Chapter 3: The Lagrange Method

Elements of Decision: Lecture Notes of Intermediate Microeconomics

Charles Z. Zheng
Tepper School of Business, Carnegie Mellon University

Last update: February 5, 2020

1 Constrained optimization with equality constraints

In Chapter 2 we have seen an instance of constrained optimization and learned to solve it by exploiting its simple structure, with only one constraint and two dimensions of the choice variable. In general, however, there may be many constraints and many dimensions to choose. We need a method general enough to be applicable to arbitrarily many constraints and choice dimensions, and systematic enough for machines to be programed to carry out the computation. That is the Lagrange method.

This course focuses on constrained optimization problems of the following form:

$$\begin{aligned} \max_{(x_1,\ldots,x_n)\in S} \quad & u(x_1,\ldots,x_n) \\ \text{subject to} \quad & g_1(x_1,\ldots,x_n)=0, \\ & \vdots \\ & g_m(x_1,\ldots,x_n)=0, \end{aligned} \tag{1}$$

where the domain S for the choice variable is assumed *open* in the sense that it contains no boundary points with respect to the space of all n-vectors. (For example, when $n=1$, the entire set $\mathbb{R}$ of real numbers is open, whereas the set $\mathbb{R}_+$ of nonnegative real numbers is not, as the latter contains the boundary point zero.) Here the choice variable $(x_1,\ldots,x_n)$ has n dimensions and is subject to m constraints, each in the form of an equation $g_k(x_1,\ldots,x_n)=0$. For example, the problem

$$\begin{aligned} \min_{(x_1,x_2)\in S} \quad & w_1x_1+w_2x_2 \\ \text{subject to} \quad & f(x_1,x_2)=y \end{aligned} \tag{2}$$

is equivalent to a problem in the form of (1):

$$\begin{aligned} \max_{(x_1,x_2)\in S} \quad & -(w_1x_1+w_2x_2) \\ \text{subject to} \quad & f(x_1,x_2)-y=0, \end{aligned} \tag{3}$$

with $n=2$, $m=1$, $u(x_1,x_2)=-(w_1x_1+w_2x_2)$ and $g_1(x_1,x_2)=f(x_1,x_2)-y$.

2 The procedure

The Lagrange method to solve Problem (1) proceeds in three steps. First, write down the *Lagrangian*, a function defined by

$$L(x_1,\ldots,x_n;\lambda_1,\ldots,\lambda_m):=u(x_1,\ldots,x_n)+\sum_{k=1}^{m}\lambda_k g_k(x_1,\ldots,x_n) \tag{4}$$

for any n-vector $(x_1, \ldots, x_n)$ and m-vector $(\lambda_1, \ldots, \lambda_m)$. Recall from basic math the summation notation $\sum_{k=1}^{m}$; for example, $\sum_{k=1}^{m} a_k$ is just the shorthand for $a_1 + \ldots + a_m$. Note from Eq. (4) that we form the Lagrangian by summing the objective u with all the constraint functions g_1, ..., and g_m, multiplied respectively by the coefficient λ_1, ..., and λ_m. For each k, the coefficient λ_k for g_k is called *Lagrange multiplier* for the kth constraint.

Second, write down the first-order condition for the Lagrangian to attain its local maximum. In other words, calculate all the partial derivatives of the Lagrangian and set each of them to zero:

$$\frac{\partial}{\partial x_i} L = \frac{\partial}{\partial x_i} u(x_1, \ldots, x_n) + \sum_{k=1}^{m} \lambda_k \frac{\partial}{\partial x_i} g_k(x_1, \ldots, x_n) = 0 \quad \text{for all } i = 1, \ldots, n; \tag{5}$$

$$\frac{\partial}{\partial \lambda_k} L = g_k(x_1, \ldots, x_n) = 0 \quad \text{for all } k = 1, \ldots, m. \tag{6}$$

Note that Eqs. (5)–(6) constitute an equation system of $n + m$ equations and $n + m$ unknowns.

Third, solve Eqs. (5)–(6) for $(x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_m)$. The $(x_1, \ldots, x_n)$ obtained thereof is a candidate for the solution of Problem (1) as long as the method described above is applicable.

