Mathematics for Economists
Aalto BIZ
Spring 2021
Juuso Välimäki

Minima and maxima of functions

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 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.

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 at $\hat{\boldsymbol{x}}$ if there exists an $\varepsilon > 0$ such that for all $\boldsymbol{x} \in B^\varepsilon(\hat{\boldsymbol{x}})$, we have:

$$f(\hat{\boldsymbol{x}}) \geq f(\boldsymbol{x})$$

.

A local minimum is defined analogously.

The main questions to be addressed in these notes are:

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?

First-order necessary conditions for local 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

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

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:

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

We call points where $\nabla f(\hat{\boldsymbol{x}}) = 0$ the *critical points* of f. The fact that $\hat{\boldsymbol{x}}$ is a critical point does not imply that $\hat{\boldsymbol{x}}$ is a maximum or a minimum. In other words, $\nabla f(\hat{\boldsymbol{x}}) = 0$ it is not a *sufficient condition* for an extremum. Just consider the function $f(x) = x^3$ at $\hat{x} = 0$. In order to classify the critical points, we must find better approximations to f at $\hat{\boldsymbol{x}}$.

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 $f \in C^k(\mathbb{R})$.

Taylor's theorem

Higher order derivatives are useful when one tries to find more accurate approximations to functions that are k times differentiable. We have already seen that differentiable functions are well approximated around $\hat{x}$ by $f(\hat{x}) + f'(\hat{x})(x - \hat{x})$. Linear approximations are good enough to identify critical points, but they are of no use for deciding whether the critical points are minima or maxima.

For example, 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 minimum.

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. The following theorem allows us to determine minima and maxima based on the sign of the second derivative at a critical point.

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

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

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

An illustration of the approximations of different orders is given in Figure 1.

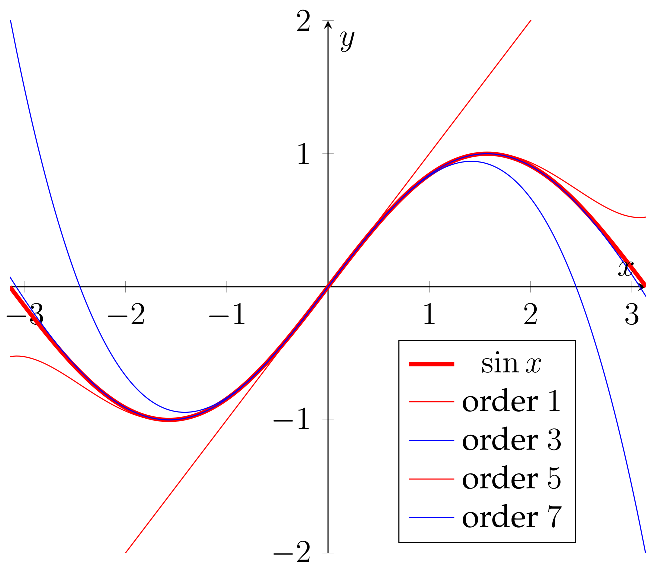

Figure 1: Approximating $f(x) = \sin(x)$.

For local analysis around $\hat{x}$, i.e. for h arbitrarily small, we need to look for the first term with a non-zero coefficient in the Taylor approximation. The other terms vanish much more quickly when $h \to 0$ (since they involve the multiplier h^k for $k > 1$. For twice (or more times) continuously differentiable functions, Taylor's theorem gives a precise reason why we called the remainder term as higher-order terms in the first-order approximation by derivatives.

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.

Higher order derivatives of multivariate functions

The gradient of a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$ 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 $Hf(\hat{\boldsymbol{x}})$:

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

Young's theorem guarantees that the Hessian matrix is symmetric:

Theorem 2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Then for all $i, j \in \{1, ..., n\}$ and all x, we have

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

Multivariate Taylor approximation

One can also define k^{th} order derivatives for multivariate functions, but there is little use for higher orders than the second order derivative defined above. 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 3. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and assume that it is 3 times continuously differentiable at $\hat{x}$. Then

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

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

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 $\boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x}$ for a symmetric matrix $\boldsymbol{A}$.

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.

