Leveling with Lagrange: An Alternate View of Constrained Optimization

DAN KALMAN
American University
Washington, D.C. 20016
kalman@american.edu

In most calculus books today [**11, 14, 15**], Lagrange multipliers are explained as follows. Say that we wish to find the maximum value of f subject to the condition that $g = 0$. Under certain assumptions about f and g, the Lagrange multipliers theorem asserts that at the solution point, the gradient vectors ∇f and ∇g are parallel. Therefore, either $\nabla f = \lambda \nabla g$ for some real number λ, or $\nabla g = 0$. Combined with the equation $g = 0$, this gives necessary conditions for a solution to the constrained optimization problem. We will refer to this as the standard approach to Lagrange multipliers.

An earlier tradition approaches this subject far differently. It defines a new function, $F = f + \lambda g$, that incorporates both the objective function and the constraint, and in which λ is considered to be an additional variable. Here, F is referred to as a *Lagrangian function*. The conditions for F to achieve an unconstrained extremum are then determined, and these become necessary conditions for a solution to the original problem. This is the Lagrangian function approach to Lagrange multipliers.

Both approaches produce the same necessary conditions, and lead to identical solutions of constrained optimization problems. The second approach is closer to the original spirit of Lagrange's work, and is popular in introductory works on mathematical methods in economics, as well as calculus texts with an applied or business emphasis. Unfortunately, it is often presented with an attractive, but fundamentally incorrect, intuitive justification [**1, 2, 6, 7**]—that the problem of finding a maximum (say) of f subject to a constraint is transformed into a search for a maximum of F without constraint. The problem is, a constrained maximum of f need not correspond to a local maximum of F. The Lagrange theorem asserts the existence of a corresponding *critical point* for F, but says nothing about whether this critical point is actually an extremum. And as we will see, further examination reveals an unexpected result: the critical point of F that satisfies the Lagrange condition is *never* a local extremum. Therefore, the idea that a constrained optimum of f corresponds to an unconstrained optimum of the Lagrangian F is never correct. For ease of reference, this mistaken idea will be termed the *transformation fallacy*, highlighting the intuition that incorporating the constraint into the objective function transforms a constrained optimization problem into an unconstrained optimization problem.

The perpetuation of a flawed intuitive explanation of the Lagrangian approach is a shame, not least because it represents a dilution of the power of a true intuition. At its heart, the idea that the constrained problem is transformed into an unconstrained problem doesn't make any sense. (How could it? It is incorrect!) And yet, on casual consideration, the idea of imposing the constraint implicitly is seductively plausible. There is even an informal proof that seems to justify this idea.

Happily, there is an alternate justification of the Lagrangian function approach to constrained optimization. It provides a memorable geometric intuition and has a catchy name, *Lagrangian leveling*. In contrast to the central idea of the transformation fallacy, which is necessarily vague, Lagrangian leveling directly confronts (and remedies) the true source of difficulty: at a constrained maximum, the partial derivatives of the objective function need not vanish.

The goal of this paper is to draw attention to the transformation fallacy, and to the idea of Lagrangian leveling. A specific example illustrating the fallacy and a critique of the rationale behind the fallacy will be presented. We will also see that leveling leads to both a proof of the Lagrange multiplier theorem and an interpretation of the value of the multiplier λ at the extremum.

The Lagrangian function approach For the sake of concreteness, let us review the standard approach to Lagrange multipliers in a specific context. Suppose that $f(x, y)$ and $g(x, y)$ are differentiable functions on a domain in $\mathbb{R}^2$, and that we wish to find the maximum of f subject to the constraint $g = 0$. As outlined in the introduction, if the solution occurs at a point (x, y) where $\nabla g \neq 0$, then there must exist a real λ such that $g(x, y) = 0$ and $\nabla f(x, y) = \lambda \nabla g(x, y)$, giving necessary conditions for the solution.

In the Lagrangian function approach we arrive at essentially the same conditions but follow a different route. Define $F(x, y, \lambda) = f(x, y) + \lambda g(x, y)$. Now F is regarded as a function of three variables. Under the influence of the transformation fallacy, we observe that a necessary condition for F to achieve a local maximum is that all of its partial derivatives vanish. (The fallacy may be avoided at this stage by asking for critical points, rather than local maxima.) Thus we arrive at the system of equations

$$\begin{aligned} \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} &= 0 \\ \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} &= 0 \\ g(x, y) &= 0. \end{aligned} \tag{1}$$

The last equation, which is imposed as a constraint in the standard approach, arises here from the partial derivative $\partial F / \partial \lambda$. It shows that for any critical point (x, y, λ) of F, the components x and y satisfy the constraint equation $g = 0$.

