Mathematics for Economists: Lecture 6

Juuso Välimäki

Aalto University School of Business

Spring 2020

This lecture covers

Quadratic functions

- A multivariate quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ takes the form:

$$f(\boldsymbol{x}) = c + \boldsymbol{b} \cdot \boldsymbol{x} + \boldsymbol{x} \cdot \boldsymbol{A}\boldsymbol{x},$$

  where $c \in \mathbb{R}$ is the constant term, the inner product $\boldsymbol{b} \cdot \boldsymbol{x}$ is the linear term for some $\boldsymbol{b} \in \mathbb{R}^n$, and $\boldsymbol{A}$ is a non-zero symmetric matrix defining a quadratic form.
- Note that for all $n \times n$ matrices $\boldsymbol{B}$, the matrix $\frac{1}{2}(\boldsymbol{B}^\top + \boldsymbol{B})$ is a symmetric matrix, and

$$\boldsymbol{x} \cdot \boldsymbol{B}\boldsymbol{x} = \frac{1}{2}\boldsymbol{x} \cdot (\frac{1}{2}(\boldsymbol{B}^\top + \boldsymbol{B}))\boldsymbol{x}.$$

- Writing out the inner products and matrix products, we see that:

$$f(\boldsymbol{x}) = c + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

Derivatives of quadratic functions

- The partial derivative of f with respect to x_k is:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} = b_k + \sum_{i=1}^{n} a_{ik} x_i + \sum_{j=1}^{n} a_{kj} x_j.$$

- Since $\boldsymbol{A}$ is symmetric, $\sum_{i=1}^{n} a_{ik} x_i = \sum_{j=1}^{n} a_{kj} x_j$ and:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_k} = b_k + 2 \sum_{i=1}^{n} a_{ik} x_i.$$

- This means that we can write the gradient of f as

$$\nabla f(\boldsymbol{x}) = \boldsymbol{b} + 2\boldsymbol{A}\boldsymbol{x}.$$

Quadratic functions

- Therefore we can solve for the critical points (by finding the inverse matrix $\boldsymbol{A}^{-1}$, by Gaussian elimination or by Cramer's rule) from the linear system :

$$2\boldsymbol{A}\boldsymbol{x} = -\boldsymbol{b}.$$

- Because of this linearity in the first-order necessary conditions, quadratic functions are manageable.
- The Hessian matrix of f is $2\boldsymbol{A}$. Hence classifying the critical points depends on the definiteness of $\boldsymbol{A}$.
- Quadratic models in economics: mean-variance preferences in finance, interdependent markets with linear demand curves, capacity expansion with quadratic adjustment costs, incentive problems with Normally distributed noise, ordinary least squares...

Application of quadratic optimization: ordinary least squares

- Statistics Finland has register data on N individuals living in Finland.
- Let y_i denote the income of individual i.
- Let $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iK})$ be a vector of numerical covariates that characterize individual i (e.g. age, years of schooling, years in continuous employment, etc.)
- Your total data: vector $\boldsymbol{y} = (y_1, ..., y_N)$ and $N \times K$ matrix of observables $\boldsymbol{X}$ with element x_{ik} for individual $i's$ characteristic k.
- How would you find the best linear model to predict y_i if you only know $\boldsymbol{x}_i$?

Application of quadratic optimization: ordinary least squares

- If the number of individuals N is large in comparison to the number of observable characteristics, K, you will not be able to find a perfect linear fit i.e. a vector $\boldsymbol{\beta} = (\beta_1, ..., \beta_K)$ such that:

$$y_i = \sum_{k=1}^{K} \beta_k x_{ki} = \boldsymbol{x}_i \cdot \boldsymbol{\beta} \text{ for all } i.$$

- Allow an individual random term ϵ_i that accounts for the discrepancy and find the vector $\boldsymbol{\beta}$ that 'minimizes the size' of the error vector $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_N)$.
- How to measure the size? Ordinary least squares minimizes norm:

$$\boldsymbol{\epsilon} \cdot \boldsymbol{\epsilon} = \sum_{i=1}^{N} \epsilon_i^2.$$

- But why not, say $\sum_{i=1}^{N} |\epsilon_i|$?