For illustration, consider the cost-minimization problem (2) with nonzero parameters w_1 and w_2 and differentiable production function f such that the partial derivatives are nonzero. Rewrite the problem in the form of (1) thereby to obtain Problem (3), based on which we construct the Lagrangian

$$L(x_1, x_2; \lambda) := -w_1 x_1 - w_2 x_2 + \lambda \left(f(x_1, x_2) - y \right).$$

Then write down the first-order conditions for this Lagrangian, as if we were seeking a local maximum of L without constraint:

$$\begin{aligned} \frac{\partial}{\partial x_1} L &= -w_1 + \lambda \frac{\partial}{\partial x_1} f(x_1, x_2) = 0, \\ \frac{\partial}{\partial x_2} L &= -w_2 + \lambda \frac{\partial}{\partial x_2} f(x_1, x_2) = 0, \\ \frac{\partial}{\partial \lambda} L &= f(x_1, x_2) - y = 0. \end{aligned}$$

Finally, solve the three equations for $(x_1, x_2; \lambda)$: the three equations are equivalent to

$$\lambda \frac{\partial}{\partial x_1} f(x_1, x_2) = w_1, \tag{7}$$

$$\lambda \frac{\partial}{\partial x_2} f(x_1, x_2) = w_2, \tag{8}$$

$$f(x_1, x_2) = y; \tag{9}$$

divide the first equation by the second to cancel out λ (which can be done because the partial derivatives are nonzero by assumption, and $\lambda \neq 0$, otherwise Eq. (7) would say that zero is equal to a nonzero number w_1) and obtain

$$\frac{\partial}{\partial x_1} f(x_1, x_2) \Big/ \frac{\partial}{\partial x_2} f(x_1, x_2) = w_1 / w_2, \tag{10}$$

which coupled with Eq. (9) gives a solution for (x_1, x_2); plug this solution into Eq. (7) or (8) to obtain λ. Note that Eqs. (9)–(10) are exactly the equation system in Chapter 2 that determines the cost-minimizing input bundle in the case where the production function is differentiable and has diminishing TRS.

3 An example with multiple constraints

Consider the problem

$$\begin{aligned} \min_{(x_1,x_2,x_3)\in(0,\infty)^3} \quad & 3x_1 + 2x_2 + 4x_3 \\ \text{subject to} \quad & \ln x_1 + 5\ln x_2 = 100, \\ & x_1 = 2x_3. \end{aligned}$$

Here a firm chooses between three kinds of inputs to deliver 100 units of output, though according to the first constraint only inputs 1 and 2 can contribute to production. The second constraint, $x_1 = 2x_3$, may be due to an environmental protection legislation that requires hiring input 3 in a certain proportion of another input that the firm hires. Note that the domain $(0,\infty)^3$, the space of 3-vectors whose coordinates are all positive, is an open set. To solve this problem, rewrite it in the form of (1):

$$\begin{aligned} \max_{(x_1,x_2,x_3)\in(0,\infty)^3} \quad & -(3x_1 + 2x_2 + 4x_3) \\ \text{subject to} \quad & \ln x_1 + 5\ln x_2 - 100 = 0, \\ & x_1 - 2x_3 = 0. \end{aligned}$$

Thus the Lagrangian is

$$L(x_1,x_2,x_3;\lambda_1,\lambda_2) := -(3x_1 + 2x_2 + 4x_3) + \lambda_1(\ln x_1 + 5\ln x_2 - 100) + \lambda_2(x_1 - 2x_3).$$

Hence the first-order conditions are

$$\begin{aligned} \frac{\partial}{\partial x_1}L &= -3 + \lambda_1/x_1 + \lambda_2 = 0, \\ \frac{\partial}{\partial x_2}L &= -2 + 5\lambda_1/x_2 = 0, \\ \frac{\partial}{\partial x_3}L &= -4 - 2\lambda_2 = 0, \\ \frac{\partial}{\partial \lambda_1}L &= \ln x_1 + 5\ln x_2 - 100 = 0, \\ \frac{\partial}{\partial \lambda_2}L &= x_1 - 2x_3 = 0. \end{aligned}$$

These equations are equivalent to

$$\begin{aligned} \lambda_1/x_1 &= 3 - \lambda_2, \\ 5\lambda_1/x_2 &= 2, \\ \lambda_2 &= -2, \\ \ln\left(x_1 x_2^5\right) &= 100, \\ x_1 &= 2x_3. \end{aligned}$$

The first and third equations together imply

$$\lambda_1/x_1 = 3 - (-2) = 5,$$

which coupled with the second equation gives $x_1 = 2x_2/25$. Plug it into the fourth equation to get

$$\frac{2}{25} x_2 x_2^5 = e^{100},$$

i.e., $x_2 = \left(25e^{100}/2\right)^{1/6}$. Hence $x_1 = 2x_2/25 = \frac{2}{25}\left(25e^{100}/2\right)^{1/6}$, which, plugged into the last equation, gives

$$\frac{2}{25}\left(25e^{100}/2\right)^{1/6} = 2x_3,$$

i.e., $x_3 = \frac{1}{25}\left(25e^{100}/2\right)^{1/6}$. Thus we obtain the only candidate for a solution of (x_1, x_2, x_3):

$$\left(\frac{2}{25}\left(25e^{100}/2\right)^{1/6}, \left(25e^{100}/2\right)^{1/6}, \frac{1}{25}\left(25e^{100}/2\right)^{1/6}\right).$$

Compared to the above procedure, it would be more cumbersome to solve the problem with the two-dimensional methods in Chapters 2 and 3, as the choice variable here lives in the 3-dimensional space, in which the two constraints are each a surface rather than a curve.

