Mathematics for Economists: Lecture 5

Juuso Välimäki

Aalto University School of Business

Spring 2021

Content of Lecture 5

- In Lecture 4, Implicit function theorem
- This Lecture:
  1. Minima and maxima of functions
  2. Necessary and sufficient conditions
  3. Taylor's theorem and quadratic approximations
  4. Quadratic forms

Minima and Maxima

- Global minima and maxima
  - We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ has a global maximum at point $\hat{\boldsymbol{x}}$ if for all $\boldsymbol{x} \in \mathbb{R}^n$,
    $$f(\hat{\boldsymbol{x}}) \geq f(\boldsymbol{x}).$$
  - Function f has a global minumum at $\hat{\boldsymbol{x}}$ if for all $\boldsymbol{x} \in \mathbb{R}^n$,
    $$f(\hat{\boldsymbol{x}}) \leq f(\boldsymbol{x}).$$
  - Minimum and maximum points are called extrema or optimum points.
- Local extrema
  - We define $B^{\varepsilon}(\hat{\boldsymbol{x}}) := \{\boldsymbol{x} \in \mathbb{R}^n | \quad \|\boldsymbol{x} - \hat{\boldsymbol{x}}\| < \varepsilon\}$.
  - The function f has a local maximum (minimum) at $\hat{\boldsymbol{x}}$ if there exists an $\varepsilon > 0$ such that for all $\boldsymbol{x} \in B^{\varepsilon}(\hat{\boldsymbol{x}})$, we have:
    $$f(\hat{\boldsymbol{x}}) \geq (\leq) f(\boldsymbol{x}).$$

Overall goal for this lecture

1. How do we know whether f has a maximum or a minimum at $\hat{\boldsymbol{x}}$?
2. How to find local minima and maxima?
3. When are local extrema also global extrema?

Necessary and sufficient conditions

- Suppose that we have the following statement $'A \Rightarrow B'$
- We say that B is necessary for A.
- We say that A is sufficient for B.
- If $'A \iff B'$, we say that B is necessary and sufficient for A.
- In this last case, we also say that A is true if and only if B is true, or that A and B are equivalent statements.

Necessary and sufficient conditions: Examples

1. Consider the statements $A =$'$f(x)$ is a polynomial' and $B = 'f(x)$ is differentiable'.
2. Then A is sufficient for B and B is necessary for A.
3. We know that if A is true, then B is true.
4. We also know that if B is not true, then A is not true.
5. '$\boldsymbol{A}$ has a nonzero determinant' is necessary and sufficient for $\boldsymbol{A}$ is invertible.

First-order necessary conditions for extrema

- Consider the partial derivatives of f at $\hat{\boldsymbol{x}}$:

$$\frac{\partial f(\hat{\boldsymbol{x}})}{\partial x_i} = \lim_{h \to 0} \frac{f(\hat{x}_1, ..., \hat{x}_{i-1}, \hat{x}_i + h, \hat{x}_{i+1}, ..., \hat{x}_n) - f(\hat{\boldsymbol{x}})}{h}.$$

- If $\frac{\partial f(\hat{\boldsymbol{x}})}{\partial x_i} > 0$, then for $|h|$ small, then

$$f(\hat{x}_1, ..., \hat{x}_{i-1}, \hat{x}_i + h, \hat{x}_{i+1}, ..., \hat{x}_n) > f(\hat{\boldsymbol{x}}) \text{ for } h > 0,$$

and

$$f(\hat{x}_1, ..., \hat{x}_{i-1}, \hat{x}_i + h, \hat{x}_{i+1}, ..., \hat{x}_n) < f(\hat{\boldsymbol{x}}) \text{ for } h < 0.$$

