Functions of Several Variables

The Unconstrained Minimization Problem

In n dimensions the unconstrained problem is stated as

$$\boxed{\text{minimize } f(x),\ x \in \mathbb{R}^n}$$

where $f(x)$ is a scalar objective function of vector argument x, and x is a column vector of n real variables.

We can take the *gradient* of $f(x)$, as

$$\boxed{\nabla f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T}$$

which is a column vector of n functions.

The $n \times n$ matrix of second derivatives is called the *Hessian matrix* and can be found by taking the gradient again, thus

$$\boxed{H(x_1, \ldots, x_n) = \nabla\nabla^T f = \nabla^2 f}$$

or more specifically, the ∇^2 operator can be written as

$$\nabla^2 = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} & \cdot & \cdot & \frac{\partial^2}{\partial x_1 \partial x_n} \\ \cdot & \frac{\partial^2}{\partial x_i \partial x_j} & & \cdot \\ \cdot & & & \cdot \\ \frac{\partial^2}{\partial x_n \partial x_1} & \cdot & \cdot & \frac{\partial^2}{\partial x_n^2} \end{bmatrix} = \left[\frac{\partial^2}{\partial x_i \partial x_j}\right]_{n \times n}$$

where the subscript i denotes the row and j denotes the column.

The Hessian matrix is *symmetric* because of the properties of the second derivatives of continuous functions.

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{\partial^2}{\partial x_j \partial x_i} f(x)$$

Some Definitions

Definition a point x^{**} is said to be the *strong global minimum* of a function $f(x)$ if for all $x \in \mathbb{R}^n$, $f(x) - f(x^{**}) > 0$. ❐

Definition a point x^* is said to be a *strong local minimum* of a function $f(x)$ if there exists a $\delta > 0$ such that $f(x) - f(x^*) > 0$ for all x such that $\|x - x^*\| < \delta$. ❐

If the inequalities in $f(x) - f(x^{**}) > 0$ and $f(x) - f(x^*) > 0$ are replaced by $<$ then we have strong global and local *maxima*. If the inequalities in $f(x) - f(x^{**}) > 0$ and $f(x) - f(x^*) > 0$ are replaced by $\geq$ then the minima are called *weak* global and *weak* local minima.

Taylor Expansion

We can expand $f(x)$ in a Taylor expansion about x as

$$f(x + \Delta x) = f(x) + \Delta x^T \nabla f(x) + \frac{1}{2}\Delta x^T \nabla^2 f(x)\, \Delta x + O(\|\Delta x\|^3)$$

where the notation $O(\|\Delta x\|^3)$ represents terms of "order" $\|\Delta x\|^3$ and these can be neglected for $\|\Delta x\|$ sufficiently small.

From the definition of local minimum, for x^* to be a local minimum this implies that $f(x^* + \Delta x) - f(x^*) > 0$ for $\|\Delta x\| > 0$.

From the Taylor expansion

$$f(x^* + \Delta x) - f(x^*) = \Delta x^T \nabla f(x^*) + \frac{1}{2}\Delta x^T \nabla^2 f(x^*)\, \Delta x$$

must be greater than 0, where we have neglected $O(\|\Delta x\|^3)$ terms.

For this to be true it is obvious that

$$\nabla f(x^*) = 0$$

otherwise we could pick Δx appropriately so that $\Delta x^T \nabla f(x^*) < 0$.

Necessary Conditions

A necessary condition for a local minimum is the analogue of a one dimensional function having a stationary point. Thus we can state the following definition and theorem.

