Chapter 6: Positive Definite Matrices

- Minima, Maxima and Saddle Points
- Test of Positive Definiteness
- Semidefinite and Indefinite Matrices
- Minimum Principles and Rayleigh Quotient

Minima, Maxima and Saddle Points

The Signs of Eigenvalues

- We have established that Hermitian (including real symmetric) matrices have real eigenvalues. It is valid to ask about their signs.
- This is related to the problem of finding the extremal points of a function of several variables.

Examples

$$F(x, y) = 7 + 2(x + y)^2 - y \sin y - x^3$$
$$f(x, y) = 2x^2 + 4xy + y^2$$

We want to find the minima of F and f. The Taylor series around a point (x_0, y_0) for function $g(x, y)$ is

$$g(x_0 + dx, y_0 + dy) = g(x_0, y_0) + g_x dx + g_y dy$$
$$+\frac{1}{2}(g_{xx} dx^2 + 2g_{xy} dx dy + g_{yy} dy^2) + \ldots$$

So the local behavior is governed by a polynomial of the second order.

Zeroth-, First- and Second-order Terms

- The zeroth-order terms have no effect at all.

$$F(0,0) = 7, \quad f(0,0) = 0$$

- The first-order terms must be zero for extremal points.

$$F_x(0,0) = F_y(0,0) = 0, \quad f_x(0,0) = f_y(0,0) = 0$$

- The second-order terms determine if it is a minimum, maximum or saddle point.

$$F_{xx} = 4, F_{xy} = 4, F_{yy} = 2, \quad f_{xx} = 4, f_{xy} = 4, f_{yy} = 2$$

Condition for Extremum

For the function $f(x, y) = ax^2 + 2bxy + cy^2$ of two variables, the point $(0, 0)$ is a

$$\begin{cases} \text{miminum,} & \text{if } a > 0,\ ac - b^2 > 0 \\ \text{maximum,} & \text{if } a < 0,\ ac - b^2 > 0 \\ \text{saddle point,} & \text{if } ac - b^2 < 0 \end{cases}$$

Since

$$ax^2 + 2bxy + cy^2 = a(x + \frac{b}{a}y)^2 + (c - \frac{b^2}{a})y^2.$$

Second-Order Polynomial to Matrix

$$f = 2x^2 + 4xy + y^2$$

$$f = 2xy$$

$$f = 2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2$$

Generalization to n Variables

Define

$$F(x) = x^T A x = \sum_{ij} a_{ij} x_i x_j,$$

where A is a real symmetric matrix. From the Taylor expansion, every function can be approximated in this form at a point with vanishing first derivatives and ignoring the constant.

Therefore the minimum/maximum problems can be analyzed in the framework of matrix algebra.

Real Symmetric Positive Definite Matrices

Tests for Positive Definite Matrices

The following four conditions are equivalent

- $x^T Ax > 0$ for all $x \neq 0$.
- All eigenvalues satisfy $\lambda_i > 0$.
- All determinants satisfy $\det A_k > 0$, where A_k is the upper-left $k \times k$ sub-matrix of A.
- All pivots satisfy $d_i > 0$.

Proofs

- (I $\Rightarrow$ II) $0 < x_i^T A x_i = \lambda_i (x_i^T x_i) \Rightarrow \lambda_i > 0$
- (I $\Leftarrow$ II) A has a complete set of orthonormal eigenvectors. From the spectral theorem $x^T A x = x^T Q \Lambda Q^T x = \sum_i c_i^2 \lambda_i > 0.$
- (I $\Rightarrow$ III) $\begin{bmatrix} x_k^T & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^T A_k x_k > 0,$

  $\Rightarrow A_k$ p.d. $\Rightarrow \det A_k = \prod_{i=1}^k \lambda_i > 0.$
- (III $\Rightarrow$ IV) $d_k = \frac{\det A_k}{\det A_{k-1}} > 0$
- (IV $\Rightarrow$ I) $A = LDL^T \Rightarrow x^T A x = \sum_i d_i (L^T x)_i^2 > 0$

Positive Definiteness and Least Squares

A is p.d. iff there exists a matrix R with independent columns such that

$$A = R^T R.$$

"If"

$$x^T A x = x^T R^T R x = ||Rx||^2 > 0 \text{ if } x \neq 0.$$

"only if"

$$A = LDL^T = LD^{1/2}D^{1/2}L^T = R^T R.$$

Ellipsoids in n Dimensions

If A is diagonal with positive entries, then the graph of

$$x^T Ax = 1,$$

is an ellipsoid. In general, if A is p.d., then the graph is an ellipsoid in n-dimensional space, since

$$x^T Ax = x^T Q\Lambda Q^T x = y^T \Lambda y = \sum_i \lambda_i y_i^2 = 1,$$

where $y_i = q_i^T x$ is the component of x along the ith eigenvector. The axes have lengths $1/\sqrt{\lambda_i}$ and point along the direction of the eigenvectors of A.

Example

Consider the equation

$$x^T A x = 5u^2 + 8uv + 5v^2 = 1,$$

where $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ is p.d.. Solving for the eigenvalues we have $\lambda_1 = 1$ and $\lambda_2 = 9$, and the eigenvectors are $\frac{1}{\sqrt{2}}[1, -1]$ and $\frac{1}{\sqrt{2}}[1, 1]$. We can re-write the equation as

$$5u^2 + 8uv + 5v^2 = 1\left(\frac{u}{\sqrt{2}} - \frac{v}{\sqrt{2}}\right)^2 + 9\left(\frac{u}{\sqrt{2}} + \frac{v}{\sqrt{2}}\right)^2 = 1,$$

which is clearly an ellipsoid.

