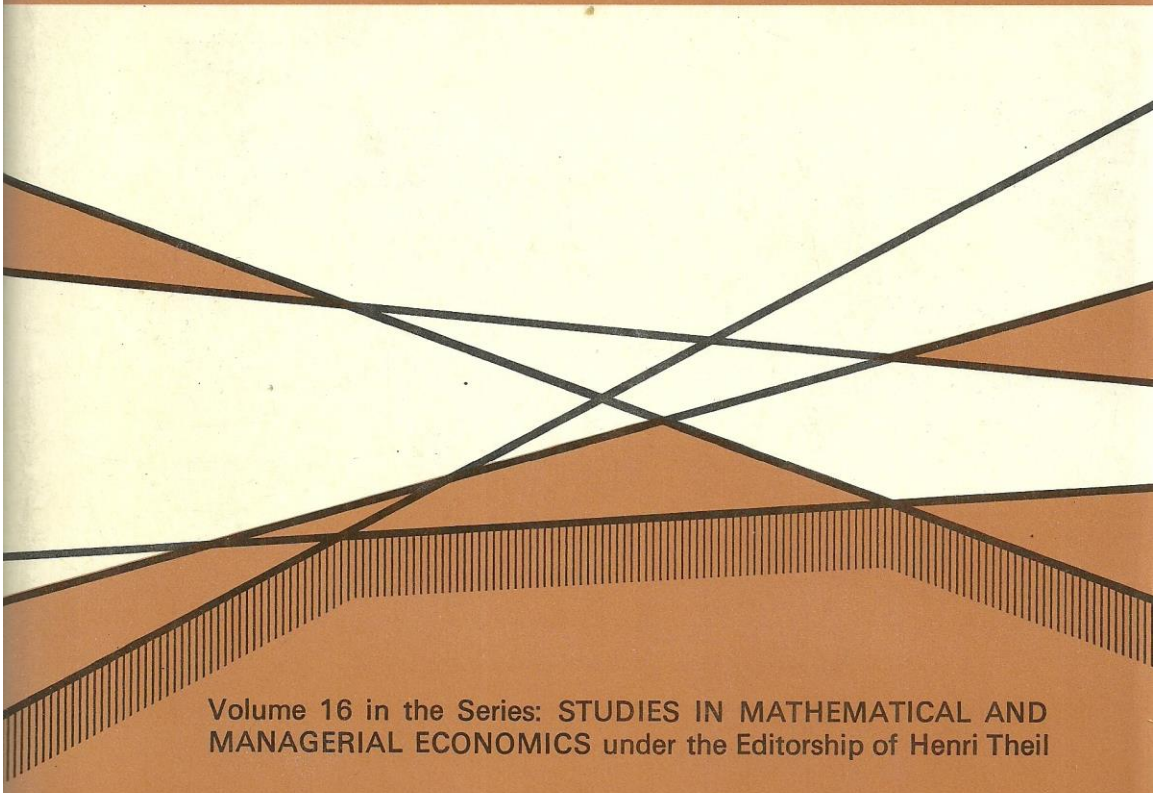
Classical Optimization: Foundations and Extensions

MICHAEL J. PANIK



Volume 16 in the Series: STUDIES IN MATHEMATICAL AND MANAGERIAL ECONOMICS under the Editorship of Henri Theil

North-Holland/American Elsevier

CLASSICAL OPTIMIZATION: FOUNDATIONS AND EXTENSIONS

MICHAEL J. PANIK
Associate Professor of Economics, Austin Dunham Barney School of Business and Public Administration University of Hartford

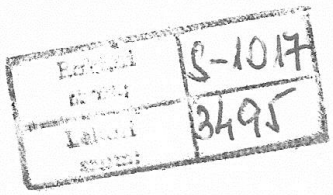
1976

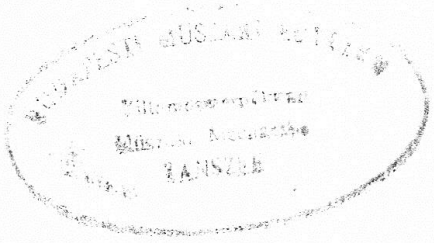
NORTH-HOLLAND PUBLISHING COMPANY–AMSTERDAM · OXFORD
AMERICAN ELSEVIER PUBLISHING COMPANY, INC.–NEW YORK

Library of Congress Catalog Card Number 73–79103
ISBN North-Holland for this series 0 7204 3300 2
ISBN North-Holland for this volume 0 7204 3315 0
ISBN American Elsevier for this volume 0 444 10568 9

PUBLISHERS:
NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM
NORTH-HOLLAND PUBLISHING COMPANY, LTD. – OXFORD

SOLE DISTRIBUTORS FOR THE U.S.A. AND CANADA:
AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE
NEW YORK, N.Y. 10017





Printed in Singapore

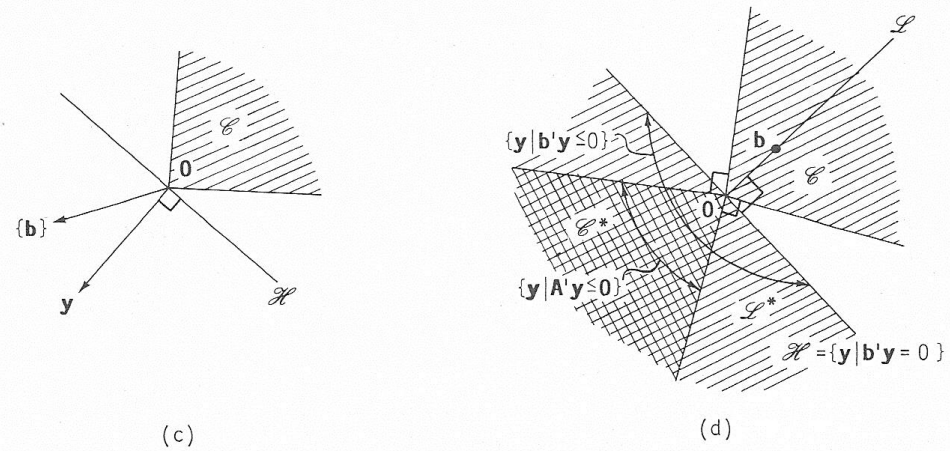


Fig. 1.10.

Our next theorem of the alternative, which is equivalent to the theorem of the separating hyperplane, is the

MINKOWSKI–FARKAS THEOREM. *If (II) above does not possess a solution, then a necessary and sufficient condition for the m-component vector* $\boldsymbol{b}$ *to lie within the finite cone spanned by the columns of the* $(m \times n)$ *matrix* $\boldsymbol{A}$ *is that* $\boldsymbol{b}'\boldsymbol{y} \leqslant 0$ *for all* $\boldsymbol{y}$ *satisfying* $\boldsymbol{A}'\boldsymbol{y} \leqslant \boldsymbol{0}$, *i.e. there exists an n-component vector* $\boldsymbol{\lambda} \geqslant \boldsymbol{0}$ *such that* $\boldsymbol{A\lambda} = \boldsymbol{b}$ *if and only if* $\boldsymbol{b}'\boldsymbol{y} \leqslant 0$ *for all* $\boldsymbol{y}$ *satisfying* $\boldsymbol{A}'\boldsymbol{y} \leqslant \boldsymbol{0}$.

Proof (sufficiency). If $\boldsymbol{A\lambda} = \boldsymbol{b}$, $\boldsymbol{\lambda} \geqslant \boldsymbol{0}$, then $\boldsymbol{\lambda}'\boldsymbol{A}' = \boldsymbol{b}'$ and $\boldsymbol{\lambda}'\boldsymbol{A}'\boldsymbol{y} = \boldsymbol{b}'\boldsymbol{y} \leqslant 0$ for all $\boldsymbol{y}$ for which $\boldsymbol{A}'\boldsymbol{y} \leqslant \boldsymbol{0}$. (Necessity.) Let $\mathscr{H} = \{\boldsymbol{y} | \boldsymbol{b}'\boldsymbol{y} = 0, \boldsymbol{y} \in \mathscr{E}^m\}$ be a hyperplane through the origin and orthogonal to the half-line $\mathscr{L} = \{\boldsymbol{y} | \boldsymbol{y} = \lambda \boldsymbol{b}, \lambda \geqslant 0, \boldsymbol{y} \in \mathscr{E}^m\}$. Then $\mathscr{L}^* = \{\boldsymbol{y} | \boldsymbol{b}'\boldsymbol{y} \leqslant 0, \boldsymbol{y} \in \mathscr{E}^m\}$. Since $\mathscr{C} = \{\boldsymbol{x} | \boldsymbol{x} = \boldsymbol{A\lambda}, \boldsymbol{\lambda} \geqslant \boldsymbol{0}, \boldsymbol{x} \in \mathscr{E}^m\}$, $\mathscr{C}^* = \{\boldsymbol{y} | \boldsymbol{A}'\boldsymbol{y} \leqslant \boldsymbol{0}, \boldsymbol{y} \in \mathscr{E}^m\}$. For each $\boldsymbol{y} \in \mathscr{C}^*$, let it also be true that $\boldsymbol{y} \in \mathscr{L}^*$. So with $\mathscr{C}^* \subset \mathscr{L}^*$, $\mathscr{L} \subset \mathscr{C}$ or $\boldsymbol{b} \in \mathscr{C}$ (fig. 1.10d). (Note that this part of the proof has employed duality properties (b) and (e) above.) Q.E.D.

1.10. Quadratic forms

DEFINITION 1.54. *Let* Q *be a real-valued function of the n variables* $x_1, \ldots, x_n$. *Then* Q *is called a quadratic form in* $x_1, \ldots, x_n$ *if*

$$Q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j,$$

where at least one of the constant coefficients $a_{ij} \neq 0$.

To determine the general properties of Q let us express the above finite

double sum explicitly as

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \sum_{j=1}^{n} a_{1j}x_1x_j + \sum_{j=1}^{n} a_{2j}x_2x_j + \cdots + \sum_{j=1}^{n} a_{nj}x_nx_j$$

$$= a_{11}x_1^2 + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + a_{21}x_2x_1 + a_{22}x_2^2 + \ldots + a_{2n}x_2x_n + \ldots + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \ldots + a_{nn}x_n^2. \tag{1.14}$$

Written in this fashion it is readily seen that Q is a homogeneous[11] polynomial of the second degree (since each term involves either the square of a variable or the product of two different variables) containing n^2 distinct terms. In addition Q is continuous for all values of the variables x_i, $i = 1, \ldots, n$, and equals zero when all of the $x_i = 0$, $i = 1, \ldots, n$.

Let us now consider an alternative mode of representing a quadratic form. Expressed in matrix form, Q equals, for all vectors $\boldsymbol{x} \in \mathscr{E}^n$, the scalar quantity

$$Q(x_1, \ldots, x_n) = Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \tag{1.15}$$

where

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \boldsymbol{A} = \begin{bmatrix} a_{11}a_{12} \ldots a_{1n} \\ a_{21}a_{22} \ldots a_{2n} \\ \vdots \quad \vdots \quad\quad \vdots \\ a_{n1}a_{n2} \ldots a_{nn} \end{bmatrix}.$$

To see this we first find

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} a_{11}\,x_1 + a_{12}\,x_2 + \ldots + a_{1n}\,x_n \\ a_{21}\,x_1 + a_{22}\,x_2 + \ldots + a_{2n}\,x_n \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}\,x_1 + a_{n2}\,x_2 + \ldots + a_{nn}\,x_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}\,x_j \\ \sum_{j=1}^{n} a_{2j}\,x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}\,x_j \end{bmatrix}.$$

[11] A form is *homogeneous of degree t* in the variables $x_1, \ldots, x_n$ if, when each variable in the form is multiplied by a scalar λ, the whole form is multiplied by λ^t, i.e. $Q(\lambda x_1, \ldots, \lambda x_n) = \lambda^t Q(x_1, \ldots, x_n)$.

Then

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = (x_1, \ldots, x_n)\begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}x_j \end{bmatrix}$$

$$= x_1 \sum_{j=1}^{n} a_{1j}x_j + x_2 \sum_{j=1}^{n} a_{2j}x_j + \ldots + x_n \sum_{j=1}^{n} a_{nj}x_j$$

$$= \sum_{j=1}^{n} a_{1j}x_1x_j + \sum_{j=1}^{n} a_{2j}x_2x_j + \ldots + \sum_{j=1}^{n} a_{nj}x_nx_j.$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j.$$

From eq. (1.14) it can be seen that $a_{ij}+a_{ji}$ is the coefficient of x_ix_j since a_{ij}, a_{ji} are both coefficients of $x_ix_j = x_jx_i$, $i \neq j$.

Example 1.18. Find the quadratic form $Q(x_1, x_2, x_3)$ associated with the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 5 & 2 \\ 1 & 1 & 2 \end{bmatrix}.$$

From eq. (1.15) we have

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\begin{bmatrix} 1 & 2 & 1 \\ 3 & 5 & 2 \\ 1 & 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (x_1, x_2, x_3)\begin{bmatrix} x_1+2x_2+x_3 \\ 3x_1+5x_2+2x_3 \\ x_1+x_2+2x_3 \end{bmatrix}$$

$$= x_1(x_1+2x_2+x_3)+x_2(3x_1+5x_2+2x_3)+x_3(x_1+x_2+2x_3)$$
$$= x_1^2+5x_1x_2+2x_1x_3+5x_2^2+3x_2x_3+2x_3^2.$$

Example 1.19. Find $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ when

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 2 \end{bmatrix}.$$

Since three elements in $\boldsymbol{A}$ are zero, Q will have $n^2-3 = 6$ individual terms.

