Contents

Chapter 5

Quadratic Forms, Positive, Negative (Semi) Definiteness

This is Chapter 5 of "Sophomore Level Self-Teaching Webbook for Computational and Algorithmic Linear Algebra and n-Dimensional Geometry" by Katta G. Murty.

5.1 Expressing a Quadratic Function in n Variables Using Matrix Notation

As defined in Section 1.7, a linear function of $x = (x_1, \ldots, x_n)^T$ is a function of the form $c_1x_1 + \ldots + c_nx_n$ where $c_1, \ldots, c_n$ are constants called the coefficients of the variables in this function. For example, when $n = 4$, $f_1(x) = -2x_2 + 5x_4$ is a linear function of $(x_1, x_2, x_3, x_4)^T$ with the coefficient vector $(0, -2, 0, 5)$.

An affine function of $x = (x_1, \ldots, x_n)^T$ is a constant plus a linear function, i.e., a function of the form $f_2(x) = c_0 + c_1x_1 + \ldots + c_nx_n$. So, if $f_2(x)$ is an affine function, $f_2(x) - f_2(0)$ is a linear function. As an example, $-10 - 2x_2 + 5x_4$ is an affine function of $(x_1, x_2, x_3, x_4)^T$.

A **quadratic form** in the variables $x = (x_1, \ldots, x_n)^T$ consists of the second degree terms of a second degree polynomial in these variables, i.e., it is a function of the form

$$Q(x) = \sum_{i=1}^{n} q_{ii} x_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} q_{ij} x_i x_j$$

where the q_{ij} are the coefficients of the terms in the quadratic form.

Define a square matrix $D = (d_{ij})$ of order n where

$$d_{ii} = q_{ii} \quad \text{for } i = 1 \text{ to } n$$

and d_{ij}, d_{ji} are arbitrary real numbers satisfying

$$d_{ij} + d_{ji} = q_{ij} \quad \text{for } j > i.$$

As an example, if $q_{12} = -10$, we could take $(d_{12} = d_{21} = -5)$, or $(d_{12} = 100, d_{21} = -110)$, or $(d_{12} = -4, d_{21} = -6)$, etc. Then

$$Dx = \begin{pmatrix} d_{11}x_1 + d_{12}x_2 + \ldots + d_{1n}x_n \\ d_{21}x_1 + d_{22}x_2 + \ldots + d_{2n}x_n \\ \vdots \\ d_{n1}x_1 + d_{n2}x_2 + \ldots + d_{nn}x_n \end{pmatrix}$$

So

$$\begin{aligned} x^T Dx &= x_1(d_{11}x_1 + \ldots + d_{1n}x_n) + \ldots + x_n(d_{n1}x_1 + \ldots + d_{nn}x_n) \\ &= \sum_{i=1}^{n} d_{ii} x_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} (d_{ij} + d_{ji}) x_i x_j \\ &= \sum_{i=1}^{n} q_{ii} x_i^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} q_{ij} x_i x_j \\ &= Q(x) \end{aligned}$$

So, for any matrix $D = (d_{ij})$ satisfying the conditions stated above, we have $x^T Dx = Q(x)$. Now define $\underline{D} = (\underline{d}_{ij})$ by

$$\begin{aligned} \underline{d}_{ii} &= q_{ii} \quad i = 1, \ldots, n \\ \underline{d}_{ij} = \underline{d}_{ji} &= (1/2) q_{ij} \quad j > i \end{aligned}$$

then $\underline{D}$ is a symmetric matrix and $Q(x) = x^T \underline{D} x$. $\underline{D}$ is known as the **symmetric coefficient matrix defining the quadratic form** $Q(x)$.

Clearly for all $D = (d_{ij})$ satisfying the conditions stated above, we have $Q(x) = x^T D x = x^T D^T x = x^T(\frac{D+D^T}{2})x$ and $\frac{D+D^T}{2} = \underline{D}$, the symmetric coefficient matrix defining the quadratic form $Q(x)$.

As an example consider

$$n = 3, x = (x_1, x_2, x_3)^T, Q(x) = 81x_1^2 - 7x_2^2 + 5x_1x_2 - 6x_1x_3 + 18x_2x_3$$

Then the following square matrices satisfy the conditions stated above for $Q(x)$.

$$D_1 = \begin{pmatrix} 81 & -10 & 100 \\ 15 & -7 & 10 \\ -106 & 8 & 0 \end{pmatrix}, D_2 = \begin{pmatrix} 81 & 200 & -1006 \\ -195 & -7 & 218 \\ 1000 & -200 & 0 \end{pmatrix},$$

$$D_3 = \begin{pmatrix} 81 & 2 & -2 \\ 3 & -7 & 15 \\ -4 & 3 & 0 \end{pmatrix}, \underline{D} = \begin{pmatrix} 81 & 5/2 & -3 \\ 5/2 & -7 & 9 \\ -3 & 9 & 0 \end{pmatrix}$$

It can be verified that $x^T D_1 x = x^T D_2 x = x^T D_3 x = x^T \underline{D} x = Q(x)$, and that

$$\underline{D} = \frac{D_1 + D_1^T}{2} = \frac{D_2 + D_2^T}{2} = \frac{D_3 + D_3^T}{2}$$

Hence a general quadratic form in n variables $x = (x_1, \ldots, x_n)^T$ can be represented in matrix notation as $x^T M x$ where $M = (m_{ij})$ is a square matrix of order n. If M is not symmetric, it can be replaced in the above formula by $(M + M^T)/2$ without changing the quadratic form, and this $(M + M^T)/2$ is known as the symmetric matrix defining this quadratic form.

