Lecture 4: Calculus and Linear Algebra

Piotr Zwiernik

Mathematics Brush-up

Barcelona School of Economics

In many economic applications, we have to deal with situations where several variables have to be included in the mathematical model.

Example: The Cobb-Douglas production function applied to an agricultural production gives the number of units produced depending on the capital invested, the labour and the area of land used for the production.

Read Chapter 11 of Werner-Sotskov and Chapters 14 and 17 of Simon-Blume

Exercises: 11.11 a), 11.21, 11.22 (Werner-Sotskov)

Definition

A function of several variables is a map $f : D \to \mathbb{R}$ where $D \subset \mathbb{R}^n$ is the domain, and for any point $(x_1, \ldots, x_n) \in D$ it assigns a number $f(x_1, \ldots, x_n)$.

We can only plot $n = 2$!

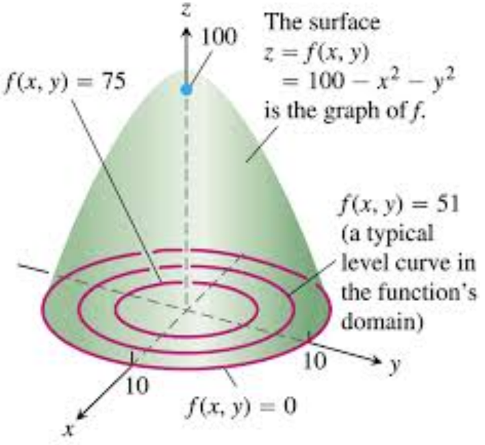
If $n = 2$, the graph of f with $z = f(x, y)$ is called a surface and it is plotted in $\mathbb{R}^3$.

If $n = 2$ a level curve of f is a curve in $\mathbb{R}^2$ given by

$$z = f(x, y) = C.$$

Example

$f(x, y) = 100 - x^2 - y^2$ the surface is a smooth cone and the level curves are circles centered at zero



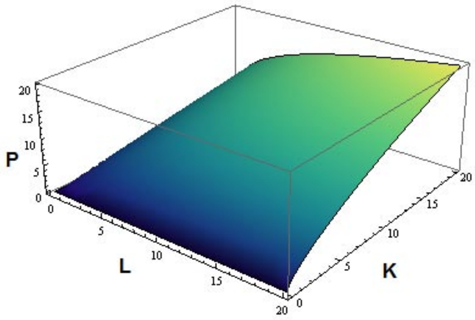
Economic example

Cobb-Douglas production function

$$P(L,K) = bL^{\alpha}K^{\beta}$$

P is the total production, L is the labor, K the capital, b the total factor productivity, and α and β are the output elasticities of labor and capital, respectively. The parameters b, α and β are constants computed by available technology.

The domain is $D = \{(L,K) \in \mathbb{R}^2 : L \geq 0, K \geq 0\}$.



Cobb-Douglas production function

The parameters α, and β measure the output change in levels of labor or capital change. For example, if $\alpha = 0.15$, a 1% increase in labor would lead to approximately a 0.15% increase in output.

If $\alpha + \beta = 1$, the production function has constant returns to scale (the output increases by the same proportional change). That is, if L and K each increase by 20%, then P increases by 20%.

If $\alpha + \beta < 1$, the returns to scale are decreasing.

If $\alpha + \beta > 1$, the returns to scale are increasing.

