# R401: Statistical and Mathematical Foundations

## Lecture 16: Introduction to the Calculus of Variations. Isoperimetric Variational Problems.

Andrey Vassilev

2016/2017

# Lecture Contents

# An abstract look at optimization problems

# What is an optimization problem, really?

- You have already seen various optimization problems.
- They required finding maxima or minima of functions.
- These functions involved one or several variables.
- In some cases these variables were constrained by (systems of) equations or inequalities.

# What is an optimization problem, really?

- You have already seen various optimization problems.
- They required finding maxima or minima of functions.
- These functions involved one or several variables.
- In some cases these variables were constrained by (systems of) equations or inequalities.

**What are the common features of the optimization problems that we have encountered?**

# The constituents of an optimization problem

From the examples of optimization problems we can extract the following:

- An optimization problem requires a function to be maximized or minimized (the *objective function*).

- In the most abstract sense, a function $f$ is a rule of association (a mapping) between the elements of a set $X$ and the elements of another set $Y$:

  $$f : X \to Y.$$

  The sets $X$ (the *domain* of $f$) and $Y$ (the *codomain* of $f$) can be of arbitrary nature.

- Examples of functions under this broadened definition include:
  - The familiar functions like $f(x) = 2x^4$ are mappings from (subsets of) $\mathbb{R}^1$ to (subsets of) $\mathbb{R}^1$.
  - Functions of $n$ real variables are mappings from $\mathbb{R}^n$ to $\mathbb{R}^1$.
  - An $m \times n$ matrix with real entries can serve to define a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$ via post-multiplication with vectors $\mathbf{x} \in \mathbb{R}^n$.
  - The determinant maps the set of square matrices to the set of real numbers.
  - The definite integral can be viewed as a mapping from a suitable set of functions to the (extended) real numbers.

# The constituents of an optimization problem

- The set of functions with domain $X$ and codomain $Y$ is denoted by $Y^X$.
  - Thus, the well-known $\mathbb{R}^n$ can be reinterpreted as a space of functions from the set $\{1, 2, \ldots, n\}$ to the real numbers.
  - Similarly, the notation $\mathbb{R}^{m \times n}$ for the real-valued matrices with $m$ rows and $n$ columns tells us that a matrix can be viewed as a function from the Cartesian product of the sets $\{1, 2, \ldots, m\}$ and $\{1, 2, \ldots, n\}$ to $\mathbb{R}$.
- In order for a function to be suitable for use as an objective function, the elements of its codomain must be ordered (that is, we must be able to tell that one is "larger" or "better" than another).
- For practical purposes this means that we'll be optimizing functions having $\mathbb{R}^1$ as their codomain.
- In contrast, the domains of the objective functions can be more diverse types of sets, as long as we have appropriate methods to choose among the different elements.

# The constituents of an optimization problem

- The elements of the domain of an objective function can be subject to various requirements – the *constraints* of the optimization problem.
- Typically, these constraints take the form of other functions (with the same domain as the objective function), which are used to define various equations or inequalities.
- This idea remains intact irrespective of the types of sets that serve as domains of the constraint functions.

# Specific classes of optimization problems

- While the above considerations are fairly broad, there are two specific classes of situations that are of special relevance for economists:
  1. The sets we are optimizing over are sequences or even bundles of sequences.
  2. The sets we are optimizing over are functions defined on a subset of the real line.
- These two classes are especially interesting because they provide a natural way to treat dynamic decisions in an optimizing framework, e.g. a consumption stream in discrete time is a sequence of nonnegative numbers and in continuous time is a function defined on an interval $[0, T]$ or $[0, \infty)$.
- The importance of such optimization problems is not confined to dynamic decisions. They can also be useful in other situations, e.g. when modelling choices with infinitesimal influence of the individual agent.

# Specific classes of optimization problems

- While formulations may differ, in the vast majority of economic cases the objective function is defined either as a *series* involving the respective sequences or an *integral* involving the functions we are optimizing over.
- Constraints may also differ but in typical cases the functions or sequences we are optimizing over are required to satisfy *differential equations* or, respectively, *difference equations*.
- Depending on the details of the formulation, such problems are called either *calculus of variations (variational) problems* or *optimal control problems* (in discrete or continuous time).
- In this lecture we will study some of the basics of the calculus of variations.