A **quadratic function** in variables $x = (x_1, \ldots, x_n)^T$ is a function which is the sum of a quadratic form in x and an affine function in x; i.e., it is of the form $x^T D x + cx + c_0$ for some square matrix D of order n, row vector $c \in R^n$ and constant term c_0.

Exercises

5.1.1: Express the following functions in matrix notation with a symmetric coefficient matrix: **(i)**: $-6x_1^2+7x_2^2-14x_4^2-12x_1x_2+20x_1x_3+6x_1x_4-7x_2x_3+8x_3x_4-9x_1+6x_2-13x_3+100$ **(ii)**: $x_1^2-x_2^2+x_3^2-18x_1x_3+12x_2x_3-7x_1+18x_2$ **(iii)**: $4x_1^2+3x_2^2-8x_1x_2$ **(iv)**: $6x_1+8x_2-11x_3+6$.

5.1.2: Express the quadratic form x^TDx+cx as a sum of individual terms in it for the following data:

$$(i): D = \begin{pmatrix} 3 & -6 & 9 & 8 \\ 10 & -2 & 0 & 12 \\ 13 & -11 & 0 & 3 \\ -4 & -9 & 9 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ -9 \\ 0 \\ 7 \end{pmatrix}.$$

$$(ii): D = \begin{pmatrix} 3 & -4 & 8 \\ -4 & 9 & 9 \\ -8 & -9 & 2 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}.$$

$$(iii): D = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad c = 0.$$

5.2 Convex, Concave Functions; Positive (Negative) (Semi) Definiteness; Indefiniteness

Consider the real valued function $f(x)$ defined over $x \in R^n$, or over some convex subset of R^n. It is said to be a **convex function** if for every x, y for which it is defined, and $0 \le \alpha \le 1$, we have

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y)$$

This inequality defining the convexity of the function $f(x)$ is called **Jensen's inequality** after the Danish mathematician who first defined it. This inequality is easy to visualize when $n = 1$. It says that if you join two points on the graph of the function by a chord, then the function itself lies underneath the chord on the interval joining these points. See Figure 5.1.

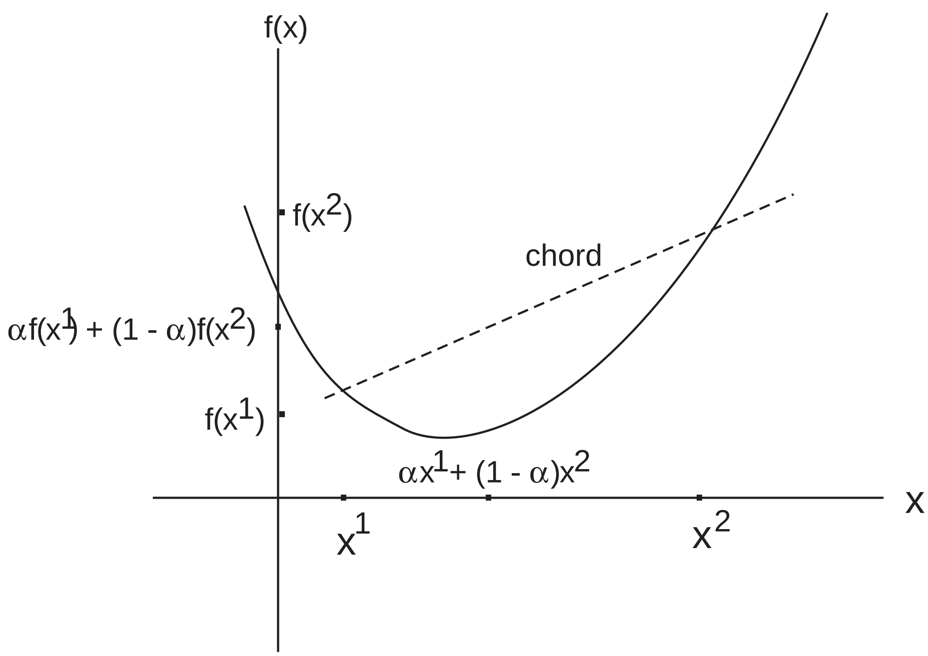

Figure 5.1: A convex function lies beneath any chord

The convex function $f(x)$ is said to be a **strictly convex function** if the above inequality holds as a strict inequality for every $x \neq y$ and $0 < \alpha < 1$.

