Università Commerciale
Luigi Bocconi

Review of Optimization Methods

Prof. Massimo Guidolin

20550– Quantitative Methods for Finance

August 2019

Outline of the Course

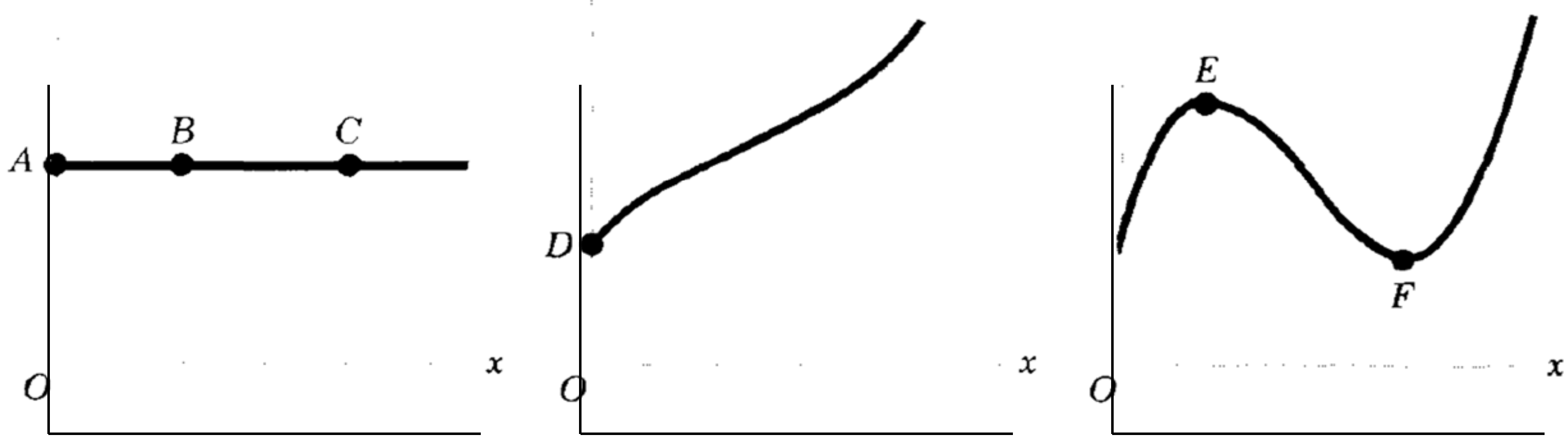
- **Lectures 1 and 2 (3 hours):**
  - Linear and non-linear functions on $\mathbb{R}$
  - Limits, continuity, differentiability, rules to compute derivatives, approximation with differentials
  - Logarithmic and exponential functions
  - Introduction to integration
- **Lecture 3 (1.5 hours):**
  - Review of matrix algebra with applications in excel
- **Lectures 4 and 5 (3 hours, 1.5 of which on your laptop):**
  - Introduction to optimization: functions of one variable
  - Generalization: functions of several variables
  - Use of Excel Solver to tackle constrained optimization

Optimization: Statement of the Problem (1/2)

- Optimization == maximizing (or minimizing) some objective function, $y = f(x)$, by picking one or more appropriate values of the control (aka choice) variable x
- The most common criterion of choice among alternatives in economics (and finance) is the goal of maximizing something (like the profit of a firm) or minimizing something (like costs or risks)
  - For instance, think of a risk-averse investor who wants to maximize a mean-variance objective by picking an appropriate set of portfolio weights
- Maxima and minima are also called extrema and may be relative (or local, that is, they represent an extremum in the neighborhood of the point only) or global
- Key assumption: $f(x)$ is n times continuously differentiable

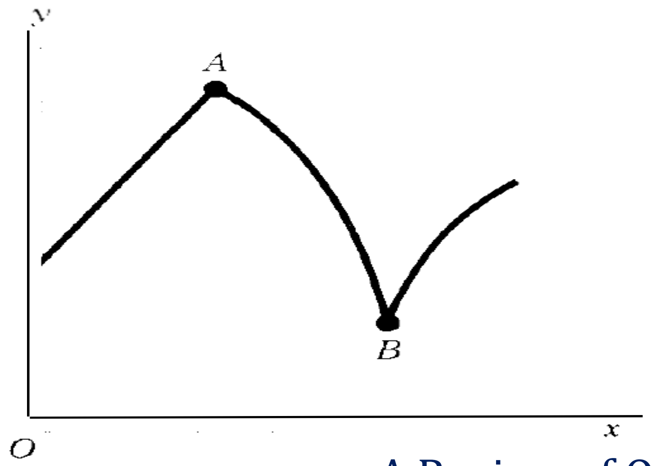
Optimization: Statement of the Problem (2/2)

