AS2051 Calculus (and Linear Algebra)

Dr Oliver Kerr

2012–13

Recall...

We were looking for a rule to determine whether a stationary point of a function of several variables is a local maximum or minimum. For the sake of simplification of notation (i.e get rid of the δs) we will move the origin to the stationary point. Then the Taylor series expansion gives

$$f(x_1, x_2, \ldots, x_n) = f(0, 0, \ldots, 0) + \frac{1}{2} \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} H_n \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \cdots$$

where

$$H_n = \begin{pmatrix} f_{x_1x_1}(0,0) & f_{x_1x_2}(0,0) & \cdots & f_{x_1x_n}(0,0) \\ f_{x_2x_1}(0,0) & f_{x_2x_2}(0,0) & \cdots & f_{x_2x_n}(0,0) \\ \vdots & \vdots & & \vdots \\ f_{x_nx_1}(0,0) & f_{x_nx_2}(0,0) & \cdots & f_{x_nx_n}(0,0) \end{pmatrix}$$

If f is smooth enough, then we can change the order of differentiation, and so the Hessian matrix H_n is a symmetric matrix.

Some Linear Algebra...

If A is a real symmetric matrix, then we can find an orthogonal matrix R such that

$$A = R^{-1}DR$$

where D is a diagonal matrix.

The orthogonal matrix has the properties that $R^T = R^{-1}$ and $|R| = 1$.

So

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T R^T DR\mathbf{x} = (R\mathbf{x})^T D(R\mathbf{x})$$

If we let

$$\begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix} = R \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

and

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

then

$$f(x_1, x_2, \dots, x_n) = f(0, 0, \dots, 0) + \frac{1}{2}\left(\lambda_1 {x_1'}^2 + \lambda_2 {x_2'}^2 + \cdots + \lambda_n {x_n'}^2\right) + \cdots$$

If

$$f(x_1, x_2, \dots, x_n) = f(0, 0, \dots, 0) + \frac{1}{2}\left(\lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \dots + \lambda_n x_n'^2\right) + \cdots$$

Then clearly if $\lambda_i > 0$ for all i then we have a minimum, and if $\lambda_i < 0$ for all i we have a maximum. (A matrix with this property is said to be positive definite).

Note that

$$|H_n| = |R^T DR| = |R||D||R| = |D| = \lambda_1 \lambda_2 \dots \lambda_n$$

If we have a non-zero H_n then we can conclude that

- if we have a minimum then we must have $|H_n| > 0$
- if we have a maximum then $|H_n| > 0$ if n is even and $|H_n| < 0$ if n is odd.

If we restrict ourselves to just looking at the first $n-1$ coordinates by setting $n_n = 0$, then

$$f(x_1, x_2, \ldots, x_{n-1}, 0) = f(0, 0, \ldots, 0) + \frac{1}{2}\left(\lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \cdots + \lambda_n x_n'^2\right) +$$

will still be a maximum or minimum as before. And so

- if we have a minimum then we must have $|H_{n-1}| > 0$
- if we have a maximum then $|H_{n-1}| > 0$ if $n-1$ is even and $|H_{n-1}| < 0$ if $n-1$ is odd.

The last condition can be re-written as

- if we have a minimum then we must have $|H_{n-1}| > 0$
- if we have a maximum then $|H_{n-1}| > 0$ if n is odd and $|H_{n-1}| < 0$ if n is even.

We repeat this exercise and conclude:

- If we have a minimum then we have

$$|H_1| > 0, \quad |H_2| > 0, \quad |H_3| > 0, \quad \ldots, \quad |H_n| > 0$$

- If we have a maximum then if n is even

$$|H_1| < 0, \quad |H_2| > 0, \quad |H_3| < 0, \quad \ldots, \quad |H_n| > 0$$

and if n is odd

$$|H_1| < 0, \quad |H_2| > 0, \quad |H_3| < 0, \quad \ldots, \quad |H_n| < 0$$

We have shown that if f has a minimum or maximum then the test must be true. We have not shown that if the test is true then f has a minimum or maximum.

This is not quite as straightforward — we need to show that if, for the case of a minimum, the determinants of all the matrices H_1, H_2, ... H_n are positive then indeed λ_1, λ_2, ... λ_n are all positive. For a flavour of a proof of this result, which is essentially Sylvester's criterion, see the Wikipedia article at

http://en.wikipedia.org/wiki/Sylvester's_criterion