Multivariate Gaussian density

A random vector $\mathbf{X}$ in $\mathbb{R}^d$ has a multivariate Gaussian distribution with a nonsingular covariance matrix Σ if Σ is positive definite and the density function of X is

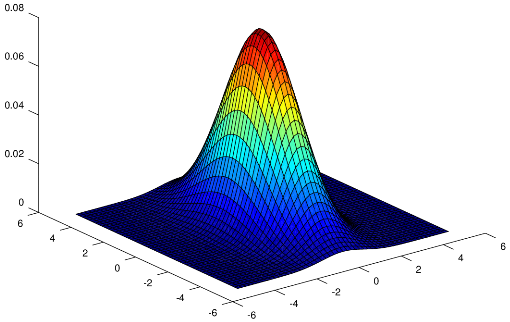
$$f(\mathbf{x}) = (2\pi)^{-d/2}\det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m})'\Sigma^{-1}(\mathbf{x}-\mathbf{m})\right)$$

where $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{m} = E(\mathbf{X})$, and

$$\Sigma = E((\mathbf{X}-\mathbf{m})(\mathbf{X}-\mathbf{m})').$$

Multivariate Gaussian density

Example: $d = 2$, $\mathbf{m} = \mathbf{0}$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$



Partial derivatives

A function $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ is partially differentiable with respect to x at (x, y) if the limit

$$\lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h} \qquad \text{exists,}$$

and is partially differentiable with respect to y at (x, y) if the limit

$$\lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h} \qquad \text{exists.}$$

Both limits are called the partial derivatives of f with respect to x and y and denoted $f_x(x, y)$, $f_y(x, y)$ or $\frac{\partial f(x,y)}{\partial x}$, $\frac{\partial f(x,y)}{\partial y}$.

Example: Total cost of a firm producing x units of product A and y units of product B is

$$C(x, y) = 200 + 22x + 16y^{3/2}.$$

Marginal costs: $C_x(x, y) = 22$, $C_y(x, y) = 24\sqrt{y}$.

Cobb-Douglas production function

The partial derivative $\frac{\partial P}{\partial L}$ is the rate at which production changes with respect to the amount of labor. It is called the marginal productivity of labor.

The partial derivative $\frac{\partial P}{\partial K}$ is the rate at which production changes with respect to the amount of capital. It is called the marginal productivity of capital.

The assumptions made by Cobb and Douglas are:

1. If either labor or capital vanish then so will the production.
2. The marginal productivity of labor is proportional to the amount of production per unit of labor, that is,
$$\frac{\partial P}{\partial L} = \alpha \frac{P}{L} \quad \text{for some } \alpha.$$
3. The marginal productivity of capital is proportional to the amount of production per unit of capital, that is,
$$\frac{\partial P}{\partial K} = \beta \frac{P}{K} \quad \text{for some } \beta.$$

(In Chapter 10 we will show that these assumptions imply that $P = bL^{\alpha}K^{\beta}$)

Tangent plane and total derivative

Geometric interpretation: $f_x(x_0, y_0)$ (resp. $f_y(x_0, y_0)$) is the slope of the tangent line $f_T(x)$ to the curve of intersection of the surface $z = f(x, y)$ and the plane $y = y_0$ (resp. $x = x_0$) at (x_0, y_0, z_0), where $z_0 = f(x_0, y_0)$.

Both tangent lines $f_T(x)$ and $f_T(y)$ span a plane $f_T(x, y)$ which is called the tangent plane to $z = f(x, y)$ at (x_0, y_0, z_0), and the equation of the tangent plane is

$$f_T(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

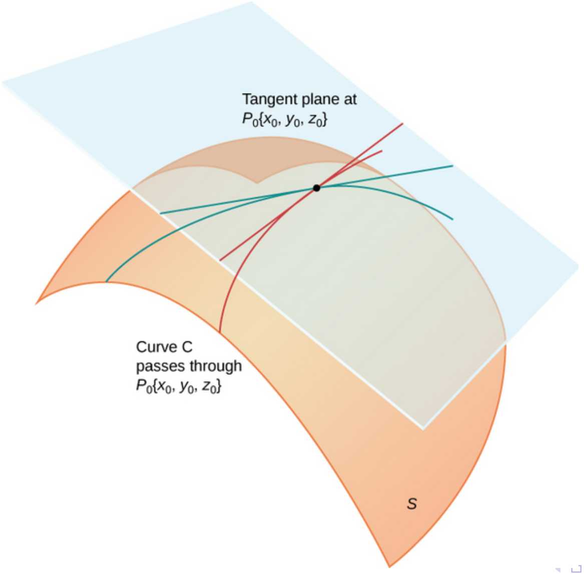
The tangent plane gives and approximation of the total change of the function from (x_0, y_0) to (x, y), when both points are close:

$$f(x, y) - f(x_0, y_0) \approx f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0),$$

which can be seen as a total derivative. In differential notation:

$$df = f_x(x_0, y_0)dx + f_y(x_0, y_0)dy.$$

Tangent plane



Higher Partial derivatives

We can consider higher-order partial derivatives of $f : D \subset \mathbb{R}^n \to \mathbb{R}$, when they exists. For example

$$f_{xy}(x, y) = \frac{\partial}{\partial y}\left(\frac{\partial f(x, y)}{\partial x}\right) = \frac{\partial^2 f(x, y)}{\partial y \partial x}.$$

Theorem (Young): If f_{xy} and f_{yx} are continuous, they coincide.

Example: $f(x, y) = \sin(3x - y)$, $f_{xy} = f_{yx} = 3\sin(3x - y)$.

The gradient

Let $\mathbf{x} = (x_1, \ldots, x_n) \in D$.The vector

$$\nabla f(\mathbf{x}) = (f_{x_1}(\mathbf{x}), \ldots, f_{x_n}(\mathbf{x})).$$

is called the gradient of f at $\mathbf{x}$, when all partial derivatives exist.

Example: $f(x, y) = 2x^2 + \frac{1}{2}y^2$, $\nabla f(x, y) = (4x, y)$, $\nabla f(1, 2) = (4, 2)$

The gradient gives the direction which the function f at $\mathbf{x}$ increases the most, since it is orthogonal to the tangent line to the level curve at this point.

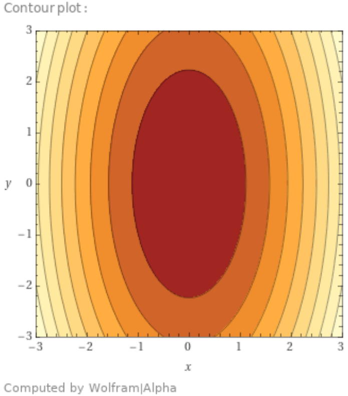
Example: production function: $f(x, y) = 3x^2y + 0.5xe^y$, $x =$labour, $y =$capital,
$\nabla f(x, y) = (6xy + 0.5e^y, 3x^2 + 0.5xe^y)$

$\nabla f(10, \log 12) = (155.09, 360)$, so to increase at maximum the production, the firm should increase labour and capital in the ratio $155.09 : 360$, that is, if we increase the labor by 1 we increase the capital by 2.32.

Example: $f(x,y) = 2x^2 + \frac{1}{2}y^2$

$\nabla f(x,y) = (4x, y)$, $\nabla f(1,2) = (4,2)$, therefore the maximum direction of increase at point $(1,2)$ is the direction $(4,2)$.

The point $(1,2)$ is at the level curve $2x^2 + \frac{1}{2}y^2 = 4$.



Differentiation rules for matrices

Consider two vectors $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_m)$ in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, where each y_i depends on $\mathbf{x}$. Then we denote by $\frac{d\mathbf{y}}{d\mathbf{x}}$ the Jacobian $m \times n$ matrix with entries $\frac{\partial y_i}{\partial x_j}$.

If $m = 1$, then the Jacobian is just the gradient.

Examples:

1. If A is a $m \times n$ matrix, then

$$\frac{d(A\mathbf{x})}{d\mathbf{x}} = A$$

   If $m = 1$ the we deduce that $\frac{d(\mathbf{a}'\mathbf{x})}{d\mathbf{x}} = \mathbf{a}$.