A real valued function $g(x)$ defined over R^n or a convex subset of R^n is said to be a **concave function** if $-g(x)$ is a convex function as defined above, i.e., if for every x, y for which it is defined, and $0 \leq \alpha \leq 1$, we have

$$g(\alpha x + (1 - \alpha)y) \geq \alpha g(x) + (1 - \alpha)g(y)$$

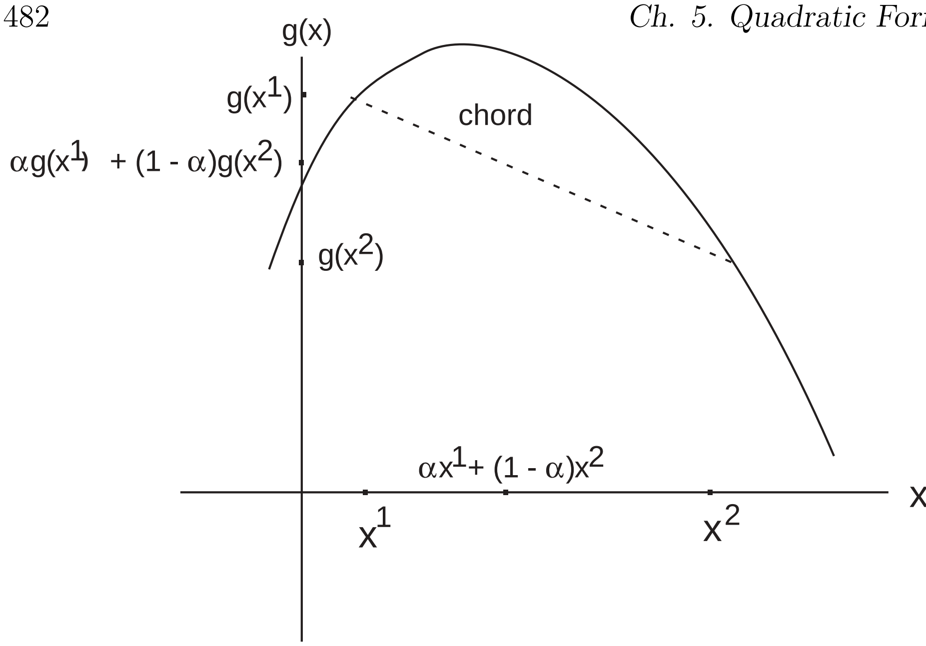

Figure 5.2: A concave function stays above any chord

The concave function $g(x)$ is said to be a **strictly concave function** if $-g(x)$ is a strictly convex function, i.e., if the above inequality holds as a strict inequality for every $x \neq y$ and $0 < \alpha < 1$.

The properties of convex and concave functions are of great importance in optimization theory. From the definition it can be verified that a function $f(x_1)$ of a single variable x_1 is

> Convex iff its slope is nondecreasing, i.e., iff its 2nd derivative is ≥ 0 when it is twice continuously differentiable

> Concave iff its slope is nonincreasing, i.e., iff its 2nd derivative is ≤ 0 when it is twice continuously differentiable.

As examples, some convex functions of a single variable x_1 are:

$$x_1^2, x_1^4, e^{-x_1}, -\log(x_1)(\text{in the region } x_1 > 0).$$

Among functions of several variables, linear and affine functions are both convex and concave, but they are not strictly convex or strictly concave. For classifying quadratic functions into the convex and concave classes, we need the following definitions.

A square matrix M of order n, whether symmetric or not, is said to be

Positive semidefinite (PSD)	iff $y^T My \geq 0$ for all $y \in R^n$
Positive definite (PD)	iff $y^T My > 0$ for all $y \in R^n$, $y \neq 0$
Negative semidefinite (NSD)	iff $y^T My \leq 0$ for all $y \in R^n$
Negative definite (ND)	iff $y^T My < 0$ for all $y \in R^n$, $y \neq 0$
Indefinite	if it is neither PSD nor NSD, i.e., iff there are points $x, y \in R^n$ satisfying $x^T Mx > 0$, and $y^T My < 0$.

We have the following results.

Result 5.2.1: Conditions for a quadratic function to be (strictly) convex: *The quadratic function $f(x) = x^T Dx + cx + c_0$ defined over R^n is a convex function over R^n iff the matrix D is PSD; and a strictly convex function iff the matrix D is PD.*

To see this, take any two points $x, y \in R^n$, and a $0 < \alpha < 1$. It can be verified that

$$\alpha f(x) + (1-\alpha)f(y) - f(\alpha x + (1-\alpha)y) = \alpha(1-\alpha)(x-y)^T D(x-y)$$

by expanding the terms on both sides. So Jensen's inequality for the convexity of $f(x)$ holds iff

$$(x-y)^T D(x-y) \geq 0 \quad \text{for all } x, y \in R^n$$

i.e., iff $z^T Dz \geq 0$ for all $z \in R^n$, i.e., iff D is PSD. Likewise for strict convexity of $f(x)$ we need

$$(x-y)^T D(x-y) > 0 \quad \text{for all } x \neq y \in R^n$$

i.e., iff $z^T Dz > 0$ for all $z \in R^n$, $z \neq 0$, i.e., iff D is PD.

Result 5.2.2: Conditions for a quadratic function to be (strictly) concave: *The quadratic function $f(x) = x^T Dx + cx + c_0$ defined over R^n is a concave function over R^n iff the matrix D is NSD; and a strictly concave function iff the matrix D is ND.*

This follows by applying Result 5.2.1 to $-f(x)$.

