Lecture 5
Principal Minors and the Hessian

Eivind Eriksen

BI Norwegian School of Management
Department of Economics

October 01, 2010

Principal minors

Let A be a symmetric $n \times n$ matrix. We know that we can determine the definiteness of A by computing its eigenvalues. Another method is to use the principal minors.

Definition

A minor of A of order k is principal if it is obtained by deleting $n-k$ rows and the $n-k$ columns with the same numbers. The leading principal minor of A of order k is the minor of order k obtained by deleting the last $n-k$ rows and columns.

For instance, in a principal minor where you have deleted row 1 and 3, you should also delete column 1 and 3.

Notation

We write D_k for the leading principal minor of order k. There are $\binom{n}{k}$ principal minors of order k, and we write Δ_k for any of the principal minors of order k.

Two by two symmetric matrices

Example

Let $A = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ *be a symmetric* 2×2 *matrix. Then the leading principal minors are* $D_1 = a$ *and* $D_2 = ac - b^2$*. If we want to find all the principal minors, these are given by* $\Delta_1 = a$ *and* $\Delta_1 = c$ *(of order one) and* $\Delta_2 = ac - b^2$ *(of order two).*

Let us compute what it means that the leading principal minors are positive for 2×2 matrices:

Example

Let $A = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ *be a symmetric* 2×2 *matrix. Show that if* $D_1 = a > 0$ *and* $D_2 = ac - b^2 > 0$*, then* A *is positive definite.*

Leading principal minors: An example

Solution

If $D_1 = a > 0$ *and* $D_2 = ac - b^2 > 0$*, then* $c > 0$ *also, since* $ac > b^2 \geq 0$*. The characteristic equation of* A *is*

$$\lambda^2 - (a+c)\lambda + (ac - b^2) = 0$$

and it has two solutions (since A *is symmetric) given by*

$$\lambda = \frac{a+c}{2} \pm \frac{\sqrt{(a+c)^2 - 4(ac - b^2)}}{2}$$

Both solutions are positive, since $(a+c) > \sqrt{(a+c)^2 - 4(ac - b^2)}$*. This means that* A *is positive definite.*

Definiteness and principal minors

Theorem

Let A be a symmetric $n \times n$ matrix. Then we have:

- *A is positive definite $\Leftrightarrow D_k > 0$ for all leading principal minors*
- *A is negative definite $\Leftrightarrow (-1)^k D_k > 0$ for all leading principal minors*
- *A is positive semidefinite $\Leftrightarrow \Delta_k \geq 0$ for all principal minors*
- *A is negative semidefinite $\Leftrightarrow (-1)^k \Delta_k \geq 0$ for all principal minors*

- In the first two cases, it is enough to check the inequality for all the leading principal minors (i.e. for $1 \leq k \leq n$).
- In the last two cases, we must check for all principal minors (i.e for each k with $1 \leq k \leq n$ and for each of the $\binom{n}{k}$ principal minors of order k).

Definiteness: An example

Example

Determine the definiteness of the symmetric 3×3 matrix

$$A = \begin{pmatrix} 1 & 4 & 6 \\ 4 & 2 & 1 \\ 6 & 1 & 6 \end{pmatrix}$$

Solution

One may try to compute the eigenvalues of A. However, the characteristic equation is

$$\det(A - \lambda I) = (1 - \lambda)(\lambda^2 - 8\lambda + 11) - 4(18 - 4\lambda) + 6(6\lambda - 16) = 0$$

This equations (of order three with no obvious factorization) seems difficult to solve!

Definiteness: An example

Solution (Continued)

Let us instead try to use the leading principal minors. They are:

$$D_1 = 1, \quad D_2 = \begin{vmatrix} 1 & 4 \\ 4 & 2 \end{vmatrix} = -14, \quad D_3 = \begin{vmatrix} 1 & 4 & 6 \\ 4 & 2 & 1 \\ 6 & 1 & 6 \end{vmatrix} = -109$$

Let us compare with the criteria in the theorem:

- *Positive definite:* $D_1 > 0, D_2 > 0, D_3 > 0$
- *Negative definite:* $D_1 < 0, D_2 > 0, D_3 < 0$
- *Positive semidefinite:* $\Delta_1 \geq 0, \Delta_2 \geq 0, \Delta_3 \geq 0$ *for all principal minors*
- *Negative semidefinite:* $\Delta_1 \leq 0, \Delta_2 \geq 0, \Delta_3 \leq 0$ *for all principal minors*

The principal leading minors we have computed do not fit with any of these criteria. We can therefore conclude that A is indefinite.