First-order necessary conditions for extrema

- Similarly, if $\frac{\partial f(\hat{\boldsymbol{x}})}{\partial x_i} < 0$, then for small $|h|$:

$$f(\hat{x}_1, ..., \hat{x}_{i-1}, \hat{x}_i + h, \hat{x}_{i+1}, ..., \hat{x}_n) < f(\hat{\boldsymbol{x}}) \text{ for } h > 0,$$

  and

$$f(\hat{x}_1, ..., \hat{x}_{i-1}, \hat{x}_i + h, \hat{x}_{i+1}, ..., \hat{x}_n) > f(\hat{\boldsymbol{x}}) \text{ for } h < 0.$$

- We conclude that to have any kind of an extremum at $\hat{\boldsymbol{x}}$, we must have for all i:

$$\frac{\partial f(\hat{\boldsymbol{x}})}{\partial x_i} = 0.$$

- We say that the *first-order necessary condition* for an extremum at $\hat{\boldsymbol{x}}$ is that all partial derivatives are zero at $\hat{\boldsymbol{x}}$. This can be written with the gradient of f as:

$$\nabla f(\hat{\boldsymbol{x}}) = 0.$$

Critical points of functions

- We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ has a critical point at $\hat{\boldsymbol{x}}$ if $\frac{\partial f(\hat{\boldsymbol{x}})}{\partial x_i} = 0$ for all $i \in \{1, ..., n\}$.
- Example: $f(x, y, z) = x^3 y + y^2 z - z^3 x - 2x - 3y + 2z$.

$$\frac{\partial f(x,y,z)}{\partial x} = 3x^2 y - z^3 - 2,$$

$$\frac{\partial f(x,y,z)}{\partial y} = x^3 + 2yz - 3,$$

$$\frac{\partial f(x,y,z)}{\partial z} = y^2 - 3z^2 x + 2.$$

- This function has a critical point at $(\hat{x}, \hat{y}, \hat{z}) = (1, 1, 1)$.
- Can you see if it is a minimum or a maximum? (I can't)
- Some critical points are neither minima nor maxima (think about $f(x) = x^3$ at $\hat{x} = 0$).

Higher order derivatives: Functions of a real variable

- Consider now the derivative $f'(x)$ as a function of $x \in \mathbb{R}$. If f' is has a derivative at $\hat{x}$, we can form the difference quotient as before:

$$\lim_{h \to 0} \frac{f'(\hat{x} + h) - f'(\hat{x})}{h}.$$

- If this limit exists, we call this derivative of the derivative the second derivative of f at $\hat{x}$. We denote the second derivative $f''(\hat{x})$.
- For any k, define the k^{th} derivative at $\hat{x}$ as the derivative of the $(k-1)^{st}$ derivative.
- We denote this by $f^{(k)}(\hat{x})$. We say that f is k times continuously differentiable if $f^{(k)}(x)$ is a continuous function on the domain of f. We write $f \in C^k(\mathbb{R})$.

Higher order derivatives: Functions of a real variable

- If $f(x) = \ln x^2 - 1$ for $x > 1$, then we can compute the first two derivatives:

$$f'(x) = \frac{2x}{x^2 - 1}, f''(x) = \frac{2(x^2 - 1) - 2x2x}{(x^2 - 1)^2}.$$

- If $f(x) = e^{\frac{-1}{x^2}}$ for $x \neq 0$ and $f(x) = 0$ for $x = 0$, we have:

$$f'(x) = 2x^{-3}e^{\frac{-1}{x^2}}, f''(x) = -6x^{-4}e^{\frac{-1}{x^2}} + 4x^{-6}e^{\frac{-1}{x^2}}.$$

- As you can see, these computations get somewhat heavy very quickly.

Higher order approximations: Motivation

- Both $f(x) = x^2$ and $f(x) = -x^2$ have a critical point at $\hat{x} = 0$.
- For the first of these functions, the critical point is the global minimum since $x^2 \geq 0$ for all x and $x^2 > 0$ for $x \neq 0$.
- For the second, $\hat{x} = 0$ is the global maximum.
- To get more accurate information, we must look at the second derivatives of f. In the example above, $f''(0) = 2$ in the first case and $f''(0) = -2$ in the second.
- Taylor's theorem allows us to determine minima and maxima based on the sign of the second derivative at a critical point.

