Marinozzi, Tomás; Nallar, Leandro; Pernice, Sergio A.

Working Paper

Intuitive mathematical economics series: General equilibrium models and the Gradient Field Method

Serie Documentos de Trabajo, No. 820

Provided in Cooperation with:
University of CEMA, Buenos Aires

Suggested Citation: Marinozzi, Tomás; Nallar, Leandro; Pernice, Sergio A. (2021) : Intuitive mathematical economics series: General equilibrium models and the Gradient Field Method, Serie Documentos de Trabajo, No. 820, Universidad del Centro de Estudios Macroeconómicos de Argentina (UCEMA), Buenos Aires

This Version is available at:
https://hdl.handle.net/10419/260514

UNIVERSIDAD DEL CEMA
Buenos Aires
Argentina

Serie
DOCUMENTOS DE TRABAJO

Área: Matemática y Economía

Intuitive Mathematical Economics Series. General Equilibrium Models and the Gradient Field Method

Tomás Marinozzi, Leandro Nallar, Sergio A. Pernice

Diciembre 2021
Nro. 820

www.cema.edu.ar/publicaciones/doc_trabajo.html
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding jae@cema.edu.ar

Intuitive Mathematical Economics Series
General Equilibrium Models and the Gradient Field Method

Tomás Marinozzi[1] Leandro Nallar[2] Sergio A. Pernice[3]

Universidad del CEMA
Av. Córdoba 374, Buenos Aires, 1054, Argentina

December, 2021

Abstract:

General equilibrium models are typically presented with mathematical methods, such as the Edgeworth Box, that do not easily generalize to more than two goods and more than two agents. This is fine as a conceptual introduction, but it may be insufficient in the "Big Data−Machine Learning−Era", with gigantic databases filled with data of extremely high dimensionality that are already changing the practice, and perhaps even the conceptual basis, of economics and other social sciences. In this paper present what we call the "Gradient Field Method" to solve these problems. It has the advantage of being, 1) as intuitive as the Edgeworth Box, 2) easily generalizes to far more complex situations, and 3) nicely mesh with the data friendly techniques of the new Era. In addition, it provides a unified framework to present both, partial equilibrium, and general equilibrium problems.

KEY WORDS: microeconomics, general equilibrium, gradient, gradient field, machine learning.

[1]Email: tomasmarinozzi1996@gmail.com
[2]Email: lean_nallar@hotmail.com
[3]Email: sp@ucema.edu.ar
The points of view of the authors do not necessarily represent the position of Universidad del CEMA.

Contents

1 Introduction

One could argue that microeconomics is divided into two fundamental blocks, on the one hand the competitive general equilibrium theory, and on the other the market failures. The basic principles of the general equilibrium theory delve into the idea that, under certain basic assumptions[4], decentralized economic agents can improve each other's well being, reaching an optimal "equilibrium" through the ecosystem called the market. An omnipresent and benevolent planner could replace the market, reaching an efficient distribution of resources, thus achieving the same "optimum". The two fundamental welfare theorems unite the two concepts of general equilibrium; the first welfare theorem states that "any competitive equilibrium is Pareto optimal"; the second theorem indicates that "any Pareto optimum can be reached as a competitive equilibrium if the initial endowments are altered." The equivalence between the two problems allows us to focus on one of them, understanding that the tools used for one are applicable to the other without loss of generality.

As is well known, the economic assumptions that guarantee competitive equilibria (the first part of microeconomics, as described above), mathematically guarantee the solvability of the optimization problem by ensuring convexity.

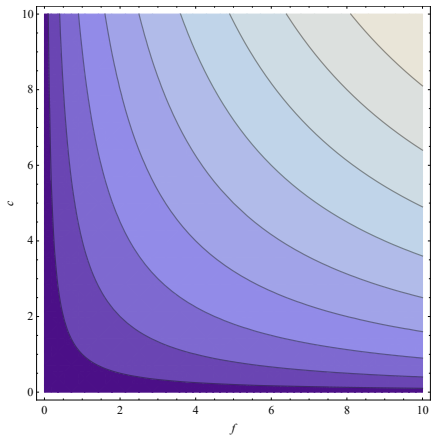
The purpose of this paper is *not* to propose a new technique for solving the optimization problems described above, where extremely powerful techniques already exist (see for example [Boyd and Vandenberghe (2004)]). Rather, the purpose is pedagogical. We present classic problems that any undergraduate student of economics learns in microeconomics courses, with mathematics that: 1) make the solution of general equilibrium problems, and related fundamental welfare theorems, "obvious", once the student learns how to solve simple partial equilibrium problems. 2) Reduces to almost zero the marginal cost of generalizing these problems, say, from 2 variables to 2 million variables. In other words, the generalization difficulties are reduced to a data processing challenge, but it doesn't require any conceptual leap. 3) It frames general equilibrium problems in a "machine-learning ready"[5] language, that very likely economic students will have to learn at some point anyways [Athey and Imbens (2019)], and it may help to learn it in the context of familiar problems.

2 Constrained maximization in microeconomics

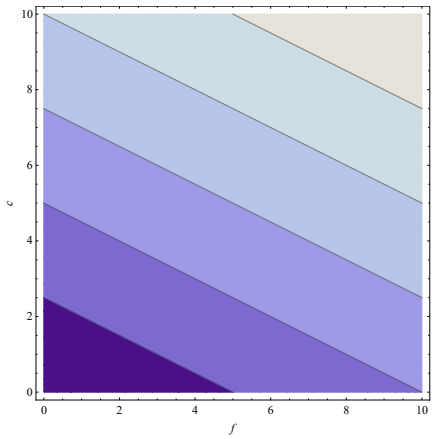
Part of the content and notation in sections 2.1 and 2.2 is in [Pernice (2018 b)]. The reader may consult that reference for more details.

[4] Among them, the absence of: 1) Market power, 2) Satiety in preferences, 3) Informative externalities, 4) Real externalities.

[5] We will implement machine learning general equilibrium problems in a different work.



(a) Utility function (2.1) level curves.



(b) Budget function (2.2) level curves.

Figure 1: 2-D representation of the utility function and the budget function by their level curves. $\alpha = \beta = 1/2$, $P_X = 1$, $P_Y = 2$, $I = 10$. These, and all other figures in this work were generated with [Mathematica 9.0.1.0].

2.1 Tangency of the level curves method to solve constrained maximization problems

2.1.1 Utility maximization subject to a budget (linear) constraint

Suppose the utility function of a person for two goods, X and Y, is given by:

$$U(x, y) = x^{\alpha} y^{\beta}, \qquad 0 < \alpha, \beta < 1 \tag{2.1}$$

where x is the number of the units to be consumed of good X and y the units to be consumed of good Y. Let us review the standard solution of the classic problem of maximization of U under the budget constraint

$$I(x, y) = xP_X + yP_Y = i \tag{2.2}$$

where P_X is the price of good X, P_Y the price of good Y, both prices are assumed externally determined, and i is the actual numerical value of the budget constraint. In figure 1 we can see the level curves of the utility function (2.1) and the budget function (2.2) for specific values of the parameters.

The standard way of solving this maximization with constraints problem is presented in figure 2 for $i = 10$ and corresponds to variants of the following line of reasoning.

The "marginal utilities" of goods X and Y are the respective partial derivatives of the utility with respect to the number of each good:

$$MU_X \equiv \frac{\partial U}{\partial x} \tag{2.3}$$

$$MU_Y \equiv \frac{\partial U}{\partial y} \tag{2.4}$$

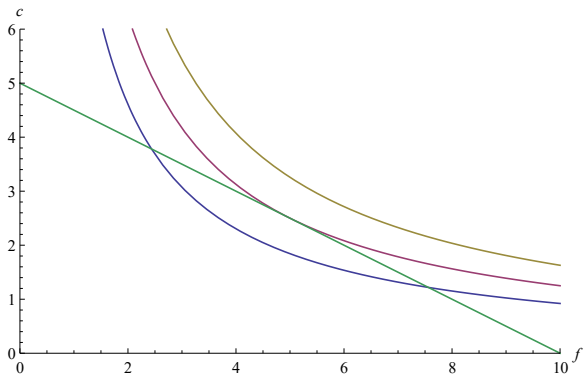


Figure 2: $U = 5/\sqrt{2} - 0.5$ (blue), $U = 5/\sqrt{2}$ (purple), $U = 5/\sqrt{2} + 0.5$ (yellow) and $I = 10$ (green).

or *how much the utility changes when one consumes one additional unit of a good.*

The economic intuition for the optimality condition is that "the marginal utility per dollar spent on good X must equal the marginal utility per dollar spent on good Y", otherwise the person would spend the marginal dollar in the good which increases her utility more:

$$\frac{MU_{X_1}}{P_1} = \frac{MU_{X_2}}{P_2} \tag{2.5}$$

$$MRS = \frac{MU_{X_1}}{MU_{X_1}} = \frac{\frac{\partial U}{\partial x}}{\frac{\partial U}{\partial y}} = \frac{P_X}{P_Y} \tag{2.6}$$

(2.6), which trivially follows from (2.5), provides an equivalent economic intuition. It says that the "marginal rate of substitution" MRS, i.e., the ratio of the marginal utilities, must equal the ratio of the prices.

To make a slightly more rigorous derivation of equations (2.5) or (2.6) it is typically pointed out that the convex form of the level curves of the utility function, makes it obvious that the maximization of utility compatible with the budget constraint will happen at the point in the (x, y) plane in which the budget line is tangent to the level curve of the utility function, intercepting it only once, see figure 2.

Indeed, if the level curve of the utility function intercepts the budget line at two points, as in the blue curve, it is clear that we can increase utility by choosing higher level curves. If it does not intercept the budget line, as in the yellow curve, then it is not compatible with our budget constraint. Therefore the optimum is the level curve of the utility function that intercepts the budget line just once, as in the purple curve. And since the level curves are smooth, at that point the budget line must coincide with the tangent line of the level curve of the utility function.

Translating this graphical intuition into equations, the level curves of the utility function in the (x, y) plane, are curves $y(x)$ implicitly given by the equation $U(x, y(x)) = u$, where u is just a number. By the rules of differentiation of implicit functions [Pernice (2018 a)], the slope dy/dx

of such curve, is given by $\partial U/\partial x + (\partial U/\partial y)(dy/dx) = 0$, or

$$\frac{dy}{dx} = -\frac{\frac{\partial U}{\partial x}}{\frac{\partial U}{\partial y}} \tag{2.7}$$

The intuitive graphical argument above indicates that at the optimum, this slope coincides with the slope of the budget line (2.2)

$$\frac{dy}{dx} = -\frac{P_X}{P_Y} \tag{2.8}$$

From (2.7) and (2.8) we arrive at the equation

$$\frac{\frac{\partial U}{\partial x}}{\frac{\partial U}{\partial y}} = \frac{P_X}{P_Y} \tag{2.9}$$

which is equation (2.6), equivalent to (2.5), derived with a more geometric argument.

For the specific U in equation (2.1), $\partial U/\partial x = \alpha x^{\alpha-1} y^{\beta}$ and $\partial U/\partial y = x^{\alpha}\beta y^{\beta-1}$, so (2.9) becomes

$$\frac{\alpha y}{\beta x} = \frac{P_X}{P_Y}, \quad \text{or} \quad y = \frac{\beta P_X}{\alpha P_Y} x \tag{2.10}$$

This is the *optimal consumption curve* (straight line in this case), for any possible budget constraint. To determine the specific amounts of X and Y that will be consumed, we have to specify what the actual budget constraint is, as shown in figure 3 for $i = 8, 10$ and 12.