Semi-Definite and Indefinite Matrices

Positive Semi-Definite Matrices

The following statements are equivalent

- $x^T Ax \geq 0$ for all $x \neq 0$ (definition for p.s.d.)
- All eigenvalues satisfy $\lambda_i \geq 0$
- All principle sub-matrices have non-negative determinants
- No pivots are negative
- $A = R^T R$ for some R, possibly with dependent columns

Example

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

Apparently there are eigenvalues of 0 and 3, the other is 3.

$$x^T Ax = 2(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3 - x_1x_3) \geq 0$$

The pivots are 2, $\frac{3}{2}$. The determinants of principle sub-matrices are $2, 3$ and 0.

Congruence Transformations

Many quantities are invariant with respect to row operations. Similarity transformations do not alter the eigenvalues. What transformations leave the signs of eigenvalues unchanged?

Answer: $A \to C^T AC$, where C is non-singular.

Let $C(t) = tQ + (1-t)QR$, where Q is the orthogonal matrix yielded by Gram-Schmidt process on C. Then $C(0) = C$ and $C(1) = Q$. By varying t from 0 to 1, $C(t)$ remains non-singular. The eigenvalues of $C(t)^T AC(t)$ also vary but they never change signs. So $C^T AC$ $(t=0)$ and $Q^T AQ$ $(t=1)$ have the same number of positive (and negative) eigenvalues.

Signs of Pivots and Eigenvalues

Since $A = LDL^T$, A is a congruence transformation of D. Therefore, the signs of the pivots agree with the signs of the eigenvalues (Λ and D look similar). This is useful in looking for eigenvalues: pivots are much simpler to find then the eigenvalues when the dimension is large. Example

$$A = \begin{bmatrix} 3 & 3 & 0 \\ 3 & 10 & 7 \\ 0 & 7 & 8 \end{bmatrix}, \quad A - 2I = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 8 & 7 \\ 0 & 7 & 6 \end{bmatrix}$$

By looking at the pivots, one can know that there is an eigenvalue between 0 and 2.

The Generalized Eigenvalue Problem

$$Ax = \lambda M x,$$

where M is assumed positive definite. Let $M = R^T R$ and $y = Rx, x = R^{-1}y \triangleq Cy$. Then

$$Ax = \lambda M x \Rightarrow AR^{-1}y = \lambda R^T y \Rightarrow C^T A C y = \lambda y$$

- λs real; y_is can be orthonormal
- they have the same signs as A
- $y_i^T y_j = x_i^T M x_j = \delta_{ij}$ (M-orthonormal)
- $x_i^T A x_j = \lambda_j \delta_{ij}$ (simultaneous diagonalization)

Minimum Principles

Minimum Principle for Linear Equations

- If A is p.d., then $P(x) = \frac{1}{2}x^T Ax - x^T b$ reaches its minimum at the point $Ax = b$, and $P_{\min} = -\frac{1}{2}b^T A^{-1}b$
- Proof: suppose $Ax = b$, then

$$P(y) - P(x) = \frac{1}{2}(y - x)^T A(y - x) \geq 0.$$

- Example: minimize
  $P(x) = x_1^2 - x_1 x_2 + x_2^2 - b_1 x_1 - b_2 x_2.$

Minimum Principle for Eigenvalue Problem

- Rayleigh's quotient

$$R(x) = \frac{x^T A x}{x^T x}.$$

- Rayleigh's principle:
  The quotient $R(x)$ is minimized by the first eigenvector x_1 corresponding to the smallest eigenvalue λ_1.

- Proof: suppose Q is an orthonormal eigenvector matrix for A and $y = Q^T x$, then

$$\frac{x^T A x}{x^T x} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum_i \lambda_i y_i^2}{\sum_i y_i^2} \geq \lambda_1$$

The Eigenvalues in the Middle

- From the above arguments,

$$\lambda_1 \leq R(x) \leq \lambda_n.$$

  The minimum (maximum) of $R(x)$ is the smallest (largest) eigenvalue of A. How about the eigenvalues in the middle?

- Idea: If we look in the subspace orthogonal to the first eigenvector and minimize the Rayleigh quotient, we get the second smallest eigenvalue. If the subspace is orthogonal to an arbitrary vector z, then the minimum of the Rayleigh quotient is not greater then λ_2.

The Minimax Principle

- For the second smallest eigenvalue,

$$\lambda_2 \geq \lambda(z) = \min_{x^T z=0} R(x) \quad \text{and} \quad \lambda_2 = \max_{z} \left[\min_{x^T z=0} R(x) \right].$$

  The maximum occurs when z is the first eigenvector x_1.

- Quite similarly, when S_2 is a $2-D$ subspace,

$$\lambda_2 \leq \max_{x \in S_2} R(x) \quad \text{and} \quad \lambda_2 = \min_{S_2} \left[\max_{x \in S_2} R(x) \right].$$

  The minimum occurs when S_2 is spanned by x_1, x_2.
  Generally,

$$\lambda_j = \min_{S_j} \left[\max_{x \in S_j} R(x) \right].$$