Definition If $\nabla f(x^*) = 0$ then x^* is called a *stationary point* of $f(x)$.❐

Theorem necessary condition for local minimum based on $\nabla f(x)$

x^* is a local minimum of $f(x)$ only if x^* is a stationary point of $f(x)$, that is $\nabla f(x^*) = 0$.❐

Sufficient Conditions

In order to obtain sufficient conditions from the Taylor expansion we need to study the second term:

$$\boxed{Q(x) = \frac{1}{2}\Delta x^T \nabla^2 f(x^*)\, \Delta x}$$

This is called a *quadratic form* which can be written more generally as

$$Q(z) = z^T A z$$

where $z \in \mathbb{R}^n$ is an arbitrary column vector,
$A \in \mathbb{R}^{n \times n}$ is an arbitrary matrix and
the quadratic form is a scalar.

The properties of the quadratic form are governed by the properties of the matrix A.

Definition The matrix A is called:

1. positive definite if $Q(z) > 0$, $\forall z$, $z \neq 0$,
2. positive semidefinite if $Q(z) \geq 0$, $\forall z$, $z \neq 0$,
3. negative definite if $Q(z) < 0$, $\forall z$, $z \neq 0$,
4. negative semidefinite if $Q(z) \leq 0$, $\forall z$, $z \neq 0$,and
5. indefinite if the sign of $Q(z)$ depends on z. ❐

Theorem Sufficient condition based on $(\nabla f, \nabla^2 f)$

if $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite then x^* is a strong local minimum of $f(x)$. ❐

Note this is not a necessary condition since if $\nabla^2 f(x^*) = 0$ and the $O(\Delta x^3)$ terms are positive then x^* is still a strong local minimum. Therefore a necessary condition involving $\nabla^2 f(x^*)$ can be stated as follows:

Theorem Necessary condition based on $(\nabla f, \nabla^2 f)$
x^* is a local minimum only if $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite. ❐

Definition a stationary point which is not a local minimum or local maximum is called a *saddle point*. ❐

Example: Consider the function

$$f(x) = x_1^2 - x_1 x_2 + x_2^2$$

the gradient is easily found as

$$\nabla f(x) = \begin{bmatrix} 2x_1 - x_2 & 2x_2 - x_1 \end{bmatrix}^T$$

and thus

$$\nabla f(x^*) = 0 \Rightarrow x_1^* = x_2^* = 0$$

$x^* = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$ is a stationary point. The Hessian

$$\nabla^2 f(x^*) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = H(x^*) = H$$

is independent of x since *f(x)* is a quadratic function. We now determine whether the Hessian is positive definite by looking at the quadratic form

$$z^T H z = \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2z_1 - z_2 & -z_1 + 2z_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 2z_1^2 - z_1 z_2 - z_1 z_2 + 2z_2^2$$

which can easily be written as

$$z^T H z = 2[(z_1 - z_2)^2 + z_1 z_2] = 2[(z_1 + z_2)^2 - 3z_1 z_2].$$

Now we note that the first form is positive for z_1 and z_2 both negative or positive and that the second form is always positive for z_1 and z_2 of opposite sign. Thus $z^T H z > 0$, $\forall z, z \neq 0$ and thus $H(x^*) = H$ is positive definite $\Rightarrow x^*$ is a local minimum.■

Examples of quadratic forms

A quadratic form can always be made so that the matrix is symmetric, for consider the following examples:

a) $x_1^2 + x_1x_2 + 3x_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & .5 \\ .5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

b) $x_1^2 + x_1x_2 + 4x_1x_3 + 2x_3^2 = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & .5 & 2 \\ .5 & 0 & 0 \\ 2 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

c) $x_1^2 + 2x_2^2 + 3x_3^2 = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

d) consider the general two dimensional case, Q can always be chosen symmetric:

$$x_1^2 + x_1x_2 + 3x_2^2 \Rightarrow Q = \begin{bmatrix} 1 & b \\ a & 3 \end{bmatrix}$$

such that $a + b = 1$, or the symmetric form

$$Q = \begin{bmatrix} 1 & c \\ c & 3 \end{bmatrix}$$

such that $c = \frac{(a+b)}{2} = 0.5$. ■

Tests for Definiteness

Definition the *principal minors* of a matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}$$

are written as $\Delta_1 = a_{11}$, $\Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, $\Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$, $\dots\Delta_n = |A|$. ❒

Definition the eigenvalues of a matrix A are defined as all the scalars, λ_i which satisfy the characteristic equation $|A - \lambda_i I| = 0$. ❒

An nxn real symmetric matrix will have n real eigenvalues.