Upon performing the indicated matrix multiplication we obtain

$$\begin{aligned} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &= x_1^2+0x_1x_2+x_1x_3+3x_2x_1+0x_2^2+0x_2x_3+x_3x_1 \\ &\quad +x_3x_2+2x_3^2 \\ &= x_1^2+3x_1x_2+2x_1x_3+x_2x_3+2x_3^2. \end{aligned}$$

1.11. Symmetric quadratic forms

If the matrix $\boldsymbol{A}$ is symmetric so that $\boldsymbol{A} = \boldsymbol{A}'$, then $a_{ij} = a_{ji}$, $i \neq j$. Thus we have

DEFINITION 1.55. *A quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is symmetric if the matrix $\boldsymbol{A}$ is symmetric, i.e. if $a_{ij} = a_{ji}$, $i \neq j$.*

Hence $a_{ij}+a_{ji} = 2a_{ij}$ is the coefficient of x_ix_j since $a_{ij} = a_{ji}$ and a_{ij}, a_{ji} are both coefficients of $x_ix_j = x_jx_i$, $i \neq j$.

If $\boldsymbol{A}$ is not a symmetric matrix so that $a_{ij} \neq a_{ji}$, we can transform it into a symmetric matrix $\boldsymbol{B}$ by defining new coefficients:

$$b_{ij} = b_{ji} = \frac{a_{ij}+a_{ji}}{2} \quad \text{for all } i, j. \tag{1.16}$$

Then $b_{ij}+b_{ji} = 2b_{ij}$ is the coefficient of x_ix_j, $i \neq j$, in

$$\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{a_{ij}+a_{ji}}{2} x_ix_j.$$

But, by definition, $b_{ij}+b_{ji} = a_{ij}+a_{ji}$. Hence the redefinition of coefficients leaves the value of Q unchanged. That is, if eq. (1.16) holds, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ for any $\boldsymbol{x} \in \mathscr{E}^n$. In sum, given any quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, the matrix $\boldsymbol{A}$ may be assumed to be symmetric; if it is not, it can always be transformed into a symmetric matrix.

Example 1.20. Given that

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 2 \\ 1 & -1 & 6 \\ 3 & 5 & 4 \end{bmatrix}$$

transforms $\boldsymbol{A}$ into a symmetric matrix $\boldsymbol{B}$. From eq. (1.16) we set

$$\begin{aligned} b_{11} &= a_{11} = 1, \\ b_{12} &= b_{21} = \frac{a_{12}+a_{21}}{2} = \frac{3+1}{2} = 2, \\ b_{13} &= b_{31} = \frac{a_{13}+a_{31}}{2} = \frac{2+3}{2} = \frac{5}{2}, \end{aligned}$$

$$b_{22} = a_{22} = -1,$$
$$b_{23} = b_{32} = \frac{a_{23}+a_{32}}{2} = \frac{6+5}{2} = \frac{11}{2},$$
$$b_{33} = a_{33} = 4.$$

Hence

$$\boldsymbol{B} = \begin{bmatrix} 1 & 2 & \frac{5}{2} \\ 2 & -1 & \frac{11}{2} \\ \frac{5}{2} & \frac{11}{2} & 4 \end{bmatrix}.$$

Example 1.21. Find the matrix A associated with the quadratic form

$$x'Ax = 2x_1^2 - 3x_1x_2 + \tfrac{7}{2}x_1x_3 + x_1x_4 + x_2^2 + 6x_2x_3 - 8x_3x_4 + 2x_4^2.$$

Since A may be assumed to be symmetric we have

$$A = \begin{bmatrix} 2 & -\frac{3}{2} & \frac{7}{4} & \frac{1}{2} \\ -\frac{3}{2} & 1 & 3 & 0 \\ \frac{7}{4} & 3 & 0 & -4 \\ \frac{1}{2} & 0 & -4 & 2 \end{bmatrix}.$$

1.12. Classification of quadratic forms

In all there are five mutually exclusive and collectively exhaustive categories of quadratic forms. First,

DEFINITION 1.56. *A quadratic form is said to be positive definite (negative definite) if it is positive (negative) at every point $x \in \mathscr{E}^n$ except $x = 0$, i.e.*

(a) $x'Ax$ is positive definite if $x'Ax > 0$ for every $x \neq 0$;

(b) $x'Ax$ is negative definite if $x'Ax < 0$ for every $x \neq 0$.

It is evident that a form which is either positive or negative definite cannot assume both positive and negative values. To see this let us assume that a definite form is positive at a point $x_1 (x_1'Ax_1 > 0)$ and negative at $x_2 (x_2'Ax_2 < 0)$. Then, because of the continuity of the form, there must exist some point $x_3 \neq 0$ between x_1, x_2 such that $x_3'Ax_3 = 0$ (see theorem 6.3, p. 99). However, this contradicts the definition of definiteness given above. Hence definite forms must be either positive or negative. We now state

DEFINITION 1.57. *A quadratic form is said to be positive semi-definite (negative semi-definite) if it is non-negative (non-positive) at every point $x \in \mathscr{E}^n$, and there exist points $x \neq 0$ for which it equals zero, i.e.*

(a) $x'Ax$ is positive semi-definite if $x'Ax \geqslant 0$ for every x and $x'Ax = 0$ for some points $x \neq 0$;

(*b*) $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ *is negative semi-definite if* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leqslant 0$ *for every* $\boldsymbol{x}$ *and* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = 0$ *for some points* $\boldsymbol{x} \neq \boldsymbol{0}$.

It is obvious that if the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is positive definite (semi-definite), then $\boldsymbol{x}'(-\boldsymbol{A})\boldsymbol{x}$ is negative definite (semi-definite) and conversely. In addition, we have

DEFINITION 1.58. *A quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ *is said to be indefinite if it is positive for some points* $\boldsymbol{x} \in \mathscr{E}^n$ *and negative for others.*

Example 1.22. From def. 1.56 it is clear that the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1^2 + x_2^2$ is positive definite while $\boldsymbol{x}'(-\boldsymbol{A})\boldsymbol{x} = -x_1^2 - x_2^2$ is negative definite since both vanish only at the point $\boldsymbol{x} = \boldsymbol{0}$.

Example 1.23. It is evident, from def. 1.57, that $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1^2 - 2x_1x_2 + x_2^2 = (x_1 - x_2)^2$ is positive semi-definite and $\boldsymbol{x}'(-\boldsymbol{A})\boldsymbol{x} = -x_1^2 + 2x_1x_2 - x_2^2 = -(x_1 - x_2)^2$ is negative semi-definite since the former is never negative while the latter is never positive, yet both equal zero for $x_1 = x_2 \neq 0$.

Example 1.24. That the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1x_2 + x_2^2$ is indefinite can be verified from def. 1.58 by noting that, on the one hand, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$ for $x_1 = -2$, $x_2 = 1$ and, on the other, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for $x_1 = 2$, $x_2 = 1$.

We shall often find it convenient to classify matrices in terms of the sign 'definiteness' or 'semi-definiteness' of their associated quadratic forms. Specifically, we have

DEFINITION 1.59. *An nth-order symmetric matrix* $\boldsymbol{A}$ *is positive definite (negative definite) if and only if* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0 (< 0)$ *for all* $\boldsymbol{x} (\neq \boldsymbol{0}) \in \mathscr{E}^n$

DEFINITION 1.60. *An nth-order symmetric matrix* $\boldsymbol{A}$ *is positive semi-definite (negative semi-definite) if and only if* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geqslant 0 (\leqslant 0)$ *for all* $\boldsymbol{x} \in \mathscr{E}^n$.

Some of the essential features of definite matrices are:

(a) if $\boldsymbol{A}$ is an nth-order positive (negative) definite matrix, then $|\boldsymbol{A}| > 0 (< 0)$ and thus $\rho(\boldsymbol{A}) = n$;

(b) if $\boldsymbol{A}$ is an nth-order positive (negative) definite matrix, then so is $\boldsymbol{A}^{-1}$.

1.13. Necessary conditions for the definiteness and semi-definiteness of quadratic forms

In this section and the next our aim will be to provide a set of theorems with which to identify the various types of quadratic forms. We state first

THEOREM 1.3.[12] *If a quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x}\in\mathscr{E}^n$, *is positive (negative) definite, all the terms involving second powers of the variables must have positive (negative) coefficients.*

Note that the theorem does not provide a sufficient condition for the definiteness of a quadratic form because its converse does not hold, i.e. a quadratic form may have positive (negative) coefficients on all its terms involving second powers yet not be definite. A case in point is provided by example 1.23 above. Similarly,

THEOREM 1.4. *If a quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x}\in\mathscr{E}^n$, *is positive (negative) semi-definite, all of the terms involving second powers of the variables must have non-negative (non-positive) coefficients.*

In this case, too, the converse of the theorem does not hold, since the quadratic form $x_1^2 - x_1 x_2$ has non-negative coefficients associated with its second-degree terms, yet happens to be indefinite.

In the next section we shall find it useful to express conditions which are simultaneously necessary and sufficient for the definiteness or semi-definiteness of a quadratic form in terms of determinants. As a prelude to this type of reasoning we cite

THEOREM 1.5.[13] *If the quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x}\in\mathscr{E}^n$, *is definite, the naturally ordered principal minors of* $\boldsymbol{A}$ *are all different from zero. In this case*

$$M_k = \begin{vmatrix} a_{11} \dots a_{1k} \\ \vdots \quad\quad \vdots \\ a_{k1} \dots a_{kk} \end{vmatrix} \neq 0, \quad k = 1, \dots, n$$

or

$$M_{11} = a_{11} \neq 0, \qquad M_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0,$$

$$M_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \neq 0, \dots, \qquad M_n = |\boldsymbol{A}| \neq 0.$$

So if any $M_k = 0$, $k = 1, \dots, n$, the form is not definite; it may be semi-definite or indefinite. That this theorem does not provide a sufficient condition for definiteness is illustrated by the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1^2 - x_2^2$.

---

[12] This theorem represents a special case of a more general theorem provided by Bushaw and Clower (1957), p. 264.

[13] Bushaw and Clower (1957), pp. 279–280.

Here

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Although $M_1 = 1 \neq 0$, $M_2 = -1 \neq 0$, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is not definite but indefinite since it is positive for some values of $\boldsymbol{x}$ and negative for others.

Example 1.25. Use theorem 1.5 to verify that the quadratic forms

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1^2 - 2x_1x_2 + x_2^2, \qquad \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1x_2 + x_2^2$$

are non-definite. First, for $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1^2 - 2x_1x_2 + x_2^2$,

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Now $M_1 = 1 \neq 0$ while $M_2 = 0$. Hence the form is not definite. It is, in fact, semi-definite, as indicated in example 1.23. Next, for $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = x_1x_2 + x_2^2$,

$$A = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}.$$

Here $M_1 = 0$ and we need not proceed further. In this case $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ happens to be indefinite, as seen in example 1.24.

1.14. Necessary and sufficient conditions for the definiteness and semi-definiteness of quadratic forms

Let us modify theorem 1.5 to get

THEOREM 1.6.[14] *The quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x} \in \mathscr{E}^n$, is positive definite if and only if the naturally ordered principal minors of $\boldsymbol{A}$ are all positive, i.e.*

$$M_k = \begin{vmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{vmatrix} > 0, \quad k = 1, \dots, n,$$

or

$$M_1 = a_{11} > 0, \qquad M_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0,$$

$$M_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \dots, \qquad M_n = |A| > 0.$$

[14] A proof of this theorem is provided by Hadley (1964), pp. 260–261.

We noted above that if the quadratic form $x'Ax$ is negative definite, then $x'(-A)x$ is positive definite. But if $x'(-A)x$ is positive definite, it follows from the preceding theorem that the naturally ordered principal minors of

$$-A = \begin{bmatrix} -a_{11} & -a_{12} \dots & -a_{1n} \\ -a_{21} & -a_{22} \dots & -a_{2n} \\ \vdots & \vdots & \vdots \\ -a_{n1} & -a_{n2} \dots & -a_{nn} \end{bmatrix}$$

are all positive or

$$M_1 = -a_{11} > 0,$$

$$M_2 = \begin{vmatrix} -a_{11} & -a_{12} \\ -a_{21} & -a_{22} \end{vmatrix} = (-1)^2 \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0,$$

$$M_3 = \begin{vmatrix} -a_{11} & -a_{12} & -a_{13} \\ -a_{21} & -a_{22} & -a_{23} \\ -a_{31} & -a_{32} & -a_{33} \end{vmatrix} = (-1)^3 \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0,$$

. . . . . .

$$M_n = |-A| = (-1)^n |A| > 0.$$

For all these principal minors M_k, $k = 1, \dots, n$, to be positive, it must be true that

$$a_{11} < 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} < 0, \dots, \quad (-1)^n |A| > 0.$$

But this last sequence of determinants represents the naturally ordered principal minors of A. Hence

THEOREM 1.7. *The quadratic form $x'Ax$, $x \in \mathcal{E}^n$, is negative definite if and only if the naturally ordered principal minors of A alternate in sign, the first being negative, i.e.*

$$(-1)^k M_k = (-1)^k \begin{vmatrix} a_{11} \dots a_{1k} \\ \vdots \quad \vdots \\ a_{k1} \dots a_{kk} \end{vmatrix} > 0, \quad k = 1, \dots, n$$

or

$$M_1 = a_{11} < 0, \qquad M_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0,$$

$$M_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} < 0, \dots, \qquad M_n = (-1)^n |A| > 0.$$

A similar set of theorems holds for semi-definite forms. In this regard,

THEOREM 1.8. *The quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x}\in\mathscr{E}^n$, *is positive semi-definite if and only if the naturally ordered principal minors of* $\boldsymbol{A}$ *are all non-negative.*

Now if $\boldsymbol{x}'(-\boldsymbol{A})\boldsymbol{x}$ is positive semi-definite, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is negative semi-definite and thus we have

THEOREM 1.9. *The quadratic form* $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{x}\in\mathscr{E}^n$, *is negative semi-definite if and only if the naturally ordered principal minors of* $\boldsymbol{A}$ *alternate in sign, the first being non-positive.*

Example 1.26. Is the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = 2x_1^2+2x_1x_2+6x_2^2+4x_2x_3+x_3^2$ positive definite? To answer this question we first find

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 6 & 2 \\ 0 & 2 & 1 \end{bmatrix}.$$

Employing theorem 1.6 we have $M_1 = 2$, $M_2 = 11$, and $M_3 = 3$. Hence $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is positive definite since all the naturally ordered principal minors of $\boldsymbol{A}$ are positive.