# Basic variational problems

# From simple optimization to variational problems

The objective "functions" we'll be working with are called *functionals*. A functional is a quantity which is determined by the choice of one or more functions.

Example 1

The utility functional is a popular one in economics. Let us look at a time interval from a starting time $0$ up to a final time $T$ and let the consumption of an economic agent at time $t$ be denoted by $c(t) \geq 0$. Then the function $t \mapsto c(t)$ for $t \in [0, T]$ represents the consumption stream of the agent. Assume for simplicity that this function is continuous. The utility functional of the agent is defined as

$$\int_0^T e^{-\rho t} U(c(t))\, dt,$$

where $U$ is a given utility function and $\rho > 0$ is the time discount factor. One can try to maximize this functional by choosing from a given set of functions $c(t)$ defined on $[0, T]$ and taking values in $\mathbb{R}_+$.

# From simple optimization to variational problems

- In the familiar optimization problems the (implicit) approach is to see what would happen with the objective function if we vary the argument slightly. If we expect no change, then we are on a flat part of the graph, possibly an extremum.
- This is the rationale behind equating the first derivative to zero.
- In variational problems we again want to change the argument slightly and see what happens with the objective functional.
- This raises the question of how one can define "varying a function slightly".

**Note:** The considerations presented below are meant to be suggestive and are therefore informal. See e.g. the Elsgolts book for more details.

# From simple optimization to variational problems

- One way of "varying" a function, say $x : [t_0, t_1] \to \mathbb{R}$ which is $C^1$, is to take another $C^1$ function $\eta : [t_0, t_1] \to \mathbb{R}$ and construct the following:

$$x(t) + \alpha\eta(t), \quad \alpha \in \mathbb{R}.$$

- The resulting function can be viewed as a perturbed version of $x(t)$.
- For $\alpha = 0$ we obtain exactly $x(t)$ and as $\alpha \to 0$ the function is close (in some sense) to $x(t)$.
- This approach is convenient as it allows us to reduce the task of perturbing a function to that of varying the scalar $\alpha$.

# From simple optimization to variational problems

- One way of "varying" a function, say $x : [t_0, t_1] \to \mathbb{R}$ which is $C^1$, is to take another $C^1$ function $\eta : [t_0, t_1] \to \mathbb{R}$ and construct the following:

$$x(t) + \alpha\eta(t), \quad \alpha \in \mathbb{R}.$$

- The resulting function can be viewed as a perturbed version of $x(t)$.
- For $\alpha = 0$ we obtain exactly $x(t)$ and as $\alpha \to 0$ the function is close (in some sense) to $x(t)$.
- This approach is convenient as it allows us to reduce the task of perturbing a function to that of varying the scalar $\alpha$.

**We are now in a position to define a simple variational problem.**

# A basic variational problem

Take the class of $C^1$ functions $x : [t_0, t_1] \to \mathbb{R}$ for which the *boundary conditions* $x(t_0) = x_0$ and $x(t_1) = x_1$ are fulfilled.

Consider the functional

$$J(x) = \int_{t_0}^{t_1} F(t, x, \dot{x})\, dt, \tag{1}$$

where the function $F$ is assumed differentiable in its arguments, as many times as needed.

The basic variational problem consists in finding a maximum or a minimum of the functional $J(x)$ for functions in the above class.

# The Euler equation

# Preliminary considerations

- A first step in solving the basic variational problem is to find necessary conditions for optimality.
- These NCs will be the counterpart of the condition $f'(x) = 0$ for a function of one variable.
- The result of the procedure we will develop is called the *Euler equation*.
- The Euler equation is a differential equation whose solutions are *extremals*, i.e. candidates for optimality.

# The Euler equation: an informal derivation

Suppose that $x(t)$ is an admissible function for which the functional $J(x)$ attains, for instance, a (local) maximum. Then, the value of $J$ will be smaller for all admissible functions that are sufficiently close to $x(t)$.

