Unit 1. Introduction to Operational Research. Non linear optimization

Jordi Villà i Freixa

Universitat de Vic - Universitat Central de Catalunya
Study Abroad. Operations Research

jordi.villa@uvic.cat

14-28/02, 2023

Preliminary

This course is strongly based on the monography on Operations Research by Carter, Price and Rabadi [Carter, 2019], and in material obtained from different sources (quoted when needed through the slides).

Learning outcomes

- Learn about the origins and applications of Operations Research
- Understand system modelling principles
- Understand algorithm efficiency and problem complexity
- Contrast between the optimality and practicality
- Learn about software for operations Research
- Introduction to the Python/Colab environment

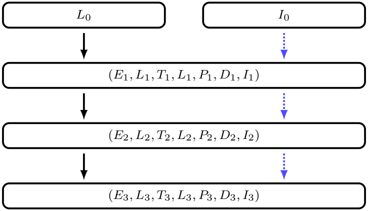
Summary

1. Nonlinear optimization
   - Concavity/convexity
   - Iterative methods
   - Constrained optimization

Example

Production cost

A company aims at minimizing the production cost during a series of production periods of time T_i, characterised by the demand $D_i i$, equipment capacities and material limitations E_i, labor force L_i (which cost depends quadratically on the difference of labor force between two periods $C_L(L_i - L_{i-1})^2$), productivity of each worker P_i, the number of units of inventory at the end of each production period I_i and the cost to bring them to the next production period C_I.



Example

The problem, as stated, aims at minimizing the function:

$$f(\vec{L}, \vec{I}) = \sum_{i=1}^{T} C_L (L_i - L_{i-1})^2 + C_I I_i$$

subject to:

$$\begin{cases} L_i P_i \leq E_i \\ I_{i-1} + L_i P_i \geq D_i \\ I_i = I_{i-1} + L_i P_i - D_i \\ L_i, I_i \geq 0, \ \forall i = 1, \ldots, T \end{cases}$$

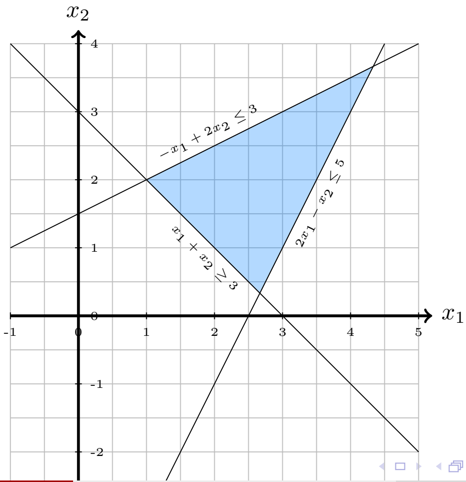
Quadratic objective function with linear constraints.

Introduction

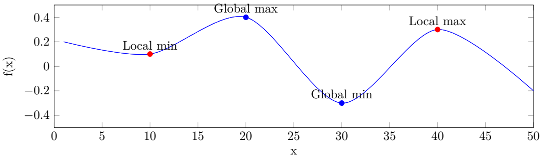
- We want to obtain the best solution to a mathematical programming problem in which both objective function and constraints have general non-linear forms.
- Most of the problems are non-linear indeed!
  - Unconstrained Problems (often dealt with differential calculus)
  - Constrained Problems (may include systems of equations to be solved)
- Classical underlying mathematical theories do not necessarily provide practical methods suitable for efficient numerical computation.
- Points of optimality can be anywhere inside the problem boundaries
- No methods applicable to all non-linear problems

Feasible region

Set of points satisfying all the constraints (the area between constraint boundaries).



Global and local extremes



A local maximum of the function $f(x)$ exists in x^* if there is a small positive number ϵ such that

$$f(x^*) > f(x),\ \forall x \in \mathbf{R} : \|x - x^*\| < \epsilon$$

A global maximum of $f(x)$ exists in x^* if

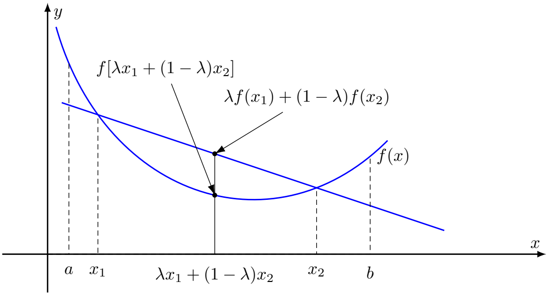
$$f(x^*) > f(x),\ \forall x \in \mathbf{R}$$