Checking whether a given quadratic function is strictly convex, or convex, is a task of great importance in optimization, statistics, and many other sciences. From Result 5.2.1, this task is equivalent to checking whether a given square matrix is PD, PSD. We discuss efficient algorithms for PD, PSD checking based on Gaussian pivot steps in the next section.

5.3 Algorithm to Check Positive (Semi) Definiteness Using G Pivot Steps

Submatrices of a Matrix

Let A be a matrix of order $m \times n$. Let $S \subset \{1, \ldots, m\}$ and $T \subset \{1, \ldots, n\}$. If we delete all the rows of A not in S, and all the columns not in T, what remains is a smaller matrix denoted by $A_{S\times T}$ known as **the submatrix of A corresponding to the subset of rows S and columns T**. As an example, consider the following matrix

$$A = \begin{pmatrix} 6 & 0 & -7 & 8 & 0 & -9 \\ -25 & -1 & 2 & 14 & -15 & -4 \\ 0 & 3 & -2 & -7 & 0 & 10 \\ 11 & 16 & 17 & -5 & 4 & 0 \end{pmatrix}.$$

Let $S_1 = \{1,4\}, S_2 = \{2,3\}, T_1 = \{1,3,4,6\}, T_2 = \{2,5,6\}$. Then the submatrices $A_{S_1 \times T_1}, A_{S_2 \times T_2}$ are

$$A_{S_1 \times T_1} = \begin{pmatrix} 6 & -7 & 8 & -9 \\ 11 & 17 & -5 & 0 \end{pmatrix}, A_{S_2 \times T_2} = \begin{pmatrix} -1 & -15 & -4 \\ 3 & 0 & 10 \end{pmatrix}.$$

Principal Submatrices of a Square Matrix

Square matrices have special submatrices called principal submatrices. Let M be a square matrix of order n, and $S \subset \{1, \dots, n\}$ with r elements in it. Then the submatrix $M_{S \times S}$ obtained by deleting from M all rows and all columns not in the set S is known as the **principal submatrix of the square matrix M corresponding to the subset S**, it is a square matrix of order r. The determinant of the principal submatrix of M corresponding to the subset S is called the **principal minor of M corresponding to the subset S**. As an example, let

$$M = \begin{pmatrix} 4 & 2 & 3 & 5 \\ 1 & 22 & 12 & -1 \\ -4 & -2 & -5 & 6 \\ 0 & -3 & 16 & -7 \end{pmatrix}.$$

Let $S = \{2,4\}$. Then the principal submatrix of M corresponding to S is

$$M_{S \times S} = \begin{pmatrix} 22 & -1 \\ -3 & -7 \end{pmatrix}.$$

and the principal minor of M corresponding to S is determinant$(M_{S \times S})$ which is -157.

In this same example, consider the singleton subset $S_1 = \{3\}$. Then the principal submatrix of M corresponding to S_1 is (-5) of order 1, i.e., the third diagonal element of M. In the same way, all diagonal elements of a square matrix are its principal submatrices of order 1.

Some Results on PSD, PD, NSD, ND Matrices

Result 5.3.1: Properties shared by all principal submatrices: *If M is a square matrix of order n which has any of the properties PSD, PD, NSD, ND; then all principal submatrices of M have the same property.*

To see this, consider the subset $S \subset \{1, 2\}$ as an example, and the principal submatrix $M_{S\times S}$ of order 2. Let $\bar{y} = (y_1, y_2, 0, \dots, 0)^T \in R^n$ with $y_i = 0$ for all $i \notin S$. Since y_3 to y_n are all 0 in $\bar{y}$, we verify that $\bar{y}^T M\bar{y} = (y_1, y_2)M_{S\times S}(y_1, y_2)^T$. So, for all y_1, y_2, if

$\bar{y}^T M\bar{y} \geq 0$ so is $(y_1, y_2)M_{S\times S}(y_1, y_2)^T$

$\bar{y}^T M\bar{y} > 0$ for all $(y_1, y_2) \neq 0$, so is $(y_1, y_2)M_{S\times S}(y_1, y_2)^T$

$\bar{y}^T M\bar{y} \leq 0$ so is $(y_1, y_2)M_{S\times S}(y_1, y_2)^T$

$\bar{y}^T M\bar{y} < 0$ for all $(y_1, y_2) \neq 0$ so is $(y_1, y_2)M_{S\times S}(y_1, y_2)^T$.

Hence, if M is PSD, PD, NSD, ND, $M_{S\times S}$ has the same property. A similar argumant applies to all principal submatrices of M.

Result 5.3.2: Conditions for a 1×1 matrix: *A 1×1 matrix (a) is*

PSD iff $a \geq 0$

PD iff $a > 0$

NSD iff $a \leq 0$

ND iff $a < 0$.

The quadratic form defined by (a) is ax_1^2; and it is ≥ 0 for all x_1 iff $a \geq 0$; it is > 0 for all $x_1 \neq 0$ iff $a > 0$; etc. Hence this result follows from the definitions.

Result 5.3.3: Conditions satisfied by diagonal entries: *Let $M = (m_{ij})$ be a square matrix. If M is a*

PSD matrix, all its diagonal entries must be ≥ 0

PD matrix, all its diagonal entries must be > 0

NSD matrix, all its diagonal entries must be ≤ 0

ND matrix, all its diagonal entries must be < 0.