4 The idea behind the method

The above procedure is encapsulated by the equation system (5)–(6). Among them Eqs. (6) are obviously necessary for a solution of the constrained optimization problem, as they are simply restatements of the constraints. What we still need to understand is Eqs. (5).

4.1 The idea illustrated by an example

To understand Eqs. (5), start with a simple example

$$\begin{aligned} \max_{(x_1,x_2)\in(0,\infty)^2} \quad & px_2 - wx_1 \\ \text{subject to} \quad & f(x_1) = x_2, \end{aligned} \tag{11}$$

where $f(x_1)$ a differentiable function of x_1 and, when x_1 increases, $f(x_1)$ increases and the derivative $\frac{d}{dx_1}f(x_1)$ decreases. A simple way to solve this problem is to plug the constraint $x_2 = f(x_1)$ into the objective so that the problem becomes

$$\max_{x_1\in(0,\infty)} pf(x_1) - wx_1.$$

Then apply the technique in Chapter 1. Since the domain $(0, \infty)$ of the choice variable x_1 is an open set, any maximum is an interior solution and hence satisfies the first-order condition $\frac{d}{dx_1}f(x_1) - \frac{w}{p} = 0$, i.e.,

$$\frac{d}{dx_1}f(x_1) = \frac{w}{p}. \tag{12}$$

Since by assumption $\frac{d}{dx_1}f(x_1)$ is decreasing in x_1, the second-order condition is always satisfied (verify that yourself). Thus a solution to Eq. (12) is the same as a solution to Problem (11). Geometrically, Eq. (12) means: On the x_1-x_2 plane, draw the graph of f, which is upward sloping and whose slope is decreasing in x_1; draw the straight line that has the slope w/p and is tangent to the graph of f; the tangent point is the solution. (Rember Exercise 4 of Chapter 2?)

The question is how to generalize this method to cases with multiple constraints and more than two choice dimensions. In those cases, the graph of a constraint equation (e.g., $f(x_1, x_2) = x_3$) is no longer just a curve but rather a surface in a higher-dimension space, and likewise for the tangent "line" (e.g., the plane $px_3 - w_1x_1 - w_2x_2 = 300$). While we can still imagine that the solution is the tangent point between the two surfaces, it makes little sense to say that the two surfaces have the same "slope," as a surface may have different slopes along different directions.

Rather than slopes, a better way to look at surfaces tangent to each other is to compare their *gradients* (which we shall define in the next subsection): at any point of a surface, the gradient of the surface is a vector perpendicular to the surface at that point. In the above example, the gradient of the tangent line is the vector $(-w, p)$, which plotted on the x_1-x_2 diagram is perpendicular to the tangent line, whose slope is constantly equal to w/p.[1] The gradient of the constraint $f(x_1) = x_2$ varies with the coordinates (x_1, x_2) on the curve, because the slope of the curve varies with (x_1, x_2), and given x_1 this gradient is the vector $\left(\frac{d}{dx_1}f(x_1), -1\right)$. Note that this vector is perpendicular to the curve $f(x_1) = x_2$ at the point with horizontal coordinate x_1.[2] Thus, the gradients of the tangent line and the constraint curve, one perpendicular to the tangent line and the other perpendicular to the constraint curve, must be *aligned*, i.e., belonging to a single straight line. We have now arrived at a viewpoint elegantly suitable for higher dimensions: *two surfaces are tangent to each other iff their gradients are aligned*. Thus, when the choice variable has higher dimensions, instead of thinking of a solution as a tangent point, think of the solution as the point on the constraint surface (the set of points that satisfy all constraints) such that at this point the gradient of this surface and the gradient of the objective are aligned.