Minimizing the sum of squared errors

- If $y_i = \sum_{k=1}^{K} \beta_k x_{ki} + \epsilon_i$, then $\epsilon_i = y_i - \sum_{k=1}^{K} \beta_k x_{ki} = y_i - \boldsymbol{x}_i \cdot \boldsymbol{\beta}$. But then:

$$\boldsymbol{\epsilon} \cdot \boldsymbol{\epsilon} = \sum_{i=1}^{N} \epsilon_i^2 = \sum_{i=1}^{N} (y_i - \boldsymbol{x}_i \boldsymbol{\beta})^2.$$

- Writing in vector form, we have:

$$\boldsymbol{\epsilon} \cdot \boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{\beta} - (\boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{y} + (\boldsymbol{\beta}\boldsymbol{X})^\top \boldsymbol{X}\boldsymbol{\beta}$$

$$= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta} \cdot \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y} \cdot \boldsymbol{y} - 2\boldsymbol{X}^\top \boldsymbol{y} \cdot \boldsymbol{\beta} + \boldsymbol{\beta} \cdot \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}.$$

Minimizing the sum of squared errors

- We see that this is a quadratic function of $\boldsymbol{\beta}$ and therefore we can use our result from above to conclude that its critical points are found at the solution to:

$$-2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = 0,$$

or the critical point $\hat{\boldsymbol{\beta}}$ satisfies:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}.$$

- This is the OLS-estimator for the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$.
- In Problem Set 3, you are asked to prove that $\boldsymbol{X}^\top \boldsymbol{X}$ is positive definite so $\hat{\boldsymbol{\beta}}$ is indeed a global minimum.

Profit maximization with CES - production function

- Consider profit maximization

$$\max_{k,l>0} f(k,l) - \frac{r}{p}k - \frac{w}{p}l,$$

with the production function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(k,l) = \left(k^\rho + l^\rho\right)^{\frac{1}{\rho}}.$$

- Form the gradient of profit:

$$\begin{pmatrix} \frac{\partial f(k,l)}{\partial k} - \frac{r}{p} \\ \frac{\partial f(k,l)}{\partial l} - \frac{w}{p} \end{pmatrix} = \begin{pmatrix} \left(k^\rho + l^\rho\right)^{\frac{1}{\rho}-1} k^{\rho-1} - \frac{r}{p} \\ \left(k^\rho + l^\rho\right)^{\frac{1}{\rho}-1} l^{\rho-1} - \frac{w}{p} \end{pmatrix}.$$

- The the Hessian matrix is the Hessian matrix of the production function:

$$Hf(k,l) = \begin{pmatrix} \frac{\partial^2 f(k,l)}{\partial k \partial k} & \frac{\partial^2 f(k,l)}{\partial k \partial l} \\ \frac{\partial^2 f(k,l)}{\partial l \partial k} & \frac{\partial^2 f(k,l)}{\partial l \partial l} \end{pmatrix}.$$

Example: CES -function

- By the product rule:

$$\frac{\partial^2 f(k,l)}{\partial k \partial k} = (\rho - 1) k^{\rho-2} (k^\rho + l^\rho)^{\frac{1}{\rho}-1} + \left(\frac{1}{\rho} - 1\right)(k^\rho + l^\rho)^{\frac{1}{\rho}-2} \rho k^{2\rho-2},$$

$$\frac{\partial^2 f(k,l)}{\partial k \partial l} = \left(\frac{1}{\rho} - 1\right)(k^\rho + l^\rho)^{\frac{1}{\rho}-2} \rho l^{\rho-1} k^{\rho-1},$$

$$\frac{\partial^2 f(k,l)}{\partial l \partial l} = (\rho - 1) l^{\rho-2} (k^\rho + l^\rho)^{\frac{1}{\rho}-1} + \left(\frac{1}{\rho} - 1\right)(k^\rho + x_2^\rho)^{\frac{1}{\rho}-2} \rho l^{2\rho-2}.$$