(analogous definitions for local/global minimum)

For a continuous convex function, given any two points x_1 and x_2:

$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda)f(x_2),\ 0 \leq \lambda \leq 1 \tag{1}$$

(analogous situation for a continuous concave function)



Convexity

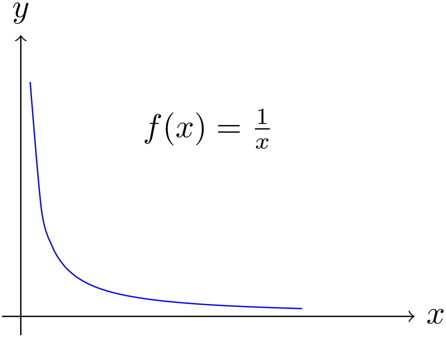
The term "convex" can be applied both to sets and functions. A set $S \in \mathbf{R}^n$ is a *convex set* if the straight line segment connecting any two points in S lies entirely inside S.

Note that $f(x)$ is a convex function if its domain S is a convex set and if for any two points x_1 and x_2 in S Eq. 1 holds.

Exercise 1

Can you draw a convex function that depends on two variables, $f(x, y)$? Can you generalize Eq. 1?

- If a convex nonlinear function is to be optimized without constraints, a global minimum may occur when $f'(x) = 0$. Not always this is the case:



- If the feasible region for a nonlinear programming problem is convex, each of the constraint functions is convex and are of the form $g_i(x) \leq b_i$

Exercise 2

Is the function $f(x) = x^2$ convex? Are the regions defined by $x^2 = 4$ or $x^2 \geq 9$ convex?

- A local minimum is guaranteed to be a global minimum for a convex objective function in a convex feasible region, and
- a local maximum is guaranteed to be a global maximum for a concave objective function in a convex feasible region.

Many functions in nonlinear programming problems are neither concave nor convex!

The Hessian matrix I

$$(\mathrm{Hess}\, f)_{ij} \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} \tag{2}$$

$$\mathrm{Hess}\left(\frac{x^2}{y}\right) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix} \tag{3}$$

Let f be a function of n variables and let $1 \leq i \leq n$ and $1 \leq j \leq n$. If the partial derivatives f_i' and f_j' of f exist in an open set S containing $(x_1, \ldots, x_n)$ and both of these partial derivatives are differentiable at $(x_1, \ldots, x_n)$ then $(\mathrm{Hess}\, f)_{ij} = (\mathrm{Hess}\, f)_{ji}$.

- If the Hessian at a given point has all positive eigenvalues, it is said to be a positive-definite matrix. This is the multivariable equivalent of “concave up” (we called it ”convex” in this course).

FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES

The Hessian matrix II

- If all of the eigenvalues are negative, it is said to be a negative-definite matrix. This is like “concave down”. (we call it ”concave” in this course).
- It the eigenvalues are mixed, you have a saddle point.

Exercise 3

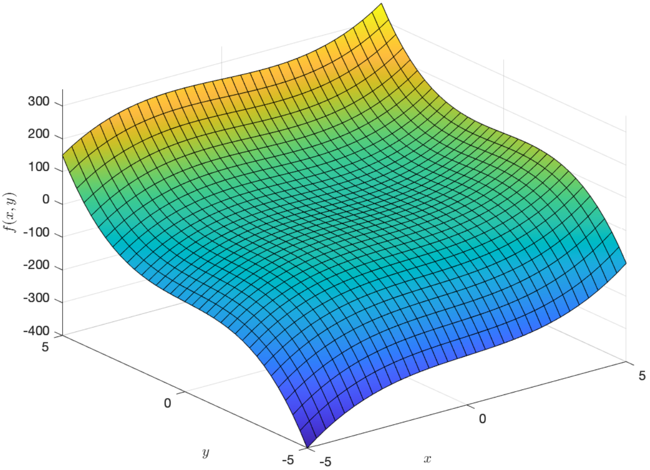
Study the critical points of the function $f(x, y) = 4x + 2y - x^2 - 3y^2$

Exercise 4

Check the character of the stationary point in $(0, 0)$ for the functions:

- $f(x, y) = x^2 + y^2$
- $g(x, y) = -x^2 - y^2$
- $h(x, y) = x^2 - y^2$
- $k(x, y) = x^3 + 2y^3 - xy$ (is it concave or convex in $(3, 3)$? what about in $(-3, -3)$?)