Algebraically, two vectors are *aligned* iff you can turn one vector into the other by multiplying all coordinates of the former by some common number. That is, in our example, vector $(-w, p)$ being aligned with vector $\left(\frac{d}{dx_1}f(x_1), -1\right)$ means that there exists a real number λ for which

$$\begin{bmatrix} -w \\ p \end{bmatrix} = -\lambda \begin{bmatrix} \frac{d}{dx_1}f(x_1) \\ -1 \end{bmatrix}. \tag{13}$$

From the viewpoint described in the previous paragraph, we reach an alternative method to solve Problem (11): instead of solving for (x_1, x_2) by coupling the tangency equation (12) with the constraint $f(x_1) = x_2$, solve for (x_1, x_2, λ) by coupling Eq. (13) with the constraint $f(x_1) = x_2$.

But are the two methods consistent to each other? To see that the answer is Yes, rearrange Eq. (13) to obtain

$$\begin{aligned} -w + \lambda\frac{d}{dx_1}f(x_1) &= 0, \\ p - \lambda &= 0. \end{aligned}$$

These two equations are simply Eq. (5) applied to our example, with the Lagrangian

$$L(x_1, x_2, \lambda) := px_2 - wx_1 + \lambda(f(x_1) - x_2).$$

[1] The two are perpendicular to each other because the slope of the vector is $p/(-w)$ and so $p/(-w) \cdot (w/p) = -1$.

[2] To see that, note that the slope of the gradient is equal to $-1/\frac{d}{dx_1}f(x_1)$, while the slope of the constraint curve $f(x_1) = x_2$ is equal to $\frac{d}{dx_1}f(x_1)$; multiply the two to obtain $\left(-1/\frac{d}{dx_1}f(x_1)\right) \cdot \frac{d}{dx_1}f(x_1) = -1$.

In the mean time, Eq. (13) is equivalent to

$$\begin{aligned} -w &= -\lambda \frac{d}{dx_1} f(x_1), \\ p &= \lambda. \end{aligned}$$

Dividing the first equation by the second to cancel out λ and obtain

$$-w/p = -\frac{d}{dx_1} f(x_1),$$

which is exactly the tangency equation (12). Hence the two methods are consistent, except that the one with gradients works also in higher dimensions.

4.2 Gradients and the general idea

The *gradient* of any differentiable function f of n variables, at any point $(x_1, \ldots, x_n)$, is defined to be the n-vector consisting of the partial derivatives of f at $(x_1, \ldots, x_n)$:

$$\nabla f(x_1, \ldots, x_n) := \begin{bmatrix} \frac{\partial}{\partial x_1} f(x_1, \ldots, x_n) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x_1, \ldots, x_n) \end{bmatrix}.$$

Intuitively speaking, the gradient of f at any point $(x_1, \ldots, x_n)$ points to the direction along which we perturb $(x_1, \ldots, x_n)$ and achieve the steepest rise of f; the gradient at $(x_1, \ldots, x_n)$ is also a vector perpendicular to the level surface consisting all the points $(x'_1, \ldots, x'_n)$ such that $f(x'_1, \ldots, x'_n) = f(x_1, \ldots, x_n)$. Now rewirte Eqs. (5) into a single equation in vector format:

$$\begin{bmatrix} \frac{\partial}{\partial x_1} u(x_1, \ldots, x_n) \\ \vdots \\ \frac{\partial}{\partial x_n} u(x_1, \ldots, x_n) \end{bmatrix} = -\lambda_1 \begin{bmatrix} \frac{\partial}{\partial x_1} g_1(x_1, \ldots, x_n) \\ \vdots \\ \frac{\partial}{\partial x_n} g_1(x_1, \ldots, x_n) \end{bmatrix} - \cdots - \lambda_m \begin{bmatrix} \frac{\partial}{\partial x_1} g_m(x_1, \ldots, x_n) \\ \vdots \\ \frac{\partial}{\partial x_n} g_m(x_1, \ldots, x_n) \end{bmatrix},$$

which, by the gradient notation just introduced, is equivalent to

$$\nabla u(x_1, \ldots, x_n) = -\sum_{k=1}^{m} \lambda_k \nabla g_k(x_1, \ldots, x_n). \tag{14}$$

This equation says that, if we scale up the gradient of each constraint by its Lagrange multiplier, then the aggregate of such gradients is aligned with the gradient of the objective. The situation is illustrated in Figure 1 (Luenberger [1]), where the gradients h'_1 and h'_2 of the two constraints span a plane that contains the gradient f' of the objective. Hence you can scale up or down h'_1 and h'_2 so that the sum of the scaled vectors is exactly f'; the precise proportions of the scalers are the Lagrange multipliers.