Justifying Lagrange multipliers

In the standard formulation, the key idea is that ∇f and ∇g are parallel at the solution point (x^*, y^*). This can be justified in a variety of ways. For example, suppose $\mathbf{r}(t)$ parameterizes the constraint curve, with $(x^*, y^*) = \mathbf{r}(t^*)$. Then since $g(\mathbf{r}(t))$ is constant and $f(\mathbf{r}(t))$ has a maximum at t^*, both of their derivatives vanish at t^*. Applying the chain rule there, both ∇f and ∇g are orthogonal to $\mathbf{r}'$ and so are parallel. A less formal argument uses the directional derivative. At the point of the constraint curve C where f assumes a maximum, the derivative of f in the direction of C must vanish. This shows that ∇f and the curve's tangent vector are orthogonal. On the other hand, since C is a level curve of g, we also have ∇g orthogonal to C at every point, and the result follows as before.

Our understanding of vectors, and in particular, of ∇f as pointing in the direction of steepest increase of f, provides other justifications. At a maximum, ∇f must be perpendicular to C for otherwise, its nonzero projection along C would point in the direction of increase of f. Or, in the succinct formulation of Farris [**3**],

> We think of ∇f as the "desired direction" and ∇g as the "forbidden direction." If you are at a relative maximum of f on the constraint $g = 0$, then this can only be because the direction you would like to go to get more f, namely ∇f, lines up perfectly with the forbidden direction.

These arguments all depend on properties of the gradient, and there are several others of a similar nature [**8**]. There are other arguments from viewpoints that are decidely different. One involving *penalty functions* is most natural in the context of a minimization problem. So, suppose we wish to minimize $f(x, y)$ subject to the condition $g(x, y) = 0$. We form the function $F(x, y) = f(x, y) + \sigma g(x, y)$, and seek an unconstrained minimum. We think of σ as a penalty imposed for allowing g to assume a positive value. The larger σ is, the larger the penalty. Intuitively, by minimizing F for ever larger values of σ, we should be able to drive the solution toward a point where g is zero and f is minimized. This can be used iteratively to seek numerical estimates for contrained minima. It can also be used to justify the Lagrange multipliers technique, and variants [**5**, pp. 255–261] [**12**].

Our final justification for Lagrange multipliers, like the penalty function approach, takes a dramatically different point of view from the standard gradient-based arguments. Pourciau [**13**] attributes this approach to a 1935 work of Carathéodory, and refers to it as the *Carathéodory Multiplier Rule*. It applies most generally to the case of n variables and $n - 1$ contraints, but we will consider it here in the case $n = 2$. I find it amazingly simple and beautiful.

As before, we wish to maximize $f(x, y)$ subject to $g(x, y) = 0$. However, this time we combine the two functions to define a mapping $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$, given by

$$\Phi(x, y) = (f(x, y), g(x, y)).$$

This situation is illustrated in FIGURE 1. As depicted in the figure, we assume the axes in the image plane are labeled u and v, so that the mapping has the alternate definition

$$u = f(x, y)$$
$$v = g(x, y).$$

Now observe that the constraint set $g = 0$ is characterized precisely as the set of points mapped by Φ to the u axis. Therefore, maximizing f subject to $g = 0$ corresponds to finding the point on the u axis in the range of Φ that is as far to the right as possible. If the constrained optimization problem has a solution at (x^*, y^*), then $(u^*, 0) = \Phi(x^*, y^*)$ must be on the boundary of the range. Otherwise, it would be possible to go even further to the right.

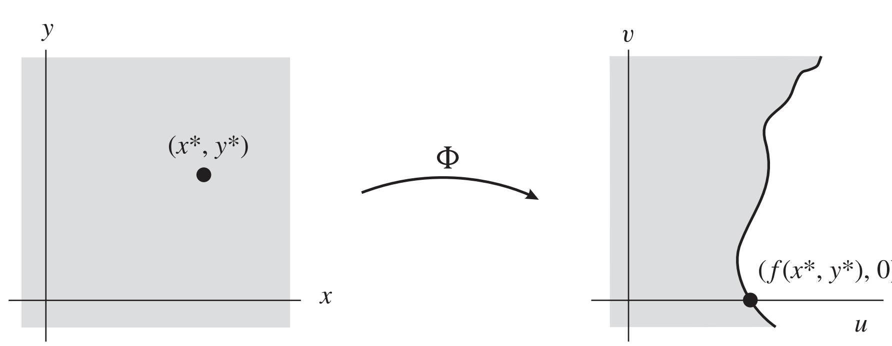

Figure 1 Action of the mapping Φ

Next, consider the derivative of Φ, which can be expressed as the matrix

$$d\Phi(x, y) = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{bmatrix}.$$

Whenever $d\Phi(x, y)$ is invertible (or nonsingular), it is known from the theory of differentiable functions that Φ maps a neighborhood of (x, y) to a neighborhood of $\Phi(x, y)$. In particular, that would put $\Phi(x, y)$ in the interior of the range. Thus, since $\Phi(x^*, y^*)$ is on the boundary of the range, we conclude that $d\Phi(x^*, y^*)$ is singular. That means the matrix has dependent rows, so $\nabla f(x^*, y^*)$ and $\nabla g(x^*, y^*)$ are parallel.