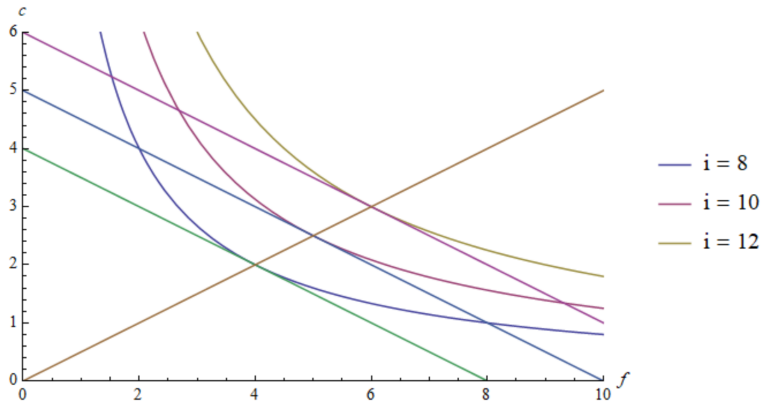


Figure 3: The optimal consumption curve (2.9) (in orange), which for the specific U in (2.1) is the straight line in (2.10), is formed by all the points in the $(x, y) \equiv (f, c)-$plane in which the level curves of U are tangent to the budget constraint.

2.1.2 Maximization problem in 2-D with 1 nonlinear constraint

In the problem solved in section 2.1.1 the constraint (2.2) is linear. The tangent of a linear function is independent of the specific value of the budget constraint, different values of i simply

parallelly displace the constraint straight line, keeping the tangent constant, as is evident in figure 3. This is why the right hand side of the optimal consumption curve equation (2.9) is independent of x and y.

However, the key idea that the local maximum of a function $U(x, y)$, subject to a constraint $V(x, y) = v$, happens at a point in the $(x, y)-$plane where a level curve of $U(x, y)$ is tangent to the curve $y = y(x)$ defined implicity by the constraint $V(x, y(x)) = v$, is valid in general, even if the constraint is nonlinear.

To convince yourself of the above statement consider the following example:

$$\text{Maximize} \quad U(x, y) = x^{\alpha} y^{\beta} \tag{2.11}$$

$$\text{subject to} \quad V(x, y) = xy - 8x - 10y + 80 = 24.45 \tag{2.12}$$

For simplicity let us consider again the case $\alpha = \beta = 1/2$. In figure 4 we see 3 level curves of the

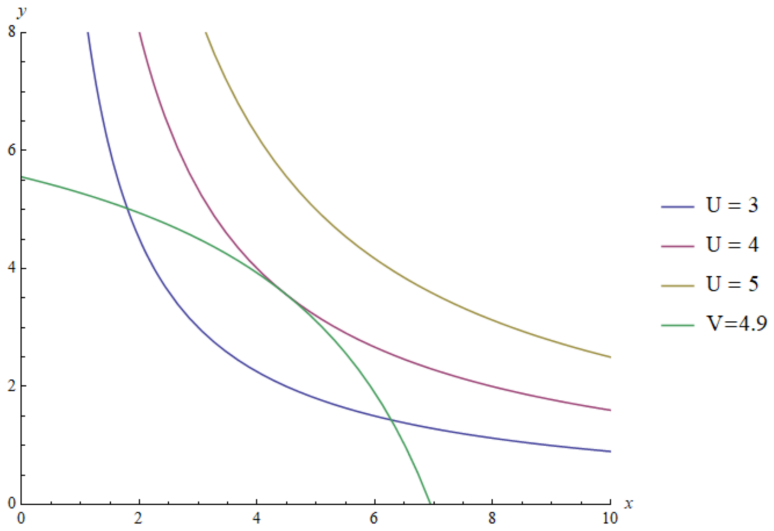


Figure 4: As the level curves of U show, this function increases as we move from the bottom-left to the top-right corner of the graph. The nonlinear constraint (2.12) corresponds to the green curve.

function U in (2.11), indicating that U increases as we move from the bottom-left to the top-right corner of the graph, and the nonlinear constraint (2.12) in the green curve.

It is obvious that the constrained maximum is at the point where the green constraint curve just touches the level curve $U = 4$, and that the tangent line of both of these curves coincide at this point. As surely the reader already knows, the same idea of common tangency that holds at the constrained maximum for linear constraints also holds for nonlinear ones. Still, to prepare for what comes in the next section, it is worth to carefully review the logic for this conclusion beyond the simple graphic observation.

One way to describe the logic is this: since our maximization is constrained, we are not allowed to visit every point in the $(x, y)-$plane, we are only allowed to visit the points that satisfy the

constraint $V(x, y) = v$, which defines an implicit curve[6] $y(x)$, such that $V(x, y(x)) = v$. Let us visit all the allowed points, starting, say, at $x = 0$ and y such that $V(0, y) = v$. In the example (2.12) this corresponds to $-10y + 80 = 24.45$, or $y = 5.55$.

So, starting at the point $(x, y) = (0, 5.55)$, we make small steps (dx, dy) so as to always satisfy the constraint. dx and dy are therefore not independent, they are related by $dV = 0$ so that V remains equal to v:

$$dV = \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy = 0 \tag{2.13}$$

therefore the relation between dx and dy is:

$$dy = -\frac{\frac{\partial V}{\partial x}}{\frac{\partial V}{\partial y}}dx \tag{2.14}$$

this condition ensures that as we explore different points, they all satisfy the constraint (2.12).

If we are at an arbitrary point (x, y) not necessarily satisfying the constraint, and we move to $(x + dx, y + dy)$ with generic (dx, dy), the function U will change its value by:

$$dU = \frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy \tag{2.15}$$

The first order condition for a maximum, $dU = 0$, imply the well known

$$\frac{\partial U}{\partial x} = 0, \quad \frac{\partial U}{\partial y} = 0 \tag{2.16}$$

(we will not worry about second order conditions here.) If the point (x, y) do satisfies the constraint, and we move to a point $(x + dx, y + dy)$ that also satisfies the constraint, then dx and dy are related by (2.14), and the change in U given by (2.15) becomes

$$dU = \left(\frac{\partial U}{\partial x} - \frac{\partial U}{\partial y}\frac{\frac{\partial V}{\partial x}}{\frac{\partial V}{\partial y}}\right)dx \tag{2.17}$$

At the constrained maximum $dU = 0$, and (2.17) implies:

$$\frac{\frac{\partial U}{\partial x}}{\frac{\partial U}{\partial y}} = \frac{\frac{\partial V}{\partial x}}{\frac{\partial V}{\partial y}} \tag{2.18}$$

This is the generalization of (2.9) to nonlinear constraints. The geometric interpretation, as before, is the equality of the tangents of both, the constraint curve, and the level curve of U that happens to be tangent at some point to the constraint curve.

[6] In this work we do not bother by the possibility that this implicit curve may not be unique, etc. We assume as valid all the necessary technical assumptions to ensure uniqueness and smoothness of these implicitly defined functions.

As was the case with (2.9), (2.18) determines the *curve of optimal constrained maxima* for every possible value of the of the constraint $V = v$. The specific solution, for a specific value v, lies at the intersection between the curve of optimal constrained maxima and the specific constraint, i.e. the solution of the system of equations.

$$\frac{\frac{\partial U}{\partial x}}{\frac{\partial U}{\partial y}} = \frac{\frac{\partial V}{\partial x}}{\frac{\partial V}{\partial y}} \tag{2.19}$$

$$V(x, y) = v \tag{2.20}$$

In figure (5) we see various level curves of the function U in (2.11) for $\alpha = \beta = 1/2$, various constraint curves corresponding to different values of v in (2.12), and the curve of optimal constrained maxima (2.18), generalizing for nonlinear constraints the optimal consumption curve (2.9) in figure 3.

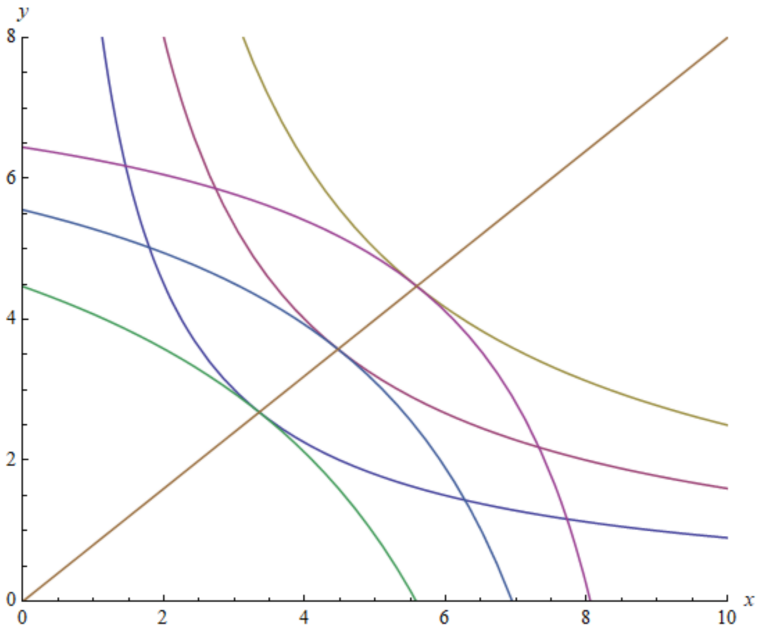


Figure 5: Curve of optimal constrained maxima (2.18) in orange.

2.2 Gradient field method to solve constrained maximization problems in 2-D with 1 constraint

The tangent method of equations (2.5-2.6) is directly motivated by economic intuition. This intuition grows into the economic student mind from early undergraduate courses, where they learn simple, low dimensional models, that stress what is viewed as the conceptual gist of the subject. It is also geometrically very intuitive to solve constrained maximization problems.

Unfortunately, as we will see in section 2.3, it does not generalizes nicely to more than two dimensions and more than one constraint. And we happen to be living in the "Big Data−Machine

Learning−Artificial Intelligence−Era", with gigantic databases filled with data of extremely high dimensionality that promises to change the practice, and perhaps even the conceptual basis, of economics and other social sciences. It may be appropriate then to adapt the mathematics that economic students learn, from the get go, towards more powerful methods that mesh better with the data friendly techniques of the new Era.

In this section we will solve again the 2-D, 1-constraint problem solved in section 2.1.2, or, more accurately, reinterpret that solution, with the *gradient field method*, which is geometrically equally intuitive, and will be shown in section 2.3 to easily generalize to any dimensions and any number of constraints.

To be clear, what we call here the "gradient field method" is none other than the Lagrange multipliers method, that economic students also learn. The only difference is that while the method is typically taught as a recipe that happens to give the right result, we present it here emphasizing its geometrical intuition. In this way it becomes at least as intuitive as the tangent method, and this intuition doesn't fade in high dimensional problems with multiple constraints. This is crucial to bridge the connections between economic problems and machine learning techniques.

Let us briefly remaind the reader what the gradient field is (for more details see [Pernice (2018 b)]). From this point on, we will make heavy use of vector algebra, that by this point was already covered in the course. There are many excellent textbooks and YouTube videos about basic linear algebra, for example Grant Sanderson's Essence of linear algebra. For consistency with the notation of the present work the reader may also consult [Pernice (2019)], or, the more concise [Pernice (2020)]. Here we remind the reader the very basics.