Higher order approximations: Taylor's theorem

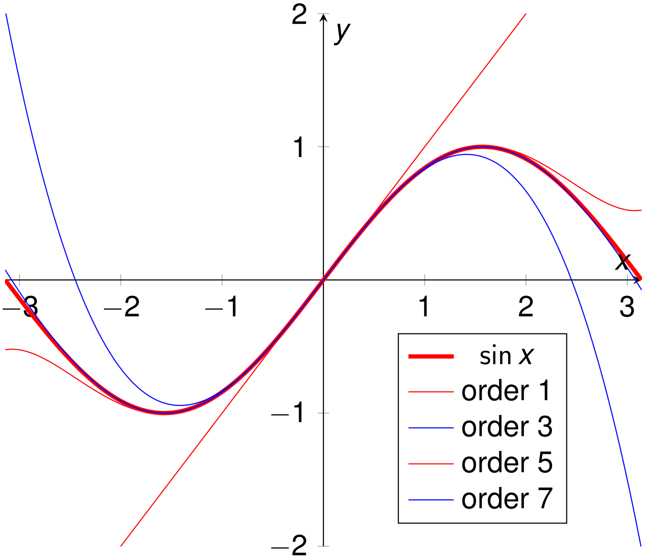
Theorem

Consider a function $f : \mathbb{R} \to \mathbb{R}$, and assume that it is $k+1$ times continuously differentiable at $\hat{x}$. Then

$$f(\hat{x}+h) = f(\hat{x}) + f'(\hat{x})\,h + \frac{1}{2}f''(\hat{x})\,h^2 + \ldots + \frac{1}{k!}f^{[k]}(\hat{x})\,h^k + \frac{1}{(k+1)!}f^{[k+1]}(x)\,h^{k+1},$$

for some x with $\hat{x} < x < \hat{x}+h$.

Taylor's theorem in practice



Taylor's theorem and the classification of critical points

- With the help of Taylor's theorem, we can classify all points with $f'(\hat{x}) = 0$:
  1. If the first l for which $f^{[l]}(\hat{x}) \neq 0$, is odd, then f does not have an extremum (i.e. minimum or maximum) at $\hat{x}$.
  2. If the first l for which $f^{[l]}(\hat{x}) \neq 0$, is even and $f^{[l]}(\hat{x}) < 0$, then f has a local maximum at $\hat{x}$.
  3. If the first l for which $f^{[l]}(\hat{x}) \neq 0$, is even and $f^{[l]}(\hat{x}) > 0$, then f has a local minimum at $\hat{x}$.
- To see why this is true, define l as above and divide the right-hand side of Taylor's theorem by h^{l-1} and let $h \to 0$.
- The requirement $f'(\hat{x}) = 0$ and $f''(\hat{x}) < 0$ is called the *second-order sufficient condition* for local maximum at $\hat{x}$.
- One more point should be kept in mind. The function f may have several local maxima and not all of them are maxima. We will have more to say about global extrema when we discuss convex and concave functions.

Second-order conditions and comparative statics

- Consider finding the $y \in \mathbb{R}$ that maximizes

$$f(y;x),$$

where $x \in \mathbb{R}$ is an exogenous variable.

- Write the problem of maximizing with respect to y as follows:

$$\max_y f(y;x)$$

- The first order necessary condition for optimum at $(\hat{y}, \hat{x})$ is:

$$\frac{\partial f}{\partial y}(\widehat{y};\widehat{x}) = 0.$$

- A sufficient condition for local maximum is obtained from Taylor's theorem:

$$f(\widehat{y}+dy;\widehat{x}) - f(\widehat{y};\widehat{x}) = \frac{\partial f}{\partial y}(\widehat{y};\widehat{x})\,dy + \frac{1}{2}\frac{\partial^2 f}{\partial y \partial y}(\widehat{y};\widehat{x})(dy)^2 + \text{ h.o.t.}$$