Example: CES -function

- By collecting the common terms, we get:

$$D^2 f(x_1, x_2) = \begin{pmatrix} \frac{\partial^2 f(k,l)}{\partial k \partial k} & \frac{\partial^2 f(k,l)}{\partial k \partial l} \\ \frac{\partial^2 f(k,l)}{\partial l \partial k} & \frac{\partial^2 f(k,l)}{\partial l \partial l} \end{pmatrix}$$

$$= (k^\rho + l^\rho)^{\frac{1}{\rho}-2} \begin{pmatrix} (\rho - 1) k^{\rho-2} l^\rho & (1-\rho) l^{\rho-1} k^{\rho-1} \\ (1-\rho) l^{\rho-1} k^{\rho-1} & (\rho - 1) l^{\rho-2} k^\rho \end{pmatrix}.$$

- When computing the determinant, we can separate the common factor:

$$\det (Hf(k,l)) =$$

$$(k^\rho + l^\rho)^{\frac{1}{\rho}-2} k^{2\rho-2} l^{2\rho-2} \det \begin{pmatrix} (\rho-1) & (1-\rho) \\ (1-\rho) & (\rho-1) \end{pmatrix} = 0.$$

- $Hf(x_1, x_2)$ is therefore negative semidefinite if $\rho < 1$ and positive semidefinite if$\rho > 1$.

Final comments on unconstrained optimization:

- Do local maxima or minima always exist?
  - Obviously you cannot find a maximum to $f(x) = 2x$.
- Are there economically meaningful cases where this could be problematic?
  - Think about production with constant returns to scale (homogenous of degree 1 production function).
- If you find all local maxima of a function, can you be sure that one of them is a global maximum?
- How do you determine which one of the local maxima is the global maximum?

Convex and concave functions: Convex sets

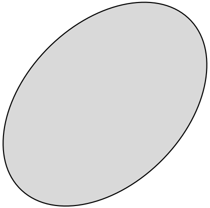
Definition

A set X *is convex* if for all $x, y \in X$ and for all $\lambda \in [0, 1]$, we have:

$$\lambda x + (1 - \lambda) y \in X.$$

We call $\lambda x + (1 - \lambda) y$ a *convex combination* of x and y.

- On the real line, convex sets are intervals $a \leq x \leq b$ for some $-\infty \leq a \leq b \leq \infty$.
- In $\mathbb{R}^n$, convex sets are sets X with the property that when you connect linearly two points in X, the entire connecting line is also in X.
- Hence a disk in the plane is convex and a cube in the three dimensional space are convex, but the circle in the plane is not, a disk with the center removed is not, a doughnut in three dimensions is not etc.

[Convex set] 

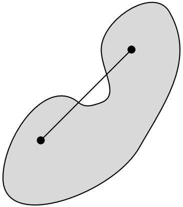
[Non-convex set] 

Figure: Illustration of convex sets.

Convex and concave functions: Definitions

- Consider a real-valued function $f : X \to \mathbb{R}$, where X is a convex set.

Definition

The function f *is convex* if for all $\boldsymbol{x}, \boldsymbol{y} \in X$ and for all $\lambda \in [0, 1]$, we have:

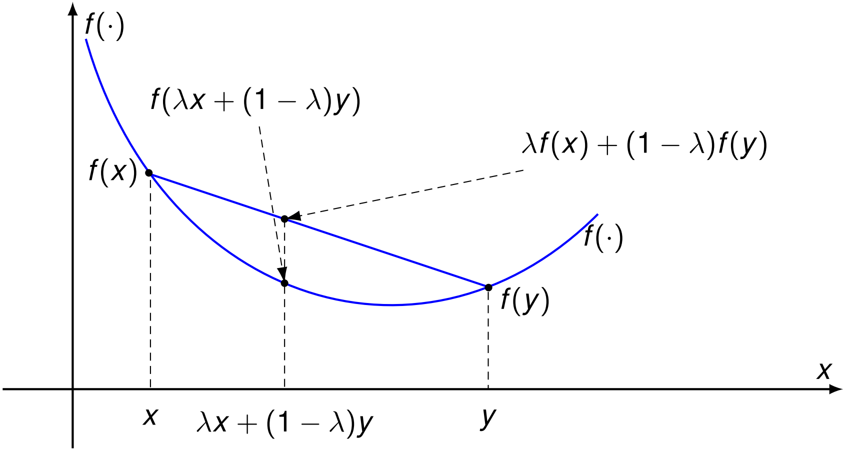
$$f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y}).$$

