 $(i=\alpha,\ \beta,\cdots,\gamma)$

independent of k (but possibly depending on the point x_k^0) such that

$$f_i^0 + \sum_{i=\alpha\cdots\gamma} B_i^0 b_{ik} = 0 \quad \text{for all } k = u, v, \cdots, w \tag{10.3.1}$$

(10.4) All the coefficients B_i^0 of (10.3.1) are *sign correct*, i. e. B_i^0 is non-negative if x_i^0 is at its lower bound $\underline{x}_i$ but non-positive if x_i^0 is at its upper bound $\bar{x}_i$ —*then* the point x_k^0 is certainly an optimum point in the sense that there exists *no other admissible* point where the preference function assumes a value that is *strictly larger than* the value f^0 which it assumes in the point x_k^0.

The fact that the ν variables $x_\alpha, \cdots, x_\gamma$ are linearly independent and the ν vectors $b_{\alpha k}, \cdots, b_{\gamma k}$ are linearly independent—or more precisely expressed: linearly independent over the field k—means that the $\nu \times n$ matrix ($\nu \leqq n$)

$$\begin{Vmatrix} b_{\alpha u} & b_{\alpha v} & \cdots & b_{\alpha w} \\ b_{\beta u} & b_{\beta v} & \cdots & b_{\beta w} \\ \cdots & \cdots & & \cdots \\ b_{\gamma u} & b_{\gamma v} & \cdots & b_{\gamma w} \end{Vmatrix} \tag{10.5}$$

is of rank ν. Obviously any set of *more* than n vectors $b_{\alpha k}, \cdots, b_{\gamma k}$, i. e. $v > n$, can *not* be linearly independent over the field k. Therefore the highest order of linear independency which can exist, is n.

The validity of the criteria (10.2)—(10.3) does in fact not depend on the linear independency of the variables $x_\alpha, \cdots, x_\gamma$ and of the vectors $b_{\alpha k}, \cdots, b_{\gamma k}$, because if a linear expression of the form (10.3.1) exists where not all the vectors $b_{\alpha k}, \cdots, b_{\gamma k}$ are linearly independent, it is easy to deduce a similar expression involving a smaller number of vectors $b_{\alpha k}, \cdots, b_{\gamma k}$ that *are* linearly independent over the field k. For simplicity of formulation we assume that this reduction has been made, and we therefore express (10.2)—(10.3) in terms of linearly independent variables $x_\alpha, \cdots, x_\gamma$ and linearly independent vectors $b_{\alpha k}, \cdots, b_{\gamma k}$.

It should be noted that (10.3) is always fulfilled if the number ν of linearly independent vectors $b_{\alpha k}, \cdots, b_{\gamma k}$ is equal to n, the number of basis variables. Indeed, in this case *any* vector ω_k (with components ω_k for $k = u, v, \cdots, w$) can be expressed linearly in terms of the vectors $b_{\alpha k}, \cdots, b_{\gamma k}$. We may, therefore, replace (10.2)—(10.3) by the single condition that there exist n linearly independent variables x_i^0 that are bound attained in the point considered. This is a correct but, of course,w eaker formulation. In any case (10.1) and (10.4) must be retained, the B_i^0 now denoting the coefficients in (10.3.1). Such an equation will certainly exist in the case considered. These coefficients are given by

$$B_i^0 = -\sum_{k=u\cdots w} f_k^0 \, \mathrm{b}_{ki}^{-1} \tag{10.6}$$

the matrix b_{ik} now being a square and non-singular matrix.

To prove the proposition expressed by (10.1)—(10.4) we note that if (10.3.1) holds for all k, we have in any point x_k, admissible or not

$$\sum_{k=u\cdots w} f_k^0 (x_k - x_k^0) = -\sum_{i=\alpha\cdots\gamma} B_i^0 (x_i - x_i^0) \tag{10.7}$$

Indeed, by inserting into the left member of (1.7) for f_k^0 its expression in terms

of $b_{\alpha k},\cdots,b_{\gamma k}$ obtained from (10.3.1), we obtain $-\sum_k \sum_i B_i^0 b_{ik}(x_k - x_k^0)$. This reduces to the expression to the right in (10.7).

If it shall be possible to find a point x_k which will make the left member of (10.7) strictly positive, there must be at least one of the terms in the right member of (10.7) that is strictly negative. This means that there must be at least one value of the affix i such that

$$(10.8)\qquad \begin{cases} \text{either } B_i^0 \text{ is strictly negative and } (x_i - x_i^0) \text{ strictly positive} \\ \text{or } \quad B_i^0 \text{ strictly positive and } (x_i - x_i^0) \text{ strictly negative.} \end{cases}$$

This, however, is impossible if all the coefficients B_i^0 are sign correct. Indeed, if x_i^0 is equal to the *upper* bound $\bar{x}_i^0$, there is no admissible point where the factor $(x_i - x_i^0)$ is strictly positive. And if we take a point x_k such that $(x_i - x_i^0)$ becomes strictly negative, the product $B_i^0(x_i - x_i^0)$ cannot become strictly negative, because by (10.4) B_i^0 is now non-positive. Similar argument if x_i^0 is at the lower bound $\underline{x}_i$. In other words there exists no admissible point x_k for which it will be possible to satisfy (10.8). Hence it is impossible to find a point x_k which will make the left member of (10.7) strictly positive.

Since by (9.11)—(9.13) the strict positivity of the left member of (10.7) is a necessary condition for a point x_k to produce a value of the preference function which is strictly larger than the value f^0 it assumes in x_k^0, we conclude that if (10.2)—(10.4) are fulfilled, it is impossible to find an admissible point x_k which produces a value of the preference function which is larger than the one which this function assumes in x_k^0. (This conclusion would follow even in the case where the quadratic form of the preference function may be singular).

This being so, the only additional condition we need to impose on x_k^0 in order to assure that x_k^0 is an optimum point, is that x_k^0 *itself* is an *admissible* point. This is taken care of by the condition (10.1).