Main take-away for this section:

Taylor's theorem for multivariate functions tells us that a critical point at $\hat{\boldsymbol{x}}$ is a local maximim (minimum) if its Hessian matrix at $\hat{\boldsymbol{x}}$ is negative (positive) definite. This is a sufficient condition for maximum (minimum). Conversely if f has a local maximum (minumum) at $\hat{\boldsymbol{x}}$, then $Hf(\hat{\boldsymbol{x}})$ is negative (positive) semi-definite.

Classifying quadratic forms

The following few subsection are long and at times cumbersome. Do not mistake the length to be a sign that it is of overwhelming importance. I discuss definiteness in some detail in the notes as it is not covered so much in the lectures.

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:

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

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) $\det \boldsymbol{A} > 0$.

For negative definiteness, assume that $a, c < 0$. Solving for the maximal x_2 for eachx_1 gives:

$$x_2 = -\frac{bx_1}{c}$$

and substituting into the quadratic form and require that:

$$ax_1^2 - 2\frac{b^2x_1^2}{c} + \frac{b^2x_1^2}{c} = \left(a - \frac{b^2}{c}\right)x_1^2 < 0.$$

We get:

$$a < \frac{b^2}{c} \text{ or } ac > b^2.$$

In other words,

$$\det \boldsymbol{A} > 0.$$

Unfortunately, the general case is tedious. I give it here for completeness, but it is not particularly illuminating. We need to consider the leading principal minors $M(k)$ of $\boldsymbol{A}$:

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

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 $M_i \left(-1\right)^i > 0$ for all i, i.e. M_i is negative for odd i and positive for even i.

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

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

and

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

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

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

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

At the end of Part II of these lectures, we will discuss the eigenvalues of a matrix. It turns out that for symmetric matrices, $\boldsymbol{A}$, there is a simple connection between definiteness and the sign of the eigenvalues. First of all, all eigenvalues of a symmetric matrix are real. If they are all positive (negative), then $\boldsymbol{A}$ is positive (negative) semidefinite. If they are all strictly positive (strictly negative), then it is positive (negative) semidefinite. $\boldsymbol{A}$ is indefinite only if it has a strictly positive and a strictly negative eigenvalue.

Definiteness with linear constraints

The definiteness of the quadratic form

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

can also be considered under linear constraints. In other words, we require that

$$\boldsymbol{b} \cdot \boldsymbol{x} = 0.$$

The restriction $\boldsymbol{b} \cdot \boldsymbol{x} = 0$ restricts the set of vectors that we consider to the plane normal to $\boldsymbol{b}$. We can ask whether $\boldsymbol{A}$ is definite for vectors in this plane.

Consider the matrix

$$\boldsymbol{B} = \begin{pmatrix} 0 & b_1 & \cdots & b_n \\ b_1 & a_{11} & & a_{1n} \\ \vdots & & & \\ b_n & a_{n1} & & a_{nn} \end{pmatrix},$$

and assume that $b_1 \neq 0$.

The matrix $\boldsymbol{A}$ consisting of elements a_{ij} is positive definite in directions $\{\boldsymbol{x} \,|\, \boldsymbol{b} \cdot \boldsymbol{x} = 0\}$ if all the leading principal minors of $\boldsymbol{B}$ except for the first one are negative. It is negative definite in directions $\{\boldsymbol{x} \,|\, \boldsymbol{b} \cdot \boldsymbol{x} = 0\}$ if all the leading principal minors of $\boldsymbol{B}$ except for the first one alternate in sign.

Examples

1. Consider the definiteness of

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

   (a) $M^1 = \det(a_{11}) = 2.$

(b) $M^2 = \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3.$

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

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

2. Consider matrix

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

This is easily seen to be indefinite (why?).

3. Consider

$$\boldsymbol{A} = \begin{pmatrix} -1 & -4 & -1 \\ -4 & 0 & 1 \\ -1 & 1 & -1 \end{pmatrix}.$$

(a) $M_1^1 = -1, M_2^1 = 0, M_3^1 = -1.$

(b) $M_{\{1,2\}}^2 = -16$, $M_{\{1,3\}}^2 = 0, M_{\{2,3\}}^2 = -1.$

We see already that $\boldsymbol{A}$ is indefinite.

4. 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

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

The Hessian matrix is given by:

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

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

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

This matrix is indefinite since $M_1^1 = 2 > 0$ and $M_{\{1,3\}}^2 = \det \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} = -1.$