Optimization

A gentle introduction

Vittorio Maniezzo
Univ. Bologna

1

Administrativia

ME
Vittorio Maniezzo
WEB
https://www.cs.unibo.it/maniezzo
EMAIL
vittorio.maniezzo@unibo.it
IOL
https://iol.unibo.it/course/view.php?id=50366
FACEBOOK
Business Analytics Unibo
DROPBOX
https://www.dropbox.com/sh/cxa4qospvbtv2ex/AAD8xpdulI8pK1WB48omXxPga?dl=0
ANSWERS
http://bit.ly/39wrBkM

2

Administrativia

- *Prerequisites*: no formal ones, but some of you will already know all what I tell (possibly better than the way I will present it) and others maybe not. ASK
- *BYOD!* I will propose and run examples (windows, excel, python). You are too many to see what's wrong if you use different configurations.
- *Programming:* python.
- *Algorithms:* basic data structures, complexity
- *Math:* basic calculus and matrices
- *Statistics:* basic descriptive statistics, normal distribution.

Administrativia

Final proof

Mine is just a module.

I will propose projects for groups of 1, 2 or 3 students.

Projects will be checked with anti-plagiarism SW.

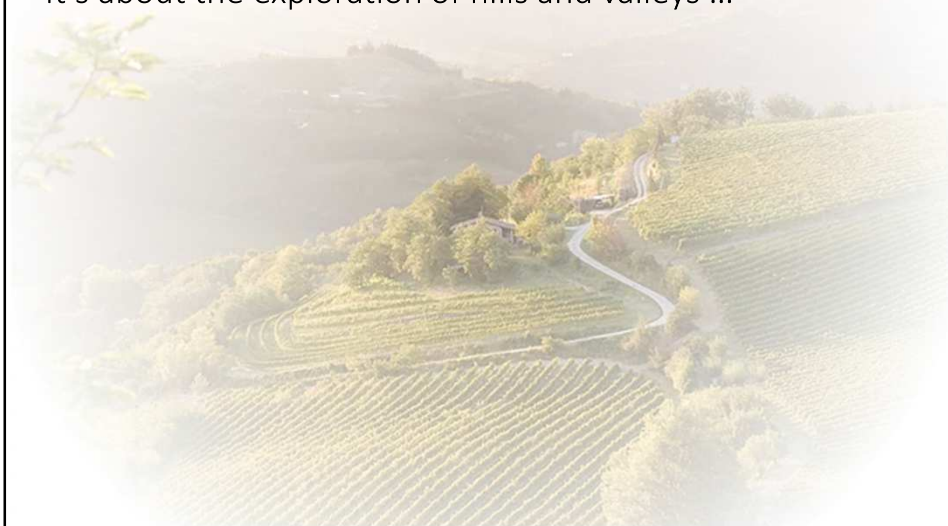
Each group must have its project accepted before starting to work.

Classroom participation will be positively evaluated.

Taking part – even unfocusedly – to the lessons will not have bad impact. In no case.

