

# R401: Statistical and Mathematical Foundations

## Lecture 14: Unconstrained Optimization. Static Optimization with Equality Constraints. Lagrange Multipliers

Andrey Vassilev

2016/2017

# General Principles and Caveats for the Optimization Module

- Emphasis on practicality over rigour
- Consequently, algorithmic approach and “recipes” rather than proofs
- Also, existence and relevant properties of various objects are often implicitly assumed
- Pathologies and mathematical peculiarities discussed only in special cases

# Lecture Contents

- 1 Warm-up: Basic Unconstrained Optimization in  $\mathbb{R}^1$
- 2 Unconstrained Optimization in  $\mathbb{R}^n$
- 3 Static Optimization with Equality Constraints. Lagrange Multipliers

# Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$

# Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$

## Fact 1

For a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  differentiable at a point  $x$ , a necessary condition for a local extreme point (i.e. a maximum or a minimum) at  $x$  is

$$f'(x) = 0.$$

## Example 1

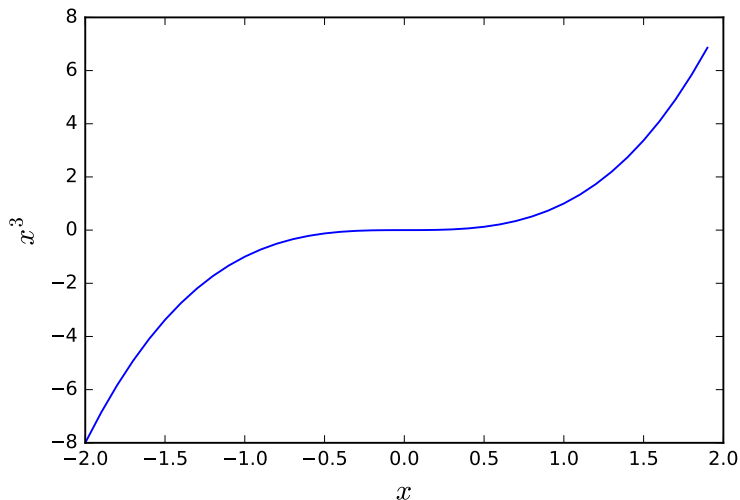
If  $f(x) = ax^2 + bx + c$ , then  $f'(x) = 2ax + b$  and the condition  $f'(x) = 0$  yields the familiar  $x = -\frac{b}{2a}$  (recall your high-school days). Depending on the sign of  $a$ , this is a maximum or a minimum (What is the relationship?).

## Example 2

If  $f(x) = x^3$ , then  $f'(x) = 3x^2$  and  $f'(x) = 0 \Rightarrow x = 0$ .

Does the function attain a maximum or a minimum at  $x = 0$ ?

# Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$



# Warm-up: Basic Unconstrained Optimization in $\mathbb{R}^1$

## Example 2 (cont.)

The answer is “neither”! The point  $x = 0$  is not a local extreme point of  $f(x) = x^3$ .

This illustrates the pitfalls of using necessary conditions – they supply only candidates that need to be checked further.

The above examples generalize in the following manner:

## Fact 2

Let a function  $f$  be  $n$  times differentiable at a point  $x$  and

$$f'(x) = f''(x) = \dots = f^{(n-1)}(x) = 0, \quad f^{(n)} \neq 0.$$

- ① If  $n$  is odd, the point  $x$  is not an extreme point of  $f(x)$ .
- ② If  $n$  is even and  $f^{(n)}(x) > 0$ , the point  $x$  is a minimum.
- ③ If  $n$  is even and  $f^{(n)}(x) < 0$ , the point  $x$  is a maximum.

# Unconstrained Optimization in $\mathbb{R}^n$



# Unconstrained Optimization in $\mathbb{R}^n$

## Necessary conditions

### Fact 3

For a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , differentiable at a point  $\mathbf{x}$ , a necessary condition for  $\mathbf{x}$  to be a local extreme point is

$$f'(\mathbf{x}) = \mathbf{0},$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{and} \quad f'(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix} (= \nabla f(\mathbf{x}))$$

**Note:** A point where the gradient of a function  $f$  vanishes is called a *critical point* or a *stationary point*. This also applies to functions on  $\mathbb{R}^1$ .