For example, consider Problem (11) in the previous subsection. In this problem, $n = 2$, $m = 1$, $u(x_1, x_2) = px_2 - wx_1$, $g(x_1, x_2) = f(x_1) - x_2$,

$$\begin{aligned} \nabla u(x_1, x_2) &= \begin{bmatrix} -w \\ p \end{bmatrix} \\ \nabla g(x_1, x_2) &= \begin{bmatrix} \frac{d}{dx_1} f(x_1) \\ -1 \end{bmatrix}. \end{aligned}$$

Hence Eq. (14) becomes the Eq. (13) in the previous subsection.

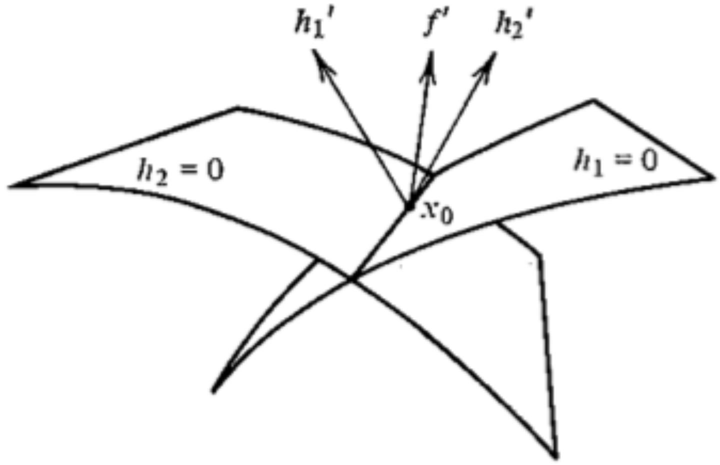

Figure 1: The f here is our objective u, and h_1 and h_2 here our constraint functions g_1 and g_2

5 When is the Lagrange method applicable?

The full answer is nuanced. The bad news is that in general we do not know if the method is applicable or not until we go through the above procedure, as the conditions available to validate the Lagrange method are mostly statements about the possible solution that one would come up with as the candidate for a solution of the original problem (1). The good news is that even without checking such conditions one is unlikely to get a wrong answer by carrying out the above procedure correctly. If Eqs. (5)–(6) produce no solution, then the Lagrange method is inapplicable. If Eqs. (5)–(6) yield a solution then, except for cases that are deemed negligible within the scope of this course, we can conclude that it is necessarily the candidate for any solution of the original Problem (1).

Still, some of the conditions we can check at the outset before plunging into the procedure: that the domain of the choice variable should be open (as we assume at the start of this chapter), that $n \geq m$ (there be no less dimensions of the choice variable than the number of constraints), and that the objective and constraint functions be all differentiable. If the domain is not open, the solution may be a boundary point, at which Eqs. (5)–(6) may admit no solution. If there are more constraints than dimensions of the choice variable, we would have more equations than unknowns and again Eqs. (5)–(6) may admit no solution. Without differentiability we cannot take derivatives, let alone obtaining the first-order conditions, Eqs. (5)–(6).

Even when the Lagrange method is applicable, its outcome is only a *necessary* condition for a solution of the original problem (1), in the sense that *if* $(x_1^*, \ldots, x_n^*)$ solves Problem (1) *then* it is equal to the one produced by the above procedure, but the converse is not guaranteed by the method. To see if a candidate produced by the Lagrange method is a solution of Problem (1), one needs to check whether the candidate satisfies the second-order condition in the manner similar to, but more complicated than, that of Chapter 1. Nevertheless, you do not need to go through such a trouble if the objective function u and the constraint functions g_k $(k = 1, \ldots, m)$ all belong to a class that are called concave functions, which we do not define for this course. In that case, the candidate produced by the Lagrange method solves Problem (1).[3]

6 The complication with inequality constraints

Constrained optimization problems with inequality constraints can be written in the form

$$\begin{aligned} \max_{(x_1,\ldots,x_n)\in S} \quad & u(x_1,\ldots,x_n) \\ \text{subject to} \quad & g_1(x_1,\ldots,x_n) \geq 0, \\ & \vdots \\ & g_m(x_1,\ldots,x_n) \geq 0. \end{aligned} \tag{15}$$

While there are methods analogous to the procedure for equality constraints, the validity of such methods require stronger conditions. At this point, just beware of a tempting danger for many—alas, including many economists!—to abuse such methods even when their validity is not warranted. When you see a "Lagrange method" solution of an optimization problem some of whose constraints are inequalities, be careful.[4]

However, some problems with inequality constraints can be turned into ones with equality constraints. For such problems we can solve by the procedure introduced above. Let us illustrate