Example 1.27. Prove that $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = -x_1^2+x_1x_2-x_2^2-x_3^2$ is negative definite. Since

$$\boldsymbol{A} = \begin{bmatrix} -1 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

it follows that $M_1 = -1$, $M_2 = 3/4$, and $M_3 = -1/2$. Thus the requirements of theorem 1.7 hold and $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is negative definite.

1.15. Quadratic functions

We shall now demonstrate that a general second-degree polynomial in the variables $x_1, \ldots, x_n$ can be written as the sum of a constant, a linear form, and a quadratic form. Specifically, let the real-valued function $y = \mathrm{f}(\boldsymbol{x})$, $\boldsymbol{x}\in\mathscr{E}^n$, appear as

$$\begin{aligned} \mathrm{f}(\boldsymbol{x}) = a &+ b_1x_1 + \ldots + b_nx_n + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n \\ &+ a_{23}x_2x_3 + \ldots + a_{2n}x_2x_n \\ &+ a_{34}x_3x_4 + \ldots + a_{3n}x_3x_n \\ &+ \ldots + a_{n-1,n}x_{n-1}x_n \\ &+ a_1x_1^2 + \ldots + a_nx_n^2. \end{aligned}$$

If

$$q_{ii} = a_i \quad \text{and} \quad q_{ij} = q_{ji} = \tfrac{1}{2}a_{ij}, \quad i \neq j, \quad i, j = 1, \ldots, n,$$

then the above expression may be rewritten as

$$\begin{aligned} \mathrm{f}(\boldsymbol{x}) &= a + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij} x_i x_j \\ &= a + \boldsymbol{b}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x}, \end{aligned}$$

where $\boldsymbol{b}$ is of order $(n \times 1)$ and $\boldsymbol{Q}$ is an nth-order symmetric matrix.

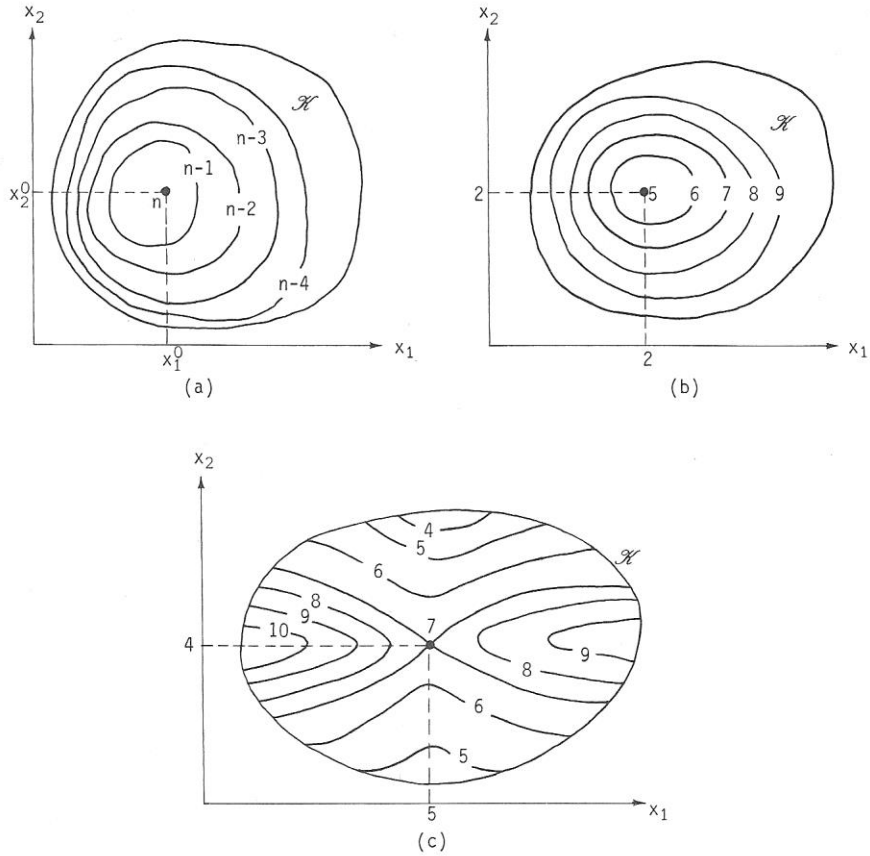


Fig. 6.22.

fig. 6.22b may be conceived as representing its contour map. And if a function hypothetically assumes a saddle value at the point $x_1' = (5, 4) \in \mathscr{K}$, with

$$f(x_1) = 7 = g \max_{x_2} f(5, x_2) = g \min_{x_1} f(x_1, 4),$$

the implied contour map may look something like the one in fig. 6.22c.

6.11. Appendix A: The Newton–Raphson, secant, and false position methods

Throughout chapter 6 we examined a variety of rather simple and straightforward examples involving the application of various theorems designed

to identify and isolate local extrema. For the most part, we may refer to these contrived examples as 'classroom exercises' whose singular purpose is pedagogic exposition, rather than the development of computational prowess. However, the vast majority of functions that one usually encounters do not lend themselves to any such clear-cut analysis. That is, we are usually not fortunate enough in most applications to compute f′, equate it to zero, and then find that we are faced with simply solving a linear or quadratic equation for the critical value(s) of x which give rise to local extrema. Indeed, one may have to solve a polynomial of rather high order or even an equation of a more complex nature. How are we to handle such contingencies? In this section we shall explore three common methods of finding the roots of non-linear equations: the Newton–Raphson method; the secant method; and the method of false position.

Basically, each of the said techniques for computing the root(s)[11] of a non-linear equation involves a process of successive approximations, i.e. each is an iterative procedure. As with any iterative technique, we begin the various rounds of iteration by choosing an initial point (near the desired solution) from which a new point is computed by an appropriate algorithm or computational rule. The process is continued (until the desired degree of accuracy is attained) by successively calculating additional points that yield improved approximations to the solution.

To see exactly how the *Newton–Raphson method* works, let us assume that we have taken the first derivative of the real-valued function $y = f(x)$ and we desire to find a value of x, x^*, for which it vanishes. That is, we desire to find a root x^* of $f'(x) = 0$. Without loss of generality, let us further assume that f attains a strong local minimum at $x = x^*$ (fig. A.1). If we choose an initial estimate of x^*, x_0, to the right (or left) of x^* (with $|x^* - x_0|$ small) and approximate f′ at the point $(x_0, f'(x_0))$ by the linear portion of Taylor's expansion of a function (eq. (5.12)), we obtain

$$f'(x) = f'(x_0) + f''(x_0)(x - x_0). \tag{A.1}$$

Here, eq. (A.1) represents the equation of the line tangent to f′ at $(x_0, f'(x_0))$. It crosses the x-axis at the point $(x_1, 0)$. A substitution of the coordinates of this point into eq. (A.1) yields

$$0 = f'(x_0) + f''(x_0)(x_1 - x_0)$$

or

$$x_1 = x_0 - f'(x_0)/f''(x_0),$$

[11] By a *root* of the real-valued function $y = f(x) = 0$ is meant a real number r such that, when r is substituted into f for the unknown, the equation is satisfied. That is, $f(x) = 0$ reduces to the identity $0 \equiv 0$ for $x = r$.

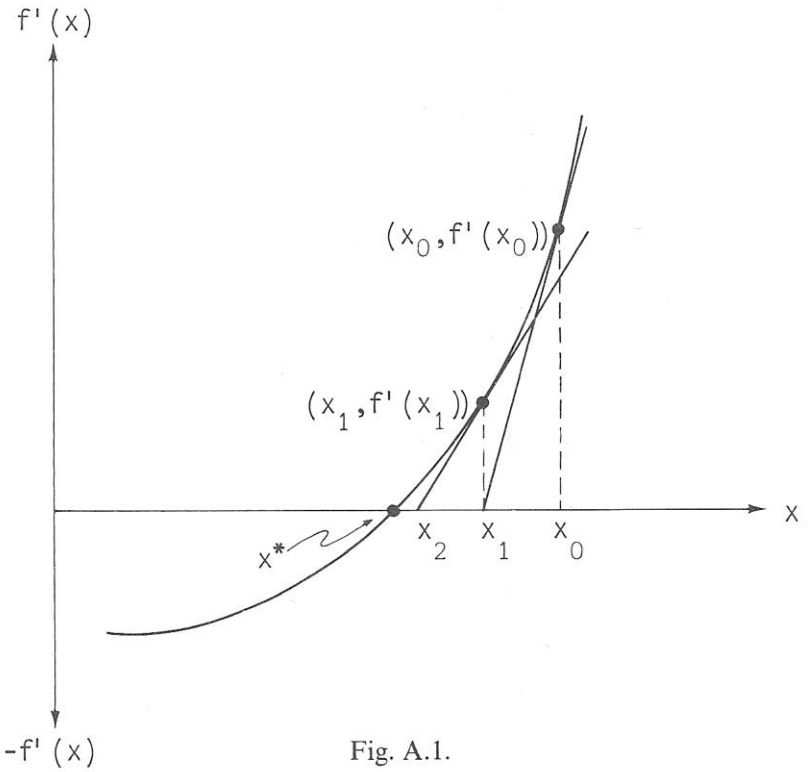


Fig. A.1.

where x_1 represents our second approximation to x^*. (Note that x_1 is closer to x^* than is x_0.) This completes the first round or iteration of the Newton–Raphson technique. To start the second iteration we again approximate f′ by Taylor's expansion, but this time at the point $(x_1, f'(x_1))$. Then

$$f'(x) = f'(x_1)+f''(x_1)(x-x_1). \tag{A.2}$$

In this instance the tangent to f′ at $(x_1, f'(x_1))$(A.2) intersects the x-axis at a point even closer to x^* than before, namely x_2. Our third approximation to x^*, x_2, is obtained by substituting the coordinates of this new point $(x_2, 0)$ into eq. (A.2). Hence

$$\begin{aligned} 0 &= f'(x_1)+f''(x_1)(x_2-x_1), \\ x_2 &= x_1-f'(x_1)/f''(x_1). \end{aligned}$$

Our second iteration is now complete. This successive-approximation technique may be repeated until the difference $|x_i-x_{i-1}|$ is as small as

one desires. In general, after i iterations,

$$x_i = x_{i-1} - \frac{f'(x_{i-1})}{f''(x_{i-1})}, \quad i = 1, 2, \ldots. \tag{A.3}$$

A summary of the various steps involved in implementing the Newton–Raphson method proceeds as follows:

(1) draw a graph of $f'(x) = 0$;

(2) choose an initial estimate x_0 near the point x^* where the curve crosses the x-axis;

(3) from x_0 move vertically to $f'(x_0)$;

(4) construct the tangent to f' at $(x_0, f'(x_0))$;

(5) find the point where the tangent to f' at $(x_0, f'(x_0))$ crosses the x-axis. This yields

$$x_1 = x_0 - f'(x_0)/f''(x_0);$$

(6) repeat the process, involving steps (1)–(5) i times in succession to obtain

$$x_i = x_{i-1} - f'(x_{i-1})/f''(x_{i-1}).$$

We note briefly that, with the Newton–Raphson (or any) successive-approximation technique, a proof of convergence to the desired solution and an estimate of the error incurred upon completing the ith iteration are needed. In addition, one must demonstrate the existence and uniqueness of the solution (see Saaty and Bram, 1964, pp. 58–62).

Example A.1. Determine the extreme values of the real-valued function

$$y = f(x) = \tfrac{1}{4}x^4 - \tfrac{2}{3}x^3 - x^2 - 5x + 2$$

over the x-axis. Setting $f'(x) = 0$ we obtain

$$x^3 - 2x^2 - 2x - 5 = 0.$$

A graph of f' (fig. A.2) indicates that this function possesses a critical root x^* somewhere between $x = 3$ and $x = 4$. To find it let us approximate x^* by the Newton–Raphson method for $i = 1, 2, 3$. From eq. (A.3) we have, for $x_0 = 3.5$,

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = 3.5000 - \frac{f'(3.5000)}{f''(3.5000)} = 3.1900,$$

$$x_2 = x_1 - \frac{f'(x_1)}{f''(x_1)} = 3.1900 - \frac{f'(3.1900)}{f''(3.1900)} = 3.1440,$$

$$x_3 = x_2 - \frac{f'(x_2)}{f''(x_2)} = 3.1440 - \frac{f'(3.1440)}{f''(3.1440)} = 3.1427.$$

Since $f''(3.1427) > 0$, f has a strong local minimum at $x^* \sim 3.1427$.