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



FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES

Unconstrained optimization of continuous $f : \mathbf{R} \to \mathbf{R}$

For unconstrained problems with just one variable x, if first and second derivatives exist in x^*:

Necessary conditions If $\frac{df}{dx} = 0$ at $x = x^*$

- $\frac{d^2 f}{dx^2} \geq 0$, for a local minimum at $x = x^*$
- $\frac{d^2 f}{dx^2} \leq 0$, for a local maximum at $x = x^*$

Sufficient conditions If $\frac{df}{dx} = 0$ at $x = x^*$

- $\frac{d^2 f}{dx^2} > 0$, for a local minimum at $x = x^*$
- $\frac{d^2 f}{dx^2} < 0$, for a local maximum at $x = x^*$

Unconstrained optimization of continuous $f : \mathbf{R}^n \to \mathbf{R}$ I

For unconstrained problems with several variables $\mathbf{x} = (x_1, \ldots, x_n)$:

Sufficient condition If $\frac{\partial f}{\partial x_i} = 0$ at $\mathbf{x} = \mathbf{x}^*$:

- $\nabla f = \mathbf{0}$, for a local minimum at $x = x^*$ if the function is convex;
- $\nabla f = \mathbf{0}$, for a local maximum at $x = x^*$ if the function is concave.

Decision conditions The function f is a convex function if H_f is positive definite or positive semidefinite for all $\mathbf{x}$; and f is concave if H_f is negative definite or negative semidefinite for all $\mathbf{x}$. For $f : \mathbf{R}^2 \to \mathbf{R}$

$$\mathbf{H}_f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} A & C \\ C & B \end{pmatrix}$$

FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES

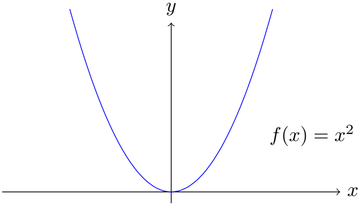
Unconstrained optimization of continuous $f : \mathbf{R}^n \rightarrow \mathbf{R}$ II

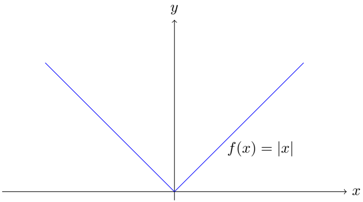
Let f be a twice-differentiable function of many variables on the convex open set S and denote the Hessian of f at the point x by $H(x)$. Then

- f is concave if and only if $H(x)$ is negative semidefinite $\forall x \in S$
- if $H(x)$ is negative definite $\forall x \in S$ then f is strictly concave (see Eq. 1)
- f is convex if and only if $H(x)$ is positive semidefinite $\forall x \in S$
- if $H(x)$ is positive definite $\forall x \in S$ then f is strictly convex (see Eq. 1)

Continuity vs derivability

Things are not so simple if, for example, functions are not derivable in x^*:





Or, if the function is not convex (nor concave), the sufficient condition is lost and we may find local optimal points, instead of global.
For constrained optimization problems, the shape of the feasible region adds difficulty.

One-Dimensional line search I

Let us assume a concave function. First, establish error tolerance ϵ. Then:

1. Set up lower and upper bounds for x by finding x_l, x_u such that $\frac{df}{dx_l} \geq 0$ and $\frac{df}{dx_u} \leq 0$
2. Get new trial intermediate solution (eg, $x = \frac{x_l + x_u}{2}$)
3. If $x_u - x_l \leq \epsilon$, the terminate
4. If $\frac{df}{dx} \geq 0$, set $x_l = x$; if $\frac{df}{dx} \leq 0$, set $x_u = x$
5. Go to step 1

For an implementation, see this link

Golden search method I

For a continuous function it guarantees to find an extremum.

Golden ratio

Two quantities are in the golden ratio if their ratio is the same as the ratio of their sum to the larger of the two quantities:

$$\frac{a+b}{a} = \frac{a}{b} = \psi \Rightarrow \psi = \frac{1+\sqrt{5}}{2} = 1.616033988...$$

Procedure of the Golden search method, given a function to be minimized, $f(x)$, the interval to be searched as $x \in (x_1, x_4)$, and their functional values $f(x_1)$ and $f(x_4)$:

1. Calculate an interior point and its functional value $f(x_2)$. To get the value of x_2, ensure the interval lengths are in the ratio $c : r$ or $r : c$ where $r = \phi 1$; and $c = 1r$, with ϕ being the golden ratio.