Since all diagonal entries of a square matrix are its 1×1 principal submatrices, this result follows from Results 5.3.1 and 5.3.2.

Result 5.3.4: Conditions satisfied by the row and column of a 0-diagonal entry in a PSD matrix: *Let* $M = (m_{ij})$ *be a square PSD matrix. If one of its diagonal entries, say* $m_{pp} = 0$, *then for all* j *we must have* $m_{pj} + m_{jp} = 0$.

Suppose $m_{11} = 0$, and $m_{12} + m_{21} = \alpha \neq 0$. Let the principal submatrix of M corresponding to the subset $\{1, 2\}$ be

$$\bar{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \begin{pmatrix} 0 & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

The quadratic form defined by $\bar{M}$ is $(y_1, y_2)\bar{M}(y_1, y_2)^T = m_{22}y_2^2 + \alpha y_1 y_2$. By fixing

$$y_1 = \frac{-1 - m_{22}}{\alpha}, \quad y_2 = 1$$

we see that this quadratic form has value -1, so $\bar{M}$ is not PSD, and by Result 5.3.1 M cannot be PSD, a contradiction. Hence α must be 0 in this case. The result follows from the same argument.

Result 5.3.5: Conditions satisfied by the row and column of a 0-diagonal entry in a symmetric PSD matrix: *Let* $M = (m_{ij})$ *be a square matrix. Symmetrize* M, *i.e., let* $D = (d_{ij}) = (M + M^T)/2$. M *is PSD iff* D *is, because the quadratic forms defined by* M *and* D *are the same. Also, since* D *is symmetric, we have* $d_{ij} = d_{ji}$ *for all* i, j. *Hence, if* D *is PSD, and a diagonal entry in* D, *say* $d_{ii} = 0$, *then all* *the entries in the ith row and the ith column of D must be zero, i.e.,* $D_{i.} = 0$ *and* $D_{.i} = 0$.

This follows from Result 5.3.4.

Example 1: Consider the following matrices.

$$M_1 = \begin{pmatrix} 10 & 3 & 1 \\ -2 & 0 & 0 \\ 1 & 0 & 4 \end{pmatrix}, M_2 = \begin{pmatrix} 100 & 0 & 2 \\ 0 & 10 & 3 \\ 2 & 3 & -1 \end{pmatrix}$$

$$M_3 = \begin{pmatrix} 10 & 3 & -3 \\ 3 & 10 & 6 \\ 3 & -6 & 0 \end{pmatrix}, M_4 = \begin{pmatrix} 2 & 0 & -7 \\ 0 & 0 & 0 \\ 10 & 0 & 6 \end{pmatrix}$$

The matrix M_1 is not symmetric, and in it $m_{22} = 0$ but $m_{12}+m_{21} = 3-2 = 1 \neq 0$, hence this matrix violates the condition in Result 5.3.4, and is not PSD.

In the matrix M_2, the third diagonal entry is -1, hence by Result 5.3.3 this matrix is not PSD.

In the matrix M_3, $m_{33} = 0$, and $m_{13}+m_{31} = m_{23}+m_{32} = 0$, hence this matrix satisfies the condition in Result 5.3.4.

The matrix M_4 is symmetric, its second diagonal entry is 0, and its 2nd row and 2nd column are both zero vectors. Hence this matrix satisfies the condition in Result 5.3.5.

Result 5.3.6: Conditions satisfied by a principal submatrix obtained after a G pivot step: *Let* $D = (d_{ij})$ *be a symmetric matrix of order* n *with its first diagonal entry* $d_{11} \neq 0$. *Perform a Gaussian pivot step on* D *with* d_{11} *as the pivot element, and let* $\bar{D}$ *be the resulting matrix; i.e. transform*

$$D = \begin{pmatrix} d_{11} & \dots & d_{1n} \\ d_{21} & \dots & d_{2n} \\ \vdots & & \vdots \\ d_{n1} & \dots & d_{nn} \end{pmatrix} \quad \textit{into} \quad \bar{D} = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ 0 & \bar{d}_{22} & \dots & \bar{d}_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & \bar{d}_{n2} & \dots & \bar{d}_{nn} \end{pmatrix}$$

Let D_1 be the matrix of order $(n-1) \times (n-1)$ obtained by deleting row 1 and column 1 from $\bar{D}$. Then D_1 is symmetric too, and D is PD (PSD) iff $d_{11} > 0$ and D_1 is PD (PSD).

Since the original matrix D is symmetric, we have $d_{ij} = d_{ji}$ for all i, j. Using this and the formula coming from the Gaussian pivot step, that for $i, j = 2$ to n, $\bar{d}_{ij} = d_{ij} - d_{i1}(d_{1j}/d_{11})$ it can be verified that $\bar{d}_{ij} = \bar{d}_{ji}$, hence D_1 is symmetric.