Definiteness and comparative statics

- If
$$\frac{\partial^2 f}{\partial y \partial y}\left(\widehat{y};\widehat{x}\right) < 0,$$
then f has a local maximum at $\left(\widehat{y};\widehat{x}\right)$.
- Note that then also the function
$$\frac{\partial f}{\partial y}\left(\widehat{y};\widehat{x}\right)$$
has a non-zero derivative w.r.t. the endogenous variable at $\left(\widehat{y};\widehat{x}\right)$ and we can apply the implicit function theorem y to get the optimal y as a function of x.

Definiteness and comparative statics

- Since
$$\frac{\partial f}{\partial y}\left(y\left(x\right);x\right)=0.$$
for all x near $\widehat{x}$, we get:
$$\frac{\partial^2 f\left(\widehat{y};\widehat{x}\right)}{\partial y \partial y}dy+\frac{\partial^2 f\left(\widehat{y};\widehat{x}\right)}{\partial y \partial x}dx=0,$$
or
$$\frac{dy}{dx}=-\frac{\frac{\partial^2 f\left(\widehat{y};\widehat{x}\right)}{\partial y \partial x}}{\frac{\partial^2 f\left(\widehat{y};\widehat{x}\right)}{\partial y \partial y}}.$$
- Since $\frac{\partial^2 f\left(\widehat{y};\widehat{x}\right)}{\partial y \partial y}<0$ by second-order condition for a maximum, we see that $\frac{dy}{dx}$ has the same sign as $\frac{\partial^2 f\left(\widehat{y};\widehat{x}\right)}{\partial y \partial x}$.

Approximating multivariate functions

- Gradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $\hat{\boldsymbol{x}}$ is the column vector of its partial derivatives $\frac{\partial f(\hat{\boldsymbol{x}})}{\partial x_i}$.
- If these partial derivatives are differentiable, we can evaluate all the partial derivatives of the partial derivatives at $\hat{\boldsymbol{x}}$.
- We define the second derivative of f to be the derivative of its gradient. Hence the second derivative at point $\hat{\boldsymbol{x}}$ is given by the matrix $Hf(\hat{\boldsymbol{x}})$ called the Hessian matrix of f:

$$Hf(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_n} \end{pmatrix}.$$

- Young's theorem: If f is twice continuously differentiable, then $\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i}$ for all $\boldsymbol{x}$ and all i, j. In words, the Hessian matrix is symmetric.

Computing the Hessian: An Example

- Consider the function

$$f(x_1, x_2, x_3) = x_1^2 - x_2^3 + x_1 x_3$$

around $(x_1, x_2, x_3) = (0, 0, 0)$. The gradient is

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 + x_3 \\ -3x_2^2 \\ x_1 \end{pmatrix}$$

Compute

$$\nabla f(0, 0, 0) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Computing the Hessian: An Example

- The Hessian matrix is given by:

$$Hf(x_1, x_2, x_3) = \begin{pmatrix} 2 & 0 & 1 \\ 0 & -6x_2 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Evaluate at $(0, 0, 0)$:

$$Hf(0, 0, 0) = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Multivariate Taylor's theorem for second order approximations

Taylor's theorem is also valid for functions $f : \mathbb{R}^n \to \mathbb{R}$. Most useful for us is the second order approximation:

Theorem

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and assume that it is 3 times continuously differentiable at $\hat{x}$. Then

$$f(\boldsymbol{x}) = f(\hat{\boldsymbol{x}}) + \nabla f(\hat{x}) \cdot (\boldsymbol{x} - \hat{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{x}}) \cdot Hf(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}) + R(\boldsymbol{x}),$$

where $\lim_{\boldsymbol{x} \to \hat{\boldsymbol{x}}} \frac{R(\boldsymbol{x})}{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2} = 0.$