# Unconstrained Optimization in $\mathbb{R}^n$

## Example 3

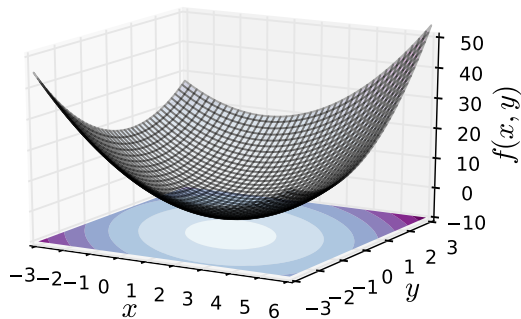
$$f(x, y) = x^2 + 2y^2 - 3x + xy$$

$$\frac{\partial f}{\partial x} = 2x - 3 + y = 0 \quad \Rightarrow \quad x = \frac{3 - y}{2}$$

$$\frac{\partial f}{\partial y} = 4y + x = 0 \quad \Rightarrow \quad y = -\frac{x}{4}$$

$$x = \frac{12}{7}, \quad y = -\frac{3}{7}$$

# Unconstrained Optimization in $\mathbb{R}^n$



# Unconstrained Optimization in $\mathbb{R}^n$

The necessity of the condition  $f'(\mathbf{x}) = \mathbf{0}$  has implications that are similar to the univariate case:

## Example 4

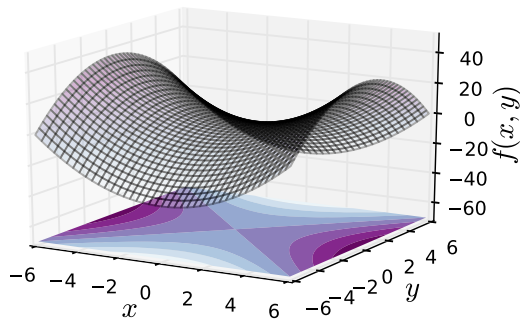
Consider the function  $f(x, y) = x^2 - y^2$ . The NCs yield the following candidate:

$$\frac{\partial f}{\partial x} = 2x = 0 \quad \Rightarrow \quad x = 0,$$

$$\frac{\partial f}{\partial y} = -2y = 0 \quad \Rightarrow \quad y = 0.$$

Let's look at the graph of the function in a neighbourhood of the point  $(0,0)'$ .

# Unconstrained Optimization in $\mathbb{R}^n$



# Unconstrained Optimization in $\mathbb{R}^n$

## Example 4 (cont.)

The critical point  $\mathbf{x} = (0,0)'$  is an example of a *saddle point*. The function  $f$  (obviously) does not attain an extremum at  $\mathbf{x}$ .

Example 4 illustrates the need to refine the approach for checking candidate points in the  $n$ -dimensional case. To this end, we have to review several concepts.

A symmetric square matrix  $A$  is called *positive semidefinite* if, for any vector  $\mathbf{x}$ , we have

$$\mathbf{x}'A\mathbf{x} \geq 0.$$

If the inequality is strict for any non-zero vector  $\mathbf{x}$ , the matrix is called *positive definite*.

Similarly, a symmetric square matrix  $A$  is called *negative semidefinite* if, for any vector  $\mathbf{x}$ , we have  $\mathbf{x}'A\mathbf{x} \leq 0$ , and *negative definite* in case of strict inequality for  $\mathbf{x} \neq \mathbf{0}$ .

# Unconstrained Optimization in $\mathbb{R}^n$

Incidentally, for a given square symmetric matrix  $A$ , the function  $Q(\mathbf{x}) = \mathbf{x}' A \mathbf{x}$  is called a *quadratic form*. Quadratic forms are also referred to as “positive/negative (semi)definite”, depending on the properties of the respective matrix.

Recall that, for an  $n \times n$  matrix  $A$ , a *principal minor* of order  $k$  ( $1 \leq k \leq n$ ), denoted by  $\Delta_k$ , is the determinant of the submatrix obtained by deleting  $n - k$  rows of the matrix and the correspondingly numbered columns, e.g.

$$\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{pmatrix}$$

**Note:** The notation  $\Delta_k$  does not identify a unique principal minor of order  $k$ .

# Unconstrained Optimization in $\mathbb{R}^n$

Incidentally, for a given square symmetric matrix  $A$ , the function  $Q(\mathbf{x}) = \mathbf{x}' A \mathbf{x}$  is called a *quadratic form*. Quadratic forms are also referred to as “positive/negative (semi)definite”, depending on the properties of the respective matrix.