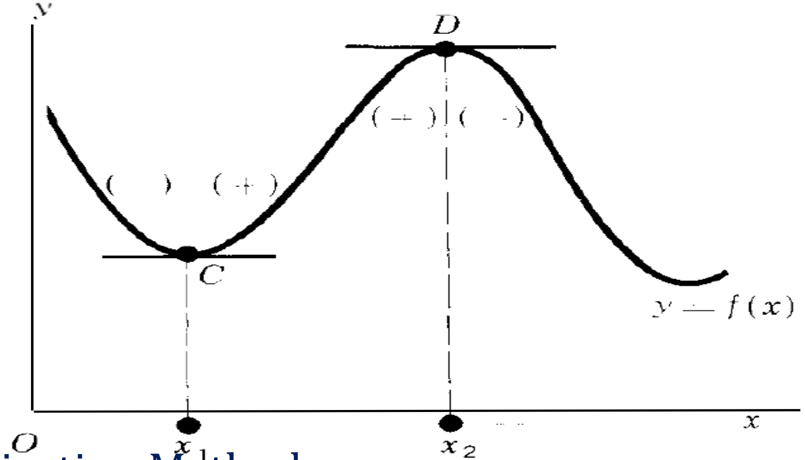
- In the leftmost graph, optimization is trivial: the function is a constant and as such all points are at the same time maxima and minima, in a relative sense
- In the second plot, $f(x)$ is monotonically increasing, there is no finite maximum, if the set of nonnegative real numbers is the domain (as the picture implies)
- The points E and F on the right are examples of a relative (local) extrema
- A function can well have several relative extrema, some of which may be maxima while others are minima



Candidate points: The First-Derivative Test (1/2)

- As a first step we want to identify the "candidate" points to solve the optimization problem, i.e., all the local extrema
- Indeed, global extrema must also be local extrema or end points of $f(x)$ on its domain
  - If we know all the relative maxima, it is necessary only to select the largest of these and compare it with the end points in order to determine the absolute maximum
- Key Result 1 (First-Derivative Test): If a relative extremum of the function occurs at $x = x_0$, then either $f'(x_0) = 0$, or $f'(x_0)$ does not exist; <u>this is a necessary condition</u> (but NOT sufficient)





Candidate points: The First-Derivative Test (2/2)

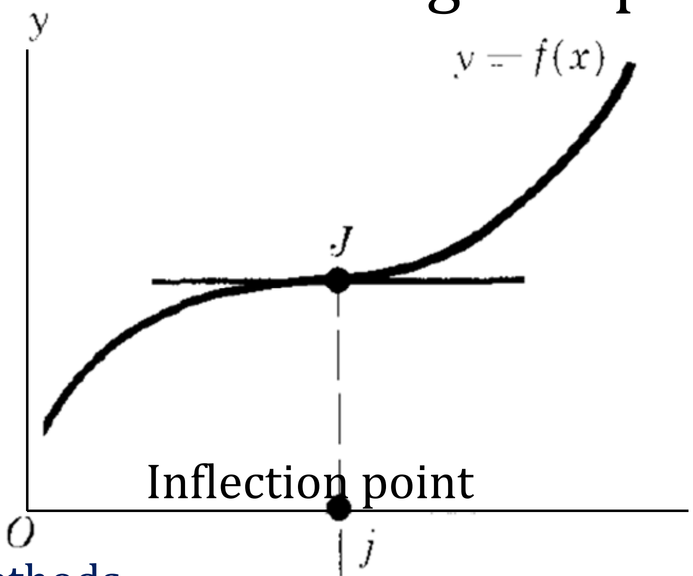
- **Key Result 1 Qualified**: If $f'(x_0) = 0$ then the value of $f(x_0)$ will be:

  (a) A **relative maximum** if the derivative $f'(x)$ changes its sign from > 0 to <0 from the immediate left of the point x_0 to its immediate right

  (b) A **relative minimum** if $f'(x)$ changes its sign from negative to positive from the immediate left of x_0 to its immediate right

  (c) Neither a relative maximum nor a relative minimum if $f'(x)$ has the same sign on both the immediate left and right of point x_0 (**inflection point**)

> **NOTE:** we are assuming that the function is continuous and possesses continuous derivatives => for smooth functions, relative extreme points can occur only when the first derivative has a zero value



One Example

Example Find the relative extrema of the function

$$y = f(x) = x^3 - 12x^2 + 36x + 8$$

First, we find the derivative function to be

$$f'(x) = 3x^2 - 24x + 36$$

To get the critical values, i.e., the values of x satisfying the condition $f'(x) = 0$, we set the quadratic derivative function equal to zero and get the quadratic equation

$$3x^2 - 24x + 36 = 0$$

By factoring the polynomial or by applying the quadratic formula, we then obtain the following pair of roots (solutions):

$$\bar{x}_1 = 2 \quad [\text{at which we have } f'(2) = 0 \text{ and } f(2) = 40]$$