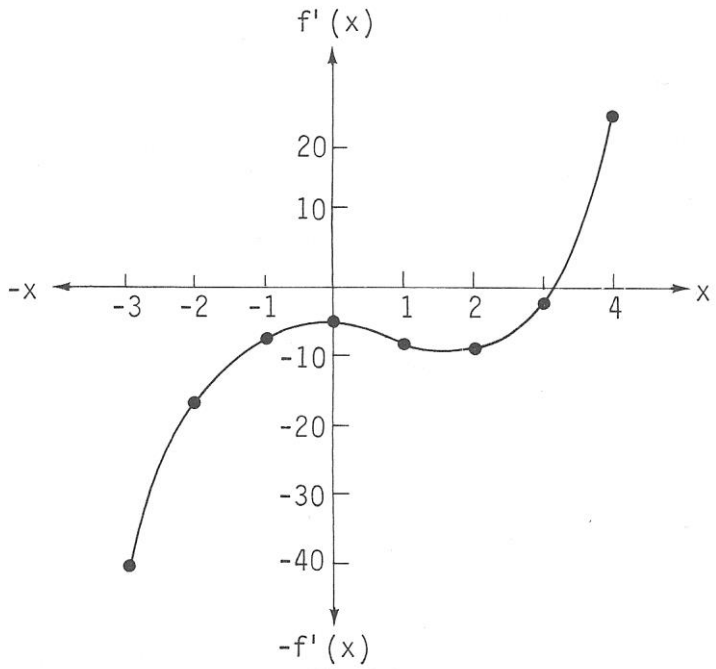


Fig. A.2.

The Newton–Raphson method may be modified for computational convenience by replacing $f''(x_{i-1})$ in eq. (A.3) by $f''(x_0)$ (Saaty and Bram, 1964, p. 62). This substitution enables us to avoid computing $f''(x_{i-1})$ at each round of the process. Hence, the iterations are now described by

$$x_i = x_{i-1} - \frac{f'(x_{i-1})}{f''(x_0)}, \quad i = 1, 2, \ldots \tag{A.3.1}$$

Example A.2. Using $f' = 0$ from example A.1 above, demonstrate that, for $x_0 = 3.5$, the successive terms of the modified Newton–Raphson process described by eq. (A.3.1) approach 3.1427 for $i = 1, \ldots, 6$. In this case

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = 3.19000, \qquad x_4 = x_3 - \frac{f'(x_3)}{f''(x_0)} = 3.14360,$$

$$x_2 = x_1 - \frac{f'(x_1)}{f''(x_0)} = 3.15000, \qquad x_5 = x_4 - \frac{f'(x_4)}{f''(x_0)} = 3.14290,$$

$$x_3 = x_2 - \frac{f'(x_2)}{f''(x_0)} = 3.14600, \qquad x_6 = x_5 - \frac{f'(x_5)}{f''(x_0)} = 3.14274.$$

Note that while eq. (A.3.1) is easier to handle from a computational viewpoint than eq. (A.3), the rapidity of convergence is substantially greater for eq. (A.3) than for eq. (A.3.1).

At times it may be desirable to further modify the Newton–Raphson process (eq. (A.3)) so as to accelerate convergence to x^*. One such modification, leading to what is called the *generalized Newton–Raphson method*, is developed as follows. If eq. (A.1) is replaced by

$$f'(x) = f'(x_0) + \tau f''(x_0)(x + x_0), \quad \tau \neq 1. \tag{A.1.1}$$

(Here eq. (A.1.1) represents the equation of a line through $(x_0, f'(x_0))$ which intersects the x-axis at a point closer to x^* than the previously specified x_1 value in fig. A.1.), then it is easily shown that

$$x_1 = x_0 - \frac{1}{\tau} \frac{f'(x_0)}{f''(x_0)}, \quad \tau \neq 1.$$

In general, the ith iteration is determined from

$$x_i = x_{i-1} - \lambda \frac{f'(x_{i-1})}{f''(x_{i-1})}, \quad i = 1, 2, \ldots, \tag{A.3.2}$$

where $\lambda = 1/\tau$. As an exercise the reader should rework example A.1 using eq. (A.3.2) with $\tau = 0.90$.

It is important to note that the sequences described by eqs. (A.3), (A.3.1), and (A.3.2) may not converge if x_0 is not chosen sufficiently close to x^*. Moreover, if they do, in fact, converge to some number $x \in \mathscr{D}_{f'}$, this number may not be the correct root.

The Newton–Raphson technique requires first- as well as second-order information about a function, i.e. it utilizes both first and second derivatives in its implementation. A modification of the Newton–Raphson method which uses only first-order information is the *secant method*, so named because f'' in eq. (A.3) is replaced by its finite difference approximation

$$f''(x_{i-1}) \sim \frac{f'(x_{i-1}) - f'(x_{i-2})}{x_{i-1} - x_{i-2}}, \tag{A.4}$$

where the difference quotient on the right-hand side of eq. (A.4) is simply the slope of the secant line between the two points

$$(x_{i-1}, f'(x_{i-1})), \qquad (x_{i-2}, f'(x_{i-2}))$$

on f' (fig. A.3) and x_{i-1}, x_{i-2} are any 'two' initial approximations to the root x^*. Upon substituting eq. (A.4) into (A.3) we obtain

$$\begin{aligned} x_i &= x_{i-1} - f'(x_{i-1}) \frac{x_{i-1} - x_{i-2}}{f'(x_{i-1}) - f'(x_{i-2})} \\ &= \frac{x_{i-2} f'(x_{i-1}) - x_{i-1} f'(x_{i-2})}{f'(x_{i-1}) - f'(x_{i-2})}, \quad i = 2, 3, \ldots. \end{aligned} \tag{A.5}$$

What is the geometric interpretation of eq. (A.5)? Given the two initial

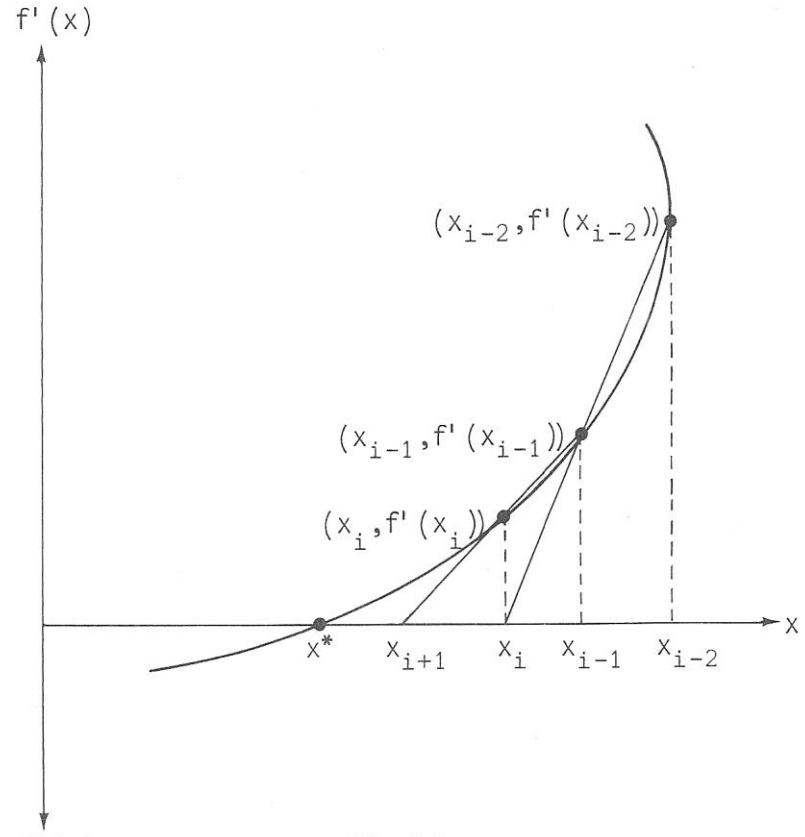


Fig. A.3.

approximations x_{i-1}, x_{i-2} to x^*, the next approximation x_i corresponds to the intersection of the chord joining points $(x_{i-1}, f'(x_{i-1}))$, $(x_{i-2}, f'(x_{i-2}))$ with the x-axis. Once x_i is obtained, x_{i+1} is determined in a similar fashion, i.e. by finding the intersection of the chord joining points

$$(x_i, f'(x_i)), \qquad (x_{i-1}, f'(x_{i-1}))$$

with the x-axis. The process is repeated until the desired degree of accuracy, in terms of $|x_i - x_{i-1}|$, is achieved.

A variation of the secant method is the *method of false position*. Let us choose the two initial approximations x_{i-2}, x_{i-1} to x^* in a fashion such that $f'(x_{i-2})$, $f'(x_{i-1})$ are of opposite sign, i.e. $f'(x_{i-2})f'(x_{i-1}) < 0$. Then x^* must lie between x_{i-2}, x_{i-1} so that if we connect points $(x_{i-1}, f'(x_{i-1}))$, $(x_{i-2}, f'(x_{i-2}))$ (fig. A.4) by a secant line, x_i in eq. (A.5) represents the intersection of this line with the x-axis. To obtain x_{i+1}, let us, in general, proceed in the following fashion:

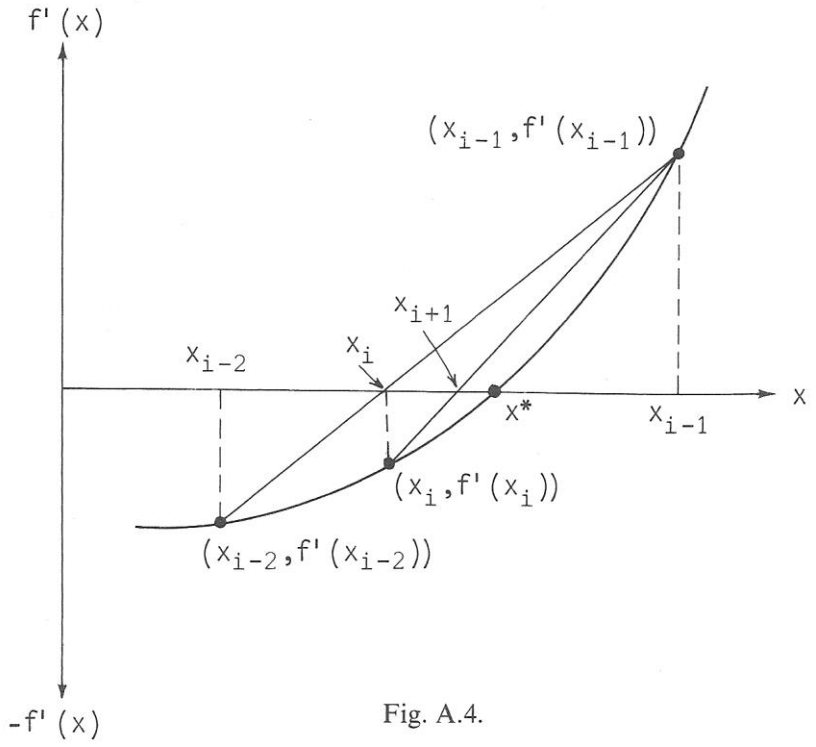


Fig. A.4.

(a) If $f'(x_{i-2})$, $f'(x_i)$ are opposite in sign ($f'(x_{i-2})f'(x_i)<0$), replace x_{i-1} by x_i (x_{i-2} is unchanged) and obtain the next approximation from eq. (A.5). (Geometrically, x_{i+1} is just the intersection of the chord joining $(x_{i-2}, f'(x_{i-2}))$, $(x_i, f'(x_i))$ and the x-axis.)

(b) Otherwise, replace x_{1-2} by x_i(x_{i-1} is invariant) and again use eq. (A.5) to determine a new approximation. (In terms of fig. A.4, if $(x_i, f'(x_i))$, $(x_{i-1}, f'(x_{i-1}))$ are connected by a secant line, the new estimate, x_{i+1}, is again taken to be the intersection of this line with the x-axis.) As always, the process is repeated until the desired degree of accuracy obtains.

When should these various iterative schemes be employed? If f'' is easily computed, the standard Newton–Raphson technique is preferred, the reason being that it converges faster to x^* than the other two methods. (In this case, the Newton–Raphson technique is said to possess *second-order convergence* or to *converge quadratically*, i.e. the error incurred at any iteration is proportional to the square of the error of the previous iteration so that, asymptotically, the number of places of accuracy doubles with each succeeding iteration beyond the first few.) Next, if the information provided

by f″ is difficult to obtain, the secant method should be used. Finally, the method of false position can be utilized if the previous two methods fail to converge to the desired critical value of x. As an exercise the reader should rework example A.1 using both the secant method and the method of false position.

6.12. Appendix B: The Newton-Raphson method, $x \in \mathscr{E}^n$

We noted above that a necessary condition for the real-valued function $y = f(x)$, $x \in \mathscr{E}^n$, to assume a local extremum at a point x_* contained within an open region $\mathscr{K} \subseteq \mathscr{E}^n$ is that $\nabla f^* = 0$. If $\nabla f = 0$ gives rise to a non-linear system of equations, then x_* may be found by a series of successive approximations using the Newton–Raphson technique. As in Appendix A, our strategy will be to obtain a series of successive approximations to x_*, with each iteration yielding an improved solution.

