Mathematics for Economists: Lecture 5

Juuso Välimäki

Aalto University School of Business

Spring 2020

This lecture covers

Unconstrained optimization

Start with some general definitions Function $f : X \to \mathbb{R}$ has a *maximum* at point x_0 if for all $y \in X$,

$$f(y) \leq f(x_0).$$

Function f has a *minimum* at x_0 if for all $y \in X$,

$$f(y) \geq f(x_0).$$

Local optima for $n = 1$

Definition

The function f has a *local maximum* at x_0 if there exists an $\varepsilon > 0$ such that for all $y \in B^{\varepsilon}(x_0) \subset X$, we have:

$$f(y) \leq f(x_0).$$

A *local minimum* is defined analogously.

- How do we know whether f has a maximum or a minimum at x_0?
- How to find minima and maxima?
- Exercise: If x_0 is a maximum of $f : X \to \mathbb{R}$ and $g : \mathbb{RR}$ is strictly increasing, then x_0 is a maximum of $g(f(x))$.

Local optima for $n = 1$: First-order condition

Consider first the case of a single variable:

$$f : \mathbb{R} \to \mathbb{R}.$$

Assume that f is differentiable. We know from the definition of the derivative that if f has a minimum or maximum at x_0, then

$$Df(x_0) = f'(x_0) = 0.$$

We call this the *first-order necessary condition*.

Local optima for $n = 1$: Second-order condition

- Assume next that $f'(x_0) = 0$.
  - If f has a maximum at x_0, then f is increasing for $x < x_0$ and decreasing for $x > x_0$.
  - If an increasing (decreasing) function has a derivative, it is positive (negative).
- In other words, we need to consider

$$f'(x_0)$$

  as a function of x_0.
- Therefore if $f'(x_0)$ has a derivative, we know that at a maximum,

$$Df'(x_0) \leq 0$$

Local optima for $n = 1$: Second-order condition

- Need to consider the derivative of the derivative. We denote:

$$f''(x_0) = \lim_{h \to 0} \frac{f'(x_0 + h) - f(x_0)}{h}$$

  and we call $f''(x_0)$ the *second derivative of f at x_0*.
- Therefore we have the *second-order necessary condition* for a maximum at x_0:

$$f''(x_0) \leq 0.$$

Higher order derivatives

- The third derivative is the derivative of the second derivative etc.
- If a function f has a k^{th} derivative at point x_0, we say that f is k times differentiable at x_0.
- We denote the k^{th} derivative at x_0 by $f^{[k]}(x_0)$.
- We say that it is k times continuously differentiable if the k^{th} derivative is continuous in x_0.

  Exercise: Consider the function f defined below and determine how many times it is differentiable at $x_0 = 0$.

$$f(x) = \begin{cases} 0 & \text{for} \quad x < 0, \\ x^2 & \text{for} \quad x \geq 0. \end{cases}$$

Taylor's theorem

Theorem (Taylor's theorem)

Suppose that $f : \mathbb{R} \to \mathbb{R}$, *is* $k+1$ *times continuously differentiable at* x_0. *Then*

$$f(x_0+h) = f(x_0) + f'(x_0)\,h + \frac{1}{2}f''(x_0)\,h^2 + \ldots + \frac{1}{k!}f^{[k]}(x_0) + \frac{1}{(k+1)!}f^{[k+1]}(x)\,h^{k+1},$$

for some x *with* $x_0 < x < x_0 + h$.

Taylor's theorem: Examples

- Consider $f(x) = \ln(x)$ and find an approximation to $\ln(1+h)$ by third order Taylor approximation. Compute around $x_0 = 1$:

$$\ln(1+h) = f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f^{(3)}(x_0)h^3$$
$$= 0 + h - \frac{1}{2}h^2 + \frac{1}{6}(x_0)h^3.$$

- Consider $f(x) = \sin(x)$ and find an approximation to $\sin(h)$ by fourth order Taylor approximation. Compute around $x_0 = 0$:

$$\ln(1+h) = f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f^{(3)}(x_0)h^3 + \frac{1}{24}f^{(4)}(x_0)h^4$$
$$= 0 + h - \frac{1}{6}h^3.$$