2. If A is a $n \times n$ matrix, then

$$\frac{d(\mathbf{x}'A\mathbf{x})}{d\mathbf{x}} = \mathbf{x}'(A + A').$$

Unconstrained optimization

A function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ has a local max (min) at $\mathbf{x}_0$ if there exists a ball $B_r(\mathbf{x_0}) \subset D$ such that

$$f(\mathbf{x}) \leq f(\mathbf{x}_0) \qquad (f(\mathbf{x}) \geq f(\mathbf{x}_0))$$

for all $\mathbf{x} \in B_r(\mathbf{x}_0)$. If the inequality holds for all $\mathbf{x} \in D$, we say it is a global max (min).

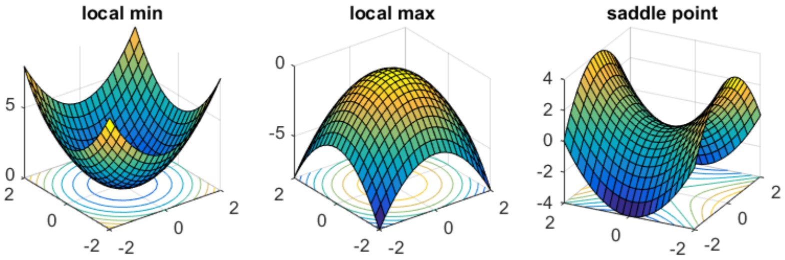
We say that $\mathbf{x}$ is an interior point if there exists a ball $B_r(\mathbf{x}) \subset D$.

Theorem (Necessary 1st order condition): If f has a local max or min at $\mathbf{x}$ and it is an interior point then $\nabla f(\mathbf{x}) = \mathbf{0}$.

Interior points $\mathbf{x}$ that $\nabla f(\mathbf{x}) = \mathbf{0}$ are called stationary points.

Stationary points that are neither a local max or min are called saddle points.

Unconstrained optimization



Local optimality conditions

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable ($f \in C^2$).
The Hessian matrix of f at point $\mathbf{x}$ is defined as:

$$H_f(\mathbf{x}) = \left(f_{x_i x_j}(\mathbf{x})\right)_{1 \leq i,j \leq n}.$$

Observe that the Hessian matrix is always symmetric, as $f \in C^2$.
Theorem (Sufficient 2nd order condition): Let $\mathbf{x}$ be a stationary point of f. Then:

1. If $H_f(\mathbf{x})$ is negative (positive) definite, then $\mathbf{x}$ is a local max (min)
2. If $H_f(\mathbf{x})$ is indefinite, then $\mathbf{x}$ is a saddle point.

   Otherwise, we don't know it from this theorem.

Theorem ($n = 2$): Let $\mathbf{x}$ be a stationary point of f. Then:

1. If $|H_f(\mathbf{x})| > 0$ (that is, $f_{xx}(\mathbf{x})f_{yy}(\mathbf{x}) > f_{xy}^2(\mathbf{x})$) , then:
   - If $f_{xx}(\mathbf{x}) < 0$ or $f_{yy}(\mathbf{x}) < 0$, then $\mathbf{x}$ is a local max.
   - If $f_{xx}(\mathbf{x}) > 0$ or $f_{yy}(\mathbf{x}) > 0$, then $\mathbf{x}$ is a local min.
2. If $|H_f(\mathbf{x})| < 0$, then $\mathbf{x}$ is a saddle point.

   Otherwise, we don't know it from this theorem.

Examples

1. Let $f(x,y) = x^2 - y^2 - xy$. Since $\nabla f(x,y) = (2x - y, -2y - x)$, the stationary points are $(0,0)$. The Hessian matrix is

$$H_f(x,y) = \begin{pmatrix} 2 & -1 \\ -1 & -2 \end{pmatrix}$$

which is indefinite, so $(0,0)$ is a saddle point.