We use **bold face** for vectors, as in $\boldsymbol{a}$. Unless otherwise specified, vectors are meant as column vectors,

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{2.21}$$

Remember that the scalar product between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is

$$\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a} = \boldsymbol{b}^\top \boldsymbol{a} = (b_1, b_2) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = b_1 a_1 + b_2 a_2 = \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \cos(\theta) \tag{2.22}$$

The dot "$\cdot$" means scalar product. It is commutative (first equality). In the second equality, $\boldsymbol{b}^\top$ is the transpose of the vector $\boldsymbol{b}$, so $\boldsymbol{b}^\top$ is a row vector, and the scalar product is written as a *matrix product*, where $\boldsymbol{b}^\top$ is viewed as a 1×2 matrix and $\boldsymbol{a}$ as a 2×1 matrix.

In the last equality, $\|\boldsymbol{a}\| \equiv \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}}$, is the *modulus*, or *length*, of the vector $\boldsymbol{a}$, and the same for the vector $\boldsymbol{b}$, and θ is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$. Since $-1 \leq \cos(\theta) \leq 1$, (2.22) indicates that keeping the length of $\boldsymbol{a}$ and $\boldsymbol{b}$ constant, the absolute value of $\boldsymbol{a} \cdot \boldsymbol{b}$ is maximum when the two vectors are parallel ($\theta = 0$) or anti-parallel ($\theta = \pi$). It also says that two no-null vectors are perpendicular to each other, $\theta = \pm\pi/2$, if and only if their scalar product is zero ($\cos(\pi/2) = \cos(-\pi/2) = 0$).

A function $U(x, y) : \mathbb{R}^2 \to \mathbb{R}$ assigns a real number to each point in the (x, y)−plane. A *vector field* $\boldsymbol{F}(x, y) : \mathbb{R}^2 \to \mathbb{R}^2$ assigns a real vector to each point in the (x, y)−plane. For any sufficiently

smooth function $U(x, y)$, the *gradient* ∇U is the vector field defined by

$$\nabla U(x, y) = \frac{\partial U}{\partial x}\hat{\boldsymbol{x}} + \frac{\partial U}{\partial y}\hat{\boldsymbol{y}} \tag{2.23}$$

$\hat{\boldsymbol{x}}$ is the unit vector in the horizontal direction and $\hat{\boldsymbol{y}}$ is the unit vector in the vertical direction. They form an orthonormal basis of $\mathbb{R}^2$: $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}} = \hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}} = 1$ and $\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = 0$.

Equation (2.15) for the change in U when x and y change, respectively, by dx and dy, can be written in vector notation in terms of the gradient field as

$$dU(x, y) = \nabla U \cdot \boldsymbol{dr} = (\nabla U)^\top \boldsymbol{dr} = \left(\frac{\partial U}{\partial x}, \frac{\partial U}{\partial y}\right)\begin{pmatrix} dx \\ dy \end{pmatrix} = \frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy \tag{2.24}$$

Equation (2.24) says that at any point (x, y), dU can be seen as the scalar product of the gradient vector field ∇U at that point, and the displacement vector $\boldsymbol{dr} = dx\,\hat{\boldsymbol{x}} + dy\,\hat{\boldsymbol{y}}$. Applying the last equality in (2.22) to (2.24), the change dU when the independent variables change by dx and dy, i.e., when the displacement vector in the $(x, y)-$plane is $\boldsymbol{dr} = dx\,\hat{\boldsymbol{x}} + dy\,\hat{\boldsymbol{y}}$, is

$$dU(x, y) = \nabla U \cdot \boldsymbol{dr} = \|\nabla U\|\,\|\boldsymbol{dr}\|\,\cos(\theta) \tag{2.25}$$

This means that, keeping the length of the displacement $\|\boldsymbol{dr}\|$ fixed, the magnitude of the change in U is maximum when the displacement is parallel ("steepest ascent") or anti-parallel ("steepest descent") to the gradient ∇U at that point. Also, if we want to move in a level curve of U, i.e., such that $dU = 0$, the displacement $\boldsymbol{dr}$ has to be perpendicular to the gradient ∇U at that point. In other words, given the function $U(x, y)$, calculate the gradient ∇U at each point, and the level curves of $U(x, y)$ at any point will be perpendicular to the gradient at that point, as can be appreciated in the right side of figures 6 and 7 for the functions $U(x, y) = x^2 + xy + 2y^2$ and $U(x, y) = x^2 - y^2$ respectively.

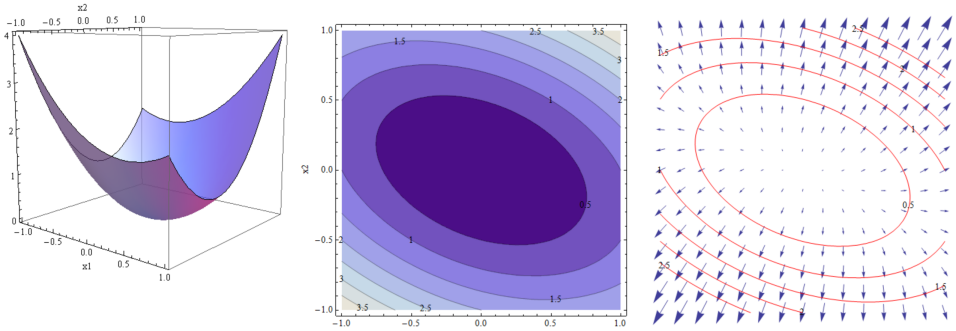


Figure 6: Left: 3-D view of $U(x, y) = x^2 + xy + 2y^2$. Center: level curves of U. Right: the gradient $\nabla U = (2x + y)\hat{\boldsymbol{x}} + (x + 4y)\hat{\boldsymbol{y}}$, and some level curves superposed.

The unconstrained first order conditions (2.16) for a maximum, a minimum or a saddle point, mean that at these points the gradient becomes zero: if $\nabla U = \boldsymbol{0}$, $dU = 0$ in equation (2.25) independently of the direction of the displacement $\boldsymbol{dr}$. In fact, provided the function is sufficiently

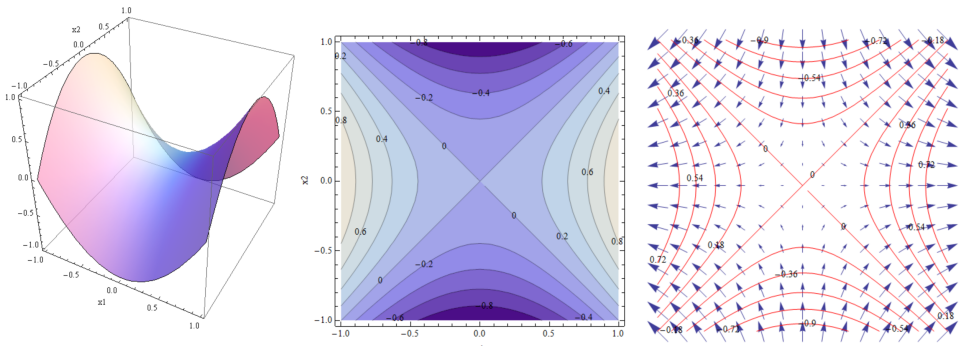


Figure 7: Left: 3-D view of $U(x, y) = x^2 - y^2$. Center: level curves of U. Right: the gradient $\boldsymbol{\nabla} U = 2x\hat{\boldsymbol{x}} - 2y\hat{\boldsymbol{y}}$, and some level curves superposed.

smooth, the magnitude of the gradient continually decreases as it approaches the critical point, becoming zero at the critical point. This can be appreciated in the right part of figure 6 for a minimum and in the right part of figure 7 for a saddle point. In these figures the gradient vector field is scaled appropriately for viewing purposes.

Incidentally, the "steepest descent method" for minimizing functions (starting at some point in the space of independent variables and moving in the direction opposite to the gradient at a velocity proportional to the length of the gradient) turns out to be spectacularly efficient in very high dimensions and it, or some stochastic variants of it, is the method of choice to minimize, or "train", cost functions of deep neural networks in machine learning.

Returning to our constrained maximization problem, let us reinterpret the equations used in section 2.1.2 in vector notation.

The relation between dx and dy in (2.14), that ensures that we move in a level curve of the constraint function V, imply that the displacement vector is:

$$\boldsymbol{dr} = \begin{pmatrix} dx \\ -\frac{\frac{\partial V}{\partial x}}{\frac{\partial V}{\partial y}} dx \end{pmatrix} \propto \begin{pmatrix} \frac{\partial V}{\partial y} \\ -\frac{\partial V}{\partial x} \end{pmatrix} \tag{2.26}$$

The symbol '$\propto$' means 'proportional to', and comes simply by multiplying the displacement vector $\boldsymbol{dr}$ by the scalar $\frac{\partial V}{\partial y}/dx$. Remember that multiplying a vector by a scalar simply changes the scale (which can even be negative) but not direction. (2.26) shows that this $\boldsymbol{dr}$ is indeed perpendicular to the gradient of V:

$$\boldsymbol{\nabla} V \cdot \boldsymbol{dr} = (\boldsymbol{\nabla} V)^{\top} \boldsymbol{dr} \propto \left(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y} \right) \begin{pmatrix} \frac{\partial V}{\partial y} \\ -\frac{\partial V}{\partial x} \end{pmatrix} = \frac{\partial V}{\partial x}\frac{\partial V}{\partial y} - \frac{\partial V}{\partial y}\frac{\partial V}{\partial x} = 0 \tag{2.27}$$

so, if we want to change the coordinates (x, y) by (dx, dy) so that we remain in the same level curve of V, the displacement vector has to be perpendicular to the gradient of V, as explained for a generic function in equation (2.25).

Let us now reinterpret equation (2.18), that embodies the fact, obvious to the naked eye in figure 4, that at the constrained optimum the constraint curve $V(x, y(x)) = v$ is tangent to the maximum level curve of the function U compatible with the constraint.

Consider the function $V(x, y)$ as function on its own right, with the same status as $U(x, y)$, i.e., a function of *independent* variables x and y, not restricted by the constraint $V(x, y) = v$. As for any other function, the gradient $\boldsymbol{\nabla} V$ is perpendicular at any point to the level curve of V passing through such point, in particular to the level curve $V(x, y) = v$ defining the constraint.

Similarly, the gradient $\boldsymbol{\nabla} U$ will be perpendicular at any point to the level curve of U passing through such point, in particular to the maximum level curve of the function U compatible with the constraint at the constrained maximum.

But since the level curve $V(x, y) = v$ and the maximum level curve of the function U compatible with the constraint are tangent to each other at the constrained maximum (this is the content of equation (2.18)), their respective gradients have to be proportional to each other (linearly dependent of each other). Let us see how this is the case:

$$\begin{aligned}
&\frac{\frac{\partial U}{\partial x}}{\frac{\partial U}{\partial y}} = \frac{\frac{\partial V}{\partial x}}{\frac{\partial V}{\partial y}} \Leftrightarrow \frac{\frac{\partial U}{\partial x}}{\frac{\partial V}{\partial x}} = \frac{\frac{\partial U}{\partial y}}{\frac{\partial V}{\partial y}} \equiv \lambda \Leftrightarrow \\
&\frac{\partial U}{\partial x} = \lambda \frac{\partial V}{\partial x} \quad \text{and} \quad \frac{\partial U}{\partial y} = \lambda \frac{\partial V}{\partial y} \Leftrightarrow \\
&\begin{pmatrix} \partial U/\partial x \\ \partial U/\partial y \end{pmatrix} = \lambda \begin{pmatrix} \partial V/\partial x \\ \partial V/\partial y \end{pmatrix} \Leftrightarrow \\
&\boldsymbol{\nabla} U = \lambda \boldsymbol{\nabla} V
\end{aligned} \tag{2.28}$$

In figure 8 we show the level curves of two functions and the linear dependence of their respective gradients at the constrained optimum.