Theorem Sylvester's Theorem
A quadratic form $z^T Qz$ is positive definite if and only if all the leading principle determinants of the matrix Q are positive. ❒

Table 1: Tests for Definiteness

definiteness	eigenvalues	principal minors
1. positive definite	all $\lambda_i > 0$	$\Delta_i > 0$, $i = 1\dots n$
2. positive semidefinite	all $\lambda_i \geq 0$	$\Delta_i \geq 0$, $i = 1\dots n$
3. negative definite	all $\lambda_i < 0$	$\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0\dots$
4. negative semidefinite	all $\lambda_i \leq 0$	$\Delta_1 \leq 0$, $\Delta_2 \geq 0$, $\Delta_3 \leq 0\dots$
5. indefinite	some $\lambda_i \geq 0$, some $\lambda_i \leq 0$	none of the above

It is also obvious from the definition of positive definite that:

Theorem the matrix A is negative definite if $-A$ is positive definite. ❒

Examples

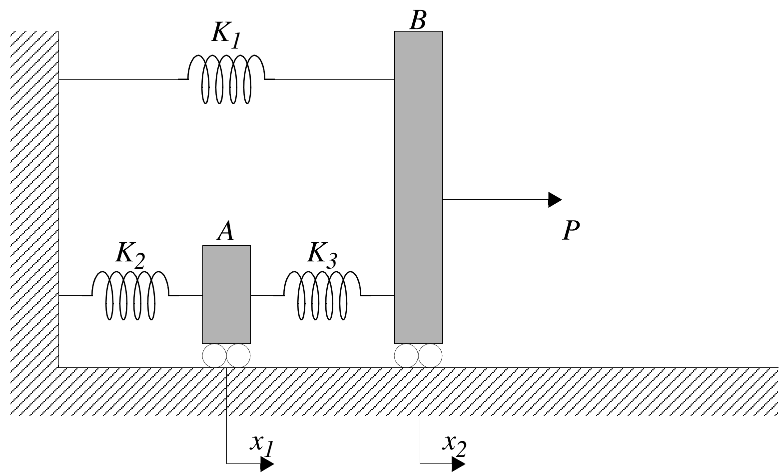
Determine the sign definiteness of the following:

1. $A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 3 & 0 \\ 2 & 0 & 5 \end{bmatrix}$; we have $\Delta_1 = 2$, $\Delta_2 = \begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} = 5$, $\Delta_3 = \begin{vmatrix} 2 & 1 & 2 \\ 1 & 3 & 0 \\ 2 & 0 & 5 \end{vmatrix} = -12 + 25 = 13$

   and therefore since all leading principal determinants are positive $\Rightarrow$ A is positive definite.

2. $A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & -3 & 0 \\ 2 & 0 & -5 \end{bmatrix}$; we have $\Delta_1 = 2$, $\Delta_2 = \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} = -7$ and $\Rightarrow A$ is indefinite. ■

Example: Static Linear Springs



System of linear springs in equilibrium

The blocks A and B are two frictionless bodies connected to 3 linear elastic springs having spring constants K_1, K_2, K_3. When the force $P = 0$, $x_1 = 0$, $x_2 = 0$ defines the natural position. We want to find new positions x_1, x_2 when a nonzero force P is applied.