We will see that in this case the function is neither concave nor convex

2. Let $f(x,y) = x^2 + y^4$. Stationary points: $(0,0)$. The Hessian matrix at $(0,0)$ is

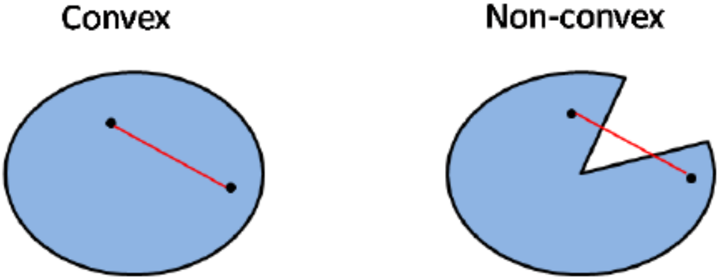
$$H_f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

which is positive semi-definite so we cannot conclude from the Hessian matrix but it is clear that $(0,0)$ is global minimum since $f \geq 0$.

3. Let $f(x,y) = x^3 + y^3$. Stationary points: $(0,0)$. Similar as 2. but it is clear that $(0,0)$ is a saddle point.

Convex domain

The domain of a function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ is said to be convex if any line segment joining two points in the domain lies completely within the domain.



Global optimality conditions

A function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ where D is convex is said to be convex (resp. concave) if the line segment between any two points on the graph of the function lies above (rep. below) or on the graph.

Theorem: $f : D \subset \mathbb{R}^n \to \mathbb{R}$ where D is convex satisfies:

1. $H_f(\mathbf{x})$ is negative (positive) semi-definite on $D \Leftrightarrow f$ is concave (convex) on D
2. $H_f(\mathbf{x})$ is negative (positive) definite on $D \Leftrightarrow f$ is strictly concave (strictly convex) on D
3. If f is convex (concave) on D, then all stationary points are global min (max).

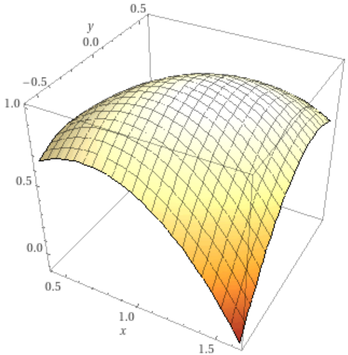
Example

Let $f(x, y) = 2x - y - x^2 + xy - y^2$.

Since $\nabla f(x, y) = (2 - 2x + y, -1 + x - 2y)$, the stationary points are $x = (0, 1)$. The Hessian matrix is

$$H_f(x, y) = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$$

which is negative definite, so f is strictly concave and $(0, 1)$ is a global maximum.



plot 2x-y-x^2+xy-y^2 | Computed by Wolfram|Alpha

Economic example

Example: A firm can sell each piece of product X/Y for 45/55 euros. The revenue is $R(x, y) = 45x + 55y$. The production cost is

$$C(x, y) = 300 + x^2 + 1.5y^2 - 25x - 35y.$$

The total profit is

$$f(x, y) = R(x, y) - C(x, y).$$

We want to know the maximum profit the firm can make. We first find the stationary points of f:

$$f_x = -2x + 70 = 0, \quad f_y = -3y + 90 = 0 \quad \implies \mathbf{x} = (35, 30).$$

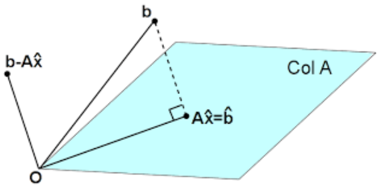
We next check the 2nd order conditions: since $H_f(\mathbf{x})$ is negative definite, $\mathbf{x}$ is a local max of f. Is it a global max? Yes, since $H_f(\mathbf{x})$ is negative definite for all $\mathbf{x}$ so strictly concave. The maximal profit is $f(35, 30) = 2275$.

Method of least squares

Consider a linear system of equations $A\mathbf{x} = \mathbf{b}$ that has no solution.