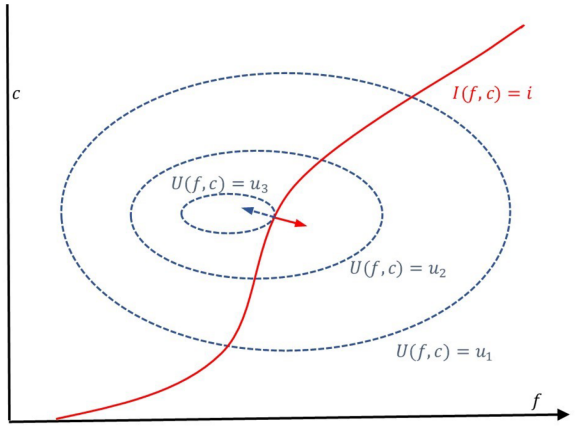


Figure 8: Linear dependence of the gradients of U and I at the maximum.

Note that the above argument does not imply that if the level curves are tangent their respective gradients are equal, it does not even imply that they point in the same direction, it only implies

that they are linearly dependent of each other. The angle between them may be 0 or it may be π, no other angle is possible. But the relation between the respective length is completely undetermined. That is why we introduce the new constant λ in (2.28), whose value has to be determined as part of the solution of the problem.

So, the above argument shows that at the maximum of U, subject to the constraint $V = v$, $\boldsymbol{\nabla} U = \lambda \boldsymbol{\nabla} V$. These are two equations, one for each component of the gradient, but now we have three variables, since we introduced the proportionality variable λ. The additional equation is the constraint itself. So, in the gradient method, the system of two equations (2.19-2.20) is replaced by the system of *three* equations

$$\boldsymbol{\nabla} U = \lambda \boldsymbol{\nabla} V \tag{2.29}$$

$$V(x, y) = v \tag{2.30}$$

as we will see, these equations, and natural generalizations based on analogous geometrical intuitions, solve the problem with any number of independent variables and any number of constraints. But before we show that, we would like to point out two things.

The first point is that you can derive equations (2.29-2.30) by the recipe of optimizing the so-called "Lagrangian function"

$$\mathcal{L}(x, y, \lambda) = U(x, y) - \lambda(V(x, y) - v) \tag{2.31}$$

with respect to the three variables x, y, and λ. This last variable is known as a *Lagrange multiplier*. The reader can easily take the first derivative of $\mathcal{L}$ with respect to each variable, equate it to zero, and see that it reproduces the three equations (2.29-2.30).

So, as we mentioned at the beginning of the section, the gradient method is really the Lagrange method for solving constrained optimization problems. We derived equations (2.29-2.30) in a different way to emphasise the geometric meaning of the method. Unfortunately, it is typically presented as a vaguely justified recipe to arrive at the right equations, blurring the richness of the Lagrange method.

The second point, that will become very significant later in the paper, is simply the observation that, to derive the constrained optimum solution by the gradient method, we were naturally led to consider the constraint function $V(x, y)$ on an equal footing than the function $U(x, y)$ to be maximized. For both functions it is true that their gradient is perpendicular to their respective level curves, and since at the optimum the level curves are tangent, their gradients must be linearly dependent at that point. One is then naturally lead to the conclusion that maximizing U under the constraint $V = v$ is equivalent to maximizing V under the constraint $U = u$, *for some* u.

One could argue that this is in fact not the case, because even though one can write equation (2.29) as $\boldsymbol{\nabla} V = \beta \boldsymbol{\nabla} U$, with $\beta = 1/\lambda$, equation (2.30) only refers to the function V, breaking the equivalence between the two problems.

This counterargument is in a literal sense obviously correct, but it misses the point that while equation (2.29), or its equivalent $\boldsymbol{\nabla} V = \beta \boldsymbol{\nabla} U$, are *intrinsic* to the constraint maximization problem, equation (2.30) merely reflects the idiosyncratic fact that we happen to know the value of

V of the final solution and not the value of U. But it is easy to conceive an equivalent situation in which we know the value of U and ignore the value of V. In another words, equation (2.30) reflects the prior knowledge of the person solving the problem, but someone else might know the value of U, since there is a one to one relation between them.

Needless to say, for some problems, a prior knowledge of V (for example the dollar value of the budget constraint) is *naturally* easier to know a priori than the value of U (for example the utility). But it is conceivable a problem with the same underlying mathematics, but where the known quantity is the value of the function U. We will see that this in fact the case in section 3. The point is that if we consider the constrained maximization problem abstractly, the equivalence becomes transparent. This observation will be very important later in the paper.

2.3 Generalization to n dimensions and m constraints, or why the gradient field method is far more versatile than the graphic method

2.3.1 n dimensions and 1 linear constraint

Consider the problem of *minimizing* the function $U : \mathbb{R}^3 \to \mathbb{R}$

$$\text{Minimize} \qquad U(x, y, z) = x^2 + y^2 + z^2 \tag{2.32}$$

$$\text{Under the constraint} \qquad V(x, y, z) = x = 0.8 \tag{2.33}$$

Since the function $U(x, y, z)$ has three independent variables, it has level "surfaces" rather than curves, which are spherical surfaces centered at the origin and of different radius, shown in figure 9.

In figure 10 we see three different level surfaces of U and the constraint surface (plane) $V(x, y, z) = x = 0.8$. The smallest level surface of U (top left) is too small and does not have any point in the surface constraint. The largest (bottom left) has infinite points in the surface constraint, forming a circle, but since we want to minimize U subject to the constraint, this is clearly not the optimum one. The level surface on the right of figure 10, with only one point in common with the surface constraint, is clearly the right one.

Note that at the minimum, the "tangent plane" of the level surface of U coincides with the tangent plane of the level surface of V (which, since V is linear, it happens to also coincide with the surface constraint itself). But how can the tangent plane be characterized? Characterizing it by curves in the level surface passing through the optimal point is not very economical, since there are infinitely many such curves: the level surface is 2-D, so there are infinitely many different 1-D curves in the surface passing through a unique point.

The easiest way to characterize the plane tangent to a smooth surface at a given point is by a vector orthogonal to the surface at that point. In 3-D, there is a *unique* direction orthogonal to a given 2-D surface at any specific point. And since the surfaces we are interested in happen to be level surfaces of a given function, and the gradient is orthogonal to the level surfaces, the gradient must point in the unique orthogonal direction to the surface of interest!

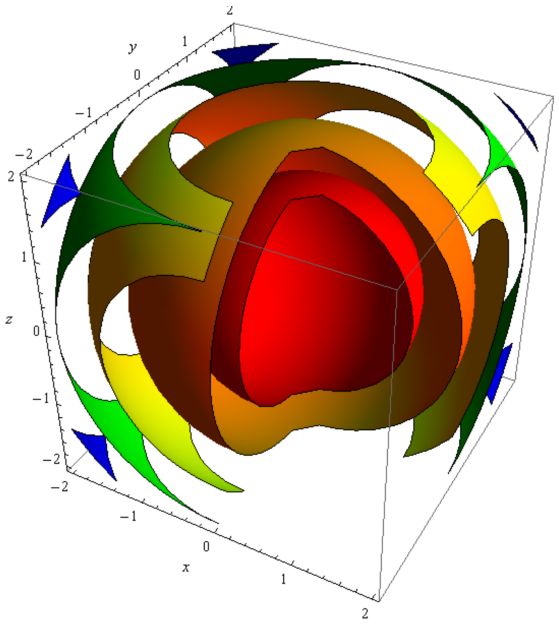


Figure 9: Level surfaces of the function $U(x, y, z) = x^2 + y^2 + z^2$.

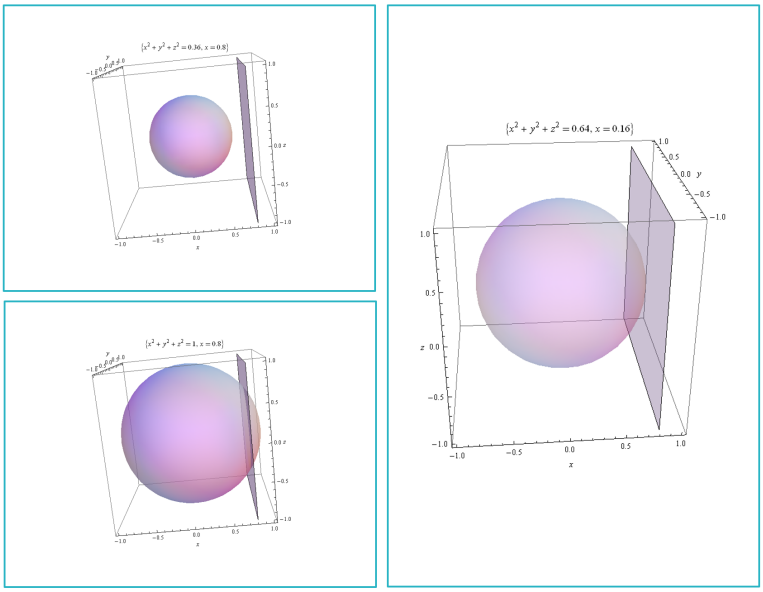


Figure 10: Three different level surfaces of U and constraint surface (plane) $V(x, y, z) = x = 0.8$.

Before we continue with the gradient, it is important to realize that the equalities in (2.22) are valid in *any* dimension: if

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \tag{2.34}$$

then

$$\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a} = \boldsymbol{b}^\top \boldsymbol{a} = \sum_{i=1}^{n} a_i b_i = \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \cos(\theta) \tag{2.35}$$

And, as in 2-D, $\|\boldsymbol{a}\| \equiv \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}}$ and the same for $\|\boldsymbol{b}\|$.

Returning to the gradient and level surfaces, (2.35) imply that the arguments in equations (2.24-2.25) remain valid in 3-D, and in fact in *any* dimension:

$$dU(x, y) = \boldsymbol{\nabla} U \cdot \boldsymbol{dr} = \sum_{i=1}^{n} (\boldsymbol{\nabla} U)_i \boldsymbol{dr}_i = \|\boldsymbol{\nabla} U\| \, \|\boldsymbol{dr}\| \cos(\theta) \tag{2.36}$$

Therefore, the qualitative arguments around equations (2.24-2.25) also generalize to any dimension: *the gradient of a any function U of n independent variables is perpendicular to the $(n-1)-$dimensional level hypersurfaces of U.*

This means that the gradient is perpendicular to everyone of the infinitely-many possible displacements $\boldsymbol{dr} = \sum_{i=1}^{n} dx_i \hat{\boldsymbol{x}}_i$ that, starting in any given point, leave U unchanged, or $dU = 0$ (there are $n-1$ linearly independent displacements that satisfy $dU = 0$). Notice how much information is efficiently encapsulated in the gradient of a function.