Recall that, for an  $n \times n$  matrix  $A$ , a *principal minor* of order  $k$  ( $1 \leq k \leq n$ ), denoted by  $\Delta_k$ , is the determinant of the submatrix obtained by deleting  $n - k$  rows of the matrix and the correspondingly numbered columns, e.g.

$$\begin{pmatrix} \cancel{a_{1,1}} & a_{1,2} & \cancel{a_{1,3}} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ \cancel{a_{3,1}} & \cancel{a_{3,2}} & \cancel{a_{3,3}} & \cdots & \cancel{a_{3,n}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{pmatrix}$$

**Note:** The notation  $\Delta_k$  does not identify a unique principal minor of order  $k$ .



# Unconstrained Optimization in $\mathbb{R}^n$

The  $k$ -th *leading principal minor* of a matrix  $A$  ( $1 \leq k \leq n$ ), denoted by  $D_k$ , is the determinant of the submatrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,k} \end{pmatrix},$$

i.e. the principal minor obtained by deleting the last  $n - k$  rows and columns and, respectively, keeping the first  $k$ .

# Unconstrained Optimization in $\mathbb{R}^n$

## Fact 4 (Sylvester's criterion)

Let  $A$  be a symmetric matrix. Then:

- ①  $A$  is positive definite if and only if  $D_k > 0$ ,  $k = 1, \dots, n$ .
- ②  $A$  is positive semidefinite if and only if  $\Delta_k \geq 0$  for all principal minors of order  $k = 1, \dots, n$ .
- ③  $A$  is negative definite if and only if  $(-1)^k D_k > 0$ ,  $k = 1, \dots, n$ .
- ④  $A$  is negative semidefinite if and only if  $(-1)^k \Delta_k \geq 0$  for all principal minors of order  $k = 1, \dots, n$ .

Note that the necessary and sufficient conditions for “semidefiniteness” involve all principal minors (and hence are cumbersome to check), not just the leading principal minors.

# Unconstrained Optimization in $\mathbb{R}^n$

Let a function  $f(\mathbf{x}) = f(x_1, \dots, x_n)$  be twice differentiable. The matrix of second partial derivatives, evaluated at a point  $\mathbf{x}$ , i.e.

$$\begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \dots & \dots & \ddots & \dots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

is called the *Hessian (matrix)* of  $f$  at  $\mathbf{x}$ .

# Unconstrained Optimization in $\mathbb{R}^n$

- The Hessian is denoted  $\mathbf{f}''(\mathbf{x})$ .
- The Hessian is symmetric.
- Sometimes the partial derivative  $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$  is written as  $f''_{ij}(\mathbf{x})$ .
- A leading principal minor of order  $k$  of the Hessian is denoted  $D_k(\mathbf{x})$ .
- An arbitrary principal minor of order  $k$  of the Hessian is denoted  $\Delta_k(\mathbf{x})$ .

# Unconstrained Optimization in $\mathbb{R}^n$

## Fact 5

Let a (twice) differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  have a critical point at  $\mathbf{x}^*$ .

- ① If the Hessian  $\mathbf{f}''(\mathbf{x}^*)$  is positive definite or, equivalently,  $D_k(\mathbf{x}^*) > 0$ ,  $k = 1, \dots, n$ , then  $\mathbf{x}^*$  is a *local minimum point*.
- ② If the Hessian  $\mathbf{f}''(\mathbf{x}^*)$  is negative definite or, equivalently,  $(-1)^k D_k(\mathbf{x}^*) > 0$ ,  $k = 1, \dots, n$ , then  $\mathbf{x}^*$  is a *local maximum point*.
- ③ If  $D_n(\mathbf{x}^*) \neq 0$  and neither 1) nor 2) is satisfied, then  $\mathbf{x}^*$  is a *saddle point*.

# Unconstrained Optimization in $\mathbb{R}^n$

## Fact 6

Let a (twice) differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  have an extreme point at  $\mathbf{x}^*$ .

- ① If  $\mathbf{x}^*$  is a local minimum point, then the Hessian  $\mathbf{f}''(\mathbf{x}^*)$  is positive semidefinite or, equivalently,  $\Delta_k(\mathbf{x}^*) \geq 0$  for all principal minors of order  $k = 1, \dots, n$ .
- ② If  $\mathbf{x}^*$  is a local maximum point, then the Hessian  $\mathbf{f}''(\mathbf{x}^*)$  is negative semidefinite or, equivalently,  $(-1)^k \Delta_k(\mathbf{x}^*) \geq 0$  for all principal minors of order  $k = 1, \dots, n$ .