f *is concave* if

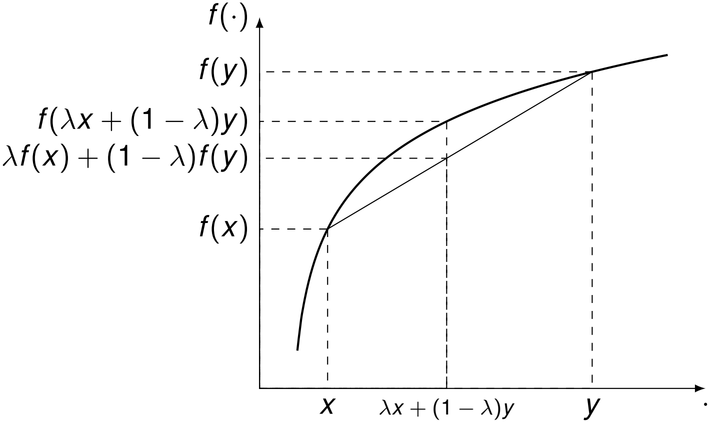
$$f(\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y}) \geq \lambda f(\boldsymbol{x}) + (1 - \lambda) f(\boldsymbol{y}).$$

- Note: If f is convex, then $-f$ is concave

A convex function on the interval $[a, b]$



A concave function of a real variable



Properties of convex functions

- If $f(\boldsymbol{x})$ is convex, then $g(\boldsymbol{x}) = -f(\boldsymbol{x})$ is concave.
- If $f(\boldsymbol{x})$ is convex, then $af(\boldsymbol{x})$ is convex if $a > 0$.
- If $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are convex, then $h(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$ is convex.
- If $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are convex, then $h(\boldsymbol{x}) = f(\boldsymbol{x})\, g(\boldsymbol{x})$ is not necessarily convex. (Give an example for both cases, i.e. where the product of convex functions is convex and where it is not).
- Exercise: Assume that $f : X \to \mathbb{R}$ is convex and $g : \mathbb{R} \to \mathbb{R}$ is also convex. Is $g(f(\boldsymbol{x}))$ convex? What if g is increasing and convex?
- (Optional Exercise): Assume that $f : X \to \mathbb{R}$ is a convex function. Show that the set

$$\{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \quad | \quad \boldsymbol{x} \in X, y \geq f(\boldsymbol{x})\}$$

  is a convex set.

Maximum of convex functions

Proposition

If $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are convex, then $h(\boldsymbol{x}) = \max\{f(\boldsymbol{x}), g(\boldsymbol{x})\}$ is convex.

Proof: By assumption, f and g are convex:

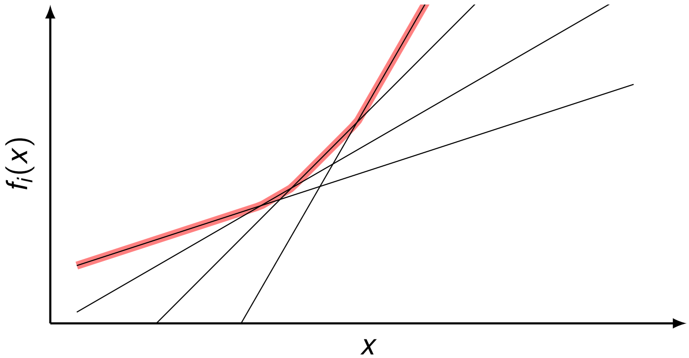
$$f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$
$$g(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda g(\boldsymbol{x}) + (1-\lambda) g(\boldsymbol{y}).$$

By definition,

$$\begin{aligned}
h(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \max\{f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}), g(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})\} \\
&\leq \max\{\lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}), \lambda g(\boldsymbol{x}) + (1-\lambda) g(\boldsymbol{y})\} \\
&\leq \lambda \max\{f(\boldsymbol{x}), g(\boldsymbol{x})\} + (1-\lambda) \max\{f(\boldsymbol{y}), g(\boldsymbol{y})\} \\
&= \lambda h(\boldsymbol{x}) + (1-\lambda) h(\boldsymbol{y}).
\end{aligned}$$

The first inequality follows from the convexity of f and g.The second follows by choosing the larger of $f(\cdot), g(\cdot)$ for $\boldsymbol{x}, \boldsymbol{y}$. The last equality is just the definition of h.