[3] For the curious mind only, here is a synopsis of the conditions for the Lagrange method. The Lagrange Multiplier Theorem says that a solution $(x_1^*,\ldots,x_n^*)$ of Problem (1) is necessarily a solution of Eqs. (5)–(6) provided that two conditions are met at $(x_1^*,\ldots,x_n^*)$: (a) the objective u and constraint functions g_k (for all $k = 1,\ldots,m$) are all continuously differentiable, and (b) the constraint functions are *regular* in the sense that their gradients span the m-dimensional vector space. The regularity condition (b), in turn, means that (i) $n \geq m$ and (ii) none of the gradients $\nabla g_1,\ldots,\nabla g_m$ at $(x_1^*,\ldots,x_n^*)$ is redundant, i.e., none is equal to a linear combination of the other $m-1$ gradients (e.g, if $\nabla g_1 = \lambda_2\nabla g_2 + \cdots + \lambda_m\nabla g_m$ for some real numbers $\lambda_2,\ldots,\lambda_m$ then ∇g_1 is redundant). There would have been a third condition (c), requiring that $(x_1^*,\ldots,x_n^*)$ not be a boundary point of its domain, which is already guaranteed because we assume at the outset that the domain S is open.

We have explained previously why the procedure needs Condition (a) and part (i) of Condition (b). To have a glimpse of what part (ii) of Condition (b) is up to, consider a case where it is violated: Suppose $n = 3$ and $m = 2$ such that the two constraints correspond to two surfaces that have exactly one common point and at that point the two surfaces are tangent to each other. That immediately pins the choice variable down to this single point, which is hence the solution of Problem (1). At this point, the gradients of the two constraints are aligned, as the corresponding surfaces are tangent to each other. Thus, no matter how we scale up or down each of them, we cannot alter the direction of the sum of the two. Consequently, if the gradient of the objective is not aligned with the two constraint gradients at the outset, there is no way to scale the two constraint gradients thereby to align them with the gradient of the objective, hence it is impossible to satisfy Eq. (14). This misalignment cannot be corrected by perturbation, because there is no wiggle room to perturb the choice variable along the direction of the gradient of the objective without violating one of the constraints (c.f. Exercise 4d.).

[4] Even the famous Kuhn and Tucker, who are credited for a main theorem handling this case (and both are played as characters in the Oscar-award-winning blockbuster "A Beautiful Mind"), made a serious mistake in the initial version of their theorem. According to the late Leo Hurwicz, they did not know of the mistake until a seminar audience pointed it out, with a counterexample, during their seminar presenting the "theorem." Then they haphazardly modified their theorem by adding a "constraint qualification" condition on the solution produced by the Lagrange method. When the late Hurwicz related the anecdote to this author, who was then a graduate student working as the former's graduate class teaching assistant, Hurwicz hastened to add a moral of the story: "One does not need to commit suicide *even* when his theorem is found wrong when he is presenting it."

with the following problem:

$$\max_{(x,y)\in\mathbb{R}^2_+} \quad py - wx \qquad (16)$$
$$\text{subject to} \quad y \le f(x),$$

where p and w are positive parameters, and f the production function that is increasing, differentiable, having derivative that is decreasing in x, and

$$\lim_{x\to 0} \frac{d}{dx} f(x) = \infty. \qquad (17)$$

i.e., when x converges to zero, the graph of f steepens to vertical. Written in the form of (15), this problem is

$$\max_{(x,y)\in\mathbb{R}^2_+} \quad py - wx$$
$$\text{subject to} \quad f(x) - y \ge 0. \qquad (18)$$

To apply the Lagrange method with equality constraints, first notice that there is no loss of generality to restrict the choice to those such that $f(x) - y = 0$, for if $f(x) - y > 0$ then the decision maker can increase the objective $py - wx$ by increasing y slightly without changing x. This change is feasible because $f(x) - y > 0$, and it brings in more profit because $p > 0$. Thus, the constraint (18) can be replaced by the equation $f(x) - y = 0$. Hence the original problem is equivalent to

$$\max_{(x,y)\in\mathbb{R}^2_+} \quad py - wx$$
$$\text{subject to} \quad f(x) - y = 0.$$

Second, note that the domain $\mathbb{R}^2_+$ of the choice variable is not open, as it contains boundary points such as $(0, y)$ and $(x, 0)$. But since the slope of the the graph of f is decreasing and because of Eq. (17), any supporting hyperplane of the graph of f touches the graph at a point where both coordinates are nonzero. It follows that at any solution (x, y) of the problem, $x > 0$ and $y > 0$. Thus, there is no loss of generality to replace the domain $\mathbb{R}^2_+$ by the open set $(0, \infty)^2$. Then we apply the Lagrange method. The Lagrangian by definition is

$$L(x, y; \lambda) := py - wx + \lambda(f(x) - y).$$

The first-order conditions are

$$\begin{aligned} \frac{\partial}{\partial x} L &= -w + \lambda \frac{d}{dx} f(x) = 0, \\ \frac{\partial}{\partial y} L &= p - \lambda = 0, \\ \frac{\partial}{\partial \lambda} L &= f(x) - y = 0. \end{aligned}$$

The first two equations together imply

$$w = \lambda \frac{d}{dx} f(x) = p \frac{d}{dx} f(x), \quad \text{i.e.,} \quad \frac{d}{dx} f(x) = \frac{w}{p},$$

which coupled with the previous equation $f(x) - y = 0$ gives exactly the equation system to determine the solution.