Definiteness: Another example

Example

Determine the definiteness of the quadratic form

$$Q(x_1, x_2, x_3) = 3x_1^2 + 6x_1x_3 + x_2^2 - 4x_2x_3 + 8x_3^2$$

Solution

The symmetric matrix associated with the quadratic form Q is

$$A = \begin{pmatrix} 3 & 0 & 3 \\ 0 & 1 & -2 \\ 3 & -2 & 8 \end{pmatrix}$$

Since the leading principal minors are positive, Q is positive definite:

$$D_1 = 3, \quad D_2 = \begin{vmatrix} 3 & 0 \\ 0 & 1 \end{vmatrix} = 3, \quad D_3 = \begin{vmatrix} 3 & 0 & 3 \\ 0 & 1 & -2 \\ 3 & -2 & 8 \end{vmatrix} = 3$$

Optimization of functions of several variables

The first part of this course was centered around matrices and linear algebra. The next part will be centered around optimization problems for functions $f(x_1, x_2, \ldots, x_n)$ in several variables.

C^2 functions

Let $f(x_1, x_2, \ldots, x_n)$ be a function in n variables. We say that f is C^2 if all second order partial derivatives of f exist and are continuous.

All functions that we meet in this course are C^2 functions, so shall not check this in each case.

Notation

We use the notation $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Hence we write $f(\mathbf{x})$ in place of $f(x_1, x_2, \ldots, x_n)$.

Stationary points of functions in several variables

The partial derivatives of $f(\mathbf{x})$ are written $f_1', f_2', \dots, f_n'$. The stationary points of f are the solutions of the first order conditions:

Definition

Let $f(\mathbf{x})$ be a function in n variables. We say that $\mathbf{x}$ is a stationary point of f if

$$f_1'(\mathbf{x}) = f_2'(\mathbf{x}) = \cdots = f_n'(\mathbf{x}) = 0$$

Let us look at an example:

Example

Let $f(x_1, x_2) = x_1^2 - x_2^2 - x_1 x_2$. Find the stationary points of f.

Stationary points: An example

Solution

We compute the partial derivatives of $f(x_1, x_2) = x_1^2 - x_2^2 - x_1 x_2$ *to be*

$$f_1'(x_1, x_2) = 2x_1 - x_2, \quad f_2'(x_1, x_2) = -2x_2 - x_1$$

Hence the stationary points are the solutions of the following linear system

$$\begin{array}{rcrcl} 2x_1 & - & x_2 & = & 0 \\ -x_1 & - & 2x_2 & = & 0 \end{array}$$

Since $\det \begin{pmatrix} 2 & -1 \\ -1 & -2 \end{pmatrix} = -5 \neq 0$*, the only stationary point is* $\mathbf{x} = (0, 0)$.

Given a stationary point of $f(\mathbf{x})$, how do we determine its type? Is it a local minimum, a local maximum or perhaps a saddle point?

The Hessian matrix

Let $f(\mathbf{x})$ be a function in n variables. The Hessian matrix of f is the matrix consisting of all the second order partial derivatives of f:

Definition

The Hessian matrix of f at the point $\mathbf{x}$ is the $n \times n$ matrix

$$\mathbf{f}''(\mathbf{x}) = \begin{pmatrix} f''_{11}(\mathbf{x}) & f''_{12}(\mathbf{x}) & \dots & f''_{1n}(\mathbf{x}) \\ f''_{21}(\mathbf{x}) & f''_{22}(\mathbf{x}) & \dots & f''_{2n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f''_{n1}(\mathbf{x}) & f''_{n2}(\mathbf{x}) & \dots & f''_{nn}(\mathbf{x}) \end{pmatrix}$$

Notice that each entry in the Hessian matrix is a second order partial derivative, and therefore a function in $\mathbf{x}$.