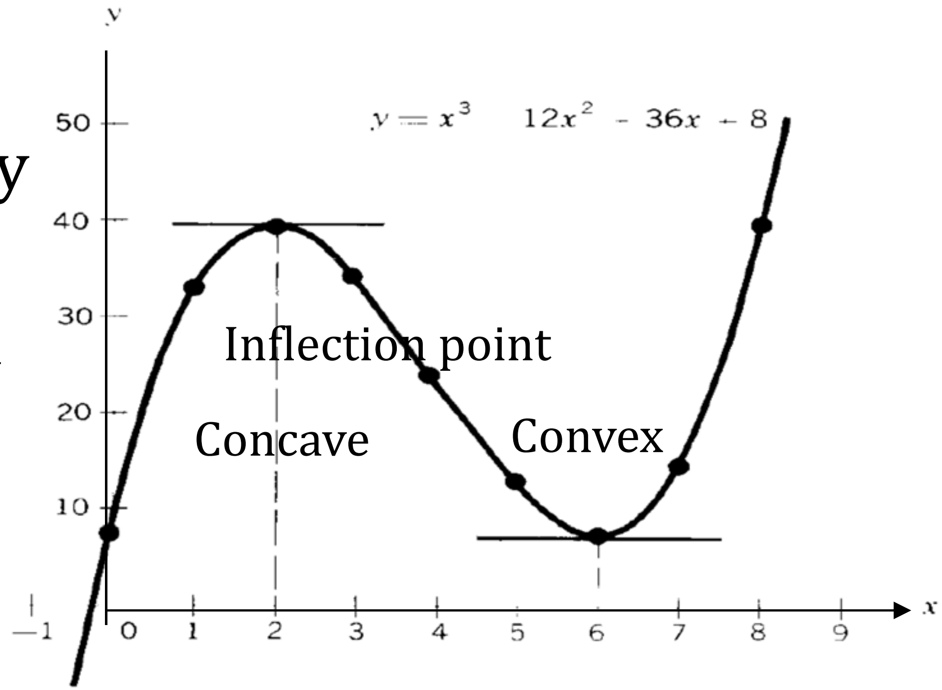
$$\bar{x}_2 = 6 \quad [\text{at which we have } f'(6) = 0 \text{ and } f(6) = 8]$$

Since $f'(2) = f'(6) = 0$, these two values of x are the critical values we desire.

It is easy to verify that $f'(x) > 0$ for $x < 2$, and $f'(x) < 0$ for $x > 2$, in the immediate neighborhood of $x = 2$; thus, the corresponding value of the function $f(2) = 40$ is established as a relative maximum. Similarly, since $f'(x) < 0$ for $x < 6$, and $f'(x) > 0$ for $x > 6$, in the immediate neighborhood of $x = 6$, the value of the function $f(6) = 8$ must be a relaive minimum.

Concavity, Convexity, and Second-Order Derivatives

- A strictly concave (convex) function is such that if we pick any pair of points M and N on the function and join them by a straight line, the line segment MN must lie entirely below (**above**) the curve, except at points M and N



$$\left.\begin{array}{l} f''(x_0) > 0 \\ f''(x_0) < 0 \end{array}\right\} \text{ means that the } \textit{slope of the curve} \text{ tends to } \left\{\begin{array}{l} \text{increase} \\ \text{decrease} \end{array}\right.$$

- If the second derivative $f''(x_0)$ is negative for all x then the function $f(x)$ is strictly concave
- If the second derivative $f''(x_0)$ is positive for all x then the function $f(x)$ is strictly convex

The Second-Order Derivative Test

- **Key Result 2 (Second-Derivative Test)**: If the first derivative of a function at $x = x_0$ is $f'(x_0) = 0$ (first-order, **necessary condition**), then $f(x_0 0)$, will be:

  (a) A relative maximum if $f''(x_0) < 0$
  (b) A relative minimum if $f''(x_0) > 0$

  Second-order, **sufficient condition**

- This test is in general more convenient to use than the first-derivative test, because it does not require us to check the derivative sign to both the left and the right of x_0
- Drawback: this test is inconclusive in the event that $f''(x_0) = 0$ when the stationary value $f(x_0)$ can be either a relative maximum, or a relative minimum, or even an inflection point
  - This is what makes the condition sufficient only

Two Examples

Example Find the relative extremum of the function

$$y = f(x) = 4x^2 - x$$

The first and second derivatives are

$$f'(x) = 8x - 1 \qquad \text{and} \qquad f''(x) = 8$$

Setting $f'(x)$ equal to zero and solving the resulting equation, we find the (only) critical value to be $\bar{x} = \frac{1}{8}$, which yields the (only) stationary value $f(\frac{1}{8}) = -\frac{1}{16}$. Because the second derivative is positive (in this case it is indeed positive for any value of x), the extremum is established as a minimum.

Example Find the relative extrema of the function

$$y = g(x) = x^3 - 3x^2 + 2$$

The first two derivatives of this function are

$$g'(x) = 3x^2 - 6x \qquad \text{and} \qquad g''(x) = 6x - 6$$

Setting $g'(x)$ equal to zero and solving the resulting quadratic equation, $3x^2 - 6x = 0$, we obtain the critical values $\bar{x}_1 = 0$ and $\bar{x}_2 = 2$, which in turn yield the two stationary values:

$$g(0) = 2 \qquad [\text{a maximum because } g''(0) = -6 < 0]$$

$$g(2) = -2 \qquad [\text{a minimum because } g''(2) = 6 > 0]$$

Functions with more than one variable

- We are now going to generalize the earlier results to optimization problems for functions of several variables, i.e.,
  - Functions $g: \mathbb{R}^n \to \mathbb{R}$, i.e., $y = g(x_1, x_2, \dots, x_n)$
- In fact, functions from $\mathbb{R}^n$ to $\mathbb{R}$ will be popping up very often in your future studies
  - For instance, the return of a portfolio is a linear function of the returns of the n assets that compose the portfolio:

$$r_p = w_1 r_1 + w_2 r_2 + \cdots + w_n r_n$$

  - Another example is a utility function $U(x_1, x_2, \dots, x_n)$ of a bundle of consumption goods
- However, we first need to generalize the concept of derivative to the case of functions of several variables
- This leads us to the introduction partial derivatives and of Jacobian derivatives