The criterion (10.3) may be replaced by the following two conditions whose fulfilment forms a necessary and sufficient condition for (10.3) to hold:

$$(10.9)\qquad \begin{vmatrix} M_{\alpha\alpha} & M_{\alpha\beta} \cdots M_{\alpha\gamma} \\ M_{\beta\alpha} & M_{\beta\beta} \cdots M_{\beta\gamma} \\ \cdots\cdots & \cdots\cdots\cdots \\ M_{\gamma\alpha} & M_{\gamma\beta} \cdots M_{\gamma\gamma} \end{vmatrix} > 0$$

$$(10.10)\qquad d_0 = 0$$

where

$$(10.11)\qquad d_0 = M_{00} + \sum_{i=\alpha\cdots\gamma} M_{0i} B_i^0 = \frac{\begin{vmatrix} M_{00} & M_{0\alpha} \cdots M_{0\gamma} \\ M_{\alpha 0} & M_{\alpha\alpha} \cdots M_{\alpha\gamma} \\ \cdots\cdots & \cdots\cdots\cdots \\ M_{\gamma 0} & M_{\gamma\alpha} \cdots M_{\gamma\gamma} \end{vmatrix}}{\begin{vmatrix} M_{\alpha\alpha} \cdots M_{\alpha\gamma} \\ \cdots\cdots \\ M_{\gamma\alpha} \cdots M_{\gamma\gamma} \end{vmatrix}}$$

$$M_{00}=\sum_{k=u\cdots w}(f_k^0)^2 \tag{10.12}$$

$$M_{0i}=M_{i0}=\sum_{k=u\cdots w}b_{ik}f_k^0 \qquad (i=\alpha,\beta,\cdots,\gamma) \tag{10.13}$$

$$M_{ij}=M_{ji}=\sum_{k=u\cdots w}b_{ik}b_{jk} \qquad \begin{pmatrix}i=\alpha,\beta,\cdots,\gamma\\ j=\alpha,\beta,\cdots,\gamma\end{pmatrix} \tag{10.14}$$

In the special case where either i or j, or both, belong to the set of affixes $k=u, v,\cdots,w$ of the basis variables, we have

$$M_{ik}=M_{ki}=b_{ik} \quad \text{if} \quad k=u, v,\cdots,w \tag{10.15}$$

In particular

$$M_{kh}=e_{kh}=\begin{bmatrix}1 & \text{if } k=h\\ 0 & \text{otherwise}\end{bmatrix} \tag{10.16}$$

when k is one of the basis affixes $u, v,\cdots,w$, and h one of the basis affixes $u, v,\cdots,w$

The numerator to the right in (10.11) depends on the point x_k^0, but the denominator does not.

The correctness of (10.9)—(10.10) as a necessary and sufficient condition for the fulfilment of (10.3), is easily recognized by noticing that the *non*-vanishing of the moment matrix (10.9) is the necessary and sufficient condition for the ν vectors $b_{\alpha k},\cdots,b_{\gamma k}$ to be linearly independent over the field k, and that the *actul* vanishing of the moment matrix in the denominator to the right in (10.11) is the necessary and sufficient conditions for the $(\nu+1)$ vectors f_k^0, $b_{\alpha k},\cdots,b_{\gamma k}$ to be linearly dependent over the field k.

The identity of the two expressions for d_0 in (10.11) follows by multiplying (10.3.1) by b_{kj} and performing a summation over k. This shows that if coefficients B_i^0 exist which will make (10.3.1) fulfilled for all k, these coefficients B_i^0 must satisfy theli near system

$$M_{h0}+\sum_{i=\alpha\cdots\gamma}M_{hi}B_i^0=0 \qquad (h=\alpha,\beta,\cdots,\gamma) \tag{10.17}$$

This system is non-singular by virtue of (10.9), and hence

$$B_i^0=-\sum_{h=\alpha\cdots\gamma}M_{ih}^{-1}M_{h0} \qquad (i=\alpha,\beta,\cdots,\gamma) \tag{10.18}$$

By inserting this value of B_i^0 in the middle part of (10.11) we get

$$M_{00}+\sum_{i=\alpha\cdots\gamma}M_{0i}B_i^0=M_{00}-\sum_{i=\alpha\cdots\gamma}\sum_{h=\alpha\cdots\gamma}M_{0i}M_{ih}^{-1}M_{h0} \tag{10.19}$$

The expression to the right in (10.19) is identical with the ratio between the two determinants to the right in (10.11).

The middle expression in (10.11) is most convenient for numerical computation, while the expression to the right has theoretical interest. It shows among other things that d_0 is always non-negative, and that it is zero when and only when an equation of the form (10.3.1) holds.

The coefficients B_i^0 computed for different points x_k^0 and different sets x_i $(i=\alpha, \beta,\cdots,\gamma)$ of bound attained variables, play an important role in the multiplex algorithm, whether it is applied to the case of a linear or to that of a quadratic (in general a

concave) preference function. In the linear case the coefficients B_i^0 are independent of the point x_k^0 and depend only on the set $(\alpha, \beta, \cdots, \gamma)$, while in the non-linear case they will depend also on the point.

For reasons not to be discussed in this connection, the coefficients B_i^0—or in a more explicit notation $B_{i(\alpha \cdots \gamma)}^0$—are termed the *regression coefficients.*

11. *Some remarks on necessary conditions for an optimum*

Obviously (10.1) is not only a part of a sufficient condition, but it is by itself a *necessary* condition for the point x_k^0 to be optimal in the sense specified in the text after (10.4).

The condition (10.2) is in essence not an actual condition but only a *definition* of a set of linearly independent variables that are bound attained in the point x_k^0 considered. This set may be empty, i. e. $\nu=0$, in which case the existence of an equation of the form (10.3.1) simply means that all the first order partial derivatives f_k^0 are zero. Incidentally, this remark shows that in the case of a linear preference function an optimum can never be realized in the *interior* of the admissible region.

The condition (10.3) is—as will appear from the reasoning in Section 19—also necessary, because if d_0 is strictly positive, it is possible to depart from x_k^0 (assumed admissible) in such a way that the preference function is actually increased. This applies regardless of whether the condition (10.4) is fulfilled or not.

On the other hand the condition (10.4) is *not necessary.* It may be that a point x_k^0 is optimal even though (10.4) is not fulfilled. But if so, it will be possible to find *another set* of ν linearly independent variables x_i, such that in the *same* point x_k^0 all the conditions (10.2)−(10.4) are fulfilled. In this particular sense we can say that a condition of the form (10.4) is necessary.