Optimization

It’s about the exploration of hills and valleys …



5

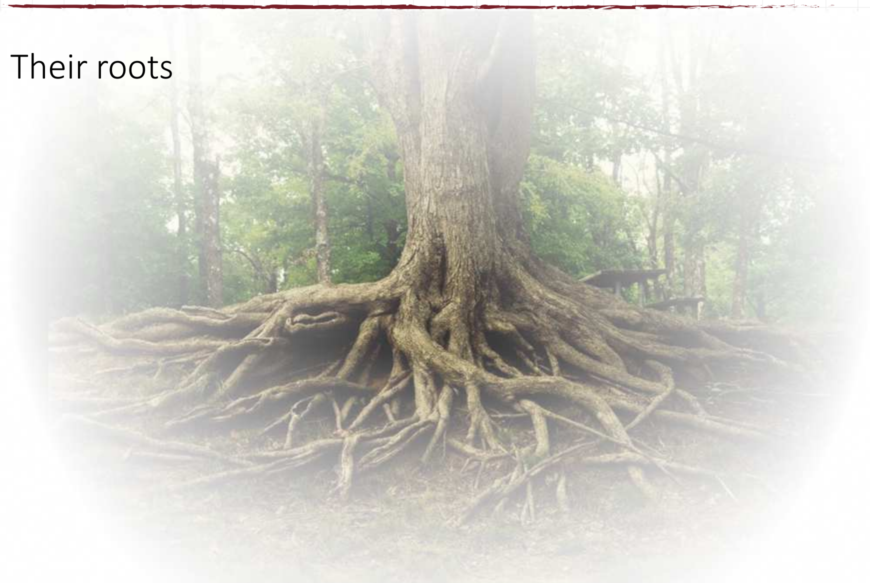
Optimization

It’s about studying trees

6

Optimization

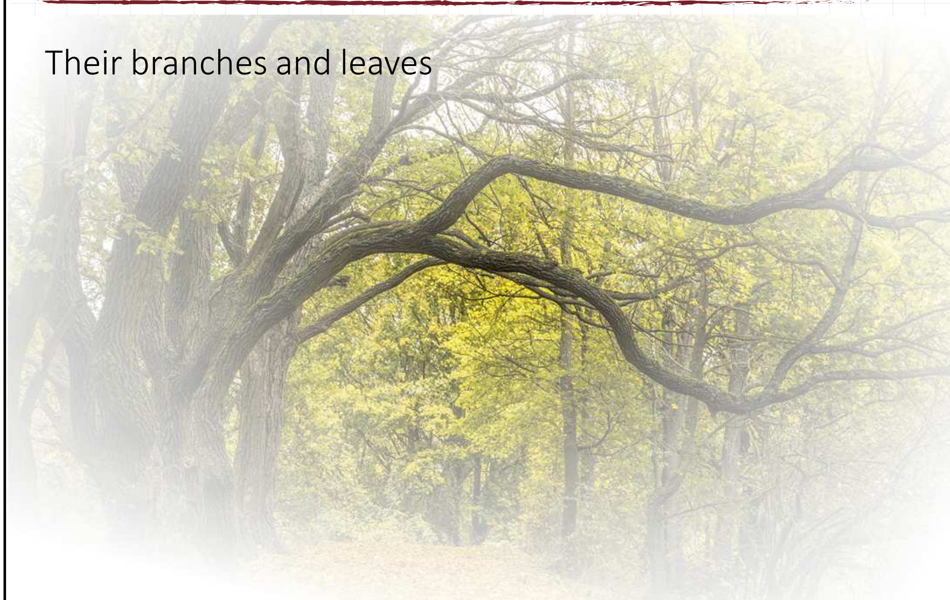
Their roots



7

Optimization

Their branches and leaves



8

Optimization

Their blossoms too!

Vittorio Maniezzo - University of Bologna 9

9

Optimization

It's about studying wonderful shapes

Vittorio Maniezzo - University of Bologna 10

10

Optimization

Mathematical optimization (aka *mathematical programming*) is the selection of a ***best*** element from some set of available alternatives.

An optimization problem consists of *minimizing* (*maximizing*) a real function by choosing input values from within an allowed set.

More generally, optimization includes finding ***best available*** values of some *objective function*, given a defined *domain* (or input), where objective functions and domains can be very diverse.

(adapted from Wikipedia)

Optimization problem

An optimization problem can be represented in the following way:

Given: a function $f : \mathrm{A} \to \mathbb{R}$ from some set A to the real numbers

Find: an element $x_0 \in \mathrm{A}$

- such that $f(x_0) \leq f(x)$ for all $x \in \mathrm{A}$ ("*minimization*") or
- such that $f(x_0) \geq f(x)$ for all $x \in \mathrm{A}$ ("*maximization*").

