Chapter 3. Bilinear forms

Lecture notes for MA1212

P. Karageorgis

pete@maths.tcd.ie

Bilinear forms

Definition 3.1 – Bilinear form

A bilinear form on a real vector space V is a function $f\colon V \times V \to \mathbb{R}$ which assigns a number to each pair of elements of V in such a way that f is linear in each variable.

- A typical example of a bilinear form is the dot product on $\mathbb{R}^n$.
- We shall usually write $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ instead of $f(\boldsymbol{x}, \boldsymbol{y})$ for simplicity and we shall also identify each 1×1 matrix with its unique entry.

Theorem 3.2 – Bilinear forms on $\mathbb{R}^n$

Every bilinear form on $\mathbb{R}^n$ has the form

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^t A \boldsymbol{y} = \sum_{i,j} a_{ij} x_i y_j$$

for some $n \times n$ matrix A and we also have $a_{ij} = \langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle$ for all i, j.

Matrix of a bilinear form

Definition 3.3 – Matrix of a bilinear form

Suppose that $\langle\, , \rangle$ is a bilinear form on V and let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be a basis of V. The matrix of the form with respect to this basis is the matrix A whose entries are given by $a_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$ for all i, j.

Theorem 3.4 – Change of basis

Suppose that $\langle\, , \rangle$ is a bilinear form on $\mathbb{R}^n$ and let A be its matrix with respect to the standard basis. Then the matrix of the form with respect to some other basis $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ is given by $B^t AB$, where B is the matrix whose columns are the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$.

- There is a similar result for linear transformations: if A is the matrix with respect to the standard basis and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ is some other basis, then the matrix with respect to the other basis is $B^{-1}AB$.

Matrix of a bilinear form: Example

- Let P_2 denote the space of real polynomials of degree at most 2. Then P_2 is a vector space and its standard basis is $1, x, x^2$.
- We can define a bilinear form on P_2 by setting

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx \qquad \text{for all } f, g \in P_2.$$

- By definition, the matrix of a form with respect to a given basis has entries $a_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$. In our case, $\boldsymbol{v}_i = x^{i-1}$ for each i and so

$$a_{ij} = \left\langle x^{i-1}, x^{j-1} \right\rangle = \int_0^1 x^{i+j-2}\,dx = \frac{1}{i+j-1}.$$

- Thus, the matrix of the form with respect to the standard basis is

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}.$$

Positive definite forms

Definition 3.5 – Positive definite

A bilinear form $\langle\, , \rangle$ on a real vector space V is positive definite, if

$$\langle \boldsymbol{v}, \boldsymbol{v} \rangle > 0 \quad \text{for all } \boldsymbol{v} \neq 0.$$

A real $n \times n$ matrix A is positive definite, if $\boldsymbol{x}^t A \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq 0$.

- A bilinear form on V is positive definite if and only if the matrix of the form with respect to some basis of V is positive definite.
- A positive definite form on $\mathbb{R}^n$ is given by the dot product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^n x_i y_i \quad \Longrightarrow \quad \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum_{i=1}^n x_i^2.$$

- A positive definite form on P_n is given by the formula

$$\langle f, g \rangle = \int_a^b f(x) g(x)\, dx \quad \Longrightarrow \quad \langle f, f \rangle = \int_a^b f(x)^2\, dx.$$

Positive definite forms: Examples

1. Consider the bilinear form on $\mathbb{R}^2$ which is defined by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1y_1 - 2x_1y_2 - 2x_2y_1 + 5x_2y_2.$$

   To check if it is positive definite, we complete the square to get

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 - 4x_1x_2 + 5x_2^2 = (x_1 - 2x_2)^2 + x_2^2.$$

   It now easily follows that the given form is positive definite.

2. Consider the bilinear form on $\mathbb{R}^2$ which is defined by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1y_1 + 2x_1y_2 + 2x_2y_1 + 3x_2y_2.$$

   Completing the square as before, one finds that

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 + 4x_1x_2 + 3x_2^2 = (x_1 + 2x_2)^2 - x_2^2.$$

   In particular, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$ is negative whenever $x_1 = -2x_2$ and $x_2 \neq 0$.

Symmetric forms

Definition 3.6 – Symmetric

A bilinear form $\langle \, , \rangle$ on a real vector space V is called symmetric, if

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle \quad \text{for all } \boldsymbol{v}, \boldsymbol{w} \in V.$$

A real square matrix A is called symmetric, if $a_{ij} = a_{ji}$ for all i, j.