That is, if we choose the point x_0 as our initial estimate of x_*, with $\|x_0 - x_*\|$ sufficiently small, and approximate ∇f at x_0 by the linear portion of Taylor's expansion of a function (eq. (5.15)), we obtain

$$\left.\begin{aligned} \nabla f &= \nabla f^0 + \left(\frac{\partial \nabla f^0}{\partial x}\right) h \\ &= \nabla f^0 + H_f(x_0)\,(x - x_0) \end{aligned}\right\} \text{ or } \begin{cases} f_1(x) = f_1^0 + \sum_{j=1}^{n} f_{1j}^0 (x_j - x_j^0), \\ f_2(x) = f_2^0 + \sum_{j=1}^{n} f_{2j}^0 (x_j - x_j^0), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_n(x) = f_n^0 + \sum_{j=1}^{n} f_{nj}^0 (x_j - x_j^0). \end{cases} \tag{B.1}$$

To undertake our first iteration let us set $x = x_1$ in system (B.1) and equate $\nabla f(x_1)$ to the null vector. Then eq. (B.1) becomes

$$0 = \nabla f^0 + H_f(x_0)(x_1 - x_0) \quad \text{or} \quad \begin{cases} 0 = f_1^0 + \sum_{j=1}^{n} f_{1j}^0 (x_j^1 - x_j^0), \\ 0 = f_2^0 + \sum_{j=1}^{n} f_{2j}^0 (x_j^1 - x_j^0), \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ 0 = f_n^0 + \sum_{j=1}^{n} f_{nj}^0 (x_j^1 - x_j^0), \end{cases}$$

or, upon transposing,

$$\boldsymbol{H}_{\mathrm{f}}(\boldsymbol{x}_0)\boldsymbol{x}_1 = \boldsymbol{H}_{\mathrm{f}}(\boldsymbol{x}_0)\boldsymbol{x}_0 - \nabla \mathrm{f}^0. \tag{B.2}$$

Here $\boldsymbol{H}_{\mathrm{f}}(\boldsymbol{x}_0)$ is the nth-order Hessian matrix of f evaluated at $\boldsymbol{x}_0$, and $\nabla \mathrm{f}^0$ is the $(n \times 1)$ gradient vector of f at $\boldsymbol{x}_0$. If $\boldsymbol{H}_{\mathrm{f}}(\boldsymbol{x}_0)$ is non-singular, we obtain, from eq. (B.2)

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - \boldsymbol{H}_{\mathrm{f}}^{-1}(\boldsymbol{x}_0)\,(\nabla \mathrm{f}^0).$$

Round one of this iterative procedure is now complete.

In general, after i iterations,

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} - \boldsymbol{H}_{\mathrm{f}}^{-1}(\boldsymbol{x}_{i-1})\,(\nabla \mathrm{f}(\boldsymbol{x}_{i-1})), \quad i = 1, 2, \ldots \tag{B.3}$$

provided $\boldsymbol{H}_{\mathrm{f}}(\boldsymbol{x}_{i-1})$ is non-singular. Additionally, this process converges to a strong local maximum (minimum) provided $\boldsymbol{H}_{\mathrm{f}}(\boldsymbol{x}_{i-1})$ is negative (positive) definite and the change in the function is less than some predetermined limit η, i.e. $|\mathrm{f}(\boldsymbol{x}_i) - \mathrm{f}(\boldsymbol{x}_{i-1})| < \eta$ for several successive values of i.

Example B.1. Does the real-valued function

$$y = \mathrm{f}(\boldsymbol{x}) = -x_1^4 + 2x_1x_2 - 3x_2^3 + 3x_1 + x_2 - 6$$

attain a local extremum over $\mathscr{E}^2$? Setting $\nabla \mathrm{f} = \boldsymbol{0}$ we obtain the system

$$\begin{aligned} -4x_1^3 + 2x_2 + 3 &= 0, \\ 2x_1 - 9x_2^2 + 1 &= 0, \end{aligned}$$

which must be solved simultaneously to find the critical point $\boldsymbol{x}_* \in \mathscr{E}^2$ where $\nabla \mathrm{f}^* = \boldsymbol{0}$. We first find

$$\boldsymbol{H}_{\mathrm{f}}(\boldsymbol{x}) = \begin{bmatrix} -12x_1^2 & 2 \\ 2 & -18x_2 \end{bmatrix}.$$

Then, from eq. (B.3) we have, for $i = 1, 2$ and $\boldsymbol{x}_0' = (1, 1)$,

$$\begin{aligned} \boldsymbol{x}_1 &= \boldsymbol{x}_0 - \boldsymbol{H}_{\mathrm{f}}^{-1}(\boldsymbol{x}_0)\,(\nabla \mathrm{f}(\boldsymbol{x}_0)) \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -0.0849 & -0.0094 \\ -0.0094 & -0.0566 \end{bmatrix} \begin{bmatrix} 1 \\ -6 \end{bmatrix} = \begin{bmatrix} 1.0285 \\ 0.6698 \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} \boldsymbol{x}_2 &= \boldsymbol{x}_1 - \boldsymbol{H}_{\mathrm{f}}^{-1}(\boldsymbol{x}_1)\,(\nabla \mathrm{f}(\boldsymbol{x}_1)) \\ &= \begin{bmatrix} 1.0285 \\ 0.6698 \end{bmatrix} - \begin{bmatrix} -0.0809 & -0.0131 \\ -0.0131 & -0.0831 \end{bmatrix} \begin{bmatrix} -0.0120 \\ -0.9804 \end{bmatrix} = \begin{bmatrix} 1.0147 \\ 0.5881 \end{bmatrix}. \end{aligned}$$

Hence, $\boldsymbol{x}_*' \sim (1.0147, 0.5881)$. Since the naturally ordered principal minors of the Hessian matrix alternate in sign (the first being negative) at $\boldsymbol{x}_*$, f assumes a strong local maximum there, i.e.

$$M_1 = -12.35 < 0, \qquad M_2 = |\boldsymbol{H}_{\mathrm{f}}(1.0147, 0.5881)| = 130.27 > 0.$$

6.13. Appendix C: The method of steepest ascent

The Newton–Raphson iteration technique may be described as an 'indirect method' of determining an extremum of a real-valued function, i.e. its implementation involves the application of a set of supplemental (necessary) conditions which must be satisfied at the maximum or minimum of the function. An alternative technique involves what may be described as a 'direct method'. This process makes use of an initial estimate from which information about the behavior of the function is used to determine locally a direction in which the function increases or decreases. The direct method which we shall consider here may generally be characterized as a process involving 'gradient search'. Specifically, it is Cauchy's *method of steepest ascent (descent)*. In what follows we shall attempt to maximize the real-valued function $y = f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathscr{K} \subseteq \mathscr{E}^n$, by the method of steepest ascent. That is to say, we seek to approximate the coordinates of a point $\boldsymbol{x}_* \in \mathscr{K}$ such that $f(\boldsymbol{x}_*) \geqslant f(\boldsymbol{x})$, $\boldsymbol{x} \in \delta(\boldsymbol{x}_*)$. The modification for handling minimization problems is obvious and will not be presented in detail.

We noted previously that the gradient vector ∇f points locally in the direction of maximum increase of f. Hence we may: (1) start at some initial point $\boldsymbol{x}_0 \in \mathscr{K}$ and compute ∇f^0; (2) take a step in the direction of steepest ascent ∇f^0, using a step length λ_0, to obtain a new point $\boldsymbol{x}_1$ (here the search parameter λ_0 is interpreted as some scalar multiple of ∇f^0); and (3) repeat the process until the desired degree of accuracy to our approximation of $\boldsymbol{x}_*$ is obtained. The iteration scheme may thus be described as

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \lambda_{i-1} \nabla f(\boldsymbol{x}_{i-1}), \quad i = 1, 2, \ldots, \tag{C.1}$$

with the process converging to a strong local maximum if the λ_{i-1} are chosen so that $f(\boldsymbol{x}_i) > f(\boldsymbol{x}_{i-1})$, i.e. the function is made to increase with each step. Since f is increasing locally in the direction of $\nabla f(\boldsymbol{x}_{i-1})$, we can be sure that there always exists a $\lambda_{i-1} > 0$ such that $f(\boldsymbol{x}_i) > f(\boldsymbol{x}_{i-1})$. The process is then stopped when the change in the function is less than some predetermined limit η or $|f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i-1})| < \eta$ for several successive values of i. Since our direction of movement throughout $\mathscr{K}$ is specified by the gradient of f at $\boldsymbol{x}_{i-1}$, it thus remains to determine λ_{i-1}.

Utilizing eq. (5.15) we may write the quadratic approximation of f at $\boldsymbol{x}_{i-1}$ as

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_{i-1}) + \nabla f(\boldsymbol{x}_{i-1})' \boldsymbol{h} + \frac{1}{2!} \boldsymbol{h}' \boldsymbol{H}_f(\boldsymbol{x}_{i-1}) \boldsymbol{h}, \tag{C.2}$$

where $\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{x}_{i-1}$. In light of the iteration scheme described by eq. (C.1) let us set

$$x = \boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \lambda_{i-1} \nabla f(\boldsymbol{x}_{i-1})$$

in eq. (C.2), thus obtaining, for

$$\begin{aligned} \boldsymbol{h} &= \boldsymbol{x}_i - \boldsymbol{x}_{i-1} = \lambda_{i-1} \nabla f(\boldsymbol{x}_{i-1}), \\ f(\boldsymbol{x}_i) &= f(\boldsymbol{x}_{i-1}) + \lambda_{i-1} \nabla f(\boldsymbol{x}_{i-1})' \nabla f(\boldsymbol{x}_{i-1}) \\ &\quad + \frac{\lambda_{i-1}^2}{2!} \nabla f(\boldsymbol{x}_{i-1})' \boldsymbol{H}_f(\boldsymbol{x}_{i-1}) \nabla f(\boldsymbol{x}_{i-1}). \end{aligned} \tag{C.3}$$

How should λ_{i-1} be chosen? Since we require a λ_{i-1} for which $f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i-1}) > 0$ is as large as possible, let us determine λ_{i-1} according to the criterion

$$f(\boldsymbol{x}_{i-1} + \lambda_{i-1} \nabla f(\boldsymbol{x}_{i-1})) = \max_{\lambda} f(\boldsymbol{x}_{i-1} + \lambda \nabla f(\boldsymbol{x}_{i-1})), \tag{C.4}$$

i.e. λ_{i-1} is chosen by maximizing f along the $\nabla f(\boldsymbol{x}_{i-1})$ direction. Then

$$\begin{aligned} &\left. \frac{df(\boldsymbol{x}_{i-1} + \lambda \nabla f(\boldsymbol{x}_{i-1}))}{d\lambda} \right|_{\lambda = \lambda_{i-1}} \\ &= \nabla f(\boldsymbol{x}_{i-1})' \nabla f(\boldsymbol{x}_{i-1}) + \lambda_{i-1} \nabla f(\boldsymbol{x}_{i-1})' \boldsymbol{H}_f(\boldsymbol{x}_{i-1}) \nabla f(\boldsymbol{x}_{i-1}) = 0 \end{aligned}$$

or

$$\lambda_{i-1} = - \frac{\nabla f(\boldsymbol{x}_{i-1})' \nabla f(\boldsymbol{x}_{i-1})}{\nabla f(\boldsymbol{x}_{i-1})' \boldsymbol{H}_f(\boldsymbol{x}_{i-1}) \nabla f(\boldsymbol{x}_{i-1})}.$$

Additionally, $f(\boldsymbol{x}_{i-1} + \lambda \nabla f(\boldsymbol{x}_{i-1}))$ will attain a maximum if

$$\frac{d^2 f(\boldsymbol{x}_{i-1} + \lambda \nabla f(\boldsymbol{x}_{i-1}))}{d\lambda^2} = \nabla f(\boldsymbol{x}_{i-1})' \boldsymbol{H}_f(\boldsymbol{x}_{i-1}) \nabla f(\boldsymbol{x}_{i-1}) < 0,$$

i.e. if $\boldsymbol{H}_f(\boldsymbol{x}_{i-1})$ is negative definite.

In general, the various iterations in the method of steepest ascent may be described as

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \lambda_{i-1} \nabla f(\boldsymbol{x}_{i-1}), \quad i = 1, 2, \ldots, \quad \text{(steepest ascent)} \tag{C.5}$$

where

$$\lambda_{i-1} = - \frac{\nabla f(\boldsymbol{x}_{i-1})' \nabla f(\boldsymbol{x}_{i-1})}{\nabla f(\boldsymbol{x}_{i-1})' \boldsymbol{H}_f(\boldsymbol{x}_{i-1}) \nabla f(\boldsymbol{x}_{i-1})}$$

is chosen to maximize f along $\nabla f(\boldsymbol{x}_{i-1})$. Moreover, convergence to a strong local maximum of f is assured if $\boldsymbol{H}_f(\boldsymbol{x}_{i-1})$ is negative definite.

If we desire to minimize f over $\mathcal{X}$, then we must employ the method of steepest descent. The modification is straightforward. Since $-\nabla f$ points locally in the direction of maximum decrease of f, $-\nabla f(\boldsymbol{x}_{i-1})$ replaces $\nabla f(\boldsymbol{x}_{i-1})$ in eq. (C.1). Hence the various rounds of the method of steepest

descent may be characterized as

$$x_i = x_{i-1} - \lambda_{i-1}\nabla f(x_{i-1}), \quad i = 1, 2, \ldots, \quad \text{(steepest descent)} \quad \text{(C.5.1)}$$

where

$$\lambda_{i-1} = \frac{\nabla f(x_{i-1})'\nabla f(x_{i-1})}{\nabla f(x_{i-1})' H_f(x_{i-1})\nabla f(x_{i-1})}$$

is chosen to minimize f along $-\nabla f(x_{i-1})$.[12] In this instance the process will converge to a strong local minimum of f if $H_f(x_{i-1})$ is positive definite.