12. *The beam variation*

Suppose that we start in any point x_k^0 admissible or not.

From this starting point x_k^0 we move along a *beam* defined by a set of given direction numbers d_k $(k=u, v, \cdots, w)$. This means that we consider the one dimensional set of points x_k generated by

$$x_k = x_k^0 + \lambda d_k \tag{12.1}$$

where λ increases from zero through positive values.

We will assume that

$$\sum_{k=u \cdots w} f_k^0 \, d_k \gtrless 0 \tag{12.2}$$

This assumption is simply a definition of the *positive direction* along the beam, the parameter λ being, as specified, confined to move from zero through positive values. If (12.2) is not fulfilled, we simply consider the opposite beam, i. e. the beam defined by $(-d_k)$.

Along the beam defined by (12.1) and (12.2) the preference function f will change as the following quadratic function of λ

$$f = f^0 + \lambda \sum_{k=u \cdots w} f_k^0 \, d_k - \frac{1}{2}\lambda^2 \sum_{K=U \cdots W} \sum_{H=U \cdots W} P_{KH} \, d_K \, d_H \tag{12.3}$$

This formula follows simply by inserting (12.1) into (9.6).

The formula (12.3) taken in conjunction with the definition (12.2) of the positive

direction along the beam and with the fact that the matrix P_{KH} is positive definite and non-singular, leads to the four cases for the beam variation of f which are classified in tab. (12.4) and described in the text following the table.

Tab. (12.4). *The four cases of beam variation of the preference function*

	All $d_K=0$ $(K=U,\cdots,W)$	*Not* all $d_K=0$ $(K=U,\cdots,W)$
$\sum_k f^0_k\, d_k=0$	I	II
$\sum_k f^0_k\, d_k>0$	III	IV

Case I. The preference function is *constant* along the beam.

Case II. The preference function *decreases* monotonically and quadratically with λ.

Case III. The preference function *increases* monotonically and linearly with λ.

Case IV. The preference function will first increase monotonically up to a well defined *maximum*, and then decrease monotonically. In other words, for a sufficiently small λ f is certainly larger than f^0.

In Case IV the value λ_0 for which the maximum occurs, is given by

$$\lambda_0=\left.\frac{\sum_k f^0_k\, d_k}{\sum_K\sum_H P_{KH}\, d_K\, d_H}\right| \quad \text{At least one } d_K\neq 0 \tag{12.5}$$

If we want to distinguish between lineo and quadro variables in the numerator of (12.5), we can write the formula

$$\lambda_0=\left.\frac{\sum_\kappa p_\kappa d_\kappa+\sum_K f^0_K\, d_K}{\sum_K\sum_H P_{KH}\, d_K\, d_H}\right| \quad \text{At least one } d_K\neq 0 \tag{12.6}$$

The formula (12.5)—or, if we like, (12.6)—can be interpreted as characterizing *all the four cases.* In Case I the ratio (12.4) is indeterminate, i. e. it indicates that the maximum occurs "everywhere along the beam". In Case II the ratio is zero, indicating that the maximum has occurred already at $\lambda=0$, In Case III the ratio is plus infinity, indicating that the maximum occurs in the positive direction but not at a finite distance. And in Case IV the ratio indicates a finite and well determined positive value of λ for which the maximum occurs.

For practical purposes it is, however, only the two cases where $\sum_K\sum_H P_{KH}\, d_K\, d_H$ is strictly positive—i. e. where at least one $d_K\neq 0$—we need to take account of. The Case III where this double sum is exactly zero while $\sum_k f^0_k\, d_k$ is strictly positive need bother us, because in this case there is no harm in continuing the beam variation (in the positive direction) as far as is possible. And the Case I where the double

sum is exactly zero and $\sum_k f_k^0\, d_k=0$ will be taken care of in another way.

Indeed, if we insert in $\sum_k f_k^0\, d_k$, $d_k=f_k^0+\sum_{i=\alpha\cdots\gamma} B_i\, b_{ik}$, we see that $\sum_k f_k^0\, d_k$ emerges as the same magnitude d_0 that was considered in (10.11), and, as will appear from the sequel, this magnitude occurs in the algorithm in such a way that Case I is covered. We therefore conclude that we only need to compute (12.5)—or, if we like, (12.6)—in the case where at least one $d_K \neq 0$.

The above remarks are formulated without regard to the *admissibility* of the points generated along the beam. Now let us consider admissibility.

Let us assume that the initial point x_k^0 is admissible, i. e. it is situated either in the interior of the admissible region or on its boundary. The latter case is the one where at least one of the N variables x_j $(j=1, 2,\cdots,N)$ is bound attained in the initial point x_k^0.

There will then exist a certain non-negative λ range for which the point x_k generated by (12. 1) *remains* in the admissible region. In extreme cases the range may consist only of the single volue $\lambda=0$ (namely if the initial point x_k^0 is on the boun dary and the positive direction of the beam points out of the admissible region), but in other cases there will exist a finite (positive) λ range for which the point that moves along the beam remains admissible.

The exact value λ_{ex} $(\gtrless 0)$ for which the point breaks out of the admissible region (ex=exodus) is determined by

$$\lambda_{ex}=\operatorname*{Min}_j \lambda_j \;\Big|\; d_j \neq 0 \tag{12.7}$$

where

$$\lambda_j=\begin{cases} \dfrac{\bar{x}_j-x_j^0}{d_j} & \Big| \; d_j>0 \\ \dfrac{x_j^0-x_j}{d_j} & \Big| \; d_j<0 \end{cases} \tag{12.8}$$

$$c_j=\sum_{k=u\cdots w} b_{jk}\, d_k \qquad (j=1, 2,\cdots,N) \tag{12.9}$$

If j is one of the basis affixes, the coefficients b_{jk} are interpreted by (1.2)

The values (12.8) are only defined for those affixes j for which the general direction numbers (12.9) are different from zero.

The formula (12.7) is easily derived by noticing that for any j we have along the beam

$$x_j=x_j^0+\lambda d_j \qquad (j=1, 2,\cdots,N) \tag{12.10}$$

where d_j is given by (12.9). Writing down the condition that all the expressions to the right in (12.10) must be situated in the interval (1.3), we get (12.7).