Putting all of the above together, since at the local optimum the $(n-1)-$dimensional level surface of the function U is tangent to the $(n-1)-$dimensional constraint surface $V = v$, and their gradients $\boldsymbol{\nabla} U$ and $\boldsymbol{\nabla} V$ are orthogonal to their respective surface, they have no choice but lying in the unique orthogonal direction to these surfaces. They must be linearly dependent for *any* number of independent variables.

Let us consider the specific example in (2.32-2.33) and compute the gradients of U and V:

$$\boldsymbol{\nabla} U = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}; \quad \boldsymbol{\nabla} V = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{2.37}$$

we want to find points in which $\boldsymbol{\nabla} U$ and $\boldsymbol{\nabla} V$ are linearly dependent. The null second and third component of $\boldsymbol{\nabla} V$ imply that the linear dependence will happen only in the x axis: $y = z = 0$. So the x axis is the "$V-$constrained optimal curve", and the constant of proportionality (also known as Lagrange multiplier) is

$$2x = \lambda * 1, \quad \text{or} \quad \lambda = 2x \tag{2.38}$$

Now we need to find the minimum for the specific value of the constraint $V = x = 0.8$. (2.37) and (2.38) imply that the minimum is at $(x, y, z) = (0.8, 0, 0)$, and the problem is finished!

Well, not quite, in fact, we still need to prove that $(0.8, 0, 0)$ is in fact a minimum, and not a maximum or, in general a critical point, with an analysis of the second order conditions. But as we mentioned earlier, we will not bother about second order conditions in this paper. So, for the purposes of this paper the problem is solved.

2.3.2 n dimensions and 1 non-linear constraint

Consider now the problem of the same function $U : \mathbb{R}^3 \to \mathbb{R}$ in (2.32) under a *nonlinear* constraint:

$$\text{Minimize} \qquad U(x, y, z) = x^2 + y^2 + z^2 \tag{2.39}$$

$$\text{Under the constraint} \qquad V(x, y, z) = (x - 1)^2 + y^2 + z^2 = 0.16 = 0.4^2 \tag{2.40}$$

Note that the level curves of $V(x, y, z) = (x - 1)^2 + y^2 + z^2$ are spherical surfaces centered at $x = 1, y = x = 0$.

Figure 11 shows that an argument similar to the one done to describe Figure 10 implies that, at the constraint minimum, the optimal level surface of U and the constraint surface (2.40) are tangent, as it was the case for a linear constraint.

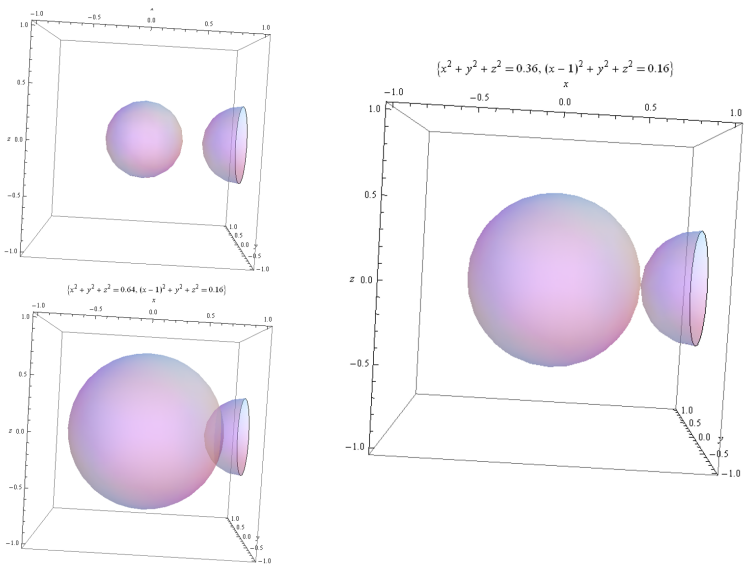


Figure 11: Three different level curves of U and *half* of the constraint nonlinear surface $V(x, y, z) = (x - 1)^2 + y^2 + z^2 = 0.16 = 0.4^2$.

So, at the constraint minimum the gradients of U and V

$$\boldsymbol{\nabla} U = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}; \quad \boldsymbol{\nabla} V = \begin{pmatrix} 2(x-1) \\ 2y \\ 2z \end{pmatrix} \tag{2.41}$$

have to be linearly dependent: $\boldsymbol{\nabla} U = \lambda \boldsymbol{\nabla} V$. This means that the proportionality has to be valid for the three components simultaneously, i.e., with the *same* constant of proportionality. The proportionality of the x component imply that

$$2x = \lambda 2(x - 1), \quad \text{or} \quad \lambda = \frac{x}{x - 1} \tag{2.42}$$

inserting this λ into the proportionality of the y component we have:

$$2y = \frac{x}{x-1}2y \quad \Rightarrow \quad \left(\frac{x}{x-1} = 1 \quad \text{if} \quad y \neq 0\right) \quad \text{or} \quad y = 0 \tag{2.43}$$

But as the reader can easily check, there is no x that satisfies $x/(x-1) = 1$! Therefore the hypothesis $y \neq 0$ is wrong and $y = 0$. With a similar argument $z = 0$. So, the "nonlinearly $V-$constrained optimal curve" is again the $x-$axis (one may except the points $x = 0$ and $x = 1$, because the level surface of U become a point at $x = 0$ and the level surface of V become a point at $x = 1$, see equation (2.42)).

For the particular case $V = 0.4^2$, since we already know that at the optimum $y = z = 0$, (2.40) becomes

$$(x-1)^2 = 0.4^2, \quad \text{or} \quad x - 1 = \pm 0.4 \tag{2.44}$$

which has two solutions: $x = 0.6$, corresponding to the constrained minimum on the right hand side of figure 11, and $x = 1.4$, which we didn't expect from figure 11! We left as an exercise to the reader to figure out what this $x = 1.4$ solution means (hint: in figure 11 only *half* of the constraint nonlinear surface $V = 0.4^2$ is shown.)

2.3.3 n dimensions and m constraints

Consider now the problem of the same function $U : \mathbb{R}^3 \to \mathbb{R}$ in (2.32) under *two* constraints:

$$\text{Minimize} \qquad U(x, y, z) = x^2 + y^2 + z^2 \tag{2.45}$$

$$\text{Under the constraints} \qquad V_1(x, y, z) = x = 0.7 \tag{2.46}$$

$$V_2(x, y, z) = y = 0.5 \tag{2.47}$$

In figure 12 we see four different level surfaces of U, each one with the constraint planes $V_1(x, y, z) = x = 0.7$ and $V_2(x, y, z) = y = 0.5$. In the top left $U = 0.09$, and the level surface is a spherical surface centered at the origin and of radius 0.3 ($0.3^2 = 0.09$.) No point of this level surface is consistent with any of the constraints. The bottom left level surface has radius 0.5 ($U = 0.5^2 = 0.25$). One point of this surface is consistent with the constraint $V_2 = 0.5$ but none with the constraint $V_1 = 0.7$. The top right level surface has radius 0.7 ($U = 0.7^2 = 0.49$). Infinite points of this surface, forming a circle, are consistent with the constraint $V_2 = 0.5$ and one point is consistent with the constraint $V_1 = 0.7$. However, no point is consistent with both constraints simultaneously, which is what we want as a solution to the problem (2.45-2.47). Finally, the bottom right level surface has radius 1 ($U = 1^2 = 1$). Infinite points of this surface, forming a circle, are consistent with the constraint $V_2 = 0.5$ and also infinite point, forming another circle, are consistent with the constraint $V_1 = 0.7$. Two points are consistent with both constraints, but clearly this level surface is not the minimum one consistent with both constraints.

In figure 13 we can see, from two different perspectives, the minimum level surface. It is tangent to the line $(0.7, 0.5, z)$, for arbitrary z, and this line is common to both surface constraints. This suggest an algorithm to solve problems with more than one constraint: find the intersection of both 2-D constraints so that, generically, one transforms the problem into a problem with only

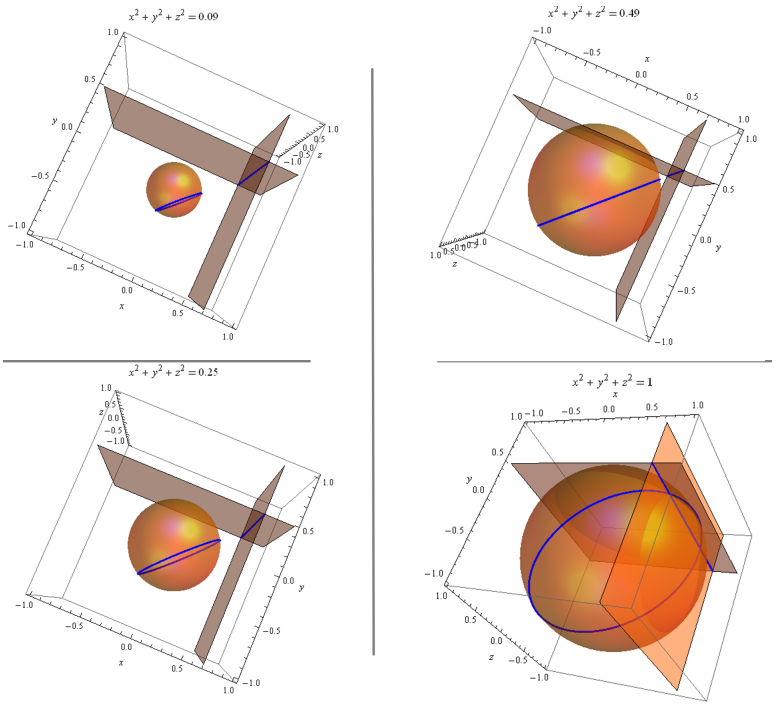


Figure 12: Four different level surfaces of U and the two constraints $V_1(x, y, z) = x = 0.7$ and $V_2(x, y, z) = y = 0.5$.

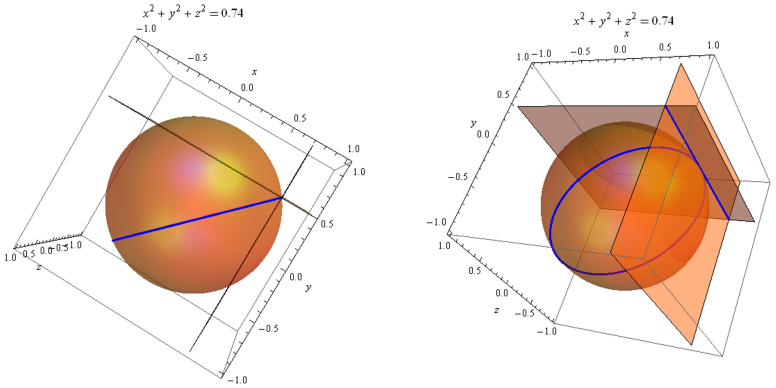


Figure 13: Two perspectives of the optimal level surface of U consistent with the two constraints $V_1(x, y, z) = x = 0.7$ and $V_2(x, y, z) = y = 0.5$.

one 1-D constraint. The problem with this algorithm is that, although for the example (2.45-2.47), cherry picked for easy visualization, it is very easy to implement in practice, for more complex, nonlinear constraints, finding the points common to both constraints tends to be as difficult as the problem we want to solve in the first place!