To construct these "nearby" functions, we require that the perturbation $\eta$ fulfils the condition $\eta(t_0) = \eta(t_1) = 0$. In this case $x(t) + \alpha\eta(t)$ will also be admissible (it will meet the boundary conditions).

With $x(t)$ and $\eta(t)$ fixed, the functional $J(x(t) + \alpha\eta(t))$ can be viewed as a function of $\alpha$. The necessary condition for an extremum is that the first derivative w.r.t. $\alpha$ vanishes and we also know that the extremum obtains for $\alpha = 0$:

$$\left.\frac{d}{d\alpha}\right|_{\alpha=0} J(x(t) + \alpha\eta(t)) = 0.$$

# The Euler equation: an informal derivation

We have

$$\frac{d}{d\alpha}\int_{t_0}^{t_1} F\left(t, x(t)+\alpha\eta(t), \dot{x}(t)+\alpha\dot{\eta}(t)\right)\,dt =$$
$$\int_{t_0}^{t_1}\Big(F'_x(t, x(t)+\alpha\eta(t), \dot{x}(t)+\alpha\dot{\eta}(t))\eta(t)+$$
$$F'_{\dot{x}}(t, x(t)+\alpha\eta(t), \dot{x}(t)+\alpha\dot{\eta}(t))\dot{\eta}(t)\Big)\,dt.$$

# The Euler equation: an informal derivation

We have

$$\begin{aligned}
&\frac{d}{d\alpha}\int_{t_0}^{t_1} F\left(t, x(t)+\alpha\eta(t), \dot{x}(t)+\alpha\dot{\eta}(t)\right)\, dt = \\
&\int_{t_0}^{t_1}\Big(F'_x(t, x(t)+\alpha\eta(t), \dot{x}(t)+\alpha\dot{\eta}(t))\eta(t)+ \\
&\qquad F'_{\dot{x}}(t, x(t)+\alpha\eta(t), \dot{x}(t)+\alpha\dot{\eta}(t))\dot{\eta}(t)\Big)\, dt.
\end{aligned}$$

Since the resulting expression has to be evaluated at $\alpha = 0$, we obtain

$$\int_{t_0}^{t_1}\Big(F'_x(t, x(t), \dot{x}(t))\eta(t) + F'_{\dot{x}}(t, x(t), \dot{x}(t))\dot{\eta}(t)\Big)\, dt.$$

# The Euler equation: an informal derivation

We have

$$\frac{d}{d\alpha}\int_{t_0}^{t_1} F\left(t, x(t) + \alpha\eta(t), \dot{x}(t) + \alpha\dot{\eta}(t)\right)\, dt =$$

$$\int_{t_0}^{t_1}\Bigg(F'_x(t, x(t) + \alpha\eta(t), \dot{x}(t) + \alpha\dot{\eta}(t))\eta(t) +$$

$$F'_{\dot{x}}(t, x(t) + \alpha\eta(t), \dot{x}(t) + \alpha\dot{\eta}(t))\dot{\eta}(t)\Bigg)\, dt.$$

Since the resulting expression has to be evaluated at $\alpha = 0$, we obtain

$$\int_{t_0}^{t_1}\Bigg(F'_x(t, x(t), \dot{x}(t))\eta(t) + F'_{\dot{x}}(t, x(t), \dot{x}(t))\dot{\eta}(t)\Bigg)\, dt.$$

We next integrate the second term by parts.

# The Euler equation: an informal derivation

$$\int_{t_0}^{t_1} \left( F'_x(t, x(t), \dot{x}(t))\eta(t) + F'_{\dot{x}}(t, x(t), \dot{x}(t))\dot{\eta}(t) \right) dt =$$

$$\int_{t_0}^{t_1} F'_x(t, x(t), \dot{x}(t))\eta(t)\, dt + \int_{t_0}^{t_1} F'_{\dot{x}}(t, x(t), \dot{x}(t))\, d\eta(t) =$$

$$\int_{t_0}^{t_1} F'_x(t, x(t), \dot{x}(t))\eta(t)\, dt + \underbrace{F'_{\dot{x}}(t, x(t), \dot{x}(t))\eta(t)\Big|_{t_0}^{t^1}}_{=0 \text{ since } \eta(t_0)=\eta(t_1)=0} - \int_{t_0}^{t_1} \eta(t)\, dF'_{\dot{x}}(t, x(t), \dot{x}(t)) =$$