If we add the condition that we shall not move so far along the beam that we go *beyond* the maximum of the preference function, we get the following formula for the value of λ where we should stop

$$\lambda_{stop}=\operatorname{Min}[\lambda_0,\ \lambda_{ex}] \tag{12.11}$$

This formula, together with (12.6) and (12.7) determines the move to be made

when any beam d_k—with positive direction (12.2)—is given from any admissible point x_k^0.

The increase in the preference function achieved during the move is obtained by inserting λ_{stop} for λ in (12.3). This increase is always non-negative.

12. *The equationally constrained beam direction. Regression analysis.*

Suppose that we start in any point x_k^0 and want to proceed along a beam d_k which is such that during the move along this beam a certain number of the variables —say x_t where t runs through a given set of affixes—*remain constant.* This set of variables we call *the constancy set.* The expression for the constancy requirement considered is

$$b_{t0}+\sum_k b_{tk}(x_k^0+\lambda d_k)=x_t^0 \qquad \text{for all } \lambda \text{ and for } t \text{ in the constancy set} \tag{13.1}$$

This condition reduces to

$$\lambda\sum_k b_{tk}\, d_k=0 \tag{13.2}$$

Since this equation is to hold for all λ—and conseqently also for some non-zero λ—the condition is equlvalent to

$$\sum_k b_{tk}\, d_k=0 \qquad t \text{ in the constancy set} \tag{13.3}$$

This can also be written

$$d_t=0 \qquad t \text{ in the constancy set} \tag{13.4}$$

where the s_t are defined by (12.9).

If not all the variables x_t are linearly independent—which is the same as to say that not all the vectors b_{tk} are linearly independent over the field k—it is not necessary to take account of all the conditions (13.3). It suffices to pick in the set t a highest linearly independent subset $i=\alpha, \beta,\cdots,\gamma$, i. e. a subset of the highest order ν (ν being the number of affixes $\alpha, \beta,\cdots,\gamma$) that is linearly independent. If there are more than one highest independent subset, any such subset chosen at random amongst the highest ones, will do in principle. But for computational reasons it is desirable to chose a subset which is "as far as possible" from being linearly dependent. This may, if necessary, be tested by a similar procedure as that discussed in Section 8.

The highest linearly independent subset chosen will be called the *operation set.* For the operation set we impose the conditions

$$\sum_k b_{ik}\, d_k=0 \qquad (i=\alpha, \beta,\cdots,\gamma) \tag{13.5}$$

which is the same as

$$d_i=0 \tag{13.6}$$

When the conditions (13.5) are imposed, all the conditions (13.3) follow automatically.

The number of conditions (13.5) can at most be equal to n, the number of basis variables, because n is the largest number of vectors b_{ik} that can be linearly independent over the field k. In the case $\nu=n$ *all* d_k *must be zero* because b_{ik} in (13.5) will

then be a square and non-singular matrix. We can interpret this as expressing the order: "Dont't proceed any further."

If ν n, the conditions (13.5) leave a certain amount of freedom in the d_k and the question arises of the perticular way in which we should dispose of these degrees of freedom.

From the discussion in Section 10 it emerges that an essential point in approaching an optimum will be to try to express the preference vector f_k^0 linearly in terms of a set of linearly independent boundary vectors b_{ik} $(i=\alpha,\ \beta,\cdots,\gamma)$. These considerations can be formalized in different ways, for instance by saying that we want to move along the projection of the vector f_k^0 on the linear manifold defined by the conditions (13.5), or to say that we want to increase as much as possible the correlation between f_k^0 and d_k over the field k subject to the conditions (13.5), or in other ways which will not be further discussed in the present connection. In whatever way we formalize this approach, we are led to assume that the direction numbers d_k should be of the form

$$d_k=f_k^0+\sum_{i=\alpha\cdots\gamma} B_i^0\, b_{ik} \qquad (k=u,\ v,\cdots,w) \tag{13.7}$$

where the B_i^0 are constants independent of k. This leaves the exact number of degrees of freedom which are needed to comply with the conditions (13.5). The problem is then only to determine the B_i^0.

Multiplying (13.7) by b_{hk} $(h=\alpha,\ \beta,\cdots,\gamma)$ and performing a summation over k, utilizing (13.5), we are led to the same equations for determining the B_i^0 as (10.17). If the B_i^0 are determined in this way, we find as a check that (13.6) is fulfilled. This is the particular way of satisfying (13.5) which we will adopt.

When the B_i^0 are determind in this way, and the $d_k(k=u,\ v,\cdots,w)$ are computed by (13.7), the other d_j may be computed by any of the two formulae (12.9) or

$$d_j=M_{j0}+\sum_{i=\alpha,\cdots\gamma} M_{ji}\, B_i^0 \qquad (j=1,\ 2,\cdots,N) \tag{13.8}$$

Computationally it will as a rule be simplest to use (12.9).

When the d_k are determined, the magnitude $\sum_k f_k^0\, d_k$—which plays an important role, as is seen from the discussion in Sections 10 and 12—will be equal to

$$\sum_{k=u\cdots w} f_k^0\ \mathrm{d}_k=d_0=M_{00}+\sum_{i=\alpha\cdots\gamma} M_{0i}\, B_i^0 \tag{13.9}$$

In determinant form d_0 can also be expressed as in (10.11). This shows that d_0—and hence $\sum_k f_k^0\, d_k$—is strictly positive except when the criterion (13.3.1) is exactly fulfilled. And in this case d_0 is zero. Compare (10.10).

In the optimality criterion (10.4) we considered the sign correctness of the B_i^0. This is important for subsequent applications. The determination of the B_i^0 and the scrutiny of their sign correctness may be called *regression analysis*. So far as the mere fulfilment of (13.5) is concerned, the sign correctness of the B_i^0 is, however, irrelevant.