Classifying critical points with Taylor's approximation

- Recall that $\nabla f(\hat{\boldsymbol{x}}) = 0$ at any critical point $\hat{\boldsymbol{x}}$, and therefore we can determine if $f(\boldsymbol{x}) \leq f(\hat{\boldsymbol{x}})$ by examining the sign of the term:

$$(\boldsymbol{x} - \hat{\boldsymbol{x}}) \cdot Hf(\hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}).$$

- Hence we have identified as the key question the determination of the sign of $\boldsymbol{x} \cdot Hf(\hat{\boldsymbol{x}})\boldsymbol{x}$ for a symmetric matrix $Hf(\hat{\boldsymbol{x}})$ for all possible $\boldsymbol{x}$.
- If it is strictly positive for all $\boldsymbol{x}$, we have a local minimum at $\hat{\boldsymbol{x}}$ (sufficient condition).
- If it is strictly negative, we have a local maximum at $\hat{\boldsymbol{x}}$.
- Conversely, if $\hat{\boldsymbol{x}}$ is a minimum (maximum), then $\boldsymbol{x} \cdot Hf(\hat{\boldsymbol{x}})\boldsymbol{x} \geq (\leq) 0$ for all $\boldsymbol{x}$ (necessary condition).

Quadratic forms and classifying extrema of $f : \mathbb{R}^n \to \mathbb{R}$

- A quadratic form is a second-degree polynomial whose terms are all of second order. They can be written as:

  $$\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x},$$

  for some symmetric matrix $\boldsymbol{A}$.
- A quadratic form is *positive definite* if for all $\boldsymbol{x} \neq 0$, $\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x} > 0$. It is *positive semidefinite* if for all $\boldsymbol{x}$, $\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x} \geq 0$.
- A quadratic form is *negative definite* if for all $\boldsymbol{x} \neq 0$, $\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x} < 0$. It is *negative semidefinite* if for all $\boldsymbol{x}$, $\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x} \leq 0$. In all other cases, we say that the quadratic form is indefinite.
- It may be helpful to write out the matrix products as summations:

  $$\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

Classifying quadratic forms

- A first observation is that $\boldsymbol{e}^i \cdot \boldsymbol{A}\boldsymbol{e}^i = a_{ii}$. Therefore a quadratic form is indefinite if it has diagonal elements with different signs.
- Another easy case is when $\boldsymbol{A}$ is a 2×2 matrix:

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

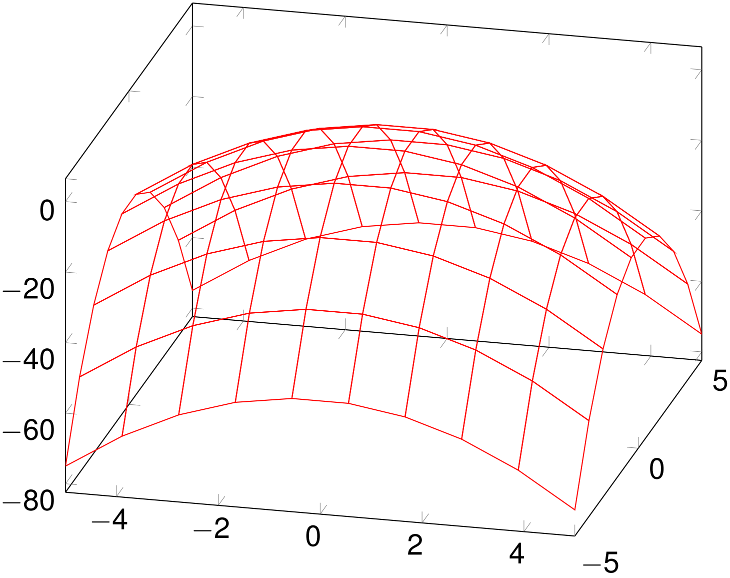
so that the quadratic form is:

$$ax_1^2 + 2bx_1x_2 + cx_2^2.$$

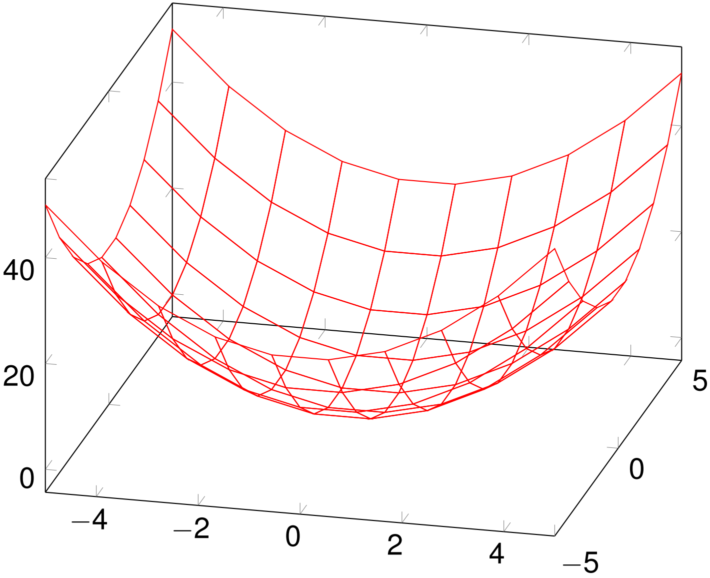
Classifying quadratic forms

- View this as a second degree function in x_2. If $c > 0$, this function has a minimum at
$$x_2 = -\frac{bx_1}{c}.$$
- Substituting into the quadratic form:
$$ax_1^2 - 2\frac{b^2x_1^2}{c} + \frac{b^2x_1^2}{c} = \left(a - \frac{b^2}{c}\right)x_1^2.$$
- This is strictly positive if
$$\begin{aligned}\left(a - \frac{b^2}{c}\right) &> 0 \text{ or} \\ ac &> b^2.\end{aligned}$$
- In other words, the quadratic form is positive definite if i) $a, c > 0$ ja ii) $\det \boldsymbol{A} > 0$.

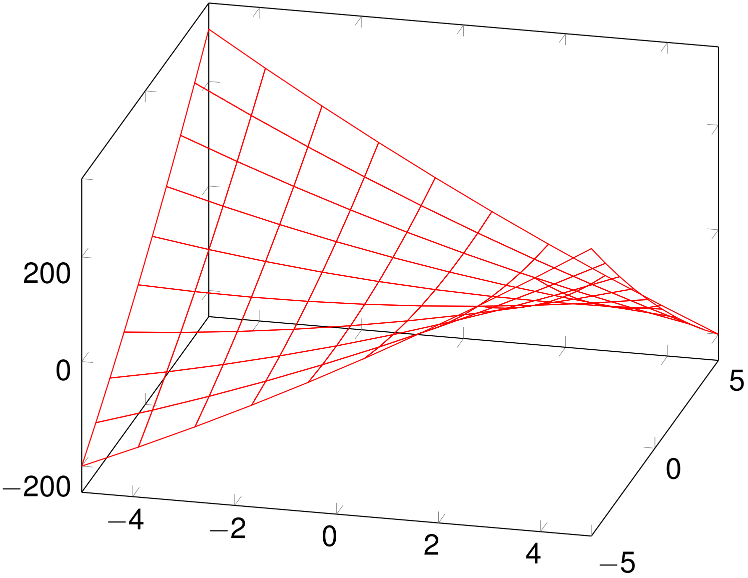
$f(x_1, x_2) = -x_2^2 - 2x_1^2$: maximum at $(0, 0)$