<u>Solution</u>:

Recall that for a linear spring *Hooke's Law* says that $F = Kd$ where K is the spring constant and d is the displacement of the spring from equilibrium. The strain energy of a spring is the work done in stretching it, that is

$$E = \int_0^x Kx\,dx = \frac{1}{2}Kx^2.$$

The work put into the system by the constant force P is given by Px_2 and thus the potential energy of the system is

$$U = \frac{1}{2}K_2x_1^2 + \frac{1}{2}K_3(x_2 - x_1)^2 + \frac{1}{2}K_1x_2^2 - Px_2 .$$

We want to minimize this potential energy U according to the principle of minimum potential energy. Thus we have

$$\frac{\partial U}{\partial x_1} = K_2x_1^* - K_3(x_2^* - x_1^*) = 0$$

$$\frac{\partial U}{\partial x_2} = K_3(x_2^* - x_1^*) + K_1x_2^* - P = 0$$

or

$$\begin{bmatrix} K_2 + K_3 & -K_3 \\ -K_3 & K_1 + K_3 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ P \end{bmatrix}$$

which gives

$$x_1^* = \frac{PK_3}{(K_1K_2 + K_1K_3 + K_2K_3)}$$

$$x_2^* = \frac{P(K_2 + K_3)}{(K_1K_2 + K_1K_3 + K_2K_3)}$$

as the only stationary point. We now determine the Hessian as

$$H(x) = \begin{bmatrix} \frac{\partial^2 U}{\partial x_1^2} & \frac{\partial^2 U}{\partial x_1 \partial x_2} \\ \frac{\partial^2 U}{\partial x_1 \partial x_2} & \frac{\partial^2 U}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} K_2 + K_3 & -K_3 \\ -K_3 & K_1 + K_3 \end{bmatrix}$$

and need to know whether $H(x)$ positive definite? The principal minors are given as

$$\Delta_1 = K_2 + K_3 > 0 , \ \Delta_2 = \begin{vmatrix} K_2 + K_3 & -K_3 \\ -K_3 & K_1 + K_3 \end{vmatrix} = (K_1K_2 + K_1K_3 + K_2K_3) > 0$$

and therefore x_1^*, x_2^* is the minimum. The value of the potential at this minimum is given as

$$U_{\min} = -0.5\frac{P^2(K_2 + K_3)}{K_2K_3 + K_2K_1 + K_1K_3} . \blacksquare$$

Quadratic Functions

In general we write a scalar quadratic function of n variables, $x \in \mathbb{R}^n$, in matrix notation as

$$q(x) = c + b^T x + \frac{1}{2} x^T A x$$

where $c \in \mathbb{R}$, $b \in \mathbb{R}^n$, and the matrix $A \in \mathbb{R}^{n \times n}$.

We note that from Taylor's expansion any function can be approximated by a quadratic function

$$f(x + \Delta x) = f(x) + \Delta x^T \nabla f(x) + \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x + O(\Delta x^3)$$

in a small enough region. Therefore what we now say about quadratic functions $q(x)$ can be used later in approximations.

In algebraic form we write $q(x)$ as

$$q(x) = c + \sum_{i=1}^{n} b_i x_i + \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} x_j a_{ij} x_i$$

where $b^T = [b_1 \ldots b_n]$, and $A = [a_{ij}]_{n \times n}$. Since $x^T A x$ is a quadratic form we can always assume it to be symmetric.

Definition If A is positive definite then $q(x)$ is called a *positive definite quadratic function*. ❐

Question Show that the gradient vector of a quadratic function $q(x) = c + b^T x + \frac{1}{2} x^T A x$ is given by $\nabla q(x) = Ax + b$.

<u>Solution</u>: To find the gradient vector we require: $\frac{\partial q}{\partial x_k}$, $k = 1, \ldots, n$. Writing $q(x)$ algebraically and rearranging we have:

$$q(x) = c + \sum_i b_i x_i + \frac{1}{2}\left[x_k \sum_j a_{kj} x_j + \sum_{i \neq k} \sum_j x_i a_{ij} x_j\right]$$

$$= c + \sum_i b_i x_i + \frac{1}{2}\left[a_{kk} x_k^2 + x_k \sum_{j \neq k} a_{kj} x_j + x_k \sum_{i \neq k} x_i a_{ik} + \sum_{i \neq k} \sum_{j \neq k} x_i a_{ij} x_j\right]$$