The gradient, again, is the superior solution. The intersection of the two constraints, by definition of intersection, necessarily lies in both surface constraints. Since the gradient of each surface constraint is perpendicular in every point to a neighborhood of that point in the surface, the

gradient of each surface constraint must be perpendicular to their intersection. In another words, the scalar product of each gradient with a vector pointing in the direction of the intersection curve must be zero. Let us see how this works in the example (2.45-2.47).

The vector $\hat{\boldsymbol{z}}$ clearly points in the direction of the line $(0.7, 0.5, z)$, which is the intersection of the surface constraints (2.46) and (2.47). The gradients of the constraint functions are

$$\boldsymbol{\nabla} V_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\nabla} V_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{2.48}$$

Clearly $\boldsymbol{\nabla} V_1 \cdot \hat{\boldsymbol{z}} = 0$ and $\boldsymbol{\nabla} V_2 \cdot \hat{\boldsymbol{z}} = 0$, which is what we wanted to show.

The vanishing of the scalar product between the gradients of the surface constraints and any vector tangent to the intersection of these constraints is valid in general, linear or nonlinear, and in any dimension. The reason is that, as explained before, the intersection lies in both (hyper)surface constraints, and each gradient is orthogonal to its corresponding (hyper)surface, therefore they must, in particular, be orthogonal to their intersection. So we have a property valid in any dimension for any number of constraints!

If $\boldsymbol{\nabla} V_1 \cdot \hat{\boldsymbol{z}} = 0$ and $\boldsymbol{\nabla} V_2 \cdot \hat{\boldsymbol{z}} = 0$, then, for any linear combination

$$\hat{\boldsymbol{z}} \cdot (\lambda_1 \boldsymbol{\nabla} V_1 + \lambda_2 \boldsymbol{\nabla} V_2) = 0 \tag{2.49}$$

This means that any linear combination of the gradients is also perpendicular to the tangent of the intersecting line.

Returning to figure 13, we noted already that the optimal level surface of U (the function being constraint-minimized) is tangent to the intersection of the two constraint surfaces, i.e., the line $(0.7, 0.5, z)$. This means that at the optimum, the gradient of U, $\boldsymbol{\nabla} U$, is perpendicular to this line. But if it is perpendicular to the line, it must be a linear combination of the gradients of the constraint functions:

$$\boldsymbol{\nabla} U = \lambda_1 \boldsymbol{\nabla} V_1 + \lambda_2 \boldsymbol{\nabla} V_2 \tag{2.50}$$

This equation (which are really three equations in 3-D) is the analogous in a two-constraint problem to equation (2.29) for one-constraint problems. In the same sense in which (2.29) determines the optimal constrained curve, (2.50) determines the "optimal constrained surface" in a two constraint problem, corresponding to optima for different values v_1, v_2 of the constraints $V_1(x, y, z) = v_1$ and $V_2(x, y, z) = v_2$.

Now we have five variables: x, y, z, λ_1 and λ_2, and (2.50) are 3 equations (in R^n (2.50) would imply n equations, and we would then have $n + 2$ variables if the problem had two constraints.) Where are the other two equations? The other two equations are the specific constraints (2.46) and (2.47). So, our system or equations is

$$\boldsymbol{\nabla} U = \lambda_1 \boldsymbol{\nabla} V_1 + \lambda_2 \boldsymbol{\nabla} V_2 \tag{2.51}$$

$$V_1(x, y, z) = v_1 \tag{2.52}$$

$$V_2(x, y, z) = v_2 \tag{2.53}$$

As it was the case for one constraint, where equations (2.29-2.30) could be derived from the Lagrangian (2.31), (2.51-2.53) can be derived from the Lagrangian

$$\mathcal{L}(x, y, z, \lambda_1, \lambda_2) = U(x, y, z) - \lambda_1(V_1(x, y, z) - v_1) - \lambda_2(V_2(x, y, z) - v_2) \tag{2.54}$$

optimizing for each of the five variables x, y, z, λ_1 and λ_2. As already mentioned, we are emphasising the geometrical intuition of the Lagrangian method for solving optimization problems with constraints.

Let us apply all this to the specific problem (2.45-2.47). The respective gradients are:

$$\boldsymbol{\nabla} U = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}, \quad \boldsymbol{\nabla} V_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\nabla} V_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{2.55}$$

applying (2.51) to (2.55) implies that $z = 0$, $2x = \lambda_1$, $2y = \lambda_2$. So the optimal constrained surface is the $(x, y)-$plane, or the $z = 0$ plane. This is the plane of constrained minima for all possible values of v_1 and v_2. For the specific values (2.46) and (2.47), the constrained minimum is $(0.7, 0.5, 0)$, as it was already obvious from figure 13.

With exactly analogous geometrical/linear algebra reasoning, if the function U being optimized has n independent variables, and m constraints $V_i = v_i, i = 1, \ldots, m$, with $n \geq m$ (otherwise, in the generic case, there won't be any solution), the optimum $(n-1)-$dimensional level hypersurface of U has to be tangent to the $(n-m)-$dimensional hypersurface of intersection of the m constraints. This means that the gradient $\boldsymbol{\nabla} U$ has to be linearly dependent of the m gradients $\boldsymbol{\nabla} V_i$, so the generalization of (2.51-2.53) is

$$\boldsymbol{\nabla} U(x_1, \ldots, x_n) = \sum_{i=1}^{m} \lambda_i \boldsymbol{\nabla} V_i \tag{2.56}$$

$$V_i(x_1, \ldots, x_n) = v_i, \qquad i = 1, \ldots, m \tag{2.57}$$

(2.56) is really n equations with $n + m$ variables, (2.57) provides that additional m equations. As promised, once we get used to the idea that, in any dimension, orthogonality between vectors is synonymous to vanishing of the scalar product between these vectors , the geometric intuition remains intact, even for this extremely general case.

Keeping the intuition while working with this level of generality is what is needed to navigate optimization problems in the Big−data−Machine−learning−Artificial−intelligence Era.

2.4 The Gradient method and the equivalence between constrained optimization problems

The gradient method of sections 2.2 and 2.3 to solve constraint optimization problems maintains the geometrical intuition of the standard tangency-of-the-level-curves-method of section 2.1, but the mathematical instruments used to concretize such intuition is far more powerful and easily generalizes to any number of independent variables and any number of constraints.

It is nothing other than the Lagrange multiplier method, but emphasising the geometrical intuitions behind it, rather than presenting it as a cooking recipe, as is done in many standard textbooks and courses in economics.

Interestingly for what comes next, the gradient method makes transparent that there is an equivalence between optimizing a function U under a constraint determined by a function V and optimizing the function V under a constraint determined by function the U. And this equivalence extends to any dimension and any number of constraints.

For example, equation (2.56) indicates that $\boldsymbol{\nabla} U$ is linearly dependent of the gradients $\boldsymbol{\nabla} V_i$, $i = 1, \ldots, m$. More deeply, it indicates that these $m + 1$ vectors belong to the same m-D subspace in the embedding n-D space, and any m of these $m + 1$ vectors can act as a basis of this subspace. So, this equation could equivalently be expressed as, say, $\boldsymbol{\nabla} V_1$ being linearly dependent of $\boldsymbol{\nabla} U$ and $\boldsymbol{\nabla} V_i$ for $i = 2, \ldots, m$, in which case we would be describing the problem as optimizing V_1 under the constraints determined by U and the functions V_i's.

As was pointed out before in the case of only one constraint, equations (2.57) clearly breaks the symmetry between U, and the V_i's, but this is not intrinsic to the *mathematical* problem, it just means that it happens to be the case that in this particular problem we know the values of the V_i's, and we don't know, and want to know, the value of U. But which functions are the known ones and which is the unknown could be different in other circumstances, or other problems.

It could happen that in some problems it is more *natural* to know a priori the values of some functions, such as the V_i's, rather than the value of some other function such as U, but what we want to point out here is that if this was the case, it would be a matter of the specific application, not an intrinsic feature of the mathematical problem. Mathematically all these problems are equivalent.

3 General equilibrium problems in microeconomics

Let us consider two apparently very different problems.

Problem 1: consider a policy maker P who has to decide how much to produce of goods X and Y, knowing that they both require labor L, and capital K, which are limited by the amounts $\bar{L}$, and $\bar{K}$ respectively, and also knowing that the technology for the production of these goods is given by the functions

$$X = f(k_x, \ell_x) \tag{3.1}$$

$$Y = g\left(k_y, \ell_y\right) \tag{3.2}$$

$$k_x + k_y = \bar{K}, \qquad \ell_x + \ell_y = \bar{L} \tag{3.3}$$

where ℓ_x and k_x are the units of labor and capital allocated to the production of good X, and ℓ_y and k_y are the units of labor and capital allocated to the production of good Y. *How should she decide this allocation for efficient production?*

The problem has two aspects, one corresponds to the more engineering-type problem of how to

produce efficiently, the other is how much to produce of each good given the preferences of the consumers and the costs of production.

It may not be obvious a priori that the problem of efficient production and the utility maximizing problem can be separated. It might be argued that the very *meaning* of "efficient production" is the production that maximizes utility. It is then appropriate to clarify this point from the get go.

Remember the notion of Pareto efficiency, a distribution of goods is Pareto efficient if no individual can be made better off without making other individual worse off. Extrapolating this notion to a production setting, given the technology (3.1-3.2) and the resources (3.3), a production of the goods X and Y is efficient if *no additional units of one good can be produced without reducing the units produced of the other good.*

Having clarified the separation between the problem of efficient production and the problem of utility maximization in problem 1, we will focus here in the first aspect only: efficient production.

Problem 2: consider a two consumer-two goods exchange economy. The consumers are Alice (A) and Bob (B) and the goods are X and Y. Alice has an initial endowment of X_a units of good X and Y_a units of good Y. Bob has X_b units of good X and Y_b units of good Y. Assume that the number of units of both goods are fixed: $X_a + X_b = \bar{X}$, $Y_a + Y_b = \bar{Y}$, $\bar{X}$ and $\bar{Y}$ fixed.

Alice and Bob both benefit from voluntarily trading goods with each other. Mathematically, both Alice and Bob can and want to increase their respective utilities $U_a(x_a, y_a)$ and $U_b(x_b, y_b)$ by trading. So the mathematical problem is:

$$\text{A wants to maximize} \quad U_a(x_a, y_a) \quad \text{with an initial endowment:} \quad x_a = X_a,\ y_a = Y_a \tag{3.4}$$

$$\text{B wants to maximize} \quad U_b(x_b, y_b) \quad \text{with an initial endowment:} \quad x_b = X_b,\ y_b = Y_b \tag{3.5}$$

$$x_a + x_b = \bar{X}, \qquad y_a + y_b = \bar{Y}, \qquad \bar{X} \text{ and } \bar{Y} \text{ fixed} \tag{3.6}$$

What is going to happen as a result of voluntary trade between Alice and Bob?

Problems 1 and 2 epitomize what microeconomics is all a about. They represent an important conceptual milestone in the formation of every economist, typically used, in a version or another, to introduce students to general equilibrium models, the first and second fundamental theorems of welfare economics, etc.

In the next section, 3.1, we will briefly review the standard textbook method for solving these problems. In 3.2 we will see how, even though problems 1 and 2 are *general* equilibrium problems while the problems in section 2 are *partial* equilibrium problems, when viewed with the powerful gradient method of sections 2.2 and 2.3, mathematically these problems are equivalent.