Subfields

Major subfields

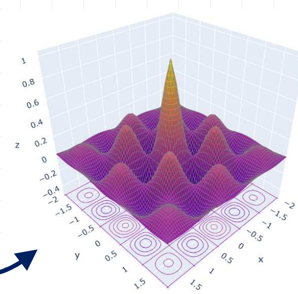
- Linear programming (LP): the objective function f is linear and the constraints are specified using only linear equalities and inequalities.
- Nonlinear programming: the objective function and/or the constraints can be nonlinear.
- Quadratic programming: the objective function can have quadratic terms, while the feasible set must be specified with linear equalities and inequalities.
- Integer programming: linear programs in which some or all variables are constrained to take on integer values (this is in general much more difficult than regular linear programming).
- Combinatorial optimization: problems where the set of feasible solutions is discrete or can be reduced to a discrete one.
- Stochastic programming: some of the constraints or parameters depend on random variables.
- Robust programming: tries to capture uncertainty in the data, finding solutions that are valid under all / a wide set of possible realizations of the uncertainties.
- Constraint programming: a programming paradigm where relations between variables are stated in the form of constraints.

. . . and many more . . . *(Adapted from wikipedia)*

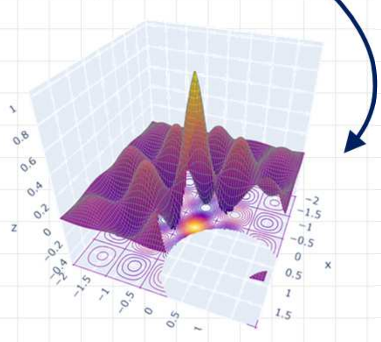
Nonlinear optimization

The objective function or the constraints or both can have nonlinear components.

Unconstrained



Constrained



Example:

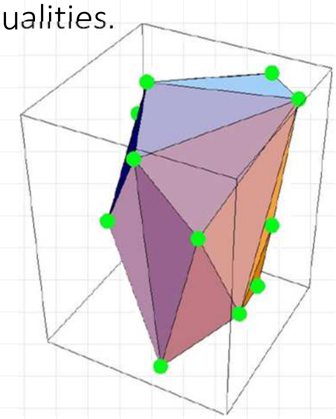
$$\begin{array}{ll} min & x^2\text{-}2x+3 \\ s.t. & x^2 \leq 9 \end{array}$$

Linear optimization

- Both the objective function and the constraints are modelled by linear functions.
- Constraints (there have to be some) are in the form of a system of equations or inequalities.

Example:

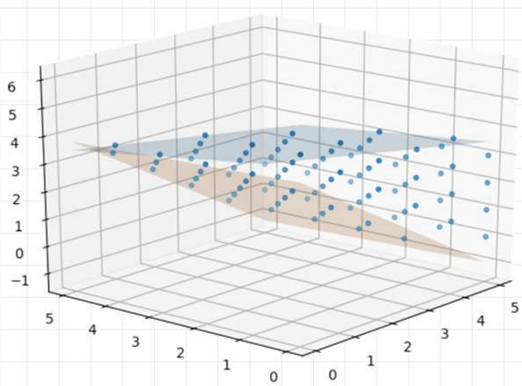
$$\begin{array}{llll} min & -x_1 & -x_2 & \\ s.t. & x_1 & -x_2 & \leq 1.5 \\ & -7x_1 & +5x_2 & \leq -17.5 \\ & & -x_2 & \leq -4.5 \\ & x_1, x_2 \geq 0 & & \end{array}$$



Integer linear optimization

Linear optimization problems where (some of) the variables are required to take on integer values.

- If all variables are integer: Integer Linear Programming (ILP)
- If some of the variables are continuous: Mixed Integer Linear programming (MILP).



Example:

$$\begin{array}{llll} min & -x_1 & -x_2 & \\ s.t. & x_1 & -x_2 & \leq 1.5 \\ & -7x_1 & +5x_2 & \leq -17.5 \\ & & -x_2 & \leq -4.5 \\ & x_1, x_2 \geq 0 \text{ and integer} & & \end{array}$$