The Hessian matrix: An example

Example

Compute the Hessian matrix of the quadratic form

$$f(x_1, x_2) = x_1^2 - x_2^2 - x_1 x_2$$

Solution

We computed the first order partial derivatives above, and found that

$$f_1'(x_1, x_2) = 2x_1 - x_2, \quad f_2'(x_1, x_2) = -2x_2 - x_1$$

Hence we get

$$\begin{aligned} f_{11}''(x_1, x_2) &= 2 & \qquad f_{12}''(x_1, x_2) &= -1 \\ f_{21}''(x_1, x_2) &= -1 & \qquad f_{22}''(x_1, x_2) &= -2 \end{aligned}$$

The Hessian matrix: An example

Solution (Continued)

The Hessian matrix is therefore given by

$$f''(\mathbf{x}) = \begin{pmatrix} 2 & -1 \\ -1 & -2 \end{pmatrix}$$

The following fact is useful to notice, as it will simplify our computations in the future:

Proposition

If $f(\mathbf{x})$ is a C^2 function, then the Hessian matrix is symmetric.

The proof of this fact is quite technical, and we will skip it in the lecture.

Convexity and the Hessian

Definition

A subset $S \subseteq \mathbb{R}^n$ is open if it is without boundary. Subsets defined by inequalities with $<$ or $>$ are usually open. Also, $\mathbb{R}^n$ is open by definition.

The set $S = \{(x_1, x_2) : x_1 > 0, x_2 > 0\} \subseteq \mathbb{R}^2$ is an open set. Its boundary consists of the positive x_1- and x_2-axis, and these are not part of S.

Theorem

Let $f(\mathbf{x})$ be a C^2 function in n variables defined on an open convex set S. Then we have:

1. $f''(\mathbf{x})$ *is positive semidefinite for all* $\mathbf{x} \in S \Leftrightarrow f$ *is convex in* S
2. $f''(\mathbf{x})$ *is negative semidefinite for all* $\mathbf{x} \in S \Leftrightarrow f$ *is concave in* S
3. $f''(\mathbf{x})$ *is positive definite for all* $\mathbf{x} \in S \Rightarrow f$ *is strictly convex in* S
4. $f''(\mathbf{x})$ *is negative definite for all* $\mathbf{x} \in S \Rightarrow f$ *is strictly concave in* S

Convexity: An example

Example

Is $f(x_1, x_2) = x_1^2 - x_2^2 - x_1x_2$ *convex or concave?*

Solution

We computed the Hessian of this function earlier. It is given by

$$f''(\mathbf{x}) = \begin{pmatrix} 2 & -1 \\ -1 & -2 \end{pmatrix}$$

Since the leading principal minors are $D_1 = 2$ *and* $D_2 = -5$*, the Hessian is neither positive semidefinite or negative semidefinite. This means that* f *is neither convex nor concave.*

Notice that since f is a quadratic form, we could also have used the symmetric matrix of the quadratic form to conclude this.

Convexity: Another example

Example

Is $f(x_1, x_2) = 2x_1 - x_2 - x_1^2 + x_1 x_2 - x_2^2$ *convex or concave?*

Solution

We compute the Hessian of f*. We have that* $f_1' = 2 - 2x_1 + x_2$ *and* $f_2' = -1 + x_1 - 2x_2$*. This gives Hessian*

$$f''(\mathbf{x}) = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$$

Since the leading principal minors are $D_1 = -2$ *and* $D_2 = 3$*, the Hessian is negative definite. This means that* f *is strictly concave.*

Notice that since f is a sum of a quadratic and a linear form, we could have used earlier results about quadratic forms and properties of convex functions to conclude this.

Convexity: An example of degree three

Example

Is $f(x_1, x_2, x_3) = x_1 + x_2^2 + x_3^3 - x_1x_2 - 3x_3$ *convex or concave?*

Solution

We compute the Hessian of f*. We have that* $f_1' = 1 - x_2$, $f_2' = 2x_2 - x_1$ *and* $f_3' = 3x_3^2 - 3$*. This gives Hessian matrix*

$$f''(\mathbf{x}) = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 6x_3 \end{pmatrix}$$

Since the leading principal minors are $D_1 = 0$, $D_2 = -1$ *and* $D_3 = -6x_3$, *the Hessian is indefinite for all* $\mathbf{x}$*. This means that* f *is neither convex nor concave.*

Extremal points

Definition

Let $f(\mathbf{x})$ be a function in n variables defined on a set $S \subseteq \mathbb{R}^n$. A point $\mathbf{x}^* \in S$ is called a global maximum point for f if

$$f(\mathbf{x}^*) \geq f(\mathbf{x}) \text{ for all } \mathbf{x} \in S$$

and a global minimum point for f if

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \text{ for all } \mathbf{x} \in S$$

If $\mathbf{x}^*$ is a maximum or minimum point for f, then we call $f(\mathbf{x}^*)$ the maximum or minimum value.