- Consider $f(x) = 3 + 4x - 2x^2 - x^3$ and find an approximation to $f(h)$ by third order Taylor approximation. Compute around $x_0 = 0$:

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \frac{1}{6}f^{(3)}(x_0)h^3$$

Taylor's theorem: Idea of proof with $k = 1$

- Consider the linear approximation:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \text{ higher order terms } r(h).$$

- We need to show that for some $x_0 < x < x_0 + h$, we have

$$r(h) = \frac{1}{2} f''(x) h^2.$$

- Let

$$g(x) = f(x) - f(x_0) - f'(x_0)h - \frac{1}{h^2}[f(x_0 + h) - f(x_0) - f'(x_0)h](x - x_0)^2.$$

- We see that $g(x_0) = g(x_0 + h) = 0$. Since g is continuously differentiable, Rolle's theorem guarantees that $g'(x') = 0$ for some $x_0 < x' < x_0 + h$.
- Since g is twice continuously differentiable, a second application of Rolle's theorem guarantees that there is an $x_0 < x < x' < x_0 + h$ such that $g''(x) = 0$.
- Since $g''(x) = f''(x) - 2\frac{1}{h^2}[f(x_0 + h) - f(x_0) - f'(x_0)h]$, the claim follows.
- The idea for the proof of the general case is based on the same idea, but involves more terms and $k + 1$ times the use of Rolle's theorem.

Taylor's theorem: significance

- For local analysis, 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.
- Note that Taylor's theorem states that differentiable functions can be locally well approximated by polynomials.
- Actually, much more is true: Any continuous function on a compact interval can be arbitrarily well approximated by polynomials (Stone - Weierstrass theorem)

Taylor's theorem: classifying critical points

- We say that x_0 is a critical point of f if $f'(x_0) = 0$.
- With the help of Taylor's theorem, we can classify all critical points
  1. If the first l for which $f^{[l]}(x_0) \neq 0$, is odd, then f does not have an extremum (i.e. minimum or maximum) at x_0.
  2. If the first l for which $f^{[l]}(x_0) \neq 0$, is even and $f^{[l]}(x_0) < 0$, then f has a local maximum at x_0.
  3. If the first l for which $f^{[l]}(x_0) \neq 0$, is even and $f^{[l]}(x_0) > 0$, then f has a local minimum at x_0.
- Make sure that you understand why this classification holds. For l defined as above, divide the Taylor approximation by $h^{l-1}$and let $h \to 0$.
- The case $f'(x_0) = 0$ and $f''(x_0) < 0$ is called the *second-order sufficient condition* for maximum at x_0.

Application: comparative statics and Taylor's theorem

- Consider the unconstrained optimization of $f(y, x)$ with respect to $y \in \mathbb{R}$.

$$\max_{y} f(y, x)$$

- The first order condition for optimum:

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

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

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

- If

$$\frac{\partial^2 f}{\partial y \partial y}(\widehat{y}, \widehat{x}) < 0,$$

then f has a local maximum at $(\widehat{y}; \widehat{x})$.

Application: comparative statics and Taylor's theorem

- Then we can apply implicit function theorem to first-order condition:

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

- 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 optimum, we see that $\frac{dy}{dx}$ has the same sign as $\frac{\partial^2 f\left(\widehat{y};\widehat{x}\right)}{\partial y \partial x}$.

Application: Optimal monopoly production

- Let x be the output by the monopolist. $P(q) = \alpha - b(q)$ is the inverse demand function and cq^2 is the cost function of the monopolist. The monopolist's maximization problem is then

$$\max_q \pi(q; \alpha, c) = q(\alpha - b(q)) - cq^2.$$

- First-order condition for optimality:

$$\frac{\partial \pi(q; \alpha, c)}{\partial q} = \alpha - b(q) - qb'(q) - 2cq = 0.$$