If we are in a point where (10.3.1) is exactly fulfilled, all the d_k $(k=u,\ v,\cdots,w)$ as determined by (13.7) are zero. That is to say the order: "Don't proceed any further",

will emerge not only in the special case where $\nu=n$ (in which case (10.3.1) must necessarily be fulfilled whatever the vector f_k^0, because any vector in the space $k=u, v,\cdots,w$ can be expressed linearly in terms of n linearly independent vectors in this space)—but the order will emerge in *any point* x_k^0 where the preference vector f_k^0 happens to be linearly expressible in terms of a certain number of boundary vectors $b_{\alpha k}$, $b_{\beta k},\cdots,b_{\gamma k}$.

Inversely: If all $d_k=0$ $(k=u, v,\cdots,w)$, when computed by (13.7), the preference vector f_k^0 is linearly expressible in terms of the ν boundary vectors $b_{\alpha k}$, $b_{\beta k},\cdots,b_{\gamma k}$ in the point considered.

Computational Note. In all cases where the B_i^0 are computed by (10.17), the d_k by (13.7) and the other d_j by (12.9), it should be *carefully checked* that all the d_i $(i=\alpha, \beta,\cdots,\gamma)$ are zero within rounding errors.

14. *The admissible beam direction*

In the preceding section we considered a beam variation which was subject to the condition that a certain number of the variables remained exactly constant. We will now consider a different type of condition: We require that some of the variables shall *not be increased* and we require that some other varibales shall *not be decreased.* In practice the former condition will be useful for varibales that are at their upper bound in the point that serves as starting point for the beam variation, while the latter condition is useful for variables that are at their lower bound. But in principle the conditions of non-increase or non-decrease can be imposed for any set of variables.

Conditions of the kind now considered will be termed *admissibility conditions.* They must not be confounded with the conditions of sign correctness for the regression coefficients discunded in (10.4)

As essential difference between the constancy conditions discussed in Section 13 and the admissibility conditions discussed now, is that the former take out degrees of freedom, while the latter do not take out degrees of freedom (except in limiting cases).

Therefore admissibility conditions may be considered for *any number of variables,* less than, equal to or larger than n, the number of basis variables. For instance, when we speak of (linearly independent) constancy conditions, a beam variation is made impossible, i. e. all d_k will be zero if the number of these conditions is equal to n, while n (linearly independent) admissibility conditions need not make a beam variation impossible. Nor does a larger number of admissibility conditions need make the beam variation impossible.

For simplicity of formulation we will in the sequel speak of the variables for which we require non-increase, as those that have reached their upper bound, and the variables for which we require non-decrease will be spoken of as those that have reached their lower bound. But it should be remembered that in principle the admissibility criterion may be imposed for any set of variables.

With this convention we can formulate the admissibility condition as follows:

$$(14.1)\qquad d_j=\begin{cases} \textit{non-positive} & \text{for the variables that have reached their upper bound in the starting point for the beam variation, i. e. } x_j^0=\bar{x}_j. \\ \textit{non-negative} & \text{for the variables the have reached their lower bound in the starting point for the beam variation, i. e. } x_j^0=\underline{x}_j. \end{cases}$$

A variable that satisfies (14.1) will be said to have an *admissible* direction number. Otherwise it will be said to have an *inadmissible* direction number.

By this definition any variable that is *not* bound attained has an admissible direction number. Hence the scrutiny for admissibility need only be made for the bound attained variables.

If we want to express the admissibility requirement in terms of the basis direction numbers $d_k (k=u, v,\cdots,w)$, we only have to substitute $\sum_k b_{jk} d_k$ for d_j in (14.1).

The condition (14.1) can be considered regardless of how the basiis direction numbers $d_k (k=u, v,\cdots,w)$ have been determined. And in particular this condition can be considered in the case where the d_k are determined by (13.7) where the B_i^0 are determined by (10.17).

In this latter case the basis direction numbers are uniquely determined and the extra condition $\sum_k f_k^0 d_k>0$ holds, except if $d_0=0$. From the programming viewpoint it is very natural to supplement the condition (14.1) by the condition $\sum_k f_k^0 d_k>0$. Indeed, we know from (9.11)—(9.13) that $\sum_k f_k^0 d_k>0$ is a necessary condition for the preference function to increase along the beam. And from IV in Section 12 we know that this condition is also a sufficient condition for the preference function to increase, provided λ is small enough.

The fact that the supplementary condition $\sum_k f_k^0 d_k>0$ is fulfilled automatically if the basis direction numbers d_k $(k=u, v,\cdots,w)$ are determined by (13.7) where the B_i^0 are determined by (10.17), applies regardless of wheter the set $i=\alpha, \beta,\cdots,\gamma$ is an empty set or not and if the set is not empty it applies regardless of whether the regression coefficients B_i^0 are sign correct or not. The only exception is the extreme case $d_0=0$, which in the algorithm will be characteristic for the case where an optimum is reached. From now on we will assume that the supplementary condition $\sum_k f_k^0 d_k>0$ is implied in the definition of admissibility. This will be taken care of automatically because we confine our attention to the case where the d_k are determined by (13.7), and $d_0>0$.

This being so, we note that admissibility is a concept that depends on the operation set considered. For precision we may therefore speak of *admissibility under the operation set* $i=\alpha, \beta,\cdots,\gamma$.

15. *The selection of the operation set*

Suppose that we are in any admissible point x_k^0 where a certain number (perhaps none) of the variables are bound attained.

A *preference move* from this point is a move along any beam d_k which is determined by (13.7) and (10.17) and is such that all direction numbers $d_j (j=1, 2,\cdots,N)$ are admissible.

The move will be termed a preference move regardless of whether the regression

coefficients B_i^0 are sign correct or not. If the regression coefficients are sign correct, we will speak of a sign correct preference move. If not all regression coefficients are sign correct, we will speak of a sign incorrect preference move. If nothing is specified with regard to sign correctness, we will simply speak of a preference move.

It seems plausible to make the move under the *smallest possible* number of constraints in the form of constancy conditions. Frequently it will be possible to make an admissible move without including in the operation set a list of (linearly independent) variables which is large enough to keep all bound attained variables constant. This is the case where we can and should proceed by *loosing contact* with one or more of the bounds which have temporarily been reached.

The logical consequence of this consideration is that we should make the further move under the smallest possible operation set. A variable should be included in the operation set only if this appears necessary to assure complete admissibility of the beam variation.

This leads to the following method of *progressive inclusion* in order to define the content of the operation set in any given point x_k^0.