- A bilinear form on V is symmetric if and only if the matrix of the form with respect to some basis of V is symmetric.
- A real square matrix A is symmetric if and only if $A^t = A$.

Definition 3.7 – Inner product

An inner product on a real vector space V is a bilinear form which is both positive definite and symmetric.

Angles and length

- Suppose that $\langle \, , \rangle$ is an inner product on a real vector space V.
- Then one may define the length of a vector $\boldsymbol{v} \in V$ by setting

$$||\boldsymbol{v}|| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

  and the angle θ between two vectors $\boldsymbol{v}, \boldsymbol{w} \in V$ by setting

$$\cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{||\boldsymbol{v}|| \cdot ||\boldsymbol{w}||}.$$

- These formulas are known to hold for the inner product on $\mathbb{R}^n$.

Theorem 3.8 – Cauchy-Schwarz inequality

When V is a real vector space with an inner product, one has

$$|\langle \boldsymbol{v}, \boldsymbol{w} \rangle| \leq ||\boldsymbol{v}|| \cdot ||\boldsymbol{w}|| \quad \text{for all } \boldsymbol{v}, \boldsymbol{w} \in V.$$

Orthogonal vectors

Definition 3.9 – Orthogonal and orthonormal

Suppose $\langle\ ,\ \rangle$ is a symmetric bilinear form on a real vector space V. Two vectors $\boldsymbol{u}, \boldsymbol{v}$ are called orthogonal, if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$. A basis $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ of V is called orthogonal, if $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0$ whenever $i \neq j$ and it is called orthonormal, if it is orthogonal with $\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle = 1$ for all i.

Theorem 3.10 – Linear combinations

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ be an orthogonal basis of an inner product space V. Then every vector $\boldsymbol{v} \in V$ can be expressed as a linear combination

$$\boldsymbol{v} = \sum_{i=1}^{n} c_i \boldsymbol{v}_i, \quad \text{where } c_i = \frac{\langle \boldsymbol{v}, \boldsymbol{v}_i \rangle}{\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle} \text{ for all } i.$$

If the basis is actually orthonormal, then $c_i = \langle \boldsymbol{v}, \boldsymbol{v}_i \rangle$ for all i.

Gram-Schmidt procedure

- Suppose that $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ is a basis of an inner product space V. Then we can find an orthogonal basis $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$ as follows.
- Define the first vector by $\boldsymbol{w}_1 = \boldsymbol{v}_1$ and the second vector by

$$\boldsymbol{w}_2 = \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1.$$

  Then $\boldsymbol{w}_1, \boldsymbol{w}_2$ are orthogonal and have the same span as $\boldsymbol{v}_1, \boldsymbol{v}_2$.
- Proceeding by induction, suppose $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k$ are orthogonal and have the same span as $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$. Once we then define

$$\boldsymbol{w}_{k+1} = \boldsymbol{v}_{k+1} - \sum_{i=1}^{k} \frac{\langle \boldsymbol{v}_{k+1}, \boldsymbol{w}_i \rangle}{\langle \boldsymbol{w}_i, \boldsymbol{w}_i \rangle} \boldsymbol{w}_i,$$

  we end up with vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{k+1}$ which are orthogonal and have the same span as the original vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k+1}$.
- Using the formula from the last step repeatedly, one may thus obtain an orthogonal basis $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$ for the vector space V.

Gram-Schmidt procedure: Example

- We find an orthogonal basis of $\mathbb{R}^3$, starting with the basis

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

- We define the first vector by $\boldsymbol{w}_1 = \boldsymbol{v}_1$ and the second vector by

$$\boldsymbol{w}_2 = \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

- Then $\boldsymbol{w}_1, \boldsymbol{w}_2$ are orthogonal and we may define the third vector by

$$\begin{aligned} \boldsymbol{w}_3 &= \boldsymbol{v}_3 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_1 \rangle}{\langle \boldsymbol{w}_1, \boldsymbol{w}_1 \rangle} \boldsymbol{w}_1 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{w}_2 \rangle}{\langle \boldsymbol{w}_2, \boldsymbol{w}_2 \rangle} \boldsymbol{w}_2 \\ &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{4}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \end{aligned}$$

Bilinear forms over a complex vector space

- Bilinear forms are defined on a complex vector space in the same way that they are defined on a real vector space. However, one needs to conjugate one of the variables to ensure positivity of the dot product.
- The complex transpose of a matrix is denoted by $A^* = \overline{A^t}$ and it is also known as the adjoint of A. One has $\boldsymbol{x}^*\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{C}^n$.