Maximum of linear functions is convex



Maxima and minima of convex functions

- The same result is true for an arbitrary set of convex functions. Let $f(\boldsymbol{x};\alpha)$ be convex in $\boldsymbol{x}$for all α. Then

$$g(\boldsymbol{x}) = \max_{\alpha} f(\boldsymbol{x};\alpha)$$

  is convex.
- Since linear functions are convex, this result holds for any set of linear functions.
- Since

$$\max\{f(\boldsymbol{x}), g(\boldsymbol{x})\} = -\min\{-f(\boldsymbol{x}), -g(\boldsymbol{x})\},$$

  and since $-f$ is concave when f is convex, we get:

$$g(\boldsymbol{x}) = \min_{\alpha} f(\boldsymbol{x};\alpha)$$

  is concave if $f(\boldsymbol{x};\alpha)$ is concave in $\boldsymbol{x}$ for all α.

Economic examples: profit maximization

- A competitive firm sells output y at price p_0 and buys inputs $\boldsymbol{x} = (x_1, ..., x_n)$ at input prices $\boldsymbol{p} = (p_1, ..., p_n)$. Its profit is

$$p_0 y - \Sigma_{i=1}^{n} p_i x_i.$$

- The maximization problem is then

$$\max_{y,x \in F} p_0 y - \Sigma_{i=1}^{n} p_i x_i,$$

  where F is the feasible set determined by technological possibilities.
- The profit function of the firm gives the maximum achievable profit amongst the feasible set at input and output prices $p_0, \boldsymbol{p}$.

$$\pi (p_0, \boldsymbol{p}) = \pi (p_0, p_1, ..., p_n) = \max_{y,x \in F} p_0 y - \Sigma_{i=1}^{n} p_i x_i$$

- Since the profit from a fixed feasible production is a linear function of the prices $(p_0, \boldsymbol{p})$, the profit function is the maximum over linear functions and therefore convex in $(p_0, \boldsymbol{p})$.

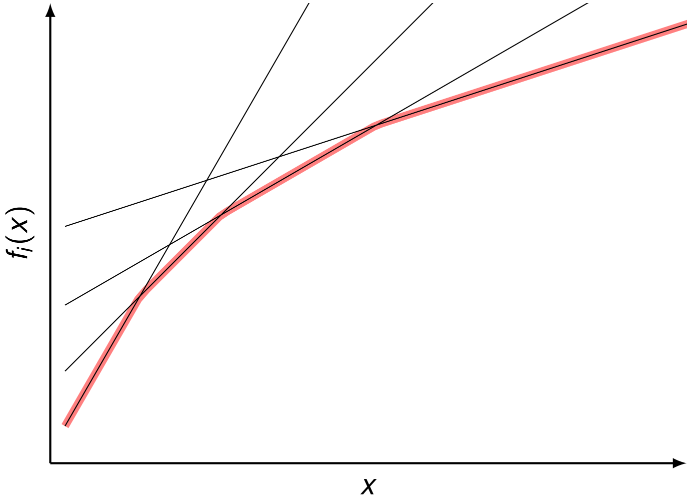
Economic examples: cost minimization

- Let X be the feasible set for inputs $\boldsymbol{x} = (x_1, ..., x_n)$ and $\boldsymbol{p} = (p_1, ..., p_n)$ be the input prices.
- The expenditure function

$$e(\boldsymbol{p}; X) = \min_{\boldsymbol{x} \in X} \boldsymbol{p} \cdot \boldsymbol{x} = \min_{\boldsymbol{x} \in X} \Sigma_{i=1}^{n} p_i x_i$$

  is a concave function by the same argument as above.
- These two examples show that convexity and concavity play a real role in economic applications.
- We shall see more applications when we discuss constrained optimization and value functions of optimization problems.

Lower envelope of linear functions is concave



Convexity and concavity of differentiable functions

- When $f : \mathbb{R} \to \mathbb{R}$, and f is convex and differentable, it it is easy to see by drawing a picture that for all x, y we have:

$$f(y) - f(x) \geq f'(x)(y - x).$$

- This just says that the graph $(x, f(x))$ of a convex function f is above all of its tangent lines.
- Similarly, the graph of a concave function lies below its tangent line

Convexity and concavity of differentiable functions

The multivariate version of this is proved in the notes:

Proposition

A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq Df(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) \text{ for all } \boldsymbol{x}, \boldsymbol{y}.$$

- Can you formulate this condition in terms of level curves and gradients?
- What is the corresponding result to concave functions?