Some math we need

Table of derivatives

$y = f(x)$	$\frac{dy}{dx} = f'(x)$
k, any constant	0
x	1
x^2	$2x$
x^3	$3x^2$
x^n, any constant n	nx^{n-1}
e^x	e^x
e^{kx}	ke^{kx}
$\ln x = \log_e x$	$\frac{1}{x}$
$\sin x$	$\cos x$
$\sin kx$	$k \cos kx$
$\cos x$	$-\sin x$
$\cos kx$	$-k \sin kx$
$\tan x = \frac{\sin x}{\cos x}$	$\sec^2 x$
$\tan kx$	$k \sec^2 kx$
$\operatorname{cosec} x = \frac{1}{\sin x}$	$-\operatorname{cosec} x \cot x$

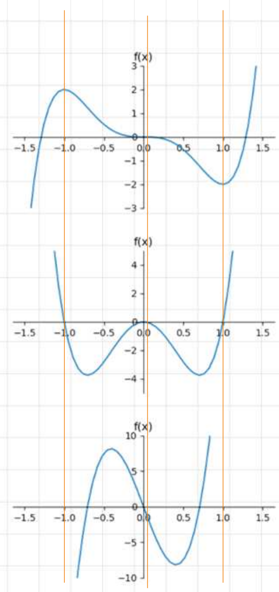
$y = f(x)$	$\frac{dy}{dx} = f'(x)$
$\sec x = \frac{1}{\cos x}$	$\sec x \tan x$
$\cot x = \frac{\cos x}{\sin x}$	$-\operatorname{cosec}^2 x$
$\sin^{-1} x$	$\frac{1}{\sqrt{1-x^2}}$
$\cos^{-1} x$	$\frac{-1}{\sqrt{1-x^2}}$
$\tan^{-1} x$	$\frac{1}{1+x^2}$
$\cosh x$	$\sinh x$
$\sinh x$	$\cosh x$
$\tanh x$	$\operatorname{sech}^2 x$
$\operatorname{sech} x$	$-\operatorname{sech} x \tanh x$
$\operatorname{cosech} x$	$-\operatorname{cosech} x \coth x$
$\coth x$	$-\operatorname{cosech}^2 x$
$\cosh^{-1} x$	$\frac{1}{\sqrt{x^2-1}}$
$\sinh^{-1} x$	$\frac{1}{\sqrt{x^2+1}}$
$\tanh^{-1} x$	$\frac{1}{1-x^2}$

… and so on

Some math we need

Let $f(x)$ be a real-valued, twice differentiable function. The value of x for which the first derivative $f'(x)$ is 0 corresponds to a maximum or a minimum of $f(x)$.

- For a maximum the second derivative $f''(x)$ is negative.
- For a minimum the second derivative $f''(x)$ is positive.
- The second derivative is 0 for an inflexion point.



Some math we need

Jacobian

The Jacobian operator is a generalization of the derivative operator to vector-valued functions. The Jacobian matrix of a vector-valued function f is the matrix of all its first-order partial derivatives.

The Jacobian matrix represents the differential of f at every point where f is differentiable.

$$\mathrm{J} = \left[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right]$$

Throughout these slides, first order derivatives of $f(\mathrm{x})$ will be denoted either as $f'(\mathrm{x})$, or as $\mathrm{J}(\mathrm{x})$, or as $\nabla(\mathrm{x})$, according to my whims.

Some math we need

Hessian

The Hessian matrix is a square matrix of second-order partial derivatives of a function f.

It describes the local curvature of a function of many variables.

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

The trace of the Hessian matrix is known as the Laplacian operator

Throughout these slides, second order derivatives of $f(\mathrm{x})$ will be denoted either as $f''(\mathrm{x})$, or as $\mathrm{H}(\mathrm{x})$, or as $\nabla^2(\mathrm{x})$, according to my whims.