Example C.1. Given $y = f(x)$, $x_0' = (1, 1)$, from example B.1, p. 141, find x_3 using the method of steepest ascent (eq. (C.5)). Since

$$\lambda_0 = -\frac{\nabla f(x_0)'\nabla f(x_0)}{\nabla f(x_0)' H_f(x_0)\nabla f(x_0)} = -\frac{(1, -6)\begin{bmatrix} 1 \\ -6 \end{bmatrix}}{(1, -6)\begin{bmatrix} -12 & 2 \\ 2 & -18 \end{bmatrix}\begin{bmatrix} 1 \\ -6 \end{bmatrix}}$$

$$= 0.0551,$$

it follows that

$$x_1 = x_0 + \lambda_0\nabla f(x_0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.0551\begin{bmatrix} 1 \\ -6 \end{bmatrix} = \begin{bmatrix} 1.0551 \\ 0.6694 \end{bmatrix}.$$

With

$$\lambda_1 = -\frac{\nabla f(x_1)'\nabla f(x_1)}{\nabla f(x_1)' H_f(x_1)\nabla f(x_1)}$$

$$= -\frac{(-0.3592, 0.9227)\begin{bmatrix} -0.3592 \\ 0.9227 \end{bmatrix}}{(-0.3592, 0.9227)\begin{bmatrix} -13.3584 & 2 \\ 2 & -12.0492 \end{bmatrix}\begin{bmatrix} -0.3592 \\ 0.9227 \end{bmatrix}}$$

$$= 0.0737,$$

$$x_2 = x_1 + \lambda_1\nabla f(x_1) = \begin{bmatrix} 1.0551 \\ 0.6694 \end{bmatrix} + 0.0737\begin{bmatrix} -0.3592 \\ 0.9227 \end{bmatrix} = \begin{bmatrix} 1.0286 \\ 0.7374 \end{bmatrix}.$$

Also, with

$$\lambda_2 = -\frac{\nabla f(x_2)'\nabla f(x_2)}{\nabla f(x_2)' H_f(x_2)\nabla f(x_2)}$$

---

[12] For a discussion on the conditions underlying the existence of a solution to an extremum problem via this method with a variety of proofs of convergence to the same, see Saaty and Bram (1964), pp. 76–88.

$$= -\frac{(0.1216, -1.8370)\begin{bmatrix} 0.1216 \\ -1.8370 \end{bmatrix}}{(0.1216, -1.8370)\begin{bmatrix} -12.6960 & 2 \\ 2 & -13.2732 \end{bmatrix}\begin{bmatrix} 0.1216 \\ -1.8370 \end{bmatrix}}$$

$$= 0.0739,$$

$$\boldsymbol{x}_3 = \boldsymbol{x}_2 + \lambda_2 \nabla f(\boldsymbol{x}_2) = \begin{bmatrix} 1.0286 \\ 0.7374 \end{bmatrix} + 0.0739 \begin{bmatrix} 0.1216 \\ -1.8370 \end{bmatrix} = \begin{bmatrix} 1.0376 \\ 0.6016 \end{bmatrix}.$$

Notice that for this particular example the successive steps of the steepest ascent method tend to be somewhat erratic, i.e. an inefficient zigzag pattern unfolds because the contours of f are non-spherical. In such instances the direction of the gradient is not coincident with the direction to the maximum (minimum) with the result that convergence is slow.

What is the relationship between the Newton–Raphson and steepest ascent methods? A glance at eqs. (B.3) and (C.5) indicates that both these equations are actually special cases of the more general expression

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \lambda_{i-1} \boldsymbol{M}(\boldsymbol{x}_{i-1}) \nabla f(\boldsymbol{x}_{i-1}), \quad i = 1, 2, \ldots, \tag{C.6}$$

where $\boldsymbol{M}(\boldsymbol{x}_{i-1})$ is termed an nth-order *deflection matrix* which serves to modify or deflect the gradient to a direction which leads to a greater total increase in the function than would be attained simply by moving locally in the direction of the gradient. In this regard, if $\boldsymbol{M}(\boldsymbol{x}_{i-1})$ equals the identity matrix $\boldsymbol{I}_n$ for all values of i, then eq. (C.5) obtains, while if $\lambda_{i-1} = 1$ for all i and $\boldsymbol{M}(\boldsymbol{x}_{i-1}) = -\boldsymbol{H}_f^{-1}(\boldsymbol{x}_{i-1})$, eq. (B.3) results. Hence it can be seen that the Newton–Raphson method is essentially a modified steepest ascent technique.[13]

---

[13] The preceding discussion has hinted at a procedure which may be used to generalize the Newton–Raphson technique (eq. (B.3)). If in eq. (C.6) we let $\boldsymbol{M}(\boldsymbol{x}_{i-1}) = -\boldsymbol{H}_f^{-1}(\boldsymbol{x}_{i-1})$, then, starting at some initial point $\boldsymbol{x}_0$, we may take a step in the direction $-\boldsymbol{H}_f^{-1}(\boldsymbol{x}_0)\nabla f(\boldsymbol{x}_0)$, using a step length λ_0 which is chosen to maximize f along $-\boldsymbol{H}_f^{-1}(\boldsymbol{x}_0)\nabla f(\boldsymbol{x}_0)$, to obtain a new point

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - \lambda_0 \boldsymbol{H}_f^{-1}(\boldsymbol{x}_0)\nabla f(\boldsymbol{x}_0).$$

In general, the ith iteration is determined as

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} - \lambda_{i-1} \boldsymbol{H}_f^{-1}(\boldsymbol{x}_{i-1})\nabla f(\boldsymbol{x}_{i-1}), \quad i = 1, 2, \ldots,$$

where λ_{i-1} is chosen so that

$$f(\boldsymbol{x}_{i-1} - \lambda_{i-1} \boldsymbol{H}_f^{-1}(\boldsymbol{x}_{i-1})\nabla f(\boldsymbol{x}_{i-1})) = \max_{\lambda} f(\boldsymbol{x}_{i-1} - \lambda \boldsymbol{H}_f^{-1}(\boldsymbol{x}_{i-1})\nabla f(\boldsymbol{x}_{i-1})).$$

6.14. Appendix D: Conjugate direction methods

6.14.1. The Fletcher–Powell (variable-metric) method

An extremely efficient deflected-gradient technique is the method of Fletcher and Powell (Fletcher and Powell, 1963, pp. 163–168). To gain some insight into the motivation underlying their technique, let us re-examine the Newton–Raphson and gradient schemes depicted by eqs. (B.3) and (C.5) respectively. Note that in these two processes the Hessian of f is computed at each iteration, a task which may at times require a considerable amount of effort. Moreover, once a particular iteration is executed, the information provided by such a calculation is discarded, i.e. none of the information previously obtained regarding the behavior of f near an extremum is stored and then used to implement further iterations which may hasten convergence to the desired critical value of $\boldsymbol{x}$. To improve upon these shortcomings Fletcher and Powell have developed a modified or accelerated steepest ascent method which relies solely upon first-order information, as generated by the gradients of f at several different points, to construct a second-order approximation to f in the neighborhood of the extreme point, i.e. a quadratic approximation to f at the said point is obtained without explicitly calculating the Hessian of f there.[14] Fletcher and Powell develop their technique for the case where the function to be maximized is quadratic in $\boldsymbol{x}$, the reason being that if an iterative process can be found which possesses quadratic convergence, i.e. it quickly and efficiently maximizes a quadratic function in a finite number of steps, then that same technique should work well on a general function which behaves like (can be closely approximated by) a quadratic in the vicinity of the extremum.

Let us assume then that the real-valued function $y = \mathrm{f}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathscr{E}^n$, is differentiable over an open region $\mathscr{K} \subseteq \mathscr{E}^n$ and quadratic in $\boldsymbol{x}$. Then

$$\mathrm{f}(\boldsymbol{x}) = a + \boldsymbol{b}'\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x},$$

where $\boldsymbol{b}$ is an $(n \times 1)$ vector and the $(n \times n)$ Hessian matrix $\boldsymbol{Q}$ is taken to be negative definite and non-singular. At an arbitrary point $\boldsymbol{x}_0 \in \mathscr{K}$, $\nabla \mathrm{f}(\boldsymbol{x}_0) = \boldsymbol{b} + \boldsymbol{Q}\boldsymbol{x}_0$. Then $\boldsymbol{x}_0 = \boldsymbol{Q}^{-1}(\nabla \mathrm{f}(\boldsymbol{x}_0) - \boldsymbol{b})$. If f attains a strong local maximum at $\boldsymbol{x}_* \in \mathscr{K}$, then $\nabla \mathrm{f}(\boldsymbol{x}_*) = \boldsymbol{0}$ so that $\boldsymbol{x}_* = -\boldsymbol{Q}^{-1}\boldsymbol{b}$. Upon subtracting $\boldsymbol{x}_0$ from $\boldsymbol{x}_*$ we obtain

$$\boldsymbol{x}_* - \boldsymbol{x}_0 = -\boldsymbol{Q}^{-1}\nabla \mathrm{f}(\boldsymbol{x}_0). \tag{D.1}$$

Here the difference $\boldsymbol{x}_* - \boldsymbol{x}_0$ simply depicts the single step traversed to the

[14] An iterative technique which does not rely upon first-order information and which performs almost as well as the Fletcher–Powell method is that developed by Powell (1964), pp. 155–162.

maximum of f from an arbitrary point x_0 at which the gradient of f is known. Let us now assume that the inverse of the Hessian Q^{-1} is unknown so that eq. (D.1) cannot be applied directly. What we would like to do is to piece together information about the curvature of f which yields something like eq. (D.1). But first let us consider

DEFINITION D.1. *Given an* $(n \times n)$ *negative-definite matrix* Q, *the directions* $s_i \in \mathscr{E}^n$, $i = 0, 1, \ldots, n-1$, *are mutually* Q*-conjugate if they are non-null and*

$$s_i' Q s_j = 0, \quad i \neq j, \quad i, j = 0, 1, \ldots, n-1. \tag{D.2}$$

In addition, an important property of such directions is stated in

THEOREM D.1. *Let* Q *be an* $(n \times n)$ *negative-definite matrix. If n non-null directions* s_i, $i = 0, 1, \ldots, n-1$, *are mutually* Q*-conjugate, then they are also linearly independent.*

Proof. Let us assume to the contrary that the s_i are linearly dependent. Then some particular s_i, say the kth, is expressible as a linear combination of the remaining s_i's, i.e.

$$s_k = \sum_{\substack{i=0 \\ i \neq k}}^{n-1} \theta_i s_i,$$

where at least one of the θ_i's, say the jth, is different from zero. Since the s_i are mutually Q-conjugate, $s_j' Q s_k = 0$. Then

$$s_j' Q s_k = s_j' Q \left(\sum_{\substack{i=0 \\ i \neq k}}^{n-1} \theta_i s_i \right) = \theta_j s_j' Q s_j \neq 0.$$

Since the expressions $s_j' Q s_k = 0$, $s_j' Q s_k \neq 0$ are contradictory, the s_i must be linearly independent. Q.E.D.

In this regard, the Fletcher–Powell method maximizes a quadratic function of n variables in exactly n iterations by generating a sequence of n mutually conjugate directions $s_0, s_1, \ldots, s_{n-1}$, the rationalization being that if we search locally in the s_k-direction and find a point x_k which maximizes f, and then search locally in the conjugate direction s_{k+1} and determine a point x_{k+1} which does the same, then the value of f at x_{k+1} cannot be increased by searching again in the s_k-direction. Hence we need search locally along each of the directions s_i, $i = 0, 1, \ldots, n-1$, only once.

Let us begin by considering the iteration scheme depicted by eq. (C.6), namely

$$x_i = x_{i-1} + \lambda_{i-1} M(x_{i-1}) \nabla f(x_{i-1}), \quad i = 1, 2, \ldots, n$$

or

$$x_i = x_{i-1} + \lambda_{i-1} s_{i-1}, \quad i = 1, 2, \ldots, n,$$

where the directions $s_{i-1} = M(x_{i-1})\nabla f(x_{i-1})$ satisfy eq. (D.2). From some initial estimate of x_*, x_0, we construct a direction $s_0 = -H_0 \nabla f(x_0)$, where the $(n \times n)$ matrix H_0 is chosen to be symmetric and negative definite so that movement locally along s_0 yields an increase in f. Hence the role of H_0 is simply to supply the current direction of motion. (Note that if $-H_0$ is the identity matrix, then our initial step is in the direction of steepest ascent. In all other instances H_0 serves to deflect the direction of ascent to a path different from that of the gradient.) Thus x_1 is determined as

$$x_1 = x_0 + \lambda_0 s_0, \qquad s_0 = -H_0 \nabla f(x_0),$$

with the search parameter λ_0 chosen to maximize $f(x_1)$ along s_0, i.e.

$$f(x_0 + \lambda_0 s_0) = \max_{\lambda} f(x_0 + \lambda s_0).$$

With H_0 negative definite,

$$\frac{d}{d\lambda} f(x_0 + \lambda s_0) \Big|_{\lambda = 0} = -\nabla f(x_0)' H_0 \nabla f(x_0) > 0.$$

Hence we are assured that f actually increases locally in the direction of s_0.