Partial derivatives and the Jacobian

Definition: Let $f: \mathbb{R}^n \to \mathbb{R}$. Then for each variable x_i at each point $\mathbf{x}^0 = (x_1^0, x_2^0, \dots, x_n^0)$ in the domain of *f*, the partial derivative with respect to x_i is

$$\frac{\partial f}{\partial x_i}(x_1^0, \dots, x_n^0) = \lim_{h \to 0} \frac{f(x_1^0, \dots, x_i^0 + h, \dots, x_n^0) - f(x_1^0, \dots, x_i^0, \dots, x_n^0)}{h}$$

if the limits exists. Only the *i*th variable changes, while the others stay constant

- The vector (more generally, matrix) $Df_{\mathbf{x}^0}$ that collects all partial derivatives

$$Df_{\mathbf{x}^0} = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}^0), \frac{\partial f}{\partial x_2}(\mathbf{x}^0), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}^0) \right)$$

is called the Jacobian derivative of f at $\mathbf{x}^0$

Partial Derivatives: One Example

- Example: consider the function $f(x_1, x_2) = 3x_1^2 x_2^2 + 4x_1 x_2^3 + 7x_2$
- Let us compute the partial derivative with respect to x_1
- Simply treat x_2 as it was a constant and apply the same rules of one-variable calculus

$$\frac{df}{dx_1} = 6x_1 x_2^2 + 4x_2^3$$

- Now let compute the partial derivative with respect to x_2

$$\frac{df}{dx_2} = 6x_1^2 x_2 + 12x_1\ x_2^2 + 7$$

- The concept can be easily generalized to a function of more than two variables

Second Order Derivatives and Hessians (1/2)

- If the n partial derivative functions of f are continuous functions at the point $\mathbf{x}^0$ in $\mathbb{R}^n$ we say that f is continuously differentiable at $\mathbf{x}^0$
- If all the n partial derivatives $\partial f / \partial x_i$ are themselves differentiable we can compute their partial derivatives
- $\frac{\partial f}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$ is called the $x_i x_j$-second order partial derivative of f and it is generally denoted as $\frac{\partial^2 f}{\partial x_i x_j}$
- When $i \neq j$ then we speak of cross (or mixed) partial derivatives
- A function of n variables has n^2 second order partial derivatives that are usually arranged into a n x n Hessian matrix

Second Order Derivatives and Hessians (2/2)

- The Hessian matrix is typically denoted as $D^2 f(\mathbf{x})$ or $D^2 f_x$ and takes the form

$$D^2 f_{\mathbf{x}} = \begin{pmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial {x_2}^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial {x_n}^2} \end{pmatrix}.$$

- **Young's theorem**: the Hessian matrix is a symmetric matrix, i.e., for each pair of indices *i* and *j*

$$\frac{\partial^2 f}{\partial x_i x_j} = \frac{\partial^2 f}{\partial x_j x_i}$$

One Example

- Consider the function $f(x_1, x_2) = 3x_1^2 x_2^2 + 4x_1 x_2^3 + 7x_2$
- Let us compute the Hessian matrix; we already computed $\frac{df}{dx_1} = 6x_1 x_2^2 + 4x_2^3,\ \frac{df}{dx_2} = 6x_1^2 x_2 + 12x_1\ x_2^2 + 7$
- Now we need to compute
- $\frac{\partial^2 f}{\partial x_1 x_1} = 6x_2^2;\ \frac{\partial^2 f}{\partial x_2 x_2} = 6x_1^2 + 24x_1 x_2;\ \frac{\partial^2 f}{\partial x_1 x_2} = 12x_1 x_2 + 12x_2^2$
- Hessian matrix is

$$\begin{pmatrix} 6x_2^2 & 12x_1 x_2 + 12x_2^2 \\ 12x_1 x_2 + 12x_2^2 & 6x_1^2 + 24x_1 x_2 \end{pmatrix}$$

- You can check that $\frac{\partial^2 f}{\partial x_1 x_2} = \frac{\partial^2 f}{\partial x_2 x_1}$

Optimization: the case of n-variable functions

- Now we are ready to generalize optimization to the case of n-variable functions
- The strategy remains looking for critical points (relative extrema) and then try to isolate global ones among them
- $\mathbf{x}^0$ is a critical point for f if it fulfills

$$Df(\mathbf{x}^0) = \mathbf{0},$$

which means that

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^0) = 0, \text{ for each } i$$

- If $\mathbf{x}^0$ is an interior point which is a local maximum or minimum then it is a critical point
- However, the reverse is not true, i.e., the condition is necessary but <u>not sufficient</u> for an interior point to be a local extremum

Checking the sign of the Hessian matrix (1/2)

- As one may guess from the one-variable case, second order conditions involve checking the sign of the Hessian matrix
- We need to add a definition to the matrix algebra review that we discussed in the last lecture
  - A principal minor of a square matrix **A** is the determinant of a submatrix obtained by eliminating some rows and the corresponding column; the order of a minor is the dimension of the considered submatrix
  - A leading principal minor A_k is a principal minor obtained by considering the first k rows and columns of the original matrix
  - For instance,

$$\begin{bmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$A_1 = a_{11}$

$A_2 = a_{11}a_{22} - a_{21}a_{12}$

$A_3 \Rightarrow$ the determinant of the 3x3 matrix itself

Checking the sign of the Hessian matrix (2/2)