7 Exercises

1. Among the sets listed below, which sets are open?

   a. $(-\infty, 0)$

   b. $[0, 3)$ (i.e., the interval between 0 and 3, including 0 but excluding 3)

   c. $(0, 5)^2$ (i.e., $(0, 5) \times (0, 5)$, with $(0, 5)$ denoting the interval between 0 and 5, excluding 0 and 5)

   d. $\{0, 1/2\}$ (i.e., the set consisting of 0 and 1/2)

   e. $(0, 1) \times (0, 1]$

2. A firm uses three kinds of inputs to produce one kind of output. If the firm employs a quantity x_1 of input 1, quantity x_2 of input 2, and quantity x_3 of input 3, with $(x_1, x_2, x_3) \in \mathbb{R}^3_+$ ($\mathbb{R}^3_+$ denotes $[0, \infty)^3$), then the maximum quantity of the output is equal to

$$f(x_1, x_2, x_3) := Ax_1^{\alpha} x_2^{\beta} x_3^{\gamma},$$

   where A, α, β and γ are each a positive parameter. For each $k = 1, 2, 3$, the market price of input k is given to be a positive number, denoted by w_k. Hence any input bundle (x_1, x_2, x_3) would cost the firm $w_1x_1 + w_2x_2 + w_3x_3$. The firm has committed to supply a quantity y of its output, with y a positive parameter.

   a. Express the firm's cost-minimization problem in a format analogous to Problem (1) in Chapter 2, with the domain of the choice variable being $\mathbb{R}^3_+$.

   b. Is the domain of the choice variable open? If not, explain why there is no loss of generality to restrict the domain into the open set $(0, \infty)^3$. (Hint: $y > 0$.)

   c. Rewrite the cost-minimization problem in a format analogous to (1) of this chapter.

   d. Use the Lagrange method, by following the procedure in Section 2, to solve the problem obtained in the previous step. (The solution should be mathematical expressions of only the parameters A, α, β, γ, y, w_1, w_2 and w_3.)

   e. Use the solution obtained in the previous step to show that the cost $C(y)$ of any output quantity y ($y > 0$) is equal to $Ky^{1/(\alpha+\beta+\gamma)}$ for some positive constant K.

   f. Use the result of the previous step to prove that the average cost is increasing in y if $\alpha + \beta + \gamma < 1$, decreasing in y if $\alpha + \beta + \gamma > 1$, and constant if $\alpha + \beta + \gamma = 1$.

3. Following is a set of 2-vectors:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \end{bmatrix}, \begin{bmatrix} -1/3 \\ -2/3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

   a. Plot these vectors on the plane and—

      i. find out the pair of vectors that are aligned to each other;

      ii. find out a pair of vectors that are perpendicular to each other; is there another such a pair?

b. Verify your observations in Step 3a. algebraically in the manner of Eq. (13)—for alignment—and Footnotes 1 and 2—for perpendicularity.

c. Find a scaler (i.e., a real number) λ_1 and a scaler λ_2 such that the linear combination

$$\lambda_1 \begin{bmatrix} -1/3 \\ -2/3 \end{bmatrix} + \lambda_2 \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$

of vectors $\begin{bmatrix} -1/3 \\ -2/3 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ -2 \end{bmatrix}$ is aligned with the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

d. Does there exist scalers λ_1 and λ_2 such that the linear combination

$$\lambda_1 \begin{bmatrix} -1/3 \\ -2/3 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

is aligned with the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$?

4. Which of the following constrained optimization problems can the Lagrange method with equality constraints be applied to, either directly or after the problem is rewritten into an equivalent form?

   a. Minimize $3x_1 + 4x_2$ among $(x_1, x_2) \in (0, \infty)^2$ subject to the constraint $\min\{x_1, 2x_2\} = y$
   b. Maximize $5y - 8x$ among $(x, y) \in [0, \infty)^2$ subject to the inequality constraint $y \leq \ln(x+1)$
   c. Maximize $5y - 8x$ among $(x, y) \in (0, \infty)^2$ subject to the constraints $y = \sqrt{x}$, $y = x - 1$ and $y = x^2 + 1$.
   d. Maximize $5y - 8x$ among $(x, y) \in (0, \infty)^2$ subject to the constraints $y = \sqrt{x}$ and $y = x/2 + 1/2$. (Hint: Condition b.ii, Footnote 3.)