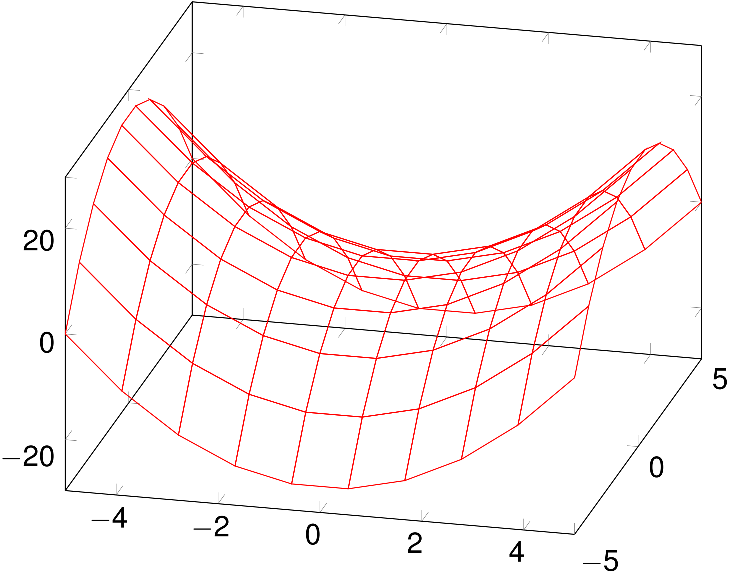
$f(x_1, x_2) = x_2^2 + x_1^2$: minimum at $(0, 0)$



$f(x_1, x_2) = x_2^2 + x_1^2 - 10x_1x_2$: saddle at $(0, 0)$



$f(x_1, x_2) = x_2^2 - x_1^2$: saddle at $(0, 0)$



Classifying quadratic forms

- The general case is tedious. We need to consider the leading principal minors $M(k)$ of $\boldsymbol{A}$:

$$\begin{aligned} M_1 &= \det a_{11}, M_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \\ M_3 &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}, \ldots \end{aligned}$$

Classifying quadratic forms

- A quadratic form
$$\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x}$$
is positive definite if $M_i > 0$ for all i. It is negative definite if $M_i (-1)^i > 0$ for all i, i.e. M_i is negative for odd i and positive for even i.

Classifying quadratic forms

- To analyze semidefiniteness of $\boldsymbol{A}$, more is needed. Define for all $1 \leq i_1 < i_2 < ... < i_n \leq n$

$$\boldsymbol{A}^n_{\{i_1,i_2,\dots,i_n\}} = \begin{pmatrix} a_{i_1 i_1} & a_{i_1 i_2} \cdots & a_{i_1 i_n} \\ \vdots & & \vdots \\ a_{i_n i_1} & a_{i_n i_2} \dots & a_{i_n i_n} \end{pmatrix}.$$

and

$$M^n_{\{i_1,i_2,\dots,i_n\}} = \det\left(\boldsymbol{A}^n_{\{i_1,i_2,\dots,i_n\}}\right).$$

Classifying quadratic forms

- The matrix $\boldsymbol{A}$ is positive semidefinite if

  $$M^n_{\{i_1,i_2,\ldots,i_n\}} \geq 0 \text{ for all } n \text{ and for all } \{i_1, i_2, \ldots, i_n\}.$$

  It is negative semidefinite if

  $$M^n_{\{i_1,i_2,\ldots,i_n\}} \leq 0 \text{ for all odd } n \text{ and for all } \{i_1, i_2, \ldots, i_n\},$$

  $$M^n_{\{i_1,i_2,\ldots,i_n\}} \geq 0 \text{ for all even } n \text{ and for all } \{i_1, i_2, \ldots, i_n\}.$$

An example

- Consider the definiteness of

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix}.$$

1. $M^1 = \det(a_{11}) = 2.$
2. $M^2 = \det\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3.$
3. $M^3 = \det\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & 1 & 1 \end{pmatrix} = (-1)^{3+3}\det\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} + (-1)^{3+2}\det\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} +$
   $(-1)^{3+1}\det\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} = 3 + 1 - 1 = 3.$

Therefore $\boldsymbol{A}$ is positive definite.

Next Lecture

- Economic applications
- Convex and concave functions
- Quasiconcave functions
- Economic applications of concavity and convexity