- A square symmetric matrix is said to be
  - Positive definite: if all its leading principal minors are **strictly positive**
  - Negative definite: if $a_{11} < 0$ and then all its leading principal minors alternate in sign (but are different from zero)
  - Indefinite: if we have a nonzero leading principal minor and at least one leading principal minor does not follow the patterns above
  - Positive semidefinite: if every principal minor is nonnegative
  - Negative semidefinite: if $a_{11} \leq 0$ and every principal minor of odd order is ≤ 0 and every principal minor of even order is ≥ 0

Sufficient second order conditions

- The sufficient second order conditions for a local extremum are as follows, given that $\mathbf{x}^0$ is an interior critical point:
  - If $D^2 f(\mathbf{x}^0)$ is negative definite => $\mathbf{x}^0$ is a local maximum point
  - If $D^2 f(\mathbf{x}^0)$ is positive definite => $\mathbf{x}^0$ is a local minimum point
  - If $D^2 f(\mathbf{x}^0)$ is indefinite => $\mathbf{x}^0$ is a saddle point
- Semidefinite cases require further investigation and we shall skip their discussion
- When the sign of the Hessian matrix does not depend on $\mathbf{x}$, the local extrema are also global because when the Hessian is positive (negative) definite over the entire domain the function is strictly convex (concave)

Example of Unconstrained Optimization (1/2)

- Study the optimization of the following function: $3x^4 + 3x^2y - y^3$
- Step 1: find the internal critical points
  - $\frac{\partial f}{\partial x} = 12x^3 + 6xy = 0$
  - $\frac{\partial f}{\partial y} = 3x^2 - 3y^2 = 0$
- Solving that is non-trivial and time consuming
- You get three critical points:
  - A(0,0); B($\frac{1}{2}, -\frac{1}{2}$); C($-\frac{1}{2}, -\frac{1}{2}$)
- Step 2: compute the Hessian matrix
- $D^2 f = \begin{pmatrix} 36x^2 + 6y & 6x \\ 6x & -6y \end{pmatrix}$
- We need to check it at A, B, and C

Example of Unconstrained Optimization (2/2)

- As an example, I will only check the sign of the Hessian at $C(-\frac{1}{2}, -\frac{1}{2})$

$$D^2 f = \begin{pmatrix} 9-3 & -3 \\ -3 & 3 \end{pmatrix} = \begin{pmatrix} 6 & -3 \\ -3 & 3 \end{pmatrix}$$

- $A_1 = 6 > 0$
- $A_2 = 6x3 - (-3)(-3) = 9 > 0$
- Then the Hessian matrix is positive definite and the point is a local minimum
- This is an easy problem and yet you see how computationally intense it is
- Sometimes the solution shall ben find numerically anyway
- Things get even worse when we introduce constraints (equality constraints, inequality constraints or both)
- We shall now introduce the Excel solver

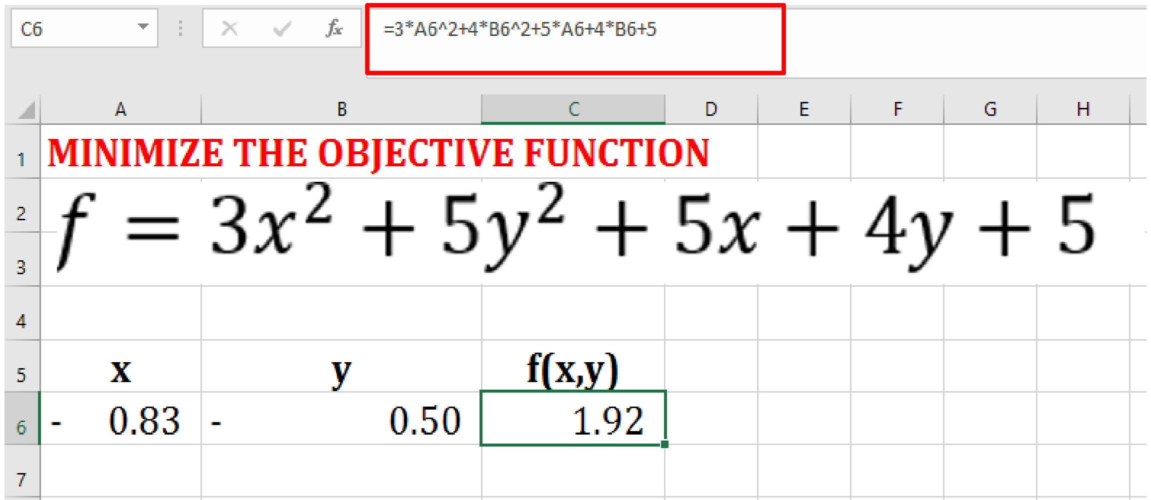
Excel Solver (1/3)

- The Solver is an analysis tool available as an additional package into Excel
- If you do not have it already installed in your Excel you can download it as Excel add-in from the Excel options
- Once you have installed it, you find it under the tab “Data”
- The Solver is able to solve optimization problems for you (even with a number of equality/inequality constraints, as we shall see later on)
- Essentially, it maximizes (minimizes) the value obtained into an objective cell in which you have to specify a certain function…
- …by changing a set of cells (control variables) that you specify elsewhere as an array in the worksheet

Excel Solver (2/3)