In general, the ith iteration is determined as

$$x_i = x_{i-1} + \lambda_{i-1} s_{i-1}, \quad i = 1, 2, \ldots, n, \tag{D.3}$$

where the directions $s_{i-1} = -H_{i-1} \nabla f(x_{i-1})$ are mutually Q conjugate, H_{i-1} is symmetric and negative definite, and, starting at x_{i-1}, λ_{i-1} is chosen so that

$$f(x_{i-1} + \lambda_{i-1} s_{i-1}) = \max_{\lambda} f(x_{i-1} + \lambda s_{i-1}). \tag{D.4}$$

As far as the quadratic convergence of eq. (D.3) to the maximum of f is concerned, we shall demonstrate that if eq. (D.2) is satisfied, $\nabla f(x_n)$ vanishes so that the maximum is attained at the nth step. From eq. (D.3) it is easily shown that $x_n = x_0 + \sum_{i=0}^{n-1} \lambda_i s_i$ or, for our purposes,

$$x_n = x_j + \sum_{i=j}^{n-1} \lambda_i s_i = x_j + \lambda_j s_j + \sum_{i=j+1}^{n-1} \lambda_i s_i, \quad j = 0, 1, \ldots, n-1.$$

Upon substituting this latter expression into $\nabla f(x) = b + Qx$ we obtain

$$\nabla f(x_n) = b + Q(x_j + \lambda_j s_j) + \sum_{i=j+1}^{n-1} \lambda_i Q s_i = \nabla f(x_{j+1}) + \sum_{i=j+1}^{n-1} \lambda_i Q s_i,$$

$$j = 0, 1, \ldots, n-1.$$

Then

$$s_j'\nabla f(x_n) = s_j'\nabla f(x_{j+1}) + \sum_{i=j+1}^{n-1} \lambda_i s_j Q s_i = \sum_{i=j+1}^{n-1} \lambda_i s_j Q s_i$$

since successive steps are mutually orthogonal, i.e. at the maximum indicated by eq. (D.4),

$$\left.\frac{df(x_{i-1}+\lambda s_{i-1})}{d\lambda}\right|_{\lambda=\lambda_{i-1}} = s_{i-1}'\nabla f(x_i) = 0, \quad i = 1, 2, \ldots, n.$$

If the directions s_i, $i = 0, 1, \ldots, n-1$, satisfy eq. (D.2), then $s_j'\nabla f(x_n) = 0$. Since the s_i, $i = 0, 1, \ldots, n-1$, are linearly independent, $\nabla f(x_n)$ is orthogonal to s_i, $i = 0, 1, \ldots, n-1$, only if $\nabla f(x_n) \equiv 0$. Hence the maximum of f is attained at the nth iteration when we search along a set of n mutually Q-conjugate directions.

Since the iterative process described by eq. (D.3) is to converge to something which resembles eq. (D.1) in n steps, we must have $H_n = Q^{-1}$. To see exactly how the sequence of deflection matrices H_i, $i = 1, \ldots, n$, actually converges to Q^{-1}, let us consider the matrix difference equation

$$H_i = H_{i-1} + A_i + B_i, \quad i = 1, \ldots, n, \tag{D.5}$$

where it is stipulated that $H_n = Q^{-1}$. We first form the sum

$$\sum_{i=1}^{n} H_i = \sum_{i=1}^{n} H_{i-1} + \sum_{i=1}^{n} A_i + \sum_{i=1}^{n} B_i$$

or

$$H_n = H_0 + \sum_{i=1}^{n} A_i + \sum_{i=1}^{n} B_i$$

so that

$$Q^{-1} = H_0 + \sum_{i=1}^{n} A_i + \sum_{i=1}^{n} B_i. \tag{D.5.1}$$

As we shall now see, the role of eq. (D.5) is twofold. As our iterations progress we desire: (1) to improve our initial estimate of Q^{-1}, H_0, by continually updating the information regarding the curvature of f at x_* obtained at each successive step; and (2) to cancel out the effects of a poor choice of H_0. If we set

$$\sum_{i=1}^{n} B_i = -H_0, \tag{D.6}$$

then

$$\sum_{i=1}^{n} A_i = Q^{-1}. \tag{D.7}$$

Thus the matrices A_i, $i = 1, \ldots, n$, serve to systematically generate Q^{-1} in n steps according to eq. (D.7), whereas the B_i, $i = 1, \ldots, n$, matrices tend to gradually eliminate the influence of H_0 by virtue of eq. (D.6).

Our final step is to determine how the A_i, B_i matrices are computed. We indicated earlier that a second-order approximation of f at x_* would be achieved by piecing together information about the curvature of f obtained from its gradients at two successive points. To this end let us form the gradient difference vector

$$d_i = \nabla f(x_i) - \nabla f(x_{i-1}), \quad i = 1, \ldots, n. \tag{D.8}$$

From our initial expression for $f(x)$ we obtain $\nabla f(x) = b + Qx$. Coupling this result with eq. (D.8) for $x = x_i, x_{i-1}$ yields

$$d_i = Q(x_i - x_{i-1}) = Q\sigma_i, \tag{D.8.1}$$

where $\sigma_i = \lambda_{i-1} s_{i-1}$ (eq. (D.3)). In addition, from eqs. (D.7) and (D.8),

$$\sigma_i = Q^{-1} Q \sigma_i = \sum_{k=1}^{n} A_k d_i = A_i d_i \tag{D.9}$$

(here $A_k d_i = 0$, $k \neq i$. For a proof on this account, see Fletcher and Powell (1963), p. 165). Then

$$\sigma_i = \sigma_i \left(\frac{\sigma_i' d_i}{\sigma_i' d_i} \right) = \left(\frac{\sigma_i' \sigma_i}{\sigma_i' d_i} \right) d_i$$

and thus, from eq. (D.9),

$$A_i = (\sigma_i \sigma_i')/(\sigma_i' d_i). \tag{D.10}$$

We now obtain, from eqs. (D.5), (D.8.1), and (D.9),

$$H_i Q \sigma_i = H_i d_i = H_{i-1} d_i + \sigma_i + B_i d_i. \tag{D.11}$$

Since the directions s_i, $i = 0, 1, \ldots, n-1$, are linearly independent, successive directions are related to H_i by $H_i Q \sigma_i = \sigma_i$, $i = 1, 2, \ldots, n$ (Fletcher and Powell, 1963, p. 165). Hence, eq. (D.11) becomes

$$B_i d_i = -H_i d_i = -H_{i-1} d_i \left(\frac{d_i' H_{i-1} d_i}{d_i' H_{i-1} d_i} \right) = -\left(\frac{H_{i-1} d_i d_i' H_{i-1}}{d_i' H_{i-1} d_i} \right) d_i,$$

whence

$$B_i = -\frac{H_{i-1} d_i d_i' H_{i-1}}{d_i' H_{i-1} d_i}. \tag{D.12}$$

The Fletcher–Powell method is now completely defined and successive iterations may be carried out as follows:

(1) to maximize f, choose an initial point x_0 and a negative-definite

matrix $\boldsymbol{H}_0$ (for convenience, let $\boldsymbol{H}_0 = -\boldsymbol{I}_n$);
(2) compute $\nabla f(\boldsymbol{x}_0)$;
(3) calculate a direction in which to move $\boldsymbol{s}_0 = -\boldsymbol{H}_0 \nabla f(\boldsymbol{x}_0)$;
(4) to move along $\boldsymbol{s}_0$, compute a step length λ_0 so that

$$f(\boldsymbol{x}_0 + \lambda_0 \boldsymbol{s}_0) = \max_{\lambda} f(\boldsymbol{x}_0 + \lambda \boldsymbol{s}_0);$$

(5) compute $\boldsymbol{\sigma}_1 = \lambda_0 \boldsymbol{s}_0$, $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{\sigma}_1$;
(6) compute $\nabla f(\boldsymbol{x}_1)$, $\boldsymbol{d}_1 = \nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_0)$;
(7) compute

$$\boldsymbol{A}_1 = \frac{\boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1'}{\boldsymbol{\sigma}_1' \boldsymbol{d}_1}, \qquad \boldsymbol{B}_1 = -\frac{\boldsymbol{H}_0 \boldsymbol{d}_1 \boldsymbol{d}_1' \boldsymbol{H}_0}{\boldsymbol{d}_1' \boldsymbol{H}_0 \boldsymbol{d}_1},$$

and thus

$$\boldsymbol{H}_1 = \boldsymbol{H}_0 + \boldsymbol{A}_1 + \boldsymbol{B}_1;$$

(8) repeat steps (1)–(7), increasing each subscript by 1 at the beginning of each round until $|f(\boldsymbol{x}_i) - f(\boldsymbol{x}_{i-1})| < \eta$, where η is set at some predetermined level.

If the minimum of f is to be determined, $\boldsymbol{H}_0$ is chosen as positive definite ($\boldsymbol{H}_0 = \boldsymbol{I}_n$ will do) and we minimize in step (4) rather than maximize.

When the Fletcher–Powell method is applied to a general function, convergence to the desired extremum will not occur in exactly n steps as in the pure quadratic case. In this instance the process should be thought of as one involving the generation of conjugate directions for a quadratic approximation of f. As the iterations progress and the sequence of $\boldsymbol{H}_n$ matrices yield increasingly better estimates of the curvature of f at the extremum, the rate of convergence accelerates as soon as the process gets reasonably close to a second-order approximation.

Example D.1. Let the real-valued function

$$y = f(\boldsymbol{x}) = \boldsymbol{b}'\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x}$$

be defined for all $\boldsymbol{x} \in \mathscr{E}^2$, where

$$\boldsymbol{b} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}, \qquad \boldsymbol{Q} = \begin{bmatrix} -1 & 0 \\ 0 & -1.5 \end{bmatrix}.$$

Maximize f using the Fletcher–Powell technique given that $\boldsymbol{x}_0' = (1, 1)$, $\boldsymbol{H}_0 = -\boldsymbol{I}_2$. We first find

$$\nabla f(\boldsymbol{x}_0) = \boldsymbol{b} + \boldsymbol{Q}\boldsymbol{x}_0 = \begin{bmatrix} 4 \\ 2.5 \end{bmatrix}, \qquad \boldsymbol{s}_0 = -\boldsymbol{H}_0 \nabla f(\boldsymbol{x}_0) = \begin{bmatrix} 4 \\ 2.5 \end{bmatrix}.$$

To determine λ_0, let us maximize

$$\begin{aligned} f(\boldsymbol{x}_0+\lambda \boldsymbol{s}_0) &= \boldsymbol{b}'(\boldsymbol{x}_0+\lambda \boldsymbol{s}_0)+\tfrac{1}{2}(\boldsymbol{x}_0+\lambda \boldsymbol{s}_0)'\boldsymbol{Q}(\boldsymbol{x}_0+\lambda \boldsymbol{s}_0) \\ &= 7.75+22.25\lambda-12.6875\lambda^2. \end{aligned}$$

From

$$\frac{df(\boldsymbol{x}_0+\lambda \boldsymbol{s}_0)}{d\lambda} = 22.25-25.375\lambda = 0$$

we obtain $\lambda_0 = 0.87$ and thus

$$\boldsymbol{\sigma}_1 = \lambda_0 \boldsymbol{s}_0 = \begin{bmatrix} 3.480 \\ 2.175 \end{bmatrix}, \quad \boldsymbol{x}_1 = \boldsymbol{x}_0+\boldsymbol{\sigma}_1 = \begin{bmatrix} 4.480 \\ 3.175 \end{bmatrix}.$$

To initiate the next iteration we compute

$$\begin{aligned} \nabla f(\boldsymbol{x}_1) &= \boldsymbol{b}+\boldsymbol{Q}\boldsymbol{x}_1 = \begin{bmatrix} 0.5200 \\ -0.7625 \end{bmatrix}, \\ \boldsymbol{d}_1 &= \nabla f(\boldsymbol{x}_1)-\nabla f(\boldsymbol{x}_0) = \begin{bmatrix} -3.4800 \\ -3.2625 \end{bmatrix}. \end{aligned}$$

From

$$\begin{aligned} \boldsymbol{A}_1 &= \frac{\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1'}{\boldsymbol{\sigma}_1'\boldsymbol{d}_1} = \begin{bmatrix} -0.6305 & -0.3941 \\ -0.3941 & -0.2463 \end{bmatrix}, \\ \boldsymbol{B}_1 &= -\frac{\boldsymbol{H}_0\boldsymbol{d}_1\boldsymbol{d}_1'\boldsymbol{H}_0}{\boldsymbol{d}_1'\boldsymbol{H}_0\boldsymbol{d}_1} = \begin{bmatrix} 0.5322 & 0.4990 \\ 0.4990 & 0.4678 \end{bmatrix} \end{aligned}$$

we obtain

$$\begin{aligned} \boldsymbol{H}_1 &= \boldsymbol{H}_0+\boldsymbol{A}_1+\boldsymbol{B}_1 = \begin{bmatrix} -1.0983 & 0.1049 \\ 0.1049 & -0.7785 \end{bmatrix}, \\ \boldsymbol{s}_1 &= -\boldsymbol{H}_1\nabla f(\boldsymbol{x}_1) = \begin{bmatrix} 0.6510 \\ -0.6481 \end{bmatrix}. \end{aligned}$$

We next compute λ_1 by maximizing

$$\begin{aligned} f(\boldsymbol{x}_1+\lambda \boldsymbol{s}_1) &= \boldsymbol{b}'(\boldsymbol{x}_1+\lambda \boldsymbol{s}_1)+\tfrac{1}{2}(\boldsymbol{x}_1+\lambda \boldsymbol{s}_1)'\boldsymbol{Q}(\boldsymbol{x}_1+\lambda \boldsymbol{s}_1) \\ &= -0.913+1.0025\lambda-1.0538\lambda^2. \end{aligned}$$

To this end we have

$$\frac{df(\boldsymbol{x}_1+\lambda \boldsymbol{s}_1)}{d\lambda} = 1.0025-2.1076\lambda = 0$$

or $\lambda_1 = 0.48$ and thus

$$\boldsymbol{\sigma}_2 = \lambda_1 \boldsymbol{s}_1 = \begin{bmatrix} 0.3125 \\ -0.3111 \end{bmatrix}, \quad \boldsymbol{x}_2 = \boldsymbol{x}_1+\boldsymbol{\sigma}_2 = \begin{bmatrix} 4.7925 \\ 2.8639 \end{bmatrix} \sim \boldsymbol{x}_*.$$

The next example is designed to provide the reader with a convenient numerical procedure for determining the value of the search parameter λ in situations where the function to be maximized or minimized is not quadratic in $\boldsymbol{x}$.