- Second-order condition:

$$\frac{\partial^2 \pi(q; \alpha, c)}{\partial q^2} < 0.$$

Application: Optimal monopoly production

- How does the optimal output change when α or c changes?
- By the previous result, the sign of the change in the endogenous variable depends on the signs of

$$\frac{\partial^2 \pi(q; \alpha, c)}{\partial q \partial \alpha}$$

and

$$\frac{\partial^2 \pi(q; \alpha, c)}{\partial q \partial c}.$$

Quadratic functions

- Quadratic functions of a real variable take the form:

$$f(x) = ax^2 + bx + c.$$

- Taylor's series around x_0 gives:

$$f(x_0 + h) = f(x_0) + (2ax_0 + b)\,h + \frac{1}{2}2ah^2.$$

- At $x_0 = -\frac{b}{2a}$, $f'(x_0) = 0$ and f has a minimum at x_0, if $a > 0$ and a maximum if $a < 0$.

Quadratic functions

- Consider next multivariate polynomial functions of second degree. These consist of a constant term, $c \in \mathbb{R}$, a first-order linear term $b \cdot x$, where

$$b = (b_1, b_2, ..., b_n), x = (x_1, x_2, ..., x_n)$$

$$b \cdot x = b^\top x = \Sigma_{i=1}^n b_i x_i$$

and a second-order term

$$\begin{aligned} &a_{11}x_1^2 + a_{12}x_1x_2 + ... + a_{1n}x_1x_n \\ &+a_{21}x_2x_1 + ... + a_{2n}x_2x_n \\ &+a_{n1}x_nx_1 + ... + a_{nn}x_n^2. \end{aligned}$$

- We can express the second order term via an $n \times n$ matrix A:

$$x \cdot Ax = x^\top Ax = (x_1, x_2, ..., x_n) A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Quadratic functions

- Since $x_i x_j = x_j x_i$ we can write A as a symmetric matrix A' by taking

$$A' = \frac{1}{2}A + A^\top.$$

- Examples: What is the long form of:

$$(x_1, x_2, x_3) \begin{pmatrix} 1 & 3 & 2 \\ 3 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}?$$

- By multiplying the second product, we get:

$$(x_1, x_2, x_3) \begin{pmatrix} x_1 + 3x_2 + 2x_3 \\ 3x_1 + 2x_2 + x_3 \\ 2x_1 + x_2 + 2x_3 \end{pmatrix}$$

$$\begin{aligned} = \quad & x_1^2 + 3x_1x_2 + 2x_1x_3 + 3x_1x_2 \\ & + 2x_2^2 + x_2x_3 + 2x_1x_3 + x_2x_3 + 2x_3^2 \\ = \quad & x_1^2 + 2x_2^2 + 2x_3^2 + 6x_1x_2 + 4x_1x_3 + 2x_2x_3. \end{aligned}$$

Quadratic functions

- Here is an example of a general quadratic function of two variables $x = (x_1, x_2)$:
$$f(x) = 6 + 7x_1 + 3x_2 + 2x_1^2 + 5x_1x_2 + 4x_2^2.$$
- How to write this in form
$$f(x) = c + b \cdot x + x^\top Ax?$$
$$c = 6, b = (7, 3), A = \begin{pmatrix} 2 & \frac{5}{2} \\ \frac{5}{2} & 4 \end{pmatrix}.$$

Local extrema of quadratic functions

- To find the local extrema of a quadratic f, compute the gradient:

$$\nabla f\left(\widehat{x}\right) = 0.$$

- Start with the partial derivatives $\frac{\partial f}{\partial x_i}$:

$$\frac{\partial f\left(\widehat{x}\right)}{\partial x_i} = b_i + 2a_{ii}\widehat{x}_i + \Sigma_{j\neq i}\left(a_{ij} + a_{ji}\right)\widehat{x}_j = b_i + 2\mathbf{a}_i \cdot \widehat{x},$$

  where $\mathbf{a}_i$ is the i^{th} row of matrix A.