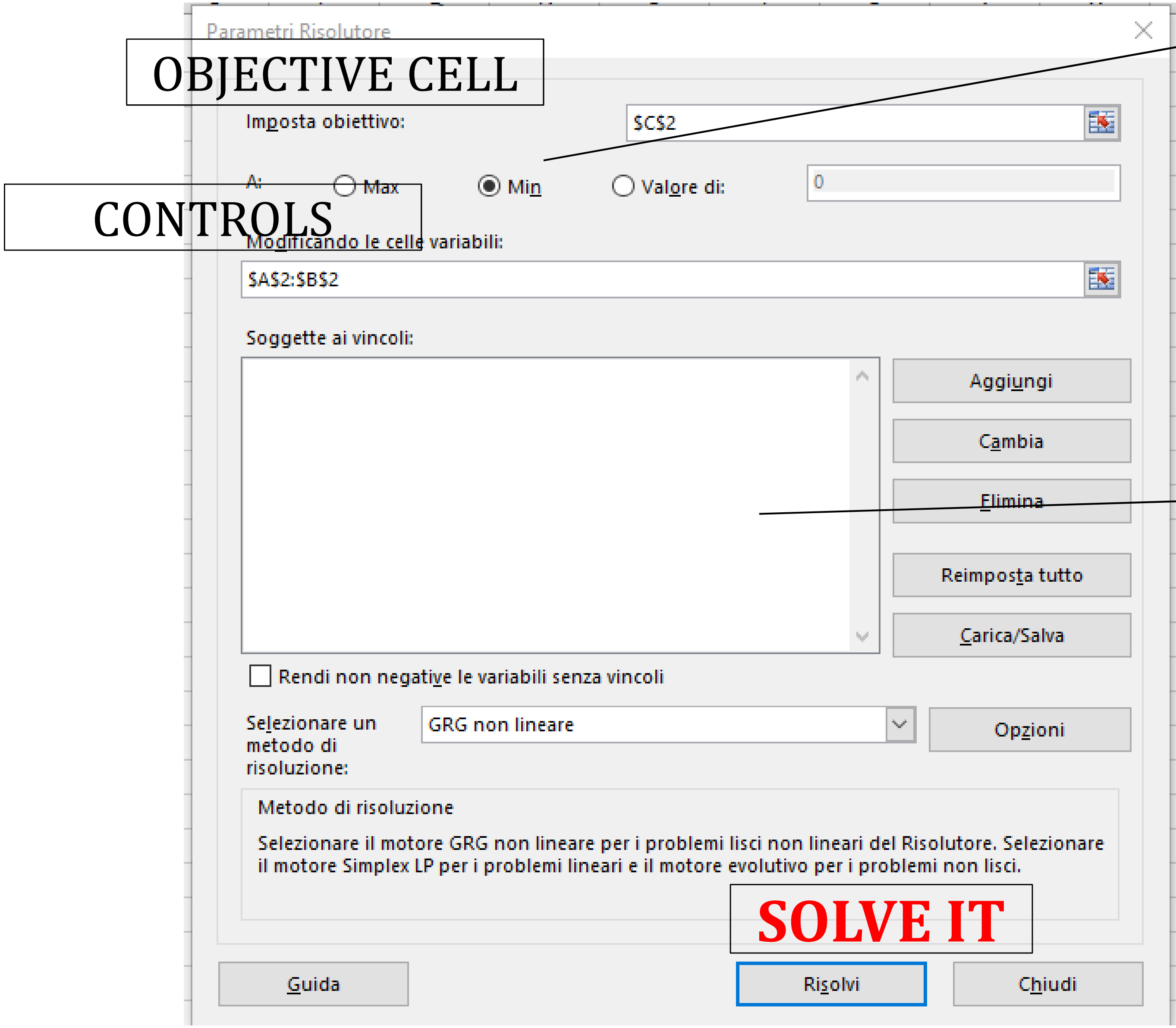
OBJECTIVE FUNCTION: $f = 3x^2 + 5y^2 + 5x + 4y + 5$

- In the cell A6 write a value (almost whatever) for *x* and in B6 write a value for *y*
- The values are only used to initialize the search (in some situations it may matter where you initialize the search, but typically the nature of the problem that you are solving suggests reasonable values)
  - For instance, when finding optimal weights for a portfolio I typically start from equal weights
- Then write the functions *f* into the cell C6



Excel Solver (3/3)

- Now open the Solver:

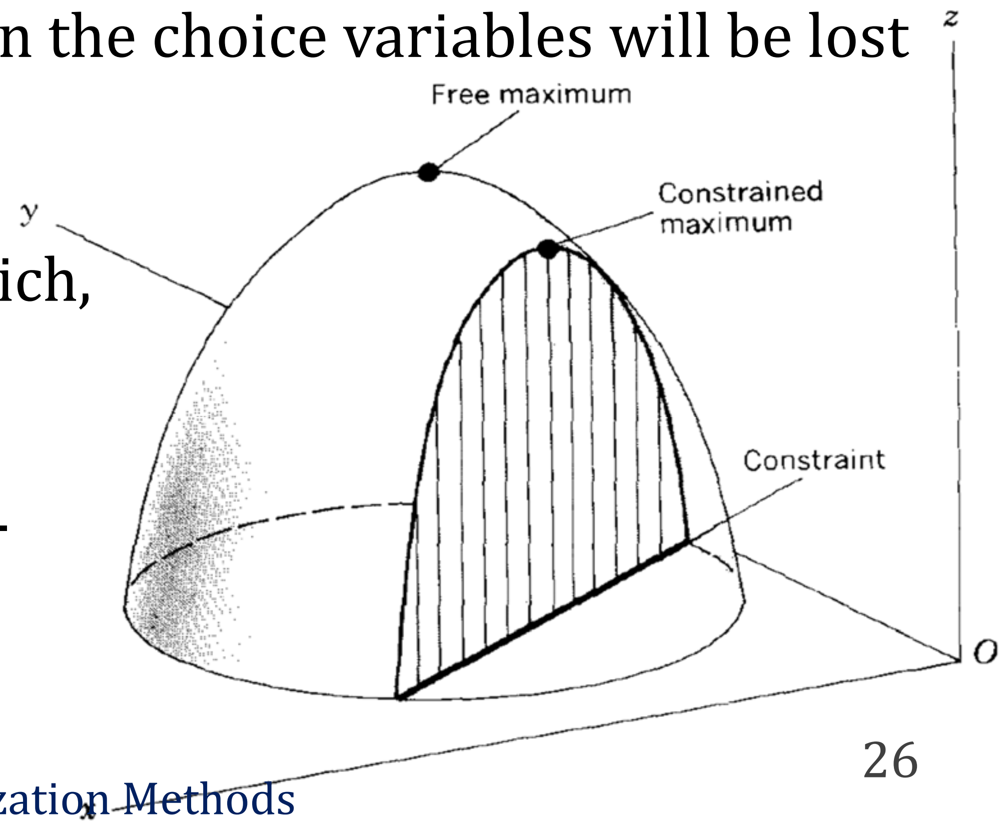


Choose if you want to look for a min or for a max

For now we are not putting constraints

Hints of Constrained Optimization (1/2)

- Up to these points, all control variables have been independent of each other: the decision made regarding one variable does not impinge upon the choices of the remaining variables
  - E.g., a two-product firm can choose any value for Q_1 and any Q_2 it wishes, without the two choices limiting each other
  - If the firm in the example is somehow required to fulfill a restriction (e.g., a production quota) in the form of $Q_1 + Q_2 = k$, however, the independence between the choice variables will be lost
  - The new optimum satisfying the production quota constitutes a constrained optimum, which, in general, may be expected to differ from the free optimum
- Key Result : A constrained maximum can never exceed the free maximum



Hints of Constrained Optimization (2/2)

- In general, a constrained maximum can be expected to achieve a lower value than the free maximum, although, by coincidence, the two maxima may happen to have the same value
    - We had added another constraint intersecting the first constraint at a single point in the *xy* plane, the two constraints together would have restricted the domain to that single point
    - Then the locating of the extremum would become a trivial matter
  - In a meaningful problem, the number and the nature of the constraints should be such as to restrict, but not eliminate, the possibility of choice
    - Generally, the number of constraints should be less than the number of choice variables
- Under C < N equality constraints, when we can write a sub-set of the choice variables as an explicit function of all others, the former can be substituted out:

$$\max_{x_1, x_2, \dots, x_N} f(x_1, x_2, \dots, x_N)$$
$$s.t. \, x_1 = g(x_2, \dots, x_N), \dots, x_C = g(x_{C+1}, \dots, x_N)$$

Hint: Lagrange Multiplier Method (1/2)

becomes: $\max\limits_{x_{C+1}, x_{C+2}, \dots, x_N} f(x_{C+1}, x_{C+2}, \dots, x_N)$,

an unconstrained problem

- However, the direct substitution method cannot be applied when the C constraints do not allow us to re-write the objective functions in N – C free control variables
  - Even if some of the variables become implicit functions of others, it would be complex to proceed because the objective would become "highly composite"
- In such cases, we often resort to the method of Lagrange (undetermined) multipliers
  - The goal is to convert a constrained extremum problem into a form such that the first-order condition of the free extremum problem can still be applied
  - For instance, consider an objective function $z = f(x.y)$ subject to the constraint $g(x,y)=c$ where c is a constant

Lagrange Multiplier Method (2/2)

- The Lagrangian problem is:

$$\max_{x,y,\lambda} \underbrace{f(x,y) - \lambda[c - g(x,y)]}_{Lagrange\ function}$$

- The necessary FOC is then:

$$Z_\lambda = c - g(x, y) = 0$$

$$Z_x = f_x - \lambda g_x = 0$$

$$Z_y = f_y - \lambda g_y = 0$$

The stationary values of the Lagrangian function Z will automatically satisfy the constraint

$$\boxed{\frac{f_x}{g_x} = \lambda \quad \text{and} \quad \frac{f_y}{g_y} = \lambda}$$

The optimal value λ^* provides a measure of the sensitivity of the Lagrangian function to a shift of the constraint

Example Find the extremum of

$$z = xy \qquad \text{subject to} \qquad x + y = 6$$

The first step is to write the Lagrangian function

$$Z = xy + \lambda(6 - x - y)$$

For a stationary value of Z, it is necessary that

$$\left.\begin{aligned} Z_\lambda &= 6 - x - y = 0 \\ Z_x &= y - \lambda = 0 \\ Z_y &= x - \lambda = 0 \end{aligned}\right\} \quad \text{or} \quad \left\{\begin{aligned} x + y &= 6 \\ -\lambda \quad + y &= 0 \\ -\lambda + x \quad &= 0 \end{aligned}\right.$$