Some math we need

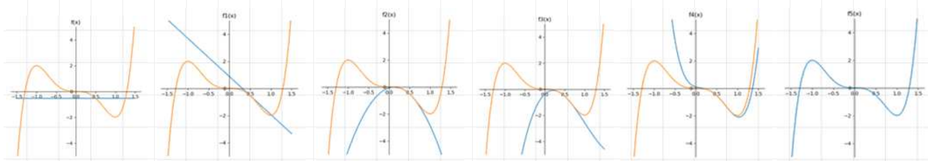
A Taylor series is a representation of an infinitely differentiable function as an infinite sum of terms that are calculated from the values of the function's derivatives at a point c.

$$f(x) = f(c) + \frac{f'(c)}{1!}(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \cdots = \sum_{n=0}^{\infty} \frac{f^n(c)}{n!}(x-c)^n$$

(The derivative of order zero of f is defined to be f itself and $(x-c)^0$ and $0!$ are both defined to be 1)

The function can be *approximated* by using a finite number of terms of its Taylor series $\bar{f}(x) = \sum_{n=0}^{\bar{n}} \frac{f^n(c)}{n!}(x-c)^n$

Ex. $f(x) = 3x^5 - 5x^3$ with $c = 0.5$



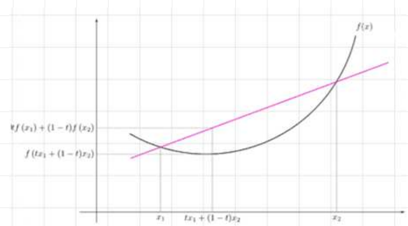
Some math we need

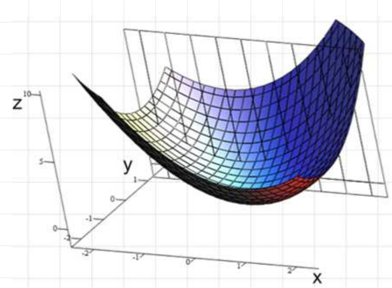
[*Wikipedia*] A real-valued function defined on an n-dimensional interval is called convex (*convex downward* or *concave upward*) if the line segment between any two points on the graph of the function lies above or on the graph.

For a twice-differentiable function of a single variable, if its second derivative is always nonnegative.

Given a convex set X, a function $f: X \to \mathbb{R}$ is called convex if

$$\forall x_1, x_2 \in X, \forall t \in [0,1]: f(tx_1 + (1-t)x_2) \le tf(x_1) + (1-t)f(x_2)$$





Some math we need

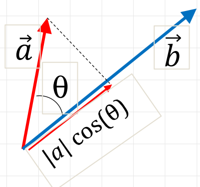
Vector dot product.

The *inner* (dot, scalar) product is an algebraic operation that takes two equal-length sequences of numbers (often coordinate vectors) and returns a single number.

(1,3,5)·(2,4,6)=(1×2+3×4+5×6)=(2+12+30)=44

Dot Product of two vectors:

- magnitudes and angle $\vec{a} \cdot \vec{b} = |a| \times |b| \times cos(\theta)$
- components $\vec{a} \cdot \vec{b} = a_x \times b_x + a_y \times b_y$

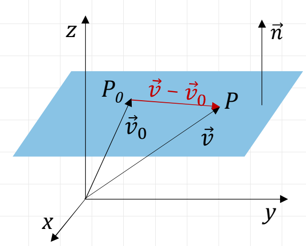


The inner product of orthogonal vectors is 0.

A vector normal to a plane is orthogonal to any vector lying on the plane.

Dot product, cont

A plane is determined by a point P_0 on the plane and a vector normal to the plane, $\vec{n}$.
P_0's position is determined by vector $\vec{v}_0$, a generic point's P by a vector $\vec{v}$.
The equation $\vec{n} \cdot (\vec{v} - \vec{v}_0)$=0 holds, since the difference vector $\vec{v} - \vec{v}_0$ lies in the plane.
We can derive the equation of the plane.
Consider the vectors components:
$\vec{n}$ =[a,b,c], $\vec{v}$ =[x,y,z] and $\vec{v}_0$ =[x_0,y_0,z_0]
Dot product orthogonality $\vec{n} \cdot (\vec{v} - \vec{v}_0)$=0 gives:
$a(x-x_0)+b(y-y_0)+c(z-z_0)=0$



Defining $d=x_0+y_0+z_0$, then the previous equation becomes:
$ax+by+cz=d$
Which is the equation of the plane.
We can thus identify the normal vector $\vec{n}$ directly from the coefficients (a,b,c) of the scalar equation of the plane.