Moreover, as we have seen, the solution through the gradient method automatically generalizes to n goods and m constraints, and are based on the same methods used in machine learning, leaving open the door to extend these models to treat them with machine learning power, which will be done in a different work.

But before doing all that, we would like to point out that, a priori, there is no reason to think that problem 1 and problem 2, mathematically, are the same problem. There are even much

less reasons to think than the problems of section 2, mathematically, are also the same problem. Focusing on problems 1 and 2, in problem 1 there is only 1 optimizer who knows everything decides the values of x_a, x_b, y_a and y_b. In problem 2, on the contrary, Alice optimizes U_a and Bob optimizes U_b. Moreover, $\bar{X}$ and $\bar{Y}$ being fixed, their incentives are misaligned, except that they both gain from trade. Alice may or may not know Bob's utility and vice versa, etc. However, none of this will matter by virtue of the equivalence discussed throughout sections 2.2 and 2.3, and specifically in section 2.4.

3.1 Graphic methods to solve general equilibrium problems, the Edgeworth Box

There are many simple ways to solve problems 1 and 2 of the previous section, as long as the numbers of goods and people are two. The Edgeworth Box is a particularly simple and visually intuitive mathematical technique. To fix ideas, let us solve problem 1 with the Edgeworth Box method and then make some comments about problem 2.

Since $0 \leq k_x, k_y \leq \bar{K}$ and $0 \leq \ell_x, \ell_y \leq \bar{L}$ we can represent these quantities in a box. To fix ideas, suppose that the production functions (3.1-3.2) are $f(k, \ell) = f(k, \ell) = k^{1/2}\ell^{1/2}$ for both goods, that $\bar{K} = 10$ and $\bar{L} = 8$:

$$X = f(k_x, \ell_x) = k_x^{1/2}\ell_x^{1/2} \tag{3.7}$$

$$Y = g\left(k_y, \ell_y\right) = k_y^{1/2}\ell_y^{1/2} \tag{3.8}$$

$$k_x + k_y = \bar{K} = 10, \qquad \ell_x + \ell_y = \bar{L} = 8 \tag{3.9}$$

We represent graphically this in figure 14. The lower left corner represents to origin of coordinates of (k_x, ℓ_x), corresponding to zero X production. The upper right corner represents to origin of coordinates of (k_y, ℓ_y), corresponding to zero Y production.

The lower-horizontal axis, from left to right, represents increasing values of k_x. The upper-horizontal axis, from right to left, represents increasing values of k_y. Both k_x and k_y are positive, and $k_x + k_y = 10$, so the horizontal axes have length 10.

The left-vertical axis, bottom up, represents increasing values of ℓ_x. The right-vertical axis, top down, represents increasing values of ℓ_y. Both ℓ_x and ℓ_y are positive, and $\ell_x + \ell_y = 8$, so the vertical axes have length 8.

Any point in the 10×8 rectangle, or *Edgeworth Box*, represents a possible allocations of capital and labor for the production of X and Y. For example, the blue dot corresponds to $k_x = 7.325, \ell_x = 2.184$ and $k_y = 10 - k_x = 2.675, \ell_y = 8 - \ell_x = 5.816$. The blue dot is clearly *not* an efficient production allocation, since, for example, ascending along the green isoquant of g we arrive at the rightmost red dot, corresponding to equal production of the Y good, and more production of the X good (the yellow isoquant of the f, to which that red dot belongs, corresponds to larger values of X than the purple isoquant of the f, to which that blue dot belongs.)

Staring at figure 14 the reader will easily convince herself that any point that is not in the brown straight line, that is, the line where the level curves of f and g are tangent to each other, cannot

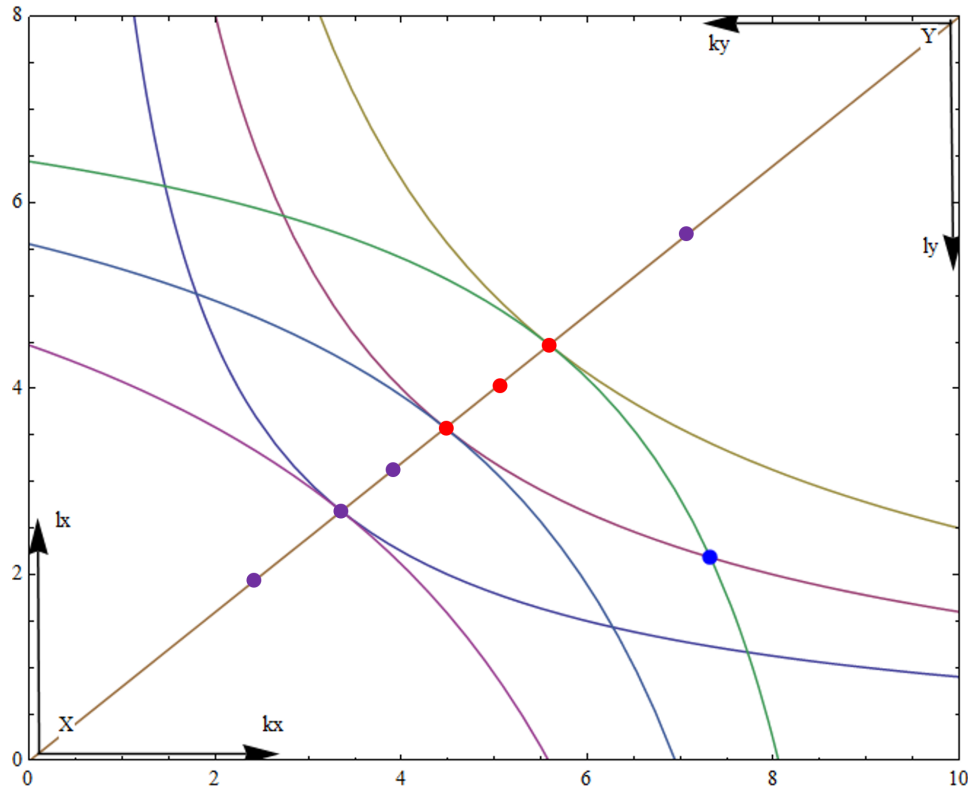


Figure 14: Edgeworth Box. The convex isoquants are level curves of $f(k_x, \ell_x) = k_x^{1/2}\ell_x^{1/2}$, with increasing values of f as we move from the lower-left corner to the upper-right corner. The concave isoquants are level curves of $g\left(k_y, \ell_y\right) = k_y^{1/2}\ell_y^{1/2}$, with increasing values of g as we move from the upper-right corner to the lower-left corner.

possibly be an efficient allocation point. But in section 2.1.2 we already solved the problem of finding the equation for the tangency of the level curves of the implicit functions determined by level lines of two general functions $f(k_x, \ell_x)$, and $g\left(k_y, \ell_y\right) = g(10 - k_x, 8 - \ell_x)$. The equation is 2.18, or, in our present notation,

$$\frac{\frac{\partial f(k_x,\ell_x)}{\partial k_x}}{\frac{\partial f(k_x,\ell_x)}{\partial \ell_x}} = \frac{\frac{\partial g(10-k_x,8-\ell_x)}{\partial k_x}}{\frac{\partial g(10-k_x,8-\ell_x)}{\partial \ell_x}} \tag{3.10}$$

with the functions f and g given in (3.7-3.8), (3.10) becomes

$$\frac{\frac{1}{2}\left(\frac{\ell_x}{k_x}\right)^{1/2}}{\frac{1}{2}\left(\frac{k_x}{\ell_x}\right)^{1/2}} = \frac{\ell_x}{k_x} = \frac{-\frac{1}{2}\left(\frac{8-\ell_x}{10-k_x}\right)^{1/2}}{-\frac{1}{2}\left(\frac{10-k_x}{8-\ell_x}\right)^{1/2}} = \frac{8-\ell_x}{10-k_x}, \quad \text{or} \quad 10\ell_x = 8\,k_x, \quad \text{or} \quad \ell_x = 0.8\,k_x \tag{3.11}$$

the last expression, $\ell_x = 0.8\,k_x$, which is the straight line in figure 14, is the "Efficient Production Set". Being the set of all the points where an isoquant of f is tangent to an isoquant of g, for all possible isoquants, any allocation outside this set can be improved by moving along an isoquant

of any of these functions towards the efficient production set. Such change in allocations would keep the production of one good fixed while increasing the production of the other good, as we did when moving from the blue dot in figure 14 to the rightmost red dot described above.

On the other hand, starting from an allocation in the efficient production set and moving along this set towards another point in the set, necessarily increases the production of one good and decreases the production of the other good. Any allocation in this set is "Pareto" efficient. Which particular point in this set the planner will finally choose depends on preferences. But this is the second part of the problem described at the beginning of section 3, which, as mentioned there, we will not address in the present work.

The relation $\ell_x = 0.8\, k_x$, together with $k_x + k_y = 10$ and $\ell_x + \ell_y = 8$, imply that in the Efficient Production Set, these four quantities are all dependent on one, say k_x:

$$\ell_x = 0.8\, k_x \tag{3.12}$$
$$k_y = 10 - k_x \tag{3.13}$$
$$\ell_y = 8 - \ell_x = 8 - 0.8\, k_x \tag{3.14}$$

with (3.12-3.14) we can find a parametric expression for the quantity of good X and Y produced in every point of the Efficient production set. From (3.7) and (3.8):

$$X = f\,(k_x, \ell_x(k_x)) = k_x^{1/2}(0.8\, k_x)^{1/2} = \sqrt{0.8}\; k_x \tag{3.15}$$
$$Y = g\left(k_y(k_x), \ell_y(k_x)\right) = (10 - k_x)^{1/2}(8 - 0.8\, k_x)^{1/2} = \sqrt{0.8}\; (10 - k_x) \tag{3.16}$$
$$0 \leq k_x \leq 10 \tag{3.17}$$

Equations (3.15-3.17) constitute a parametric from of the "Production Possibility Frontier" (PPF), in terms of the parameter k_x. One may prefer a non-parametric, explicit function $Y(X)$ for the PPF. In this case, from (3.15), $k_x = X/\sqrt{0.8}$, inserting this into (3.16), we have:

$$Y = \sqrt{0.8}\; \left(10 - \frac{X}{\sqrt{0.8}}\right) = \sqrt{80} - X \approx 8.944 - X \tag{3.18}$$

In figure 15 we draw the Production Possibility Frontier, and in the $(X, Y)-$plane we also draw the blue point in figure 14, which clearly lies below the PPF, and the rightmost red dot in figure 14, corresponding to equal production of good Y, and more production of good X, and lying in the PPF.

Two final points before we make some comments about problem 2. The first one is that, although the economic reasoning is different, the mathematics to find the Efficient Production Set was essentially identical to the one used in section 2.1 to solve constrained optimization problems with the tangency of level curves method. The last step, to go from the Efficient Production Set to the Production Possibility Frontier, is just plugging into the f and g functions (3.7-3.8) the solution of the Efficient Production Set problem.