Now since A is symmetric $a_{ij} = a_{ji}$ and we have

$$q(x) = c + \sum_i b_i x_i + \frac{1}{2}\left[a_{kk} x_k^{\ 2} + 2x_k \sum_{j \neq k} a_{kj} x_j + \sum_{i \neq k} \sum_{j \neq k} x_i a_{ij} x_j\right]$$

and therefore

$$\frac{\partial q}{\partial x_k} = b_k + \frac{1}{2}\left[2a_{kk} x_k + 2\sum_{j \neq k} a_{kj} x_i\right] = b_k + \sum_j a_{kj} x_j .$$

This result can be written in vactor notation as

$$\boxed{\nabla q(x) = b + Ax}$$

Thus we see that the gradient of a quadratic is easily found by a matrix multiplication and vector addition. ■

Question Show that the Hessian matrix of the quadratic function $q(x) = c + b^T x + \frac{1}{2} x^T A x$ is equal to A.

<u>Solution</u>: We know that the Hessian matrix is given as

$$H(x) = \nabla(\nabla q(x))^T = \left[\frac{\partial^2}{\partial x_i \partial x_j} q(x)\right]_{n \times n}$$

and from the above we see that

$$\frac{\partial q}{\partial x_j} = b_j + \sum_k a_{jk} x_k .$$

Therefore

$$\frac{\partial}{\partial x_i \partial x_j} q(x) = a_{ji} = a_{ij}$$

and in matrix form

$$\boxed{H(x) = H = A}$$

and the Hessian of a quadratic function is a constant matrix. ■

Thus we see that if $q(x)$ is a positive definite quadratic function then H is positive definite at all points which implies that x^* will be a minimum. In fact since

$$\nabla q(x^*) = b + Ax^* = 0$$

$$\boxed{x^* = -A^{-1}b}$$

is the <u>unique minimum</u> of a positive definite quadratic function.

Example

Consider the following quadratic function in two variables:

$$f(x_1, x_2) = K + dx_1 + ex_2 + cx_2^2 + bx_1x_2 + ax_1^2$$

where: $a, b, c, d, e, K \in \subsetneq$. Determine the conditions under which $f(x)$ is positive definite.

<u>Solution</u>: Rewriting $f(x)$ in matrix form as

$$f(x) = K + \begin{bmatrix} d & e \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

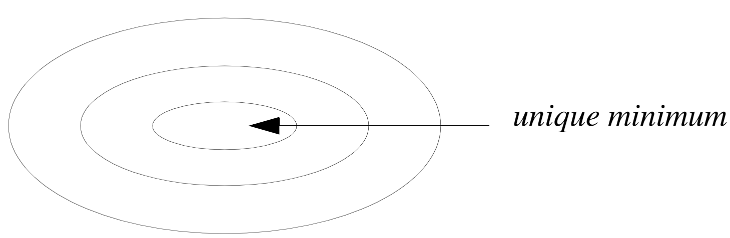
we see that the Hessian is given by

$$A = \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix}.$$

Therefore by Sylvester's theorem A is positive definite if and only if:

$$\Delta_1 = a > 0, \Delta_2 = 4ac - b^2 > 0 \quad \Rightarrow c > 0.$$

From geometry we know that if $4ac - b^2 > 0$ then the contour plots $f(x) = const.$ define concentric ellipses.



Concentric ellipses define shape of quadratic function in 2-D

As a specific example if $a = d = K = 1$, $e = -1$, $b = 0$, and $c = 2$, we have

$$f(x) = 1 + x_1 - x_2 + x_1^2 + 2x_2^2$$

$$f(x) = 1 + \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and therefore

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}, \ b = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

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

$$x^* = -A^{-1}b = -\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{-1}{2} \\ \frac{1}{4} \end{bmatrix}$$

is the unique minimum. ■