$$\int_{t_0}^{t_1} F'_x(t, x(t), \dot{x}(t))\eta(t)\, dt - \int_{t_0}^{t_1} \eta(t) \frac{d}{dt} F'_{\dot{x}}(t, x(t), \dot{x}(t))\, dt =$$

$$\int_{t_0}^{t_1} \left( F'_x(t, x(t), \dot{x}(t)) - \frac{d}{dt} F'_{\dot{x}}(t, x(t), \dot{x}(t)) \right) \eta(t)\, dt.$$

# The Euler equation: an informal derivation

Recall that the expression we obtained should be equal to zero:

$$\int_{t_0}^{t_1} \left( F'_x(t, x(t), \dot{x}(t)) - \frac{d}{dt} F'_{\dot{x}}(t, x(t), \dot{x}(t)) \right) \eta(t)\, dt = 0.$$

There is one additional complication to overcome: the presence of $\eta(t)$ under the integral sign.

The following result comes to our rescue:

**Fact 1 (The fundamental lemma of the calculus of variations)**

If a continuous function $\Phi : [t_0, t_1] \to \mathbb{R}$ is such that

$$\int_{t_0}^{t_1} \Phi(t)\eta(t)\, dt = 0$$

for all $C^1$ functions $\eta : [t_0, t_1] \to \mathbb{R}$ with $\eta(t_0) = \eta(t_1) = 0$, then $\Phi(t) \equiv 0$.

# The Euler equation: an informal derivation

Applying Fact 1, we obtain

$$F'_x(t, x(t), \dot{x}(t)) - \frac{d}{dt} F'_{\dot{x}}(t, x(t), \dot{x}(t)) = 0. \tag{2}$$

This differential equation is the celebrated *Euler equation*.

**Note:** The Euler equation is sometimes called the Euler-Lagrange equation. Here we stick to the shorter form.

# More on the Euler equation

- The Euler equation is a second-order differential equation.
- Sometimes it is written in abbreviated form as

$$F'_x = \frac{d}{dt} F'_{\dot{x}}.$$

- However, it may be more instructive to write it out in full:

$$\begin{aligned} F'_x(t, x(t), \dot{x}(t)) = & F''_{t\dot{x}}(t, x(t), \dot{x}(t)) + F''_{x\dot{x}}(t, x(t), \dot{x}(t))\dot{x}(t) + \\ & F''_{\dot{x}\dot{x}}(t, x(t), \dot{x}(t))\ddot{x}(t). \end{aligned}$$

- As indicated, the Euler equation provides NCs for optimality of variational problems.
- However, using it to obtain extremals can be more complicated since finding a solution does not involve a standard initial-value (Cauchy) problem for an ODE. Instead, the conditions are imposed on both ends of the interval on which the solution is to be determined. Such problems are called *boundary value problems*.

# Boundary value problems

# Isoperimetric variational problems

# The basic isoperimetric problem

Isoperimetric problems historically derive their name from the geometric problem of finding a figure with maximal area for a given perimeter. Today the term is used to describe a wide range of variational problems subject to constraints.

In its simplest form, an isoperimetric variational problem is to find an extremum of the functional

$$\int_{t_0}^{t_1} F(t, x, \dot{x})\, dt$$

s.t.

$$\int_{t_0}^{t_1} F_i(t, x, \dot{x})\, dt = l_i, \quad i = 1, \ldots, m,$$

where $l_i$ are given constants and

$$x(t_0) = x_0, \quad x(t_1) = x_1.$$

# Solving the basic isoperimetric problem

The solution recipe for the isoperimetric problem is the following:

## Algorithm

1. Form the auxiliary functional

$$\int_{t_0}^{t_1} \left( F(t,x,\dot{x}) + \sum_{i=1}^{m} F_i(t,x,\dot{x}) \right) dt,$$

   where $\lambda_i$ are constants.
2. Write the Euler equation for it.
3. Use the isoperimetric conditions $\int_{t_0}^{t_1} F_i(t,x,\dot{x})\,dt = l_i$ and the boundary conditions $x(t_i) = x_i,\ i = 0, 1,$ to determine the constants in the equation and $\lambda_i$.