- Since

$$\nabla f\left(\widehat{x}\right) = \begin{pmatrix} \frac{\partial f\left(\widehat{x}\right)}{\partial x_1} \\ \vdots \\ \frac{\partial f\left(\widehat{x}\right)}{\partial x_n} \end{pmatrix},$$

  we get

$$\nabla f\left(\widehat{x}\right) = b + 2A\widehat{x}.$$

Local extrema of quadratic functions

- Hence we see that the gradient is zero if

$$\nabla f\left(\widehat{x}\right) = 0 \Leftrightarrow \widehat{x} = -\frac{1}{2}A^{-1}b.$$

- For this to make sense, A must have full rank so that A^{-1} exists.
- Notice the resemblance to the single variable case: $x = \frac{b}{2a}$.
- We'll see an important application of this result next.

Application: sum of least squares

- Consider a statistical sample consisting of on N pairs of observations

$$(y_1, x_1), ..., (y_N, x_N).$$

- Suppose that we want to find a linear relation between x and y.
- We would like to find a coefficient β that rationalizes the observations as

$$y_i = \beta x_i.$$

- If we have many observations, this will not be satisfied in general.
- To account for errors in the linear relationship, specify the following statistical model

$$y_i = \beta x_i + \varepsilon_i,$$

where ε_i is an identically and independently distributed error term for all i.

Application: sum of least squares

- Our task is to infer β from the sample. One way of doing this is based on minimizing the sum of squared error terms $(\varepsilon_i)^2$, i.e. to

$$\min_{\beta} f(\beta) = \Sigma_{i=1}^{N} (y_i - \beta x_i)^2 .$$

- Compute $f'(\beta)$ and find $\widehat{\beta}$ such that:

$$f'\left(\widehat{\beta}\right) = 0.$$

- By taking the derivative, we get:

$$f'\left(\widehat{\beta}\right) = \sum_{i=1}^{N} -2x_i \left(y_i - \widehat{\beta} x_i\right).$$

- As a result, $f'\left(\widehat{\beta}\right) = 0$ if

$$\widehat{\beta} = \frac{\Sigma_{i=1}^{N} x_i y_i}{\Sigma_{i=1}^{N} x_i^2}.$$

- Since $f''(\hat{\beta}) = \sum_{i=1}^{N} x_i^2 > 0$, we have found the minimum.

Application: sum of least squares

- If we want to include a constant term α, we get:

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

- The sum of squares as a function of (α, β):

$$f(\alpha, \beta) = \Sigma_{i=1}^{N} (y_i - \alpha - \beta x_i)^2.$$

- To find $\left(\widehat{\alpha}, \widehat{\beta}\right)$ such that

$$\frac{\partial f\left(\widehat{\alpha}, \widehat{\beta}\right)}{\partial \alpha} = \frac{\partial f\left(\widehat{\alpha}, \widehat{\beta}\right)}{\partial \beta} = 0,$$

we get:

$$\begin{aligned} \Sigma_{i=1}^{N} \left(y_i - \widehat{\alpha} - \widehat{\beta} x_i\right) &= 0, \\ \Sigma_{i=1}^{N} - 2x_i \left(y_i - \widehat{\alpha} - \widehat{\beta} x_i\right) &= 0. \end{aligned}$$

Application: sum of least squares

- Solving for α from the first equation gives:

$$\widehat{\alpha} = \frac{\Sigma_{i=1}^{N} y_i - \widehat{\beta}\Sigma_{i=1}^{N} x_i}{N} := \overline{y} - \widehat{\beta}\overline{x}.$$

- Using the first equation we also see that:

$$\Sigma_{i=1}^{N} \overline{x} \left(y_i - \widehat{\alpha} - \widehat{\beta} x_i \right) = 0.$$

- By substituting into the second, we get:

$$\widehat{\beta} = \frac{\Sigma_{i=1}^{N} \left(x_i - \overline{x} \right) \left(y_i - \overline{y} \right)}{\Sigma_{i=1}^{N} \left(x_i - \overline{x} \right)^2} = \frac{Cov\left(y, x \right)}{Var\left(x \right)}.$$

Application: sum of least squares

- More generally, we can consider samples with more explanatory variables: $(y_1, x_{11}, x_{21}, ..., x_{K1}), ... (y_N, x_{1N}, ..., x_{KN})$ and a linear model