That is, $\mathbf{b}$ does not belong to the span of columns of A

The solutions $\hat{\mathbf{x}}$ that make $|\mathbf{b} - A\hat{\mathbf{x}}|$ minimal are called least square solutions.



This is equivalent to find the orthogonal projection of $\mathbf{b}$ onto the span of the columns of A

Method of least squares

The solution is $\hat{\mathbf{x}}$ is such that

$$\langle A\hat{\mathbf{x}} - \mathbf{b}, A\mathbf{x} \rangle = 0, \qquad \text{for all } \mathbf{x}.$$

This is equivalent to

$$(A\hat{\mathbf{x}} - \mathbf{b})^T A\mathbf{x} = (\hat{\mathbf{x}}'A'A - \mathbf{b}'A)\mathbf{x} = 0,$$

for all $\mathbf{x}$.

Therefore, $\hat{\mathbf{x}}$ is the solution to the linear system of equations

$$A'\mathbf{b} = A'A\hat{\mathbf{x}}.$$

This system has a unique solution if and only if the columns of A are linearly independent. In this case,

$$\hat{\mathbf{x}} = (A'A)^{-1}A'\mathbf{b}.$$

Linear fit

Assume that we have n measurements $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ of two variables x and y, respectively.

We want to find the line $y = ax + b$ that describes the relation between both variables using the least square approximation. That is,

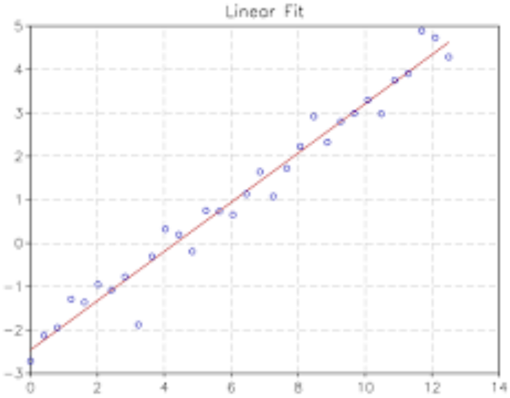
$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

In this case, the solution to the linear system $A'\mathbf{b} = A'A\hat{\mathbf{x}}$ is

$$a = \frac{n\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$b = \frac{\sum_{i=1}^n y_i - a\sum_{i=1}^n x_i}{n}$$

Linear fit



Polynomial fit

Assume that we have n measurements $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ of two variables x and y, respectively.

We want to find the polynomial of degree m, $y = c_0 + c_1 x + c_2 x^2 + \cdots + c_m x^m$ that describes the relation between both variables.

Here A is the $n \times m + 1$ matrix of a column of ones, the observations $\mathbf{x}$, $\mathbf{x}^2, \ldots, \mathbf{x}^m$. The vector $\mathbf{b}$ is the vector $\mathbf{y}$, and the least square solution is $(c_0, c_1, \ldots, c_m)$.

Again we need to solve the system of equations $A'\mathbf{b} = A'A\hat{\mathbf{x}}$.

Example: quadratic fit

We want to find the best quadratic polynomial that fits the data $\mathbf{x} = (-1, -0.5, 0, 0.5, 1)$ and $\mathbf{y} = (1, 0.5, 0, 0.5, 2)$.

Here $m = 2$, $n = 5$,

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & -0.5 & 0.25 \\ 1 & 0 & 0 \\ 1 & 0.5 & 0.25 \\ 1 & 1 & 1 \end{pmatrix}$$

and

$$A'A = \begin{pmatrix} 5 & 0 & 2.5 \\ 0 & 2.5 & 0 \\ 2.5 & 0 & 2.125 \end{pmatrix} \quad A'\mathbf{b} = \begin{pmatrix} 4 \\ 1 \\ 3.25 \end{pmatrix}$$

The solution to this system is the polynomial $0.0857 + 0.4x + 1.4286x^2$.

Example:quadratic fit

We plot the polynomial $0.0857 + 0.4x + 1.4286x^2$ and the data.