Golden search method II

2. Using the triplet, determine if convergence criteria are fulfilled. If they are, estimate the x at the minimum from that triplet and terminate/return.
3. From the triplet, calculate the other interior point and its functional value.
4. The three points for the next iteration will be the one where $f(x)$ is a minimum, and the two points closest to it in x.
5. Go to step 2

Exercise 5

Use the Golden Search method to find the minimum of function $f(x) = x^2 - 6x + 15$ in the interval $x \in (0, 10)$. Set up a spreadsheet to do so.

FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES UVIC | UVIC·UCC

E1. NLO program

Programming exercise

E2. Build a Python code that can run one-dimensional search algorithms for user defined functions, using or not first derivatives. Implement the one-dimensional line search and the golden search algorithms and find the optimal solution for the functions:

- $f(x) = x^4 - 16x^3 + 45x^2 - 20x + 203$ *within the range* $x \in (2.5, 14)$
- $g(x) = x^5 - 2x^4 - 23x^3 - 12x^2 + 36x$ *within the range* $x \in (2, 3)$

Try both a direct implementation as well as the use of solvers in scipy.

Algorithms for unconstrained optimization

As a reference (not the central aim of this course) here is a list of some algorithms/methods to explore to perform unconstrained optimization in multiple dimensions:

- Newton's method,
- Quasi-Newton methods (BFGS, SRI,...),
- Trust-region methods,
- Conjugated gradient methods.

and their many variants.

In the previous section we have considered functions of several variables that are independent from each other. What happens if the variables are, however, not independent from each other? (see excellent material in Richard Weber's web site)

What is the maximum height we can attain constrained to the trajectory of the road?

Lagrange multipliers I

Let us consider optimizing a function $f(x, y)$ (the height of the hill in the previous exemple) subject to the constrain $g(x, y) = c$ (the equation describing the path, the road).
One can simply substitute $g(x, y) = c$ in $f(x, y)$, but this may not involve trivial algebra.
Alternatively, let us consider the method of the Lagrange multipliers. To optimize $f(x, y)$ we require

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = 0$$

we saw that if dx and dy were independent, $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$. But they are not here: they are constrained by $g(x, y)$, which is a constant:

$$dg = \frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy = 0$$

FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES

Lagrange multipliers II

Let us consider a parameter λ (the *Lagrange multiplier*) and obtain an expression for which dx and dy are independent:

$$d(f+\lambda g) = (\frac{\partial f}{\partial x} + \lambda\frac{\partial g}{\partial x})dx + (\frac{\partial f}{\partial y} + \lambda\frac{\partial g}{\partial y})dy = 0$$

Now:

$$\begin{aligned} \frac{\partial f}{\partial x} + \lambda\frac{\partial g}{\partial x} &= 0 \\ \frac{\partial f}{\partial y} + \lambda\frac{\partial g}{\partial y} &= 0 \\ g(x,y) &= c \end{aligned}$$

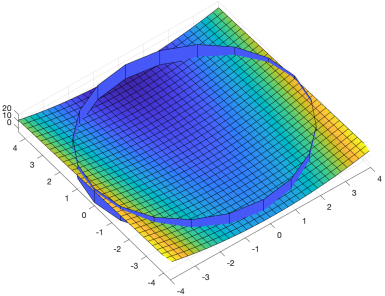
are sufficient to find λ, x^* and y^*.

Exercise 6

Find the optimal values of the function $f(x, y) = x^2 - y$ constrained within the circumference $x^2 + y^2 = 4$. (note that, interestingly, in the optimal value x^* we will get a gradient for the constraint that is parallel to the gradient of the function: $\nabla f(x^*) = \lambda^* \nabla g(x^*)$)

Exercise 7

Find the optimal values of the function $f(x,y) = x^2 - y$ constrained within the circumference $x^2 + y^2 = 4$. (note that, interestingly, in the optimal value x^* we will get a gradient for the constraint that is parallel to the gradient of the function: $\nabla f(x^*) = \lambda^* \nabla g(x^*)$)



The general problem we study takes the form

$$\begin{aligned} \text{maximize} \quad & f(x) \\ \text{subject to} \quad & \begin{cases} x \in X \\ g(x) = b \end{cases} \end{aligned}$$

where

$$\begin{aligned} x \in \mathbf{R}^n \quad & (n \text{ decision variables}) \\ f : \mathbf{R}^n \to \mathbf{R} \quad & (\text{objective function}) \\ X \subseteq \mathbf{R}^n \quad & (\text{regional constraints}) \\ \left.\begin{aligned} g : \mathbf{R}^n \to \mathbf{R}^m \\ b \subseteq \mathbf{R}^m \end{aligned}\right\} \quad & (m \text{ functional constraints}) \end{aligned}$$