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \beta_1 x_{11} + & \cdots & \beta_K x_{K1} \\ \vdots & & \vdots \\ \beta_1 x_{1N} & \cdots & \beta_K x_{KN} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

or in matrix form:

$$y = X\beta + \varepsilon.$$

Application: sum of least squares

- We can compute the sum of squares now as:

$$\begin{aligned} f(\beta) &= \varepsilon \cdot \varepsilon = (y - X\beta)^\top (y - X\beta) \\ &= y \cdot y - (X\beta)^\top y - y^\top X\beta + \beta^\top X^\top X\beta \\ &= y \cdot y - 2y^\top X\beta + \beta^\top X^\top X\beta. \end{aligned}$$

- We can now use the general formula found earlier for the quadratic functions:

$$\nabla f\left(\widehat{\beta}\right) = -2X^\top y + 2X^\top X\widehat{\beta}.$$

- Therefore we can find a candidate for the extremum by setting

$$\nabla f\left(\widehat{\beta}\right) = 0.$$

- Solving for β, we get:

$$\widehat{\beta} = \left(X^\top X\right)^{-1} X^\top y.$$

Exercise: Is there a need to add a constant term to get a more general formula?

Classifying critical points of multivariate functions

- How can we determine if a multivariate function has a minimum or a maximum at $\widehat{x}$ such that $f'(\hat{x}) = 0$?
- Denote by $D^2 f\left(\widehat{x}\right)$ the second derivative of f at $\widehat{x}$.
- Assume that Taylor's theorem is true for multivariate functions as well (as it is).
- Then we would have locally:

$$f\left(\widehat{x}+h\right) = f\left(\widehat{x}\right) + \nabla f\left(\widehat{x}\right)\cdot h + \frac{1}{2} h \cdot D^2 f\left(\widehat{x}\right) h$$

- Consider the gradient as a function of $\widehat{x}$.

$$\nabla f\left(\widehat{x}\right) : \mathbb{R}^n \to \mathbb{R}^n,$$

- Then the derivative of the gradient is a linear function from $\mathbb{R}^n$ to $\mathbb{R}^n$ and we define:

$$D^2 f\left(\widehat{x}\right) := D\left(\nabla f\left(\widehat{x}\right)\right) = \begin{pmatrix} \frac{\partial f\left(\widehat{x}\right)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial f\left(\widehat{x}\right)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f\left(\widehat{x}\right)}{\partial x_n \partial x_1} & \cdots & \frac{\partial f\left(\widehat{x}\right)}{\partial x_n \partial x_n} \end{pmatrix}.$$

Classifying critical points of multivariate functions

- We call this matrix of second order partial derivatives the Hessian matrix of f.
- Taylor's theorem tells us that for $\hat{x}$ such that

$$\nabla f\left(\widehat{x}\right) = 0,$$

we have:

$$f\left(\widehat{x} + h\right) - f\left(\widehat{x}\right) = \frac{1}{2} h \cdot D^2 f\left(\widehat{x}\right) h.$$

- Whether f has a minimum, a maximum or neither at $\widehat{x}$ depends on whether

$$h \cdot D^2 f\left(\widehat{x}\right) h \gtreqless 0$$

for all h.

Classifying critical points of multivariate functions

- An important observation relating to critical points of $f : \mathbb{R}^n \to \mathbb{R}$ is that if $Df(x_0) = 0$, then for all strictly increasing and differentiable $g : \mathbb{R} \to \mathbb{R}$, we have that $Dg(f(x_0)) = 0$. This follows immediately from the chain rule.

Theorem

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

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

As a result, the Hessian matrix of f at x is a symmetric matrix for all x.

Quadratic forms and the definiteness of matrices

- Because of the length of what follows, let me sum the important issues in this part.
  1. The definitions of definiteness and semidefiniteness
  2. Second-order Taylor approximation of multivariate functions f and the Hessian matrix of f
  3. Connecting definiteness of the Hessian to convexity and concavity discussed in the next lecture
- A quadratic form is a homogenous second-degree polynomial whose terms are all of second order. They can be written as:

$$x \cdot Ax$$

  for some symmetric matrix A.
- A quadratic form is *positive definite* if for all $x \neq 0$, $x \cdot Ax > 0$. It is *positive semidefinite* if for all x, $x \cdot Ax \geq 0$.
- A quadratic form is *negative definite* if for all $x \neq 0$, $x \cdot Ax < 0$. It is *negative semidefinite* if for all x, $x \cdot Ax \leq 0$. In all other cases, we say that the quadratic form is indefinite.