Example D.2. Minimize the real-valued function

$$y = \mathrm{f}(\boldsymbol{x}) = x_1^2 x_2 + x_2^2 + x_1 x_2, \quad \boldsymbol{x} \in \mathscr{E}^2,$$

using the Fletcher–Powell method. Upon choosing $\boldsymbol{x}_0' = (1, 1)$, $\boldsymbol{H}_0 = \boldsymbol{I}_2$, we first obtain

$$\nabla \mathrm{f}(\boldsymbol{x}_0) = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \qquad \boldsymbol{s}_0 = -\boldsymbol{H}_0 \nabla \mathrm{f}(\boldsymbol{x}_0) = \begin{bmatrix} -3 \\ -4 \end{bmatrix}.$$

Our next step is to minimize $\mathrm{f}(\boldsymbol{x}_0 + \lambda \boldsymbol{s}_0)$. Since f is not quadratic in $\boldsymbol{x}$, let us approximate $\mathrm{f}(\boldsymbol{x}_0 + \lambda \boldsymbol{s}_0)$ by a second-order polynomial in λ. Setting

$$\mathrm{f}(\boldsymbol{x}_0 + \lambda \boldsymbol{s}_0) = \mathrm{g}(\lambda) = a + b\lambda + c\lambda^2,$$

we have, for $\lambda = 0$,

$$\mathrm{g}(0) = a = \mathrm{f}(\boldsymbol{x}_0) = 3.$$

Additionally,

$$\left.\frac{\mathrm{dg}}{\mathrm{d}\lambda}\right|_{\lambda=0} = b = \left(\left.\frac{\partial \mathrm{f}}{\partial \boldsymbol{x}}\right|_{\lambda=0}\right)' \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda} = \nabla \mathrm{f}(\boldsymbol{x}_0)' \boldsymbol{s}_0 = -25.$$

To obtain c, let us arbitrarily set $\lambda = 1$. Then

$$\mathrm{g}(1) = \mathrm{f}(\boldsymbol{x}_0 + \boldsymbol{s}_0) = 3 = 3 - 25 + c \quad \text{or} \quad c = 25.$$

Hence

$$\mathrm{f}(\boldsymbol{x}_0 + \lambda \boldsymbol{s}_0) = 3 - 25\lambda + 25\lambda^2,$$
$$\frac{\mathrm{df}(\boldsymbol{x}_0 + \lambda \boldsymbol{s}_0)}{\mathrm{d}\lambda} = -25 + 50\lambda = 0$$

and thus $\lambda_0 = 0.5$. Then

$$\boldsymbol{\sigma}_1 = \lambda_0 \boldsymbol{s}_0 = \begin{bmatrix} -1.5 \\ -2 \end{bmatrix}, \qquad \boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{\sigma}_1 = \begin{bmatrix} -0.5 \\ -1 \end{bmatrix}.$$

To start the next round of calculations we first determine

$$\nabla \mathrm{f}(\boldsymbol{x}_1) = \begin{bmatrix} 0 \\ -2.25 \end{bmatrix}, \qquad \boldsymbol{d}_1 = \nabla \mathrm{f}(\boldsymbol{x}_1) - \nabla \mathrm{f}(\boldsymbol{x}_0) = \begin{bmatrix} -3 \\ -6.25 \end{bmatrix}.$$

Then

$$\boldsymbol{A}_1 = \frac{\boldsymbol{\sigma}_1 \boldsymbol{\sigma}_1'}{\boldsymbol{\sigma}_1' \boldsymbol{d}_1} = \begin{bmatrix} 0.1324 & 0.1765 \\ 0.1765 & 0.2353 \end{bmatrix},$$

$$\boldsymbol{B}_1 = -\frac{\boldsymbol{H}_0 \boldsymbol{d}_1 \boldsymbol{d}_1' \boldsymbol{H}_0}{\boldsymbol{d}_1' \boldsymbol{H}_0 \boldsymbol{d}_1} = \begin{bmatrix} -0.2043 & -0.3901 \\ -0.3901 & -0.8127 \end{bmatrix}$$

and thus

$$\boldsymbol{H}_1 = \boldsymbol{H}_0 + \boldsymbol{A}_1 + \boldsymbol{B}_1 = \begin{bmatrix} 0.9283 & -0.2136 \\ -0.2136 & 0.4226 \end{bmatrix},$$

$$\boldsymbol{s}_1 = -\boldsymbol{H}_1 \nabla \mathrm{f}(\boldsymbol{x}_1) = \begin{bmatrix} -0.4806 \\ 0.9506 \end{bmatrix}.$$

We will now minimize $\mathrm{f}(\boldsymbol{x}_1 + \lambda \boldsymbol{s}_1)$. Again a quadratic approximation to $\mathrm{f}(\boldsymbol{x}_1 + \lambda \boldsymbol{s}_1) = \mathrm{g}(\lambda)$ is in order. For $\lambda = 0$,

$$\mathrm{g}(0) = a = \mathrm{f}(\boldsymbol{x}_1) = 1.25.$$

Furthermore,

$$\left.\frac{\mathrm{dg}}{\mathrm{d}\lambda}\right|_{\lambda=0} = b = \nabla \mathrm{f}(\boldsymbol{x}_1)' \boldsymbol{s}_1 = -2.1389.$$

Finally, for $\lambda = 1$,

$$\mathrm{g}(1) = \mathrm{f}(\boldsymbol{x}_1 + \boldsymbol{s}_1) = 0.0033 = 1.25 - 2.1389 + c \quad \text{or} \quad c = 0.8922.$$

Then

$$\mathrm{f}(\boldsymbol{x}_1 + \lambda \boldsymbol{s}_1) = 1.25 - 2.1389\lambda + 0.8922\lambda^2,$$

$$\frac{\mathrm{df}(\boldsymbol{x}_1 + \lambda \boldsymbol{s}_1)}{\mathrm{d}\lambda} = -2.1389 + 1.7844\lambda = 0$$

and thus $\lambda_1 = 1.1987$. Hence

$$\boldsymbol{\sigma}_2 = \lambda_1 \boldsymbol{s}_1 = \begin{bmatrix} -0.5761 \\ 1.1395 \end{bmatrix}, \qquad \boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{\sigma}_2 = \begin{bmatrix} -1.0761 \\ 0.1398 \end{bmatrix}.$$

The above process is repeated until the successive iterations converge to $\boldsymbol{x}_*$.

6.14.2. The Fletcher–Reeves (conjugate gradient) method

We noted in the previous section that to undertake the various rounds of the Fletcher–Powell technique an $(n \times n)$ symmetric negative-definite matrix $\boldsymbol{H}_i$ had to be computed and updated at the next iteration to insure that $\boldsymbol{H}_n = \boldsymbol{Q}^{-1}$. From the viewpoint of economizing on time and effort, a method which does not store any such information but simply locates the maximum may at times be preferred. Such a process is that provided by

Fletcher and Reeves. If the function to be maximized is quadratic in $\boldsymbol{x}$, their method will also locate the extremum in n steps by generating a sequence of mutually conjugate directions.

Thus, in what follows we shall attempt to find the maximum of the differentiable real-valued function

$$y = \mathrm{f}(\boldsymbol{x}) = a + \boldsymbol{b}'\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x}, \quad \boldsymbol{x} \in \mathscr{E}^n,$$

over an open region $\mathscr{K} \subseteq \mathscr{E}^n$. Given the iteration scheme depicted by

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \lambda_{i-1}\boldsymbol{s}_{i-1}, \quad i = 1, 2, \ldots, n,$$

let us start from an arbitrary estimate of the critical point $\boldsymbol{x}_*$, $\boldsymbol{x}_0$, and search initially in the direction of steepest ascent. Then $\boldsymbol{x}_1$ is obtained as

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \lambda_0 \boldsymbol{s}_0, \qquad \boldsymbol{s}_0 = \nabla \mathrm{f}(\boldsymbol{x}_0),$$

where λ_0 is determined by maximizing $\mathrm{f}(\boldsymbol{x}_1)$ along $\boldsymbol{s}_0$, i.e.

$$\mathrm{f}(\boldsymbol{x}_0 + \lambda_0 \boldsymbol{s}_0) = \max_{\lambda} \mathrm{f}(\boldsymbol{x}_0 + \lambda \boldsymbol{s}_0).$$

At the implied maximum

$$\boldsymbol{s}_0' \nabla \mathrm{f}(\boldsymbol{x}_1) = \boldsymbol{s}_0'(\boldsymbol{b} + \boldsymbol{Q}(\boldsymbol{x}_0 + \lambda_0 \boldsymbol{s}_0)) = 0$$

or

$$\lambda_0 = -\frac{\boldsymbol{s}_0'(\boldsymbol{b} + \boldsymbol{Q}\boldsymbol{x}_0)}{\boldsymbol{s}_0'\boldsymbol{Q}\boldsymbol{s}_0} = -\frac{\boldsymbol{s}_0' \nabla \mathrm{f}(\boldsymbol{x}_0)}{\boldsymbol{s}_0'\boldsymbol{Q}\boldsymbol{s}_0}.$$

To execute succeeding iterations the search directions $\boldsymbol{s}_i$, $i > 0$, are chosen to equal the gradient at $\boldsymbol{x}_i$, $i > 0$, plus an appropriate linear combination of the previous directions. Hence the directions $\boldsymbol{s}_i$, $i = 0, 1, \ldots, n-1$, will satisfy eq. (D.2) if $\boldsymbol{s}_0 = \nabla \mathrm{f}(\boldsymbol{x}_0)$ and

$$\boldsymbol{s}_i = \nabla \mathrm{f}(\boldsymbol{x}_i) + \beta_{i-1}\boldsymbol{s}_{i-1}, \quad i = 1, 2, \ldots, n-1,$$

where, by the orthogonality of successive gradients,[15]

$$\beta_{i-1} = \frac{\nabla \mathrm{f}(\boldsymbol{x}_i)' \nabla \mathrm{f}(\boldsymbol{x}_i)}{\nabla \mathrm{f}(\boldsymbol{x}_{i-1})' \nabla \mathrm{f}(\boldsymbol{x}_{i-1})}, \quad i = 1, 2, \ldots, n-1.$$

In general, the various rounds of the Fletcher–Reeves method may be carried out as follows:

(1) to maximize f, choose an arbitrary point $\boldsymbol{x}_0$;

(2) compute $\nabla \mathrm{f}(\boldsymbol{x}_0)$;

---

[15] A detailed development of these last two expressions is provided by Beckman (1964), pp. 62–72, and Zangwill (1969), pp. 139–144.

(3) choose as the initial direction $s_0 = \nabla f(x_0)$;
(4) to move along s_0, compute the step length λ_0 so that

$$f(x_0+\lambda_0 s_0) = \max_{\lambda} f(x_0+\lambda s_0),$$

that is,

$$\lambda_0 = -\frac{s_0'\nabla f(x_0)}{s_0' Q s_0};$$

(5) compute $x_1 = x_0+\lambda_0\nabla f(x_0)$;
(6) for $i = 1, 2, \ldots, n-1$, choose λ_i so that

$$f(x_i+\lambda_i s_i) = \max_{\lambda} f(x_i+\lambda s_i),$$

that is,

$$\lambda_i = -\frac{s_i'\nabla f(x_i)}{s_i' Q s_i},$$

where

$$s_i = \nabla f(x_i)+\beta_{i-1}s_{i-1}, \qquad \beta_{i-1} = \frac{\nabla f(x_i)'\nabla f(x_i)}{\nabla f(x_{i-1})'\nabla f(x_{i-1})};$$

(7) compute

$$x_i = x_{i-1}+\lambda_{i-1}s_{i-1}, \quad i = 2, 3, \ldots, n;$$

(8) after the first iteration is completed, repeat steps (6) and (7) until $\nabla f(x_i) = 0$.

To minimize f, choose $s_0 = -\nabla f(x_0)$, $s_i = -\nabla f(x_i)+\beta_{i-1}s_{i-1}$, and minimize in steps (4) and (6) rather than maximize.[16]

If the function to be maximized or minimized is not quadratic in x, the desired extremum will not be located in exactly n steps. In this instance Fletcher and Reeves suggest restarting the process in the direction of steepest ascent or descent after every $n+1$ iterations. As an exercise the reader is asked to rework example D.2 using the algorithm developed in this section.

[16] For an evaluation of the relative performance of the Fletcher–Powell, Fletcher–Reeves techniques, among others, see Box (1966), pp. 67–77; Pearson (1969), pp. 171–178; Fiacco and McCormick (1968), pp. 162–165.

Befejezett.