Bilinear forms on $\mathbb{R}^n$	**Bilinear forms on $\mathbb{C}^n$**
Linear in the first variable	Conjugate linear in the first variable
$\langle \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$	$\langle \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
$\langle \lambda\boldsymbol{u}, \boldsymbol{v} \rangle = \lambda \langle \boldsymbol{u}, \boldsymbol{v} \rangle$	$\langle \lambda\boldsymbol{u}, \boldsymbol{v} \rangle = \overline{\lambda} \langle \boldsymbol{u}, \boldsymbol{v} \rangle$
Linear in the second variable	Linear in the second variable
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^t A \boldsymbol{y}$ for some A	$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^* A \boldsymbol{y}$ for some A
Symmetric, if $A^t = A$	Hermitian, if $A^* = A$
Symmetric, if $a_{ij} = a_{ji}$	Hermitian, if $a_{ij} = \overline{a}_{ji}$

Real symmetric matrices

Theorem 3.11 – Inner product and matrices

Letting $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^* \boldsymbol{y}$ be the standard inner product on $\mathbb{C}^n$, one has

$$\langle A\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, A^* \boldsymbol{y} \rangle \quad \text{and} \quad \langle \boldsymbol{x}, A\boldsymbol{y} \rangle = \langle A^* \boldsymbol{x}, \boldsymbol{y} \rangle$$

for any $n \times n$ complex matrix A. In fact, these formulas also hold for the standard inner product on $\mathbb{R}^n$, in which case A^* reduces to A^t.

Theorem 3.12 – Eigenvalues of a real symmetric matrix

The eigenvalues of a real symmetric matrix are all real.

Theorem 3.13 – Eigenvectors of a real symmetric matrix

The eigenvectors of a real symmetric matrix corresponding to distinct eigenvalues are necessarily orthogonal to one another.

Orthogonal matrices

Definition 3.14 – Orthogonal matrix

A real $n \times n$ matrix A is called orthogonal, if $A^t A = I_n$.

Theorem 3.15 – Properties of orthogonal matrices

1. To say that an $n \times n$ matrix A is orthogonal is to say that the columns of A form an orthonormal basis of $\mathbb{R}^n$.
2. The product of two $n \times n$ orthogonal matrices is orthogonal.
3. Left multiplication by an orthogonal matrix preserves both angles and length. When A is an orthogonal matrix, that is, one has

$$\langle A\boldsymbol{x}, A\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \quad \text{and} \quad ||A\boldsymbol{x}|| = ||\boldsymbol{x}||.$$

- An example of a 2×2 orthogonal matrix is $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Spectral theorem

Theorem 3.16 – Spectral theorem

Every real symmetric matrix A is diagonalisable. In fact, there exists an orthogonal matrix B such that $B^{-1}AB = B^tAB$ is diagonal.

- When the eigenvalues of A are distinct, the eigenvectors of A are orthogonal and we may simply divide each of them by its length to obtain an orthonormal basis of $\mathbb{R}^n$. Such a basis can be merged to form an orthogonal matrix B such that $B^{-1}AB$ is diagonal.
- When the eigenvalues of A are not distinct, the eigenvectors of A may not be orthogonal. In that case, one may use the Gram-Schmidt procedure to replace eigenvectors that have the same eigenvalue with orthogonal eigenvectors that have the same eigenvalue.
- The converse of the spectral theorem is also true. That is, if B is an orthogonal matrix and B^tAB is diagonal, then A is symmetric.

Orthogonal diagonalisation: Example 1

- Consider the real symmetric matrix

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

- Its eigenvalues $\lambda = 0, 4, -2$ are distinct and its eigenvectors are

$$\boldsymbol{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

- Since $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are orthogonal, dividing each of them by its length gives an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors. Then

$$B = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$$

is an orthogonal matrix such that $B^{-1}AB = B^t AB$ is diagonal.