Quadratic forms and the definiteness of matrices

- By Taylor's theorem, we can use definiteness to classify the local extrema of quadratic functions.
- Let $\nabla f\left(\widehat{x}\right) = 0$.
- Then if $D^2 f\left(\widehat{x}\right)$ is positive definite, then $\widehat{x}$ is a local minimum.
- If $D^2 f\left(\widehat{x}\right)$ is negative definite, then $\widehat{x}$ is a local maximum.
- When is A positive definite? The easiest case is when A is a diagonal matrix. In this case, it is positive definite if and only if all of its diagonal elements are strictly positive.
- More generally any positive definite matrix has strictly positive diagonal elements.

Quadratic forms and the definiteness of matrices

- Another easy case is when A is a 2×2 matrix:

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

  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

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

Quadratic forms and the definiteness of matrices

- Plugging 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) $\det A > 0$. For semidefiniteness, the inequalities are weak.

Quadratic forms and the definiteness of matrices

- For negative definiteness, assume that $a, c < 0$. Solving for the maximal x_2 for each x_1 gives:

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

and plugging 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,

$$\det A > 0.$$

- For semidefiniteness, the inequalities in the above are weak inequalities.

Example: CES -function

- Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x_1, x_2) = \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}}.$$

- Form the gradient:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{\partial f(x_1,x_2)}{\partial x_1} \\ \frac{\partial f(x_1,x_2)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-1} x_1^{\rho-1} \\ \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-1} x_2^{\rho-1} \end{pmatrix}.$$

- The the Hessian matrix is:

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

Example: CES -function

- By the product rule:

$$\begin{aligned}\frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_1} &= (\rho-1)\, x_1^{\rho-2}\left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-1} \\ &\quad + \left(\frac{1}{\rho}-1\right)\left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} \rho x_1^{2\rho-2},\end{aligned}$$

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2} = \left(\frac{1}{\rho}-1\right)\left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} \rho x_2^{\rho-1} x_1^{\rho-1},$$

$$\begin{aligned}\frac{\partial^2 f(x_1,x_2)}{\partial x_2 \partial x_2} &= (\rho-1)\, x_2^{\rho-2}\left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-1} \\ &\quad + \left(\frac{1}{\rho}-1\right)\left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} \rho x_2^{2\rho-2}.\end{aligned}$$

Example: CES -function

- By collecting the common terms, we get:

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

$$= \left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} \begin{pmatrix} (\rho-1)\, x_1^{\rho-2} x_2^\rho & (1-\rho)\, x_2^{\rho-1} x_1^{\rho-1} \\ (1-\rho)\, x_2^{\rho-1} x_1^{\rho-1} & (\rho-1)\, x_2^{\rho-2} x_1^\rho \end{pmatrix}.$$

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

$$\det\left(D^2 f(x_1, x_2)\right) =$$

$$\left(x_1^\rho + x_2^\rho\right)^{\frac{1}{\rho}-2} x_1^{2\rho-2} x_2^{2\rho-2} \det \begin{pmatrix} (\rho-1) & (1-\rho) \\ (1-\rho) & (\rho-1) \end{pmatrix} = 0.$$

- $D^2 f(x_1, x_2)$ is therefore negative semidefinite if $\rho < 1$ and positive semidefinite if $\rho > 1$.

Quadratic forms and the definiteness of matrices

- Unfortunately, the general case for determining definiteness is tedious. It is covered in the notes in detail and we do not cover it here.
- At the end of Part II of these lectures, we will discuss the eigenvalues of a matrix.
- It turns out that for symmetric matrices, A, there is a simple connection between definiteness and the sign of the eigenvalues.
- A symmetric matrix is positive definite if all of its eigenvalues are strictly positive. It is positive semidefinite if all eigenvalues are non-negative (and similarly for negative definiteness).