Second derivatives and convexity

- Start again with functions of a single variable. By Taylor's theorem without the remainder term,

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2}f''(x)(y-x)^2 + \frac{1}{6}f'''(x)(y-x)^3 + \ldots$$

- In order to have

$$f(y) - f(x) \geq f'(x)(y-x)$$

for $|y-x|$ small, we must have

$$f''(x) \geq 0.$$

- In other words, convex functions have a positive second derivative.

Second derivatives and convexity

- Taylor's theorem with a remainder term of second degree:

$$f(y) = f(x) + f'(x)(y - x) + \frac{1}{2}f''(z)(y - x)^2$$

  for some $z \in [x, y]$.
- If f'' is everywhere non-negative, we get:

$$f(y) - f(x) \geq f'(x)(y - x)$$

  for all y, x and f is therefore convex.
- Let's generalize now to $f : X \to \mathbb{R}$, where X is a convex subset of $\mathbb{R}^n$.
- Convexity corresponds to positive semidefiniteness of the Hessian matrix.
- Concavity corresponds to negative semidefiniteness of the Hessian matrix.
- Hence we see an immediate connection between convexity and the second order conditions for optimality.

Quasiconvex and quasiconcave functions

- Even though the name suggests something extremely technical and tedious, quasiconcavity is actually one of the most important notions for functions in economic theory.

Definition

- A function f on a convex set X is *quasiconcave* if for all $\boldsymbol{x}, \boldsymbol{y} \in X$ and for all $\lambda \in [0, 1]$
$$f\left(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\right) \geq \min\{f(\boldsymbol{x}), f(\boldsymbol{y})\}.$$
f is *quasiconvex* is for all $\boldsymbol{x}, \boldsymbol{y} \in X$ and for all $\lambda \in [0, 1]$
$$f\left(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\right) \leq \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}.$$

Exercise: f is quasiconcave, then $-f$ is quasiconvex.

Quasiconvex and quasiconcave functions: Observations

- If f is quasiconcave, then af is quasiconcave if $a > 0$.
- If f and g are quasiconcave $f + g$ is not necessarily quasiconcave.
- All monotone (i.e. all increasing and all decreasing) functions of a single variable are both quasiconcave and quasiconvex. This is NOT true for multidimensional functions
- All concave functions are quasiconcave. Show this as an exercise.
- Not all quasiconcave functions are concave.
- If f is a quasiconcave function and g is a strictly increasing function, then $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$ is a quasiconcave function.

Quasiconvex and quasiconcave functions: Observations

- An upper contour set of function f for value α is denoted by $U(f;\alpha)$ and defined as:

$$U(f;\alpha) := \{\boldsymbol{x} \in X | f(\boldsymbol{x}) \geq \alpha\}.$$

- Interpretation: if f is a utility function, $U(f;\alpha)$ is the better side of the indifference curve giving utility level α.

Proposition

A function f is quasiconcave if and only if $U(f;\alpha)$ is a convex set for all α.

Quasiconvex and quasiconcave functions: Observations

Definition

A function f on a convex set X is *strictly quasiconcave* if for all $\boldsymbol{x}, \boldsymbol{y} \in X$ and for all $\lambda \in (0, 1)$

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) > \min\{f(\boldsymbol{x}), f(\boldsymbol{y})\}.$$

- The following exercise shows why strict quasiconcavity is very useful for optimization problems.
- Exercise: show that if a strictly quasiconcave function has a maximum, then the maximum is unique.

Quasiconcavity and differentiability

- A differentiable function f on a convex set X is quasiconcave if and only if:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) \Rightarrow Df(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) \geq 0.$$

- Exercise: Compare this to the definition of concavity for differentiable functions and relate this condition to the geometry of upper contour sets and tangent planes to the upper contour sets.

Next Lecture:

- Introduction to constrained optimization
- Equality constraints and Lagrange's function
- Examples of constrained optimization problems