# Unconstrained Optimization in $\mathbb{R}^n$

## Example 5 (Verification of Example 3)

Recall that:

$$f(x, y) = x^2 + 2y^2 - 3x + xy$$

$$\frac{\partial f}{\partial x} = 2x - 3 + y, \quad \frac{\partial f}{\partial y} = 4y + x.$$

We now have:

$$\frac{\partial^2 f}{\partial x^2} = 2, \quad \frac{\partial^2 f}{\partial y^2} = 4, \quad \frac{\partial^2 f}{\partial x \partial y} = 1, \quad \frac{\partial^2 f}{\partial y \partial x} = 1.$$

$$D_1 = \det(2) = 2 > 0, \quad D_2 = \det \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} = 2 \cdot 4 - 1 \cdot 1 = 7 > 0.$$

Since  $D_1 > 0$ ,  $D_2 > 0$ , the critical point  $x = \frac{12}{7}$ ,  $y = -\frac{3}{7}$  is a minimum.

**Homework:** Apply the same procedure to Example 4.

# Static Optimization with Equality Constraints. Lagrange Multipliers



# Static Optimization with Equality Constraints

## Formulation

Now we look at problems of the form

$$f(x_1, \dots, x_n) \rightarrow \min(\max) \quad (1)$$

s.t.

$$\begin{aligned} g_1(x_1, \dots, x_n) &= b_1 \\ g_2(x_1, \dots, x_n) &= b_2 \\ &\dots \\ g_m(x_1, \dots, x_n) &= b_m \end{aligned} \quad (2)$$

where  $m < n$ . (Can you explain the last requirement?)

**Note:** In what follows, all required properties of the objects in (1) and (2) like differentiability are implicitly assumed.

# Static Optimization with Equality Constraints

## Formulation

Using vector notation for compactness, the objective function is:

$$f(\mathbf{x}) \rightarrow \min(\max)$$

We introduce

$$\mathbf{g}(\mathbf{x}) := (g_1(\mathbf{x}), \dots, g_m(\mathbf{x}))', \quad \mathbf{b} = (b_1, \dots, b_m)'$$

and the constraints are written as

$$\mathbf{g}(\mathbf{x}) = \mathbf{b}.$$

# Static Optimization with Equality Constraints

## The Lagrangian

The standard approach to solving (1)-(2) starts by defining a *Lagrangian*:

$$\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \lambda_1(g_1(\mathbf{x}) - b_1) - \cdots - \lambda_m(g_m(\mathbf{x}) - b_m).$$

The numbers  $\lambda_1, \dots, \lambda_m$  are called *Lagrange multipliers*.

This can also be written in vector notation:

$$\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \boldsymbol{\lambda}'(\mathbf{g}(\mathbf{x}) - \mathbf{b}),$$

where  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)'$  is the vector of Lagrange multipliers.

# Static Optimization with Equality Constraints

We can use the Lagrangian to produce necessary conditions for optimality in the following manner:

## Algorithm

- ① Form the Lagrangian as above
- ② Differentiate it w.r.t. the variables we are optimizing over, i.e.

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f(\mathbf{x})}{\partial x_i} - \sum_{j=1}^m \lambda_j \frac{\partial g_j(\mathbf{x})}{\partial x_i}, \quad i = 1, \dots, n$$

- ③ Set the resulting derivatives equal to zero, i.e.

$$\frac{\partial \mathcal{L}}{\partial x_i} = 0, \quad i = 1, \dots, n$$

- ④ The equations in the preceding step, together with the constraints (2), form a system of  $n + m$  equations which is solved for the unknowns  $x_i$  and  $\lambda_j$

# Static Optimization with Equality Constraints

## Remarks

- Sometimes the Lagrangian is equivalently formulated as

$$\mathcal{L}(\mathbf{x}) = f(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \cdots - \lambda_m g_m(\mathbf{x}).$$

It obviously makes no difference as to the result of the differentiation step.

- One modification of the algorithm requires to also differentiate the Lagrangian w.r.t.  $\lambda_j$  and set the resulting derivatives equal to zero. This simply reproduces the constraints (2) and is covered by the last step of our algorithm.
- Let the algorithm yield a candidate  $\mathbf{x}^*$ . Roughly, if the Lagrangian is convex in  $\mathbf{x}$ , then the candidate  $\mathbf{x}^*$  is a minimum. If the Lagrangian is concave in  $\mathbf{x}$ , then the candidate  $\mathbf{x}^*$  is a maximum. (See SHSS, p. 117, for the precise formulation.)
- A Lagrange multiplier is interpreted as a *shadow price*, i.e. the gain (or loss) arising from relaxing the associated constraint.