The other part of this result can be seen from the following.

$$\begin{aligned} x^T D x = \quad & d_{11}[x_1^2 + 2(d_{12}/d_{11})x_1x_2 + \ldots + 2(d_{1n}/d_{11})x_1x_n \\ & +(1/d_{11})\sum_{i=2}^{n}\sum_{j=2}^{n} d_{ij}x_ix_j] \end{aligned}$$

Let

$$\begin{aligned} \theta &= (d_{12}/d_{11})x_2 + \ldots + (d_{1n}/d_{11})x_n \\ \delta &= (1/d_{11})\sum_{i=2}^{n}\sum_{j=2}^{n} d_{ij}x_ix_j \end{aligned}$$

Then $x^T D x = d_{11}[x_1^2 + 2x_1\theta + \delta] = d_{11}[(x_1 + \theta)^2 + (\delta - \theta^2)]$, because $x_1^2 + 2x_1\theta = (x_1 + \theta)^2 - \theta^2$. In mathematical literature this argument is called **completing the square argument**.

After rearranging the terms it can be verified that $\delta - \theta^2 = (x_2, \ldots, x_n)D_1(x_2, \ldots, x_n)^T$. So,

$$x^T D x = d_{11}(x_1 + \theta)^2 + (x_2, \ldots, x_n)D_1(x_2, \ldots, x_n)^T$$

From this it is clear that D is is PD (PSD) iff $d_{11} > 0$ and D_1 is PD (PSD).

Based on these results, we now describe algorithms to check whether a given square matrix is PD, PSD. To check if a square matrix is ND, NSD, apply the following algorithms to check if its negative is PD, PSD.

Algorithm to Check if a Given Square Matrix M of Order n is PD

BEGIN

Step 1: First symmetrize M, i.e., compute $D = (M + M^T)/2$. The algorithm works on D.

If any of the diagonal entries in D are ≤ 0, D and M are not PD by Result 5.3.3, terminate.

Step 2: Start performing Gaussian pivot steps on D using the diagonal elements as the pivot elements in the order 1 to n. At any stage of this process, if the current matrix has an entry ≤ 0 in its main diagonal, terminate with the conclusion that D, M are not PD. If all the pivot steps are completed and all the diagonal entries are > 0, terminate with the conclusion that D, M are PD.

END

Example 2: Check whether the following matrix M is PD.

$$M = \begin{pmatrix} 3 & 1 & 2 & 2 \\ -1 & 2 & 0 & 2 \\ 0 & 4 & 4 & 5/3 \\ 0 & -2 & -13/3 & 6 \end{pmatrix}$$

Symmetrizing, we get

$$D = \begin{pmatrix} 3 & 0 & 1 & 1 \\ 0 & 2 & 2 & 0 \\ 1 & 2 & 4 & -4/3 \\ 1 & 0 & -4/3 & 6 \end{pmatrix}$$

All the entries in the principal diagonal of D are > 0. So, we apply the algorithm, and obtain the following results. The pivot elements are boxed in the following. PR, PC indicate pivot row, pivot column for each G pivot step.

	PC↓			
PR	$\boxed{3}$	0	1	1
	0	2	2	0
	1	2	4	$-4/3$
	1	0	$-4/3$	6

		PC↓		
	3	0	1	1
PR	0	$\boxed{2}$	2	0
	0	2	$11/3$	$-5/3$
	0	0	$-5/3$	$17/3$

			PC↓	
	3	0	1	1
	0	2	2	0
PR	0	0	$\boxed{5/3}$	$-5/3$
	0	0	$-5/3$	$17/3$

	3	0	1	1
	0	2	2	0
	0	0	$5/3$	$-5/3$
	0	0	0	4

The algorithm terminates now. Since all the diagonal entries in all the tableaus are > 0, D and hence M are PD.

Example 3: Check whether the following matrix M is PD.

$$M = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 2 & 4 & 0 \\ 2 & 4 & 4 & 5 \\ 0 & 0 & 5 & 3 \end{pmatrix}$$

The matrix M is already symmetric and its diagonal entries are > 0. So, we apply the algorithm on it, and obtain the following results. The pivot elements are boxed in the following. PR, PC indicate pivot row, pivot column for each G pivot step.

	PC↓			
PR	$\boxed{1}$	0	2	0
	0	2	4	0
	2	4	4	5
	0	0	5	3
	1	0	2	0
	0	2	4	0
	0	4	0	5
	0	0	5	3

Since the third diagonal entry in this tableau is 0, we terminate with the conclusion that this matrix M is not PD.

A square matrix which is not PD, could be PSD. We discuss the algorithm based on G pivot steps, for checking whether a given square matrix is PSD, next.

Algorithm to Check if a Given Square Matrix M of Order n is PSD

BEGIN

Initial step: First symmetrize M, i.e., compute $D = (M + M^T)/2$. The algorithm works on D. Apply the following General step on the matrix D.

General Step: If any of the diagonal entries in the matrix are < 0, terminate with the conclusion that the original matrix is not PSD (Results 5.3.3, 5.3.6). Otherwise continue.

Check if the matrix has any 0-diagonal entries. For each 0-diagonal entry check whether its row and column in the matrix are both completely 0-vectors, if not the original matrix is not PSD (Results 5.3.5, 5.3.6), terminate. If this condition is satisfied, delete the 0-row vector and the 0-column vector of each 0-diagonal entry in the matrix.