# Solving the basic isoperimetric problem

Remarks

- As a shortcut for step 1) of the algorithm, form directly the Lagrangian

$$\mathfrak{L} = F(t, x, \dot{x}) + \sum_{i=1}^{m} F_i(t, x, \dot{x}).$$

Then write the Euler equation for it as

$$\frac{\partial \mathfrak{L}}{\partial x} + \frac{d}{dt}\frac{\partial \mathfrak{L}}{\partial \dot{x}} = 0.$$

- There are more complex formulations of the isoperimetric problem involving several unknown functions (see the Elsgolts book), i.e.

$$\int_{t_0}^{t_1} F(t, x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)\, dt$$

s.t.

$$\int_{t_0}^{t_1} F_j(t, x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)\, dt = l_j, \quad j = 1, \ldots, m,$$

$$x_i(t_0) = x_{i,0}, \quad x_i(t_1) = x_{i,1}, \quad i = 1, \ldots, n.$$

# Solving the basic isoperimetric problem

Remarks

- The algorithm provides necessary conditions, as evidenced by the use of the Euler equation.
- Whether the candidates are minimizing or maximizing extremals (if at all) should be established separately.
- This depends on the structure of the problem e.g. concavity or convexity of the objective functional.

# Example: Optimal distribution for a continuous aggregator

Consider the problem

$$\left| \begin{array}{l} \min_{c(j)} \ \int_0^1 p(j)c(j)dj \\ \\ \int_0^1 c(j)^{\frac{\theta-1}{\theta}} dj = C^{\frac{\theta-1}{\theta}} \end{array} \right. ,$$

where $p(j) > 0$ are given prices, and $\theta > 1$ and $C > 0$ are constants.

The solution $c(j)$ of this problem is given by the formula

$$c(j) = \left(\frac{p(j)}{P}\right)^{-\theta} C,$$

with $P := \left(\int_0^1 p(j)^{1-\theta} dj\right)^{\frac{1}{1-\theta}}$. Also

$$\int_0^1 p(j)c(j)dj = PC.$$

# Example: Optimal distribution for a continuous aggregator

We set up

$$\mathfrak{L} = p(j)c(j) + \lambda c(j)^{\frac{\theta-1}{\theta}}.$$

The Euler equation becomes $\partial\mathfrak{L}/\partial c = 0$, which takes the form

$$p(j) + \lambda \frac{\theta-1}{\theta} c(j)^{-\frac{1}{\theta}} = 0.$$

Therefore,

$$p(j) = \left(-\lambda \frac{\theta-1}{\theta}\right) c(j)^{-\frac{1}{\theta}},$$

so

$$c(j) = \left(-\lambda \frac{\theta-1}{\theta}\right)^{\theta} p(j)^{-\theta}. \tag{3}$$

# Example: Optimal distribution for a continuous aggregator

Substitute this into the isoperimetric constraint

$$\left(-\lambda\frac{\theta-1}{\theta}\right)^{\theta-1}\int_0^1 p(j)^{1-\theta}dj = C^{\frac{\theta-1}{\theta}},$$

i.e.

$$\left(-\lambda\frac{\theta-1}{\theta}\right)P^{-1} = C^{\frac{1}{\theta}}.$$

Then (3) implies

$$c(j) = \left(PC^{\frac{1}{\theta}}\right)^{\theta} p(j)^{-\theta} = \left(\frac{p(j)}{P}\right)^{-\theta} C.$$

Finally, we verify that

$$\int_0^1 p(j)c(j)dj = \int_0^1 p(j)p(j)^{-\theta}P^{\theta}Cdj = CP^{\theta}P^{1-\theta} = PC.$$

# Readings

Main references:
Elsgolts. *Differential equations and the calculus of variations*. Chapters 6 and 9.

Additional readings:
Sydsæter et al. [SHSS] *Further mathematics for economic analysis*. Chapter 8.