If the constraint is $g(x) \leq b$ we introduce a **slack variable** z and write:

$$g(x) + z = b, \; z \geq 0$$

Regional and functional constraints define the **feasible set** for x

The Lagrangian I

Let us take this example:

$$\begin{array}{ll} \text{maximize} & f(x) = x_1 - x_2 - 2x_3 \\ \text{subject to} & \begin{cases} x_1 + x_2 + x_3 = 5 \\ x_1^2 + x_2^2 = 4 \\ x = (x_1, x_2, x_3) \in \mathbf{R}^3 \end{cases} \end{array}$$

1. We write the Lagrangian as (note we have two constraints):

$$\begin{aligned} L(x,\lambda) &= x_1 - x_2 - 2x_3 - \lambda_1(x_1 + x_2 + x_3 - 5) - \lambda_2(x_1^2 + x_2^2 - 4) \\ &= [x_1(1-\lambda_1) - \lambda_2 x_1^2] + [x_2(-1-\lambda_1) - \lambda_2 x_2^2] + [x_3(-2-\lambda_1 \\ &\quad +5\lambda_1 + 4\lambda_2 \end{aligned}$$

2. We want to minimize the Lagrangian subject to the regional constraints $x = (x_1, x_2, x_3) \in \mathbf{R}^3$.

The Lagrangian II

3. We find the values λ such that will define a feasible set

$$Y = \left\{ \lambda : \underbrace{\min}_{x \in \mathbf{R}^3} L(x, \lambda) > -\infty \right\}$$

   So, we only consider the values of λ for which we obtain a finite minimum. In the example

$$Y = \{\lambda : \lambda_1 = -2, \lambda_2 < 0\}$$

4. For $\lambda \in Y$, the miniumum will be obtained at some $x(\lambda)$ (that depends on λ in general). Here, $x(\lambda) = (3/2\lambda_2, 1/2\lambda_2, x_3)^T$.

The Lagrangian III

5. Finally, we find a λ value that provides a feasible value $x(\lambda)$. Here

$$x_1^2 + x_2^2 = 4 \Rightarrow \lambda_2 = -\sqrt{5/8}$$

So, x^* is optimal:

$$x^* = (-3\sqrt{2/5}, -\sqrt{2/5}, 5 + 4\sqrt{2/5})^T; \; \lambda^* = (-2, -\sqrt{5/8})^T$$

Kuhn-Tucker theorem

Let us consider a general optimization problem with constraints: let $f : \mathbf{R}^n \to \mathbf{R}$ be the objective function and $\mathbf{g} : \mathbf{R}^n \to \mathbf{R}^m$ a set of constraint functions ($\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}))$ with $\mathbf{x} = (x_1, \ldots, x_n)$. We want to obtain

$$\text{Min } f(\mathbf{x}) \text{ such that } \mathbf{g}(\mathbf{x}) \geq 0$$

Theorem

If $\boldsymbol{x}^*$ is a local minimum for the optimisation problem and the constraints $\boldsymbol{g}(\boldsymbol{x}) \geq 0$ are satisfied at $\boldsymbol{x}^*$, then $\boldsymbol{x}^*$ must satisfy:

$$\begin{aligned}
\nabla f(\boldsymbol{x}) &= \lambda_1 \nabla g_1(\boldsymbol{x}) + \cdots + \lambda_m \nabla g_m(\boldsymbol{x}) \\
0 &= \lambda_1 g_1(\boldsymbol{x}) + \cdots + \lambda_m g_m(\boldsymbol{x}) \\
0 &\leq \boldsymbol{g}(\boldsymbol{x})
\end{aligned}$$

References

Michael W. Carter, Camille C. Price, and Ghaith Rabadi (2019)
Operations Research, 2nd Edition
CRC Press.

David Harel, with Yishai Feldman (2004)
Algorithmics: the spirit of computing, 3rd Edition
Addison-Wesley.

K.F. Riley, M.P. Hobson, S.J. Bence (2002)
Mathematical Methods for Physics and Engineering (2nd Ed)
McGraw Hill.

J. Nocedal, S. J. Wright (2006)
Numerical Optimization (2nd Ed)
Springer.