If the remaining matrix is of order 1×1, its entry will be > 0, terminate with the conclusion that the original matrix is PSD.—

If the remaining matrix is of order ≥ 2, its first diagonal entry will be positive, perform a G pivot step on the matrix with this first diagonal element as the pivot element. After this pivot step, delete row 1, column 1 of the resulting matrix.

Now apply the same general step on the remining matrix, and repeat the same way.

END

Example 4: Check whether the following matrix M is PSD.

$$M = \begin{pmatrix} 0 & -2 & -3 & -4 & 5 \\ 2 & 3 & 3 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 0 & 0 & 8 & 4 \\ -5 & 0 & 0 & 4 & 2 \end{pmatrix}$$

Symmetrizing, we get

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 3 & 0 & 0 \\ 0 & 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 8 & 4 \\ 0 & 0 & 0 & 4 & 2 \end{pmatrix}$$

The first diagonal entry in D is 0, and in fact $D_{1.}, D_{.1}$ are both 0, so we eliminate them and apply the general step on the remaining matrix. Since there is a 0-diagonal entry in D, the original matrix M is not PD, but it may possibly be PSD.

	PC↓			
PR	$\boxed{3}$	3	0	0
	3	3	0	0
	0	0	8	4
	0	0	4	2
	3	3	0	0
	0	0	0	0
	0	0	8	4
	0	0	4	2

Eliminating row 1 and column 1 in the matrix resulting after the first G step leads to the matrix

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 8 & 4 \\ 0 & 4 & 2 \end{pmatrix}$$

The diagonal entries in this matrix are all ≥ 0, and the first diagonal entry is 0. Correspondingly, row 1 and column 1 are both 0-vectors, so we delete them, and apply the general step on the remaining matrix.

	PC↓	
PR	$\boxed{8}$	4
	4	2
	8	4
	0	0

.

Eliminating row 1 and column 1 in the matrix resulting after this G pivot step leads to the 1×1 matrix (0). Since the entry in it is ≥ 0, we terminate with the conclusion that the original matrix M is PSD but not PD.

Example 5: Check whether the following matrix M is PSD.

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 4 & 0 \\ 1 & 4 & 1 & 5 \\ 0 & 0 & 5 & 3 \end{pmatrix}$$

The matrix M is already symmetric, and its diagonal entries are all > 0. So, we apply the general step on it.

	PC↓			
PR	$\boxed{1}$	0	1	0
	0	2	4	0
	1	4	1	5
	0	0	5	3
	1	0	1	0
	0	2	4	0
	0	4	0	5
	0	0	5	3

Eliminating row 1 and column 1 in the matrix resulting after the first G step leads to the matrix

$$\begin{pmatrix} 2 & 4 & 0 \\ 4 & 0 & 5 \\ 0 & 5 & 3 \end{pmatrix}$$

The 2nd diagonal entry in this matrix is 0, but its 2nd row and 2nd column are not 0-vectors. So, we terminate with the conclusion that the original matrix M is not PSD.

Exercises

5.3.1: Find all the conditions that $p, q > 0$ have to satisfy for the following matrix to be PD.

$$\begin{pmatrix} \frac{1-p^2}{2p} & \frac{1-pq}{p+q} \\ \frac{1-pq}{p+q} & \frac{1-q^2}{2q} \end{pmatrix}.$$

5.3.2: Show that the following matrices are PD.

$$\begin{pmatrix} 3 & 1 & 2 & 2 \\ -1 & 2 & 0 & 2 \\ 0 & 4 & 4 & 5/3 \\ 0 & -2 & -13/3 & 6 \end{pmatrix}, \begin{pmatrix} 2 & -11 & 3 & 4 \\ 9 & 2 & -39 & 10 \\ -1 & 41 & 5 & 12 \\ -2 & -10 & -12 & 2 \end{pmatrix},$$

$$\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

5.3.3: Show that the following matrices are not PD.

$$\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 2 & 4 & 0 \\ 2 & 4 & 4 & 5 \\ 0 & 0 & 5 & 3 \end{pmatrix}, \begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 10 & -2 & -1 \\ -2 & 0 & 1 \\ -1 & 1 & 5 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 3 \\ 0 & 2 & 1 \\ 3 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 6 \\ 0 & 0 & 2 \end{pmatrix}.$$

5.3.4: Show that the following matrices are PSD.

$$\begin{pmatrix} 0 & -2 & -3 & -4 & 5 \\ 2 & 3 & 3 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 0 & 0 & 8 & 4 \\ -5 & 0 & 0 & 4 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}, \begin{pmatrix} 2 & -2 & 1 \\ -2 & 4 & 2 \\ 1 & 2 & 6 \end{pmatrix}.$$

5.3.3: Show that the following matrices are not PSD.

$$\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 2 & 4 & 0 \\ 2 & 4 & 4 & 5 \\ 0 & 0 & 5 & 3 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 4 \end{pmatrix}.$$

5.3.5: Let A be a square matrix of order n. Show that $x^T Ax$ is a convex function in x iff its minimum value over R^n is 0.

5.3.6: Consider the square matrix of order n with all its diagonal entries $= p$, and all its off-diagonal entries $= q$. Determine all the values of p, q for which this matrix is PD.