We first attempt a *free* preference move from this point, that is, a move with no equational constraints. In conformity with (13.7) it is definned by

$$d_k = f_k^0 \tag{15.1}$$

If this leads to no inadmissible direction number d_j, this is accepted as the move to make.

If one or more inadmissible direction numbers occur, one should look for "the most inadmissible one". It may be defined simply as the direction number d_j which has the largest absolute value amongst the inadmissible ones. To assure comparability of units of measurement one may normalize d_j by considering

$$\frac{d_j}{\left|\sqrt{\left(\sum_k b_{jk}^2\right)\left(\sum_k d_k^2\right)}\right|} = \frac{\sum_k b_{jk}\, d_k}{\left|\sqrt{\left(\sum_k b_{jk}^2\right)\left(\sum_k d_k^2\right)}\right|} \tag{15.2}$$

This is the correlation coefficient between the boundary vector b_{jk} and the beam vector d_k over the field k. Geometrically it is the same as the *cosine* of the angle between these two vectors.

To choose j so as to make (15.2) as *large* as possible in absolute value (i. e. as close as possible to $+1$ or -1 as the case may be) is the same thing as to pick that one amongst the inadmissible direction numbers which is such that the boundary vector for this variable forms the smallest possible angle with the beam direction. If more than one j amongst the inadmissible ones give the maximum of (15.2), we may choose one of these j at random.

The variable thus selected should then be *added* to the operation set.

In this way we may continue and add variables one by one until we reach an operation set that makes all the direction numbers d_j admissible.

The solution of the equation for the B_i^0 which are needed in this process, will not be too costly because one will add one row—and the corresponding column—at a time so that all previous work can be utilized when a new variable is added. It will in each step be found simpler to make a one way solution than to compute a full inversion.

As an extra precaution one may already in this building up of the operation set drop any variable that turns out to give a sign incorrect regression coefficient, but this occurrence will not be frequent if the additions are made by minimizing (15.2) for each new variable to be included.

The regression coefficient for a new variable that is added because it has an inadmissible direction number, must always turn out as sign correct, but some of the previously included variables may have their regression coefficients changed from sign correct to sign incorrect or vice versa.

Computationally it is not needed for the comparison in (15.2) to divide by the factor $\left|\sqrt{\sum_k \left(d_k^2\right)}\right|$ since this factor is independent of j.

We could, of course, start by normalizing all variables in such a way that by definition

$$\sum_k b_{jk}^2 = 1 \qquad \text{for all } j, \tag{15.3}$$

but this would involve some extra work which could profitably be saved, at least when the computations are made on desk machines. The sumsquare $\sum_k b_{jk}^2$ would then only be computed for those j for which it turns out to be necessary in the course of the algorithm.

The above is a perfectly general procedure—and in a sense a very effective procedure—which may be applied in any given point regardless of the way in which this point has been reached. But the procedure does involve a computational cost which is not negligeable.

It will involve less cost in each step to proceed from point to point in such a way that one only includes one variable or exclude one variable according to the criteria of Sections 17 and 18 below. Whether this simplified procedure or a complete application of the method of progressive inclusion in each point will in the long run be the most economical, can only be decided by extensive numerical tests.

[*To be continued.*]

SUMMARY

Linear and quadratic programming problems may be approached in many different ways. I feel that in quadratic progromming—or in programming with more general forms of a preference function—the only practical procedure is to work through a series of movements in n-space, being guided by some sort of steepest gradient principle. I shall not attempt to survey all attempts that have been made in this direction, but concentrate on the use of the multiplex method.

I was originally lead to develop this method for the case of a linear preference function in an attempt to cover the case where many of the variables may be linearly dependent either exactly or nearly. For the linear case the multiplex method has been successfully coded by Mr. Ole-Johan Dahl and applied on the good-sized electronic computer of the Norwegian Defence Research Organization. And it seems that the method will be equally applicable to the quadratic case.

The purpose of my approach to the quadratic case can be briefly stated as follows:

First, I am not interested in determining local maxima. I need a method that will yield a point which is a global maximums in the sense that no other admissible point exists which will yield a strictly higher value of the preference function.

Second, I need a method which can work fairly quickly and without the complications caused by a non-convex admissible region or by a preference function which is such that one needs to go through a complicated discussion in order to distinguish between a maximum and a minimum.

Within the desiderata as thus specified I want to consider the most general case that is variable. This is the case where the bounds are linear and the preference function of a mixed linear and quadratic character, the quadratic term being minus a positive definite and non singular quadratic form. This mixed case is essentially the same as the case where there are no strictly linear terms but the quadratic form is singular. For a great number of applications, and certainly in many macro-economic problems it is absolutely necessary to cover this case. In principle the method can be applied to the case of any *concave* preference function but in this paper only the mixed linear and quadratic case is considered.

Section 1 of the paper is essential because it contains definitions and symbols. This also applies to (2.15) where the nature of the preference function and the distinction between lineo and quadrato variables are discussed. A lineo variable is one that only enters linearly, while a quadrato variable is one that also enters in the non-singular quadratic form.

Most of sections 2–8 contains work on clearing the ground before a discussion of the method itself can start. These sections are partly concerned with a statement of some classical algebraic facts and partly with a discussion of some practical computational techniques. I shall not dwell on these parts of the paper except perhaps for drawing attention to section 4 which contains a practical worksheet for building up the inverse of a symmetric matrix for which the inverse of one of its principal minors is known. Also I would like to draw attention to the practical technique of determining the rank of any rectangular matrix, which is given in section 8.

The sssential part of the programming discussion starts in sections 9 and 10. Section 10 gives a cemplete formulation of a sufficient criterion for optimum. The criterion is built on the four conditions (10.1)—(10.4). In section 11 mention is made of the particular sense in which these conditions are also necessary.

Section 12 gives a discussion of what will happen if we start in any admissible point and move along a beam from this point. Compare (12.1). The positive direction along the beam is defined by (12.2), which gives the direction in which the preference function is increasing (or at least not decreasing). Along such a beam the movement should continue until we either break out of the admissible region, or reach a maximum point along the beam. Only one such maximum can exist. The case of a maximum along the beam may occur in the quadratic case but not in the linear. An exact criterion for where to stop is given by (12.7)—(12.11). The increase of the preference function, measured per unit of the move, is denoted d_0 and given by (10.11).