Orthogonal diagonalisation: Example 2

- Consider the real symmetric matrix

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

- Its eigenvalues are $\lambda = 1, 1, 4$ and its eigenvectors are

$$\boldsymbol{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

- In this case, we use the Gram-Schmidt procedure to replace $\boldsymbol{v}_1, \boldsymbol{v}_2$ by two orthogonal eigenvectors $\boldsymbol{w}_1, \boldsymbol{w}_2$. Dividing each of $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{v}_3$ by its length, we then obtain the columns of the orthogonal matrix

$$B = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}.$$

Quadratic forms

Definition 3.17 – Quadratic form

A quadratic form in n variables is a function that has the form

$$Q(x_1, x_2, \ldots, x_n) = \sum_{i \leq j} a_{ij} x_i x_j.$$

This can be written as $Q(\boldsymbol{x}) = \boldsymbol{x}^t A \boldsymbol{x}$ for some symmetric matrix A.

- Here, one needs to be careful with the off-diagonal entries a_{ij}, as the coefficient of $x_i x_j$ needs to be halved whenever $i \neq j$. For instance,

$$Q(\boldsymbol{x}) = x_1^2 + 4x_1 x_2 + 3x_2^2 = \boldsymbol{x}^t A \boldsymbol{x}, \qquad A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}.$$

- The most general quadratic function in n variables has the form

$$Q(\boldsymbol{x}) = \sum_{i \leq j} a_{ij} x_i x_j + \sum_k b_k x_k + c = \boldsymbol{x}^t A \boldsymbol{x} + \boldsymbol{b}^t \boldsymbol{x} + c.$$

Diagonalisation of quadratic forms

Theorem 3.18 – Diagonalisation of quadratic forms

Let $Q(\boldsymbol{x}) = \boldsymbol{x}^t A \boldsymbol{x}$ for some symmetric $n \times n$ matrix A. Then there exists an orthogonal change of variables $\boldsymbol{x} = B\boldsymbol{y}$ such that

$$Q(\boldsymbol{x}) = \sum_{i \leq j} a_{ij} x_i x_j = \sum_{i=1}^{n} \lambda_i y_i^2,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of the matrix A.

Definition 3.19 – Signature of a quadratic form

The signature of a quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}^t A \boldsymbol{x}$ is defined as the pair of integers (n_+, n_-), where n_+ is the number of positive eigenvalues of A and n_- is the number of negative eigenvalues of A.

Diagonalisation of quadratic forms: Example

- We diagonalise the quadratic form

$$Q(\boldsymbol{x}) = 5x_1^2 + 4x_1x_2 + 2x_2^2 = \boldsymbol{x}^t A\boldsymbol{x}, \qquad A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}.$$

- The eigenvalues $\lambda = 1, 6$ are distinct and one can easily check that

$$B = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \quad \Longrightarrow \quad B^t AB = \begin{bmatrix} 1 & \\ & 6 \end{bmatrix}.$$

  As usual, the columns of B were obtained by finding the eigenvectors of A and by dividing each eigenvector by its length.

- Changing variables by $\boldsymbol{x} = B\boldsymbol{y}$, we now get $\boldsymbol{y} = B^t\boldsymbol{x}$ and also

$$y_1^2 + 6y_2^2 = \left(\frac{x_1 - 2x_2}{\sqrt{5}}\right)^2 + 6\left(\frac{2x_1 + x_2}{\sqrt{5}}\right)^2 = Q(\boldsymbol{x}).$$

  This is the change of variables which is asserted by Theorem 3.18.

Tests for positive definiteness

Theorem 3.20 – Tests for positive definiteness

The following conditions are equivalent for a symmetric matrix A.

1. One has $\boldsymbol{x}^t A\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq 0$.
2. The eigenvalues of A are all positive.
3. One has $\det A_k > 0$ for all $k \times k$ upper left submatrices A_k.

- The last condition is known as Sylvester's criterion. When it comes to a 3×3 matrix, for instance, it refers to the three submatrices

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

- We say that A is negative definite, if $\boldsymbol{x}^t A\boldsymbol{x} < 0$ for all $\boldsymbol{x} \neq 0$. Thus, a negative definite symmetric matrix A has negative eigenvalues and its upper left submatrices A_k are such that $(-1)^k \det A_k > 0$ for all k.

Sylvester's criterion: Example

- Let a be a real parameter and consider the matrix

$$A = \begin{bmatrix} a & 1 & 1 \\ 1 & 1 & a \\ 1 & a & 5 \end{bmatrix}.$$

- By Sylvester's criterion, A is positive definite if and only if

$$a > 0, \qquad \det \begin{bmatrix} a & 1 \\ 1 & 1 \end{bmatrix} > 0, \qquad \det A > 0.$$

- The first two conditions give $a > 0$ and $a > 1$, while

$$\det A = -a^3 + 7a - 6 = -(a-1)(a-2)(a+3).$$

- It easily follows that A is positive definite if and only if $1 < a < 2$.