5.3.7: A square matrix $A = (a_{ij})$ of order n is said to be *skew-symmetric* if $A + A^T = 0$. Show that the quadratic form $x^T Ax$ is 0 for all $x \in R^n$ iff A is skew-symmetric.

Also, if A is a skew symmetric matrix of order n, and B is any matrix of order n, show that $x^T(A+B)x = x^T Bx$ for all x.

5.3.8: Let A be a square matrix of order n and $f(x) = x^T Ax$. For $x, y \in R^n$ show that $f(x+y) - f(x) - f(y)$ is a bilinear function of x, y as defined in Section 1.7.

5.3.9: Find the range of values of α for which the quadratic form $x_1^2 + 4x_1x_2 + 6x_1x_3 + \alpha x_2^2 + \alpha x_3^2$ is a strictly convex function.

5.3.10: Find the range of values of α for which the following matrix is PD.

$$\begin{pmatrix} 1 & 4 & 2 \\ 0 & 5-\alpha & 8-4\alpha \\ 0 & 0 & 8-\alpha \end{pmatrix}.$$

5.4 Diagonalization of Quadratic Forms and Square Matrices

Optimization deals with problems of the form: find a $y = (y_1, \ldots, y_n)^T \in R^n$ that minimizes a given real valued function $g(y)$ of y. This problem becomes easier to solve if $g(y)$ is separable, i.e., if it is the sum of n functions each involving one variable only, as in

$$g(y) = g_1(y_1) + \ldots + g_n(y_n).$$

In this case, minimizing $g(y)$ can be achieved by minimizing each $g_j(y_j)$ separately for $j = 1$ to n; and the problem of minimizing a function is much easier if it involves only one variable.

Consider the problem of minimizing a quadratic function

$$f(x) = cx + (1/2)x^T Mx$$

where $M = (m_{ij})$ is an $n \times n$ symmetric PD matrix. If some $m_{ij} \neq 0$ for $i \neq j$, the function involves the product term $x_i x_j$ with a nonzero coefficient, and it is not separable. To make this function separable, we need to convert the matrix M into a diagonal matrix. For this we can try to apply a linear transformation of the variables x with the aim of achieving separability in the space of new variables. Consider the transformation

$$x = Py$$

where P is an $n \times n$ nonsingular matrix, and the new variables are $y = (y_1, \ldots, y_n)^T$. Then, in terms of the new variables y, the function is

$$F(y) = f(x = Py) = cPy + (1/2)y^T P^T MPy.$$

$F(y)$ is separable if $P^T MP$ is a diagonal matrix. If $P^T MP = \text{diag}(d_{11}, \ldots, d_{nn})$, then let $cP = \bar{c} = (\bar{c}_1, \ldots, \bar{c}_n)$. In this case $F(y) = \sum_{j=1}^{n}(\bar{c}_j y_j + d_{jj} y_j^2)$. The point minimizing $F(y)$ is $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_n)^T$ where $\bar{y}_j$ is the minimizer of $\bar{c}_j y_j + d_{jj} y_j^2$. Once $\bar{y}$ is found, the minimizer of the original function $f(x)$ is $\bar{x} = P\bar{y}$.

The condition

$$P^T MP = \begin{pmatrix} d_{11} & 0 & \ldots & 0 \\ 0 & d_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{nn} \end{pmatrix}$$

requires

$$(P_{.i})^T MP_{.j} = 0 \quad \text{for all } i \neq j.$$

When$M$ is a PD symmetric matrix, the set of column vectors $\{P_{.1}, \ldots, P_{.n}\}$ of a matrix P satisfying the above condition is said to be a **conjugate set of vectors WRT** M. Optimization methods called **conjugate direction methods** or **conjugate gradient methods** are based on using this type of transformations of variables.

Given a symmetric matrix M, this problem of finding a nonsingular matrix P satisfying $P^T MP$ is a diagonal matrix is called the **problem**

of diagonalizing a quadratic form. Conjugate direction methods in nonlinear optimization have very efficient special routines for finding such a matrix P, but a discussion of these methods is outside the scope of this book.

In mathematics they often discuss other matrix diagonalization problems, which we now explain.

The **orthogonal diagonalization problem** is the same as the above, with the additional requirement that the matrix P must be an orthogonal matrix; i.e., it should also satisfy $P^T = P^{-1}$. Hence the orthogonal diagonalization problem is: given a square symmetric matrix M of order n, find an orthogonal matrix P of order n satisfying:

> $P^{-1}MP$ is a diagonal matrix D (this is also written sometimes as $MP = PD$).

The **general matrix diagonalization problem** discussed in mathematics is: given a square matrix M of order n, find a nonsingular square matrix P of order n such that $P^{-1}MP$ is a diagonal matrix.

Solving either the orthogonal diagonalization problem, or the matrix diagonalization problem involves finding the eigenvalues and eigenvectors of a square matrix, which are discussed in the next chapter.

Index

For each index entry we provide the section number where it is defined or discussed first.Clicking on the light blue section number takes you to the page where this entry is defined or discussed.

List of algorithms

List of Results

8. Result 5.3.6: Conditions satisfied by a principal submatrix obtained after a pivot step.