How do we find global maxima and global minima points, or global extremal points, for a given function $f(\mathbf{x})$?

First order conditions

Proposition

If $\mathbf{x}^*$ is a global maximum or minimum point of f that is not on the boundary of S, then $\mathbf{x}^*$ is a stationary point.

This means that the stationary points are candidates for being global extremal points. On the other hand, we have the following important result:

Theorem

Suppose that $f(\mathbf{x})$ is a function of n variables defined on a convex set $S \subseteq \mathbb{R}^n$. Then we have:

1. *If f is convex, then all stationary points are global minimum points.*
2. *If f is concave, then all stationary points are global maximum points.*

Global extremal points: An example

Example

Show that the function

$$f(x_1, x_2) = 2x_1 - x_2 - x_1^2 + x_1x_2 - x_2^2$$

has a global maximum.

Solution

Earlier, we found out that this function is (strictly) concave, so all stationary points are global maxima. We must find the stationary points. We computed the first order partial derivatives of f earlier, and use them to write down the first order conditions:

$$f_1' = 2 - 2x_1 + x_2 = 0, \quad f_2' = -1 + x_1 - 2x_2 = 0$$

Global extremal points: An example

Solution (Continued)

This leads to the linear system

$$\begin{array}{rcrcl} 2x_1 & - & x_2 & = & 2 \\ x_1 & - & 2x_2 & = & 1 \end{array}$$

We solve this linear system, and find that $(x_1, x_2) = (1, 0)$ *is the unique stationary point of* f. *This means that* $(x_1, x_2) = (1, 0)$ *is a global maximum point for* f.

Example

Find all extremal points of the function f *given by*

$$f(x, y, z) = x^2 + 2y^2 + 3z^2 + 2xy + 2xz + 3$$

Global extremal points: Another example

Solution

We first find the partial derivatives of f*:*

$$f'_x = 2x + 2y + 2z, \quad f'_y = 4y + 2x, \quad f'_z = 6z + 2x$$

This gives Hessian matrix

$$f''(\mathbf{x}) = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 4 & 0 \\ 2 & 0 & 6 \end{pmatrix}$$

We see that the leading principal minors are $D_1 = 2$, $D_2 = 4$ *and* $D_3 = 8$*, so* f *is a (strictly) convex function and all stationary points are global minima. We therefore compute the stationary points by solving the first order conditions* $f'_x = f'_y = f'_z = 0$.

Global extremal points: More examples

Solution (Continued)

The first order conditions give a linear system

$$\begin{array}{ccccccc} 2x & + & 2y & + & 2z & = & 0 \\ 2x & + & 4y & & & = & 0 \\ 2x & & & + & 6z & = & 0 \end{array}$$

We must solve this linear system. Since the 3×3 *determinant of the coefficient matrix is* $8 \neq 0$, *the system has only one solution* $\mathbf{x} = (0, 0, 0)$. *This means that* $(x_1, x_2, x_3) = (0, 0, 0)$ *is a global minimum point for* f.

Example

Find all extremal points of the function

$$f(x_1, x_2, x_3) = x_1^4 + x_2^4 + x_3^4 + x_1^2 + x_2^2 + x_3^2$$

Global extremal points: More examples

Solution

The function f has one global minimum $(0, 0, 0)$. See the lecture notes for details.

So far, all examples where functions defined on all of $\mathbb{R}^n$. Let us look at one example where the function is defined on a smaller subset:

Example

Show that $S = \{(x_1, x_2) : x_1 > 0, x_2 > 0\}$ is a convex set and that the function $f(x_1, x_2) = -x_1^3 - x_2^2$ defined on S is a concave function.

Solution

See the lecture notes for details.