5. For each function defined below, calculate the gradient at the point $(1, 1)$ and the gradient at $(4, 2)$ and plot each gradient in a two-dimensional coordinate system.

   a. $\pi(x_1, x_2) := 6x_2 - 3x_1$ for all nonnegative x_1 and x_2
   b. $g(x_1, x_2) := \sqrt{x_1} - x_2$ for all nonnegative x_1 and x_2
   c. $u(x_1, x_2) := 2x_1 + 5x_2$ for all nonnegative x_1 and x_2
   d. $h(x_1, x_2) := x_1^{1/2} x_2^{1/2}$ for all nonnegative x_1 and x_2

6. Consider Problem (16) with $f(x) := \sqrt{x}$, $p = 3$ and $w = 6$ per unit.

   a. On the x-y plane, graph the function f and the straight line line whose slope is equal to w/p (with the specific numbers given above) and passes through the point $(4, 2)$. Note that the point belongs to the graph of f.
   b. Draw an arrow to indicate the gradient of the objective function at the point $(4, 2)$.
   c. Analogously, draw an arrow to indicate the gradient of the graph of f at the point $(4, 2)$.
   d. Is Eq. (13) satisfied at the point $(4, 2)$? In other words, are the two gradients aligned?

e. Find the coordinates of the point on the graph of f at which Eq. (13) satisfied. On the above diagram draw the gradients of the objective and the graph of f at that point.

7. Which of the following optimization problems is equivalent to one where the inequality constraint is replaced by its corresponding equality, and the domain replaced by an open set?

   a. $\max_{(x,y)\in[0,\infty)^2}(10y-3x)$ subject to $y \leq 5x^{1/3}$

   b. $\max_{(x,y)\in[0,\infty)^2}(10y-3x)$ subject to $y \leq 2\ln(x+1)$

   c. $\max_{(x_1,x_2)\in[0,\infty)^2} x_1 x_2^{2/3}$ subject to $3x_1 + 2x_2 \leq 100$

   d. $\max_{(x_1,x_2)\in[0,\infty)^2}(10-(x_1-1)^2-(x_2-2)^2)$ subject to $3x_1+2x_2 \leq 100$

8. Consider the following optimization problem:

$$\begin{aligned} \max_{(x_1,x_2,x_3)\in\mathbb{R}^3_+} \quad & \ln x_1 + \beta \ln x_2 \\ \text{subject to} \quad & p_1 x_1 + x_3 \leq m \\ & r x_3 = p_2 x_2, \end{aligned}$$

with parameters $0 < \beta < 1$, $p_1 > 0$, $p_2 > 0$, $m > 0$ and $r > 1$.

   a. Explain why the domain of the choice variable can be restricted to the open set $(0,\infty)^3$ without loss of generality.

   b. Explain why the weak inequality constraint can be restricted to an equality constraint without loss of generality.

   c. Rewrite the above problem in a form analogous to (1) of this chapter.

   d. Define the Lagrangian for the problem obtained in the previous step.

   e. Write down the first-order conditions

   f. Demonstrate that the solution of the first-order conditions is:

$$(x_1^*, x_2^*, x_3^*, \lambda_1^*, \lambda_2^*) = \left(\frac{m}{p_1(1+\beta)}, \frac{\beta}{1+\beta}\frac{rm}{p_2}, \frac{\beta m}{1+\beta}, -\frac{\beta+1}{m}, -\frac{\beta+1}{rm}\right).$$

   g. Denote the objective by $u(x_1,x_2,x_3) := \ln x_1 + \beta \ln x_2$ and the two constraint functions by $g_1(x_1,x_2,x_3) := p_1x_1 + x_3 - m$ and $g_2(x_1,x_2,x_3) := p_2x_2 - rx_3$. Calculate the gradients of the objective and the two constraint functions at the solution (x_1^*, x_2^*, x_3^*) obtained in the previous step.

   h. Prove that Eq. (14) is satisfied at the solution (x_1^*, x_2^*, x_3^*), i.e.,

$$\lambda_1^* \nabla g_1(x_1^*, x_2^*, x_3^*) + \lambda_2^* \nabla g_2(x_1^*, x_2^*, x_3^*) = -\nabla u(x_1^*, x_2^*, x_3^*).$$

References

[1] David G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969. 4.2