# Static Optimization with Equality Constraints

## Example 6 (Basic intertemporal optimization)

- An economic agent lives for two periods and supplies a fixed amount of labour in the first period of his life in exchange for monetary payment  $y$ .
- In period 1 the agent consumes  $c_1$  units of a good out of his income and saves the remaining  $y - c_1$ . (For convenience we assume there is no inflation and the price of the good is normalized to one.)
- Savings are remunerated at an interest rate  $r$ . Thus, in the second period the agent has at his disposal

$$(y - c_1)(1 + r)$$

to finance consumption, denoted  $c_2$ .

- The agent obtains utility from consumption according to the utility function

$$u(c_1, c_2) = \ln c_1 + \beta \ln c_2, \quad \beta \in (0, 1).$$

- The agent seeks to maximize utility w.r.t.  $c_1, c_2$ .

# Static Optimization with Equality Constraints

## Example 6 (cont.)

The above problem can be formalized as

$$\max_{c_1, c_2} u(c_1, c_2)$$

s.t.

$$c_2 = (y - c_1)(1 + r).$$

Notice that the constraint can be written equivalently as

$$c_1 + \frac{c_2}{1 + r} = y$$

to conform to the  $\mathbf{g}(\mathbf{x}) = \mathbf{b}$  convention. (Can you interpret the last equation in terms of discounting to period 1 quantities?)

The Lagrangian for this problem is

$$\mathcal{L} = \ln c_1 + \beta \ln c_2 - \lambda \left( c_1 + \frac{c_2}{1 + r} - y \right).$$

# Static Optimization with Equality Constraints

## Example 6 (cont.)

The solution algorithm yields

$$\frac{\partial \mathcal{L}}{\partial c_1} = \frac{1}{c_1} - \lambda = 0 \quad \Rightarrow \quad c_1 = \frac{1}{\lambda}$$

$$\frac{\partial \mathcal{L}}{\partial c_2} = \beta \frac{1}{c_2} - \frac{\lambda}{1+r} = 0 \quad \Rightarrow \quad c_2 = \frac{\beta(1+r)}{\lambda}$$

Combining the above equations to eliminate  $\lambda$ , we obtain

$$c_2 = \beta(1+r)c_1.$$

Substitute the last expression in the budget constraint:

$$c_1 + \frac{\beta(1+r)c_1}{1+r} = y \quad \Rightarrow \quad c_1^* = \frac{y}{1+\beta}.$$



# Static Optimization with Equality Constraints

## Example 6 (cont.)

We then have

$$c_2^* = \beta(1+r)c_1 = (1+r)\frac{\beta}{1+\beta}y.$$

Let us check how the optimal value of the utility function  $u^* = u(c_1^*, c_2^*)$  changes with income:

$$\begin{aligned}\frac{\partial u^*}{\partial y} &= \frac{\partial}{\partial y} \left( \ln \frac{y}{1+\beta} + \beta \ln \frac{(1+r)\beta y}{1+\beta} \right) \\ &= \frac{1+\beta}{y} \frac{1}{1+\beta} + \beta \frac{1+\beta}{(1+r)\beta y} \frac{(1+r)\beta}{1+\beta} \\ &= \frac{1}{y} + \frac{\beta}{y} = \frac{1+\beta}{y}\end{aligned}$$

# Static Optimization with Equality Constraints

## Example 6 (cont.)

We then have

$$c_2^* = \beta(1+r)c_1 = (1+r)\frac{\beta}{1+\beta}y.$$

Let us check how the optimal value of the utility function  $u^* = u(c_1^*, c_2^*)$  changes with income:

$$\begin{aligned}\frac{\partial u^*}{\partial y} &= \frac{\partial}{\partial y} \left( \ln \frac{y}{1+\beta} + \beta \ln \frac{(1+r)\beta y}{1+\beta} \right) \\ &= \frac{1+\beta}{y} \frac{1}{1+\beta} + \beta \frac{1+\beta}{(1+r)\beta y} \frac{(1+r)\beta}{1+\beta} \\ &= \frac{1}{y} + \frac{\beta}{y} = \frac{1+\beta}{y} = \frac{1}{c_1^*} = \lambda. \quad \text{Interpretation?}\end{aligned}$$

# Static Optimization with Equality Constraints

**Homework:** Check that the Lagrangian in Example 6 is concave in  $c_1, c_2$  to convince yourself that the algorithm indeed produces a maximum.

**Homework:** What are the conditions on the model parameters (i.e. on  $y, r, \beta$ ) in Example 6 that ensure that consumption will be equal between the two periods? What are the conditions to have  $c_1 > c_2$ ? What's the economic intuition for that?

# Readings

## Main references:

Sydsæter et al. [SHSS] *Further mathematics for economic analysis*. Chapter 3.

## Additional readings:

Simon and Blume. *Mathematics for economists*. Chapters 17 and 18.