At first sight, it may seem unexpected that the mathematics for the solution of general equilibrium problems addressed in this section is essentially the same as the mathematics of section 2.1 for the solution of much simpler partial equilibrium problems. But this is due to the fact that the

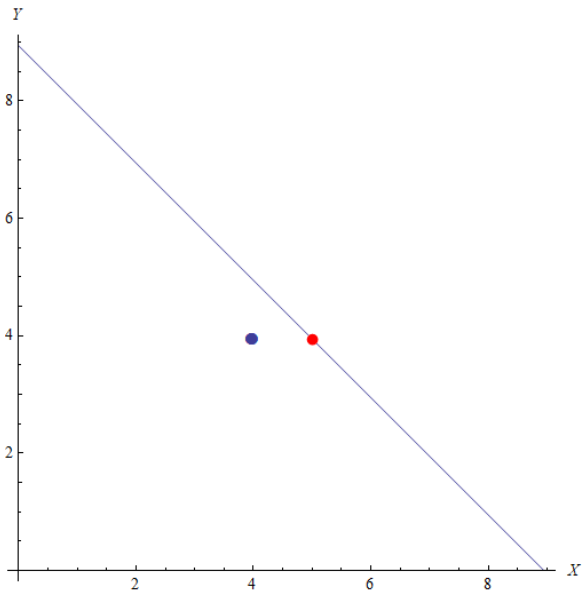


Figure 15: Production Possibility Frontier. The blue point in figure 14 is seen here below the PPF. The rightmost red dot in figure 14, corresponding to equal production of the good Y, and more production of the good X, lies in the PPF.

tangency of level curves method used in this section, is just the tip of the iceberg of the deeper and far more general gradient method, as was explained in sections 2.2 and 2.3. In the next section we will see that the equivalence stressed out in section 2.4, which is completely transparent in the gradient method, makes the use of exactly the same math for both problems completely natural.

The second point is that, since in the optimum production processes of problem 1 we are assuming that the planner knows the technology for the production of both goods (3.7-3.8), the maximum amounts of inputs for production (3.9), and *she is free to allocated these inputs in whatever way she likes*, any point in the Efficient Production Set is accessible to her. Graphically, not only the red points in figure 14, but the violet points too are accessible to the planner with the mentioned hypothesis. Which specific point she will finally choose in the Efficient Production Set depends on the "preferences" aspect of the problem that we are not addressing here.

For problem 2 of section 3, the reader can easily convince herself that the solution is essentially identical, and goes along almost the same arguments, as the ones used to solve problem 1. The only difference is that not Alice nor Bob have the freedom to take possession of the goods that the planner had in problem 1. We assumed that both, Alice and Bob want to maximize their own utility, and that the interchange is voluntary, so they will negotiate, and we assume they both are rational. This means that they will never accept an outcome that leads to less utility than what their original endowment provides. Therefore, in terms of figure 14, if their initial endowment corresponds to the blue dot, only the red points (and any point in between them) in the Pareto efficient curve is a possible outcome of the exchange.

3.2 The gradient field to solve general equilibrium problems

In the previous section we solved problems 1 and 2 by what is known as the Edgeworth Box, with the double axes in figure 14. But the constraints (3.3) (or (3.6) for problem 2) are trivially resolved as $k_y = \bar{K} - k_x$, $\ell_y = \bar{L} - \ell_x$ (or $x_b = \bar{X} - x_a$, $y_b = \bar{Y} - y_a$). Analytically, we can allow the variables k_x, k_y, ℓ_x, ℓ_y to have any real value (even negative, or greater than $\bar{K}$ and $\bar{L}$ respectively) while keeping the relations $k_y = \bar{K} - k_x$, $\ell_y = \bar{L} - \ell_x$, and at the end of the problem consider only the solutions in the proper range for these variables. Numerically it is trivial to impose the proper range.

So, for example, problem 1 becomes the problem of finding Efficient Production Set, and the Production Possibility Frontier given the technologies f and g and the total inputs $\bar{K}$ and $\bar{L}$:

$$X = f(k_x, \ell_x) \tag{3.19}$$
$$Y = g\left(\bar{K} - k_x, \bar{L} - \ell_x\right) \tag{3.20}$$

No double axis are necessary. The simple fact that g, whose level curves are convex in the $(k_y, \ell_y)-$plane, depends on $\bar{K} - k_x$ and $\bar{L} - \ell_x$, automatically makes its level curves concave in the $(k_x, \ell_x)-$plane. This is exactly as in figure 14, but without the need for the upper-horizontal nor the right-vertical axis.

What we want is an efficient production of goods X and Y, i.e., a production such that no additional units of one good can be produced without reducing the units produced of the other good. How do we achieve this?

The arguments in section 2.1.2, from equation (2.13) to (2.18) can be repeated almost word by word. Making small steps $(dk_x, d\ell_x)$ in the $(k_x, \ell_x)-$plane such that we always remain in a level curve of one product, either a level line of f or a level line of g (it is irrelevant which one, but the fact that it is irrelevant is important!) Say that we move in a level curve of f. As we make tiny displacements along this level curve, the quantities produced of X will, of course, remain constant, but the quantities of Y will in general change. But we will eventually reach a point in which the additional tiny displacements does not change the quantity produced of good Y. By the above definition, that point is an efficient production point, and since Y does not chance under the tiny displacement, it means that we are also moving in a level curve of g. This means that at the optimum the level curves of f and g are tangent. We arrive to the equivalent of formula (2.18), which in our present language is

$$\frac{\frac{\partial f}{\partial k_x}}{\frac{\partial f}{\partial k_y}} = \frac{\frac{\partial g}{\partial k_x}}{\frac{\partial g}{\partial k_y}} \tag{3.21}$$

Note that formula (2.20) doesn't have any equivalence here, because in the problem we are intereseted there is no special value of neither f nor g.

By an reasoning equivalent to (2.28), (3.21) is, in turn, equivalent to the linear dependence of the gradients at the Efficient Production Set

$$\boldsymbol{\nabla} f = \lambda \boldsymbol{\nabla} g \tag{3.22}$$

For the particular case (3.7-3.9), (3.19-3.20) becomes

$$X = f(k_x, \ell_x) = k_x^{1/2} \ell_x^{1/2} \tag{3.23}$$

$$Y = g(10 - k_x, 8 - \ell_x) = (10 - k_x)^{1/2} (8 - \ell_x)^{1/2} \tag{3.24}$$

and (3.22)

$$\begin{pmatrix} (1/2)\ell_x^{1/2}/k_x^{1/2} \\ (1/2)k_x^{1/2}/\ell_x^{1/2} \end{pmatrix} = \lambda \begin{pmatrix} -(1/2)(8 - \ell_x)^{1/2}/(10 - k_x)^{1/2} \\ -(1/2)(10 - k_x)^{1/2}/(8 - \ell_x)^{1/2} \end{pmatrix} \tag{3.25}$$

or, canceling the $1/2$'s

$$\frac{\ell_x^{1/2}}{k_x^{1/2}} = -\lambda \frac{(8 - \ell_x)^{1/2}}{(10 - k_x)^{1/2}} \tag{3.26}$$

$$\frac{k_x^{1/2}}{\ell_x^{1/2}} = -\lambda \frac{(10 - k_x)^{1/2}}{(8 - \ell_x)^{1/2}} \tag{3.27}$$

dividing (3.26) by (3.27) leads to (3.11), and the equivalence is manifest.

As mentioned in section 2.3.1 and 2.3.2, in 2-D there is no real advantage of the gradient method with respect to the standard tangency of level curves method, and the above exercise shows that much. The real practical advantage appears when we tackle problems of more independent variables and more constraints, as shown in section 2.3. We will exploit this practical advantage in future works.

But there was another, subtle, conceptual advantage of the gradient methods described in section 2.4, and it was that the method makes transparent the equivalence of optimizing f under a constraint determined by g and optimizing g under a constraint determined by f. We mentioned that in constrained optimization problems, equations like (2.57) clearly break this equivalence in partial equilibrium problems characterized by optimization with constraints. But as we see now, in general equilibrium problems this equivalence is manifest in its full glory. The gradients of f and g are exactly on equal footing. And this is the deep reason why, when attacked with the proper methods, the much more profound general equilibrium problems ultimately use the same math than the more modest partial equilibrium problems.

4 Conclusions

Throughout the paper we have discussed the advantages of presenting the problems of partial and general equilibrium with gradient fields. The geometry is quite intuitive, and, with a unified method, general equilibrium problems are handled as easily as the conceptually much simpler partial equilibrium problems. With such a simple static optimization tool, we have covered practically half an entire course in microeconomics, since any subsequent general equilibrium exercise would replicate the bases covered in the paper.

In applications to the real economy, if we knew all the n-production functions with all m-inputs, it would be possible to find, with this simple gradient based algorithm, the contract surface of any economy. One might think that the lack of knowledge of production techniques, the number of companies, individuals, products and inputs, would make the problem not feasible in practice. That the required data storage and processing capacity makes any realistic implementation just too costly. However, as machine learning practice shows, stochastic versions of the gradient descent algorithm find solutions ("train" deep neural networks in the jargon) whose generalization capacity is surprising and, in fact, better than what is "reasonable" to expect, see for example, [Zhang et al. (2017)], [Bartlett et al. (2021)]. In a future work we will try to implement a machine-learning-type general equilibrium model where the parameters of the production functions, utilities, etc. are variables to fix in the same optimization process that leads to the the contract surface of any economy.

Regardless, it seems reasonable to assume that Economics and other social sciences will continue being drastically transformed by the increasing access to gigantic databases filled with data of extremely high dimensionality. In this context, the classical two dimensional models, or, in general, simple models, while still obviously useful conceptually, would be far more useful if presented with a mathematical formalism that seamlessly extend to any dimensionality and any number of constraints. This work pretends to add our grain of sand in this direction. Economic students deserve to learn the powerful techniques of the present times from the beginning. Specially those techniques that mesh naturally with economic concepts such as gradient methods for optimization.

References

[Athey and Imbens (2019)] Athey, S. and Imbens, G. W., *Machine Learning Methods That Economists Should Know About*, Annual Review of Economics, (2019).

[Bartlett et al. (2021)] Bartlett, P., Montanari, A., and Rakhlin, A. *Deep learning: A statistical viewpoint*, Acta Numerica, 30, 87 − 201. doi:10.1017/S0962492921000027.

[Boyd and Vandenberghe (2004)] Boyd S. & Vandenberghe L. "Convex Optimization", *Cambridge University Press* (2004).

[Mathematica 9.0.1.0] Wolfram Research, Inc. "Mathematica", *Wolfram Research, Inc.* (2013).

[Pernice (2018 a)] Pernice S., (2018). "Intuitive Mathematical Economics Series: Chain Rule and Derivatives of Functions Defined Implicitly", *Ucema Working Papers* **679** (2018).

[Pernice (2018 b)] Pernice S., (2018). "Intuitive Mathematical Economics Series: Constrained MaximizationandtheMethodofLagrangeMultipliers", *Ucema Working Papers* **680** (2018).

[Pernice (2019)] Pernice S., (2019). "Intuitive Mathematical Economics Series. Linear Structures I: Linear Manifolds, Vector Spaces and Scalar Products", *Ucema Working Papers* **689** (2019).

[Pernice (2020)] Pernice S., (2020). “Serie de Machine Learning. Revisión de Algebra Lineal 1”, *Ucema Working Papers* **736** (2020).

[Pernice (2020)] Pernice S., (2020). “Serie de Machine Learning. Revisión de Algebra Lineal 1”, *Ucema Working Papers* **736** (2020).

[Zhang et al. (2017)] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. *Understanding deep learning requires rethinking generalization*. In 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net, 2017. arXiv:1611.03530.