In section 13 I discuss how we can define a beam variation that satisfies certain linearly independent equational constraints. This sort of variation is adopted to the case where a certain number of linearly independent bounds are reached and we

want to proceed in such a way that these bounds continue to be reached. Amongst all the directions that lie in this manifold of the boundary, that particular one is chosen along which the increase in the preference function is the steepest. This leads to the direction numbers (13.7) where the regression coefficients B_i^0 are given by (10.17). The set of linearly dependent variables which are thus kept constant, is called the *operation* set. Compare the text after (13.4).

As stated in (10.4) a regression coefficient B_i^0 is called sign correct if it is non-negative when the variable X_i is at its lower bound, but non-positive if X_i is at its upper bound.

In section 14 is discussed the admissibility of the beam itself. The admissibility of the direction number d_j is defined by (14.1). The admissibility discussion is made in order to eliminate beam directions along which no movement would be possible accoading to the rules of section 12. The admissibility criterion is a particularly useful device for handling the case of linear dependencies. Since the direction numbers depend on the regression coefficients, and these coefficients in turn depend on the operation set, the concept of admissibility refers to a given operation set. This is a help in choosing the operation set.

* * *

On the basis of the above discussion it is possible to indicate an algorithm which is simple enough to be applicable to desk machine computation as well as to automatic computation, and which will lead to an optimum point in a finite number of steps. Since time did not permit to include a description of the algorithm in this paper a brief summary will be given here.

Start in any point in the admissible region, not necessarily a vertex of the boundary. Different metheds are available for finding such a point. Frequently an admissible point is known a priori. Move from point to point by computing in each point the preference gradient f_k^0 as given by (9.3) and decide on the next step according to the following rule, that gives a mutually exclusive and collectively exhaustive classification of all cases.

Case I. Not all B_i^0 sign correct: Drop from the operation set all variables with sign incorrect B_i^0.

Case II. All B_i^0 sign correct, $d_0 > 0$ and not all d_j admissible: Add to the operation set the variable with the most inadmissible direction number according to (15.2).

Case III. All B_i^0 sign correct, $d_0 > 0$ and all d_j admissible: Make a preference move, i. e. proceed along the beam and stop at the point defined in section 13.

Case IV. All B_i^0 sign correct and $d=0$: An optimum is reached. Stop and give output.

If the above rules lead to a situation where a succession of rounds (perhaps two or three) indicate a preference move *without change in the operation set*, and if this stationary operation set contains a subset that is lineo-independent, i. e. linearly independent over the field of the lineo variables: Replace the determination of the direction numbers by means of the regression coefficients by a direction determination that leads directly towards the maximum point under the constraints defined by the operation set that has emerged as stationary. One and only one such point will now exist. The relevant formulae are given by the non-lineo part S_{K_0} of (6.18), X_k^{des} being given by (3.2). The maximum point considered can also be determined by another

and more direct formula. We could have continued according to the cases I—IV without shifting to this new procedure for determing the direction of the move, but this would have slower up convergency (and may indeed have lead to an infinite number of steps). Along the beam now chosen we move according to the rules of section 12. If a new bound is encountered in the course of this move, we continue according to the cases I—IV.

A final remark on machine capacity may be added. The equations (10.17) need only be considered explicitely for those variables in the operation set that are *dependent* variables. For the basis variables that belong to the operatian set, the regression coefficients can be written down at sight. If the maximum number of effective equations (10.17) which the machine can handle, is reached, a *basis clearing* can be made. This means that a number of the basis variables that are at this stage *not* included in the operation set, are removed from the set of basis variables and instead the same number of the dependent variables from the operation set introduced as new basis variables. This only involves an inversion of the order of the number of variables shifted. This being done the work can continue as before. If all the variables in the effective equation (10.17) where removed from the basis set, there will be no such equation in the next round and we are prepared for possible new inclusions of independent avriables into the operation until the machine capacity may be reached anew.

In this way we may continue and handle a virtually unlimited number of degrees of freedom and of dependent variables.

The small numerical example appended to the paper was constructed by random drawing of the coefficients It contains 12 variables and 5 degrees of freedom, of which 2 are represented by lineo variables and 3 by quadrato variables. An optimum point was reached after 5 rounds. The successive values of the preference function are indicated in the first cell in the bottom row of each round. The solution was worked out on a desk machine by a semitrained computor in about 7 hours.

The round designated No. 105 is only an illustration of what would have happened if we had not stopped before the breaking out point on the beam that leads from point 4. The round designated No. 6 was only made to improve on the rounding errors so as to get a more perfect fulfilment of the optimum criteria. It will be seen that from round 5 to round 6 the preference function does not change in the digits recorded, but the optimum criteria become fulfilled with greater accuracy.

Tab. (26.1). *Data for numerical example for quadratic programming by the multiplex method.*

	Constant term	$k=1$	2	3	4	5	Check sum	Lower bound	Upper bound	Initial point
$j=6$	−0.6641	0.7840	0.4764	−0.6293	−0.4011	0.7274	0.2933	−1	1	−0.6641
7	0.9524	−0.2835	0.1808	0.1634	0.3996	0.2404	1.6531	−1	1	0.9524
8	−0.7483	−0.1579	−0.6473	0.9296	0.7152	0.8516	0.9429	−1	1	−0.7483
9	0.5246	−0.3513	−0.4113	0.3760	−0.8985	−0.0221	−0.7826	−1	1	0.5246
10	−0.9143	−0.6891	−0.4124	0.9694	−0.2783	−0.1825	−0.6824	−1	1	−0.9143
11	−0.2859	−0.5111	−0.3709	0.9210	−0.4931	−0.2545	−0.9945	−1	1	−0.2859
12	0.2466	−0.2207	−0.7125	−0.3571	−0.3513	0.4676	−0.9274	−1	1	0.2466
Check sum	−0.8890	−1.4296	−1.0724	2.3730	−1.3075	1.8279	−0.4976			
Lower bound	−1	−1	−1	−1	−1	−1				
Upper bound	1	1	1	1	1	1				
Preference coefficient $p_k=$	0	−0.2223	−0.7723	0.9598	−0.4519	−0.6993				
Preference coefficient P_{KH} $K=3$				2.09577403	0.93063312	1.46436083	4.49076798			
Preference coefficient P_{KH} 4					0.65236979	0.63824197	2.22124488			
Preference coefficient P_{KH} 5						1.65461973	3.75722253			
Initial point, (admissible)	0	0	0	0	0	0	10.46923539			