Application 1: Second derivative test

- Given a function $f(x, y)$ of two variables, its directional derivative in the direction of a unit vector $\boldsymbol{u}$ is given by $D_{\boldsymbol{u}} f = u_1 f_x + u_2 f_y$.
- In particular, the second derivative of f in the direction of $\boldsymbol{u}$ is

$$
\begin{aligned}
D_{\boldsymbol{u}} D_{\boldsymbol{u}} f &= u_1(u_1 f_x + u_2 f_y)_x + u_2(u_1 f_x + u_2 f_y)_y \\
&= u_1^2 f_{xx} + u_1 u_2 f_{yx} + u_2 u_1 f_{xy} + u_2^2 f_{yy} \quad = \boldsymbol{u}^t A \boldsymbol{u}.
\end{aligned}
$$

- This computation allows us to classify the critical points of f. If the second derivative is positive for all $\boldsymbol{u} \neq 0$, then the function is convex in all directions and we get a local minimum. If the second derivative is negative for all $\boldsymbol{u} \neq 0$, then we get a local maximum.
- To classify the critical points, one looks at the Hessian matrix

$$
A = \begin{bmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{bmatrix}.
$$

This is symmetric, so it is diagonalisable with real eigenvalues. Once we now consider three cases, we obtain the second derivative test.

Application 2: Min/Max value on the unit sphere

- Let A be a symmetric $n \times n$ matrix and consider the quadratic form

$$Q(\boldsymbol{x}) = \sum_{i \leq j} a_{ij} x_i x_j = \boldsymbol{x}^t A \boldsymbol{x}.$$

- To find the minimum value of $Q(\boldsymbol{x})$ on the unit sphere $||\boldsymbol{x}|| = 1$, we let B be an orthogonal matrix such that $B^t AB$ is diagonal and then use the orthogonal change of variables $\boldsymbol{x} = B\boldsymbol{y}$ to write

$$Q(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \, y_i^2.$$

- Since $||\boldsymbol{y}|| = ||B\boldsymbol{y}|| = ||\boldsymbol{x}|| = 1$ by orthogonality, we find that

$$Q(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda_i \, y_i^2 \geq \sum_{i=1}^{n} \lambda_{\text{min}} \, y_i^2 = \lambda_{\text{min}}.$$

  In particular, $\min Q(\boldsymbol{x})$ is the smallest eigenvalue of A, while a similar argument shows that $\max Q(\boldsymbol{x})$ is the largest eigenvalue of A.

Application 3: Min/Max value of quadratics

- Every quadratic function of n variables can be expressed in the form

$$Q(\boldsymbol{x}) = \sum_{i \le j} a_{ij} x_i x_j + \sum_k b_k x_k + c = \boldsymbol{x}^t A \boldsymbol{x} + \boldsymbol{x}^t \boldsymbol{b} + c.$$

- Suppose A is positive definite symmetric and let $\boldsymbol{x}_0 = -\frac{1}{2} A^{-1} \boldsymbol{b}$. Then

$$Q(\boldsymbol{x}_0) = \boldsymbol{x}_0^t A \boldsymbol{x}_0 + \boldsymbol{x}_0^t \boldsymbol{b} + c = -\boldsymbol{x}_0^t A \boldsymbol{x}_0 + c$$

  is the minimum value that is attained by the quadratic because

$$\begin{aligned} 0 \le (\boldsymbol{x} - \boldsymbol{x}_0)^t A (\boldsymbol{x} - \boldsymbol{x}_0) &= \boldsymbol{x}^t A \boldsymbol{x} - 2 \boldsymbol{x}^t A \boldsymbol{x}_0 + \boldsymbol{x}_0^t A \boldsymbol{x}_0 \\ &= \boldsymbol{x}^t A \boldsymbol{x} + \boldsymbol{x}^t \boldsymbol{b} + c - Q(\boldsymbol{x}_0) \\ &= Q(\boldsymbol{x}) - Q(\boldsymbol{x}_0). \end{aligned}$$

- When A is negative definite symmetric, the inequality is reversed and thus $Q(\boldsymbol{x}_0)$ is the maximum value that is attained by the quadratic.