Thus, by Cramer's rule or some other method, we can find

$$\bar{\lambda} = 3 \qquad \bar{x} = 3 \qquad \bar{y} = 3$$

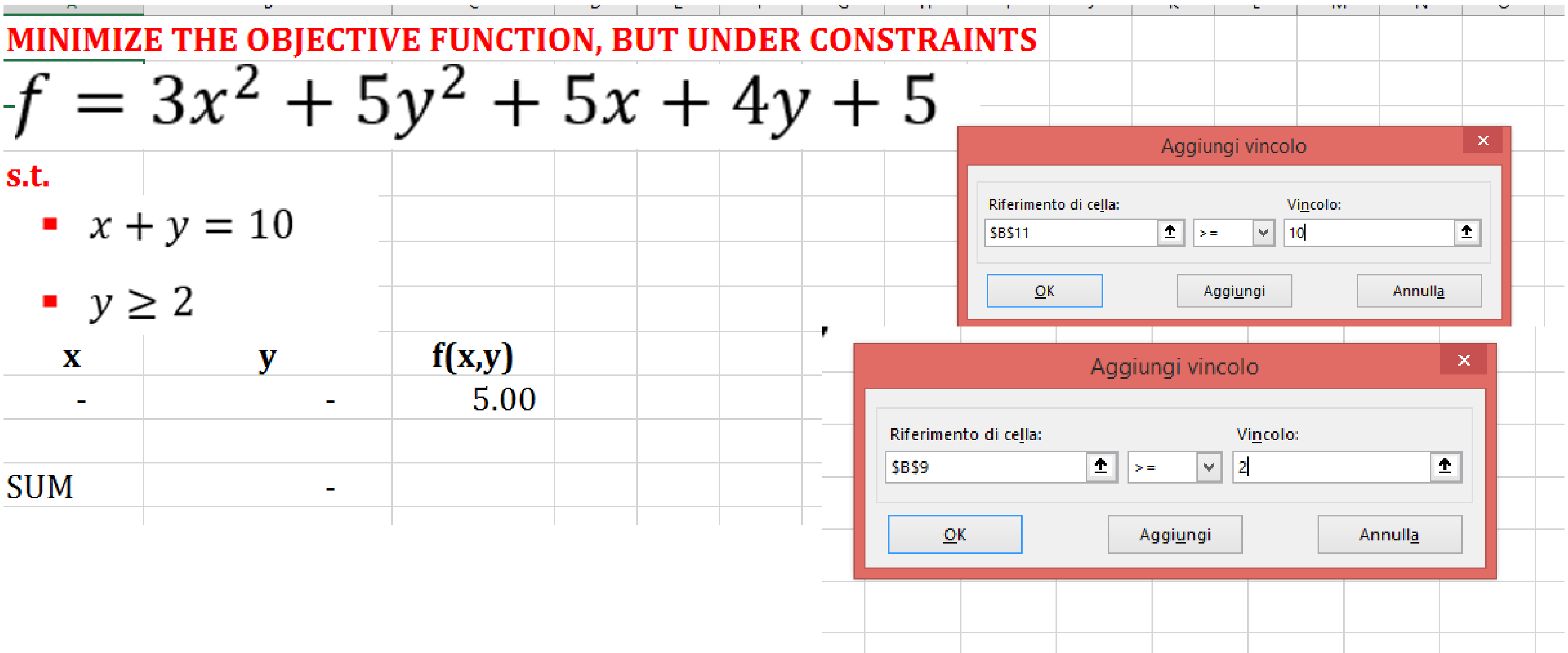
The stationary value is $\bar{Z} = \bar{z} = 9$, which needs to be tested against a second-order condition before we can tell whether it is a maximum or minimum (or neither). That will be taken up later.

Excel Solver for constrained optimization (1/2)

- Lagrange multiplier method gives us critical (candidate) points, but also in this case we need to check second order conditions
- This requires to check positive (negative) definiteness of large square matrices (bordered Hessian matrices) – we shall skip the details
- Again, the Excel Solver may help us!
- First of all, the Solver easily solves “Kuhn-Tucker” type of problems (that is, problems in which choice variables are restricted to only take positive values)
- It is sufficient to tick the box “make all unconstrained variable positive)
- Try with the problem we did before (the new minimum will be at (0,0), quite intuitively)

Excel Solver for constrained optimization (2/2)

- Now let us add constraints to our problem
  - $x + y = 10$
  - $y \geq 2$
- Click on “add” to add equality / inequality constraints



Exercises

- Solve the following exercises using Excel Solver

  - Maximize the function

$$f(x,y,z) = y \text{ under constraint } h(x,y,z) = \begin{cases} h_1(x,y,z) = x^2 + y^2 + z^2 - 2 \\ h_2(x,y,z) = y - z \end{cases}$$

  - Minimize the function

$$f(x,y) = x + y + 1 \text{ under constraint } g_1(x,y) = x^2 + y^2 - 2 \leq 0$$

  - Minimize the function

$$f(x,y,z) = x^2 + x + z \text{ under constraint } h(x,y,z) = \begin{cases} h_1(x,y,z) = x + y + z \\ h_2(x,y,z) = x^2 + y^2 - 1 \end{cases}$$