Tab. (26.2). *Round chart for numerical example*

Variable No. $j=$	Round 0			
	Initial point x_j^0	At bound	Direction numbers d_j^0	Ranking parameter λ_j^0
1	0		−0.22230000	4.49842555
2	0		−0.77230000	1.29483361
3	0		0	∞
4	0		0	∞
5	0		0	∞
6	−0.6641		−0.54220692	0.61950519
7	0.9524		−0.07660979	25.48499349
8	−0.7483		0.53501096	3.26778352
9	0.5246		0.39574098	1.20129080
10	−0.9143		−0.16530959	0.51842122
11	−0.2859		0.40006360	3.21423894
12	0.2466		0.59932536	1.25708013
Check sum	−0.8890		0.15141460	
0	0		0.64586458	∞

Variable No. $j=$	($\lambda=0.51842122$) Round 1			
	New point x_j^1	At bound	Direction numbers d_j^1	Ranking parameter λ_j^1
1	−0.11524504		−0.39893062	2.21781662
2	−0.40037671		−0.66659333	0.89953389
3	6		0	∞
4	0		0	∞
5	0		0	∞
6	−0.94519157		−0.63032667	0.08695242
7	0.91268386		−0.00742326	257.66090100
8	−0.47093897		0.49447701	2.97473682
9	0.72976052		0.41431416	0.65225741
10	−1.00000000	V	0	
11	−0.07849854		0.45113291	2.39064479
12	0.55730298		0.56299174	0.78632951
Check sum	−0.81050346		0.21964194	
0	0.33482991		0.60349231	∞

Variable No. $j=$	($\lambda=0.08695242$) New point x_j^2	At bound	Round 2 Direction numbers d_j^2	Ranking parameter λ_j^2
1	−0.14993302		5.59728436	0.20544481
2	−0.45833861		−6.33954343	0.08544171
3	0		4.51156351	0.22165265
4	0		−6.51057587	0.15359624
5	0		1.56774472	0.63785895
6	−1.00000000	∨	0	
7	0.91203839		−4.97434204	0.38438016
8	−0.42794300		1.42226943	1.00398910
9	0.76578614		8.22187567	0.02848667
10	−1.00000000	∨	0	
11	−0.03927644		7.25507061	0.14324760
12	0.60625647		3.22461258	0.12210569
Check sum	−0.79140507		10.84047010	
0	0.38730502		12.02040485	1.00000003

Variable No. $j=$	($\lambda=0.02848667$) New point x_j^3	At bound	Round 3 Direction numbers d_j^3	Ranking parameter λ_j^3
1	0.00951497		3.18661241	0.31082695
2	−0.63893109		−0.16047886	2.24994688
3	0.12851942		1.88038934	0.46345752
4	−0.18546463		−0.33908113	2.40218431
5	−0.04465983		−1.88964590	0.50556571
6	−1.00000000	∨	0	
7	0.77033595		−1.21493129	1.45714903
8	−0.38742728		−0.50301148	1.21781062
9	1.00000000	∧	0	
10	−1.00000000	∨	0	
11	0.16740136		0.81079838	1.02688740
12	0.69811494		−2.02491043	0.83861237
Check sum	−0.48259619		−0.25425896	
0	0.72484909		2.61823977	1.00000010

Variable No. $j=$	($\lambda=0.31082695$) New point x^4_j	At bound	Round 4 Direction numbers d^4_j	Ranking parameter λ^4_j
1	+1.00000000	∧	0	
2	−0.68881224		0.00218219	773.907056
3	0.71299510		−0.00229419	746.666623
4	−0.29086018		−0.00184993	383.333326
5	−0.63201270		−0.00443407	82.990863
6	−1.00000000	∨	0	
7	0.39270256		−0.00178553	779.993928
8	−0.54377680		−0.00864435	52.777039
9	1.00000000	∧	0	
10	−1.00000000	∨	0	
11	0.41941935		−0.00088167	1609.921342
12	0.06871821		−0.00215905	494.994655
Check sum	−0.56162670		−0.01986660	
0	1.41219008		0.00003312	0.35247343

Variable No. $j=$	($\lambda=0.35247343$) New point x^5_j	At bound	Round 5 Direction numbers d^5_j
1	1.00000000	∧	0
2	−0.68804309		−0.00000059
3	0.71218646		0.00000064
4	−0.29151223		0.00000051
5	−0.63357559		0.00000118
6	−0.00000000	∨	−0.00000004
7	0.39207321		0.00000049
8	−0.54682370		0.00000235
9	1.00000000	∧	0
10	−1.00000000	∨	0.00000002
11	0.41910858		0.00000026
12	0.06795720		0.00000056
Check sum	−0.56862915		0.00000364
0	1.41219592		0.00000002

Variable No. $j=$	($\lambda=0.35237781$) New point x_j^6	At bound	Round 6 Direction numbers d_j^6	Ranking parameter λ_j^6
1	+1.00000000	∧	0	
2	−0.68804328		0.00000009	
3	0.71218668		−0.00000004	
4	−0.29151205		−0.00000006	
5	−0.63357517		0.00000010	
6	−1.00000000	∨		
7	0.39207338			
8	−0.54682288			
9	1.00000000	∧		
10	−1.00000000	∨		
11	0.41910867			
12	0.06795741			
Check sum	−0.56862724			
0	0.56862724		0	

Variable No. $j=$	($\lambda=52.777039$) New point x_j^{105}	Round 105 At bound
1	1.00000000	
2	−0.57364271	
3	0.59191454	
4	−0.38849401	
5	−0.86602979	
6	−1.00000000	
7	0.29846757	
8	−1.00000000	
9	1.00000000	
10	−1.00000000	
11	0.37288742	
12	−0.04523006	
Check sum	−1.61012704	
0	1.28308265	