While this last justification is beautiful and elegant, it depends on a fairly well developed mathematical sophistication, at the level of an advanced calculus course say. The earlier arguments based on properties of the gradient are more accessible, and would be suitable for a multivariable calculus class serving math, science, and engineering students. However, in courses in mathematical methods for economics and applied or business calculus, many of which do not emphasize vector methods, the foregoing justifications for Lagrange multipliers may be out of reach. Perhaps this contributes to the popularity of the Lagrangian function formulation for these classes.

For these students, it is easy to understand the attraction of the transformation fallacy. It provides a memorable rationale for the Lagrange multipliers technique, builds on prior knowledge of unconstrained optimization, and in the end leads to correctly constituted necessary conditions for the solution. Often, it is presented with no justification beyond the vague idea that incorporating g into the objective function F implicitly imposes the constraint on the optimization process. But why should that be? Does the manner of incorporating g into F matter at all? Is there a specific argument to show why a maximum of F has anything to do with a constrained maximum of the original function f?

I have seen a justification in more than one source along the following lines. Because $\partial F/\partial\lambda = g$, a maximum of F will have to occur at a point where $g = 0$. Thus, maximizing F implicitly imposes the constraint condition. At the same time, if $F(x^*, y^*, \lambda^*)$ is a local maximum of F, then it certainly is greater than or equal to the values assumed by F at all nearby points, including those where $g = 0$. So, since F and f are identical for the points where $g = 0$, we have actually found a local maximum of f among such points.

Although this is all correct, it suffers from two major flaws. First, it turns out that trying to maximize F is destined to fail, F will not have any local maxima. Consequently, the strategy of solving the original problem by finding the maximum of F loses much of its appeal.

The second flaw is more subtle, and depends on a level of sophistication that is rare among calculus students. Correctly understood, the Lagrange multipliers technique provides a *necessary* condition for a solution of a constrained optimization problem. This justifies how the technique is usually applied. We find all solutions to the Lagrange conditions, and then choose among them the point that solves our original problem. This is valid because the Lagrange conditions are necessary: They must be satisfied by the solution of the opimization problem.

There is one logically correct way to justify a necessary condition. You must assume that a solution to the original problem is given, and then show that it also satisfies the necessary conditions. In this light, we see that the rationale above for the transformation fallacy is exactly backward. It begins with a point that is assumed to be a local extremum of F (the proposed necessary condition), and then tries to argue that such a point is a local constrained maximum of f. Even if F had local extrema (and it doesn't), there would be no assurance that they include *all* local constrained maxima of f. In particular, there is no assurance that maximizing F will find the global constrained maximum of f, which is what we ultimately wish to find.

An example The preceding argument refutes one proposed justification for the transformation fallacy, but that is not enough to establish that it is indeed a fallacy. Therefore, let us consider an example. Specifically, we wish to see that an extremum of a function f subject to a constraint need not correspond to an unconstrained local extremum of the Lagrangian function F. Indeed, any example will do, because the Lagrangian function (essentially) *never* possesses any local maxima or minima.

With that in mind, let us consider the following perfectly pedestrian application of Lagrange multipliers. The problem is to find the point of the curve $xy = 1$ that is closest to the origin. In the standard formulation, we must minimize $f(x, y) = x^2 + y^2$ subject to the constraint $g(x, y) = 0$, where $g(x, y) = xy - 1$. By symmetry, we may restrict our attention to (x, y) in the first quadrant. The Lagrange multiplier conditions (1) become

$$\begin{aligned} 2x + \lambda y &= 0 \\ 2y + \lambda x &= 0 \\ xy &= 1. \end{aligned}$$

The only solution is $(x^*, y^*, \lambda^*) = (1, 1, -2)$. Geometric considerations show that $(1, 1)$ is indeed the closest point of the curve $xy = 1$ to the origin.

Now we ask, does $F(x, y, \lambda) = x^2 + y^2 + \lambda(xy - 1)$ have a local extremum at $(1, 1, -2)$? To see that it does not, it will be enough to show that the restriction of F to a particular plane through $(1, 1, -2)$ has a saddle point there. Then any neighborhood of $(1, 1, -2)$ includes points where F assumes both greater values and lesser values than $F(1, 1, -2)$, which can therefore be neither a local minimum nor a local maximum.

Consider the plane generated by unit vectors parallel to $\nabla g(x^*, y^*) = (1, 1)$ and the λ axis. In (x, y, λ) coordinates, the unit vectors are $\mathbf{u} = (1/\sqrt{2}, 1/\sqrt{2}, 0)$ and $\mathbf{v} = (0, 0, 1)$. Therefore, a point on the plane may be expressed parametrically as

$$\begin{aligned} \mathbf{r}(s, t) &= (1, 1, -2) + s\mathbf{u} + t\mathbf{v} \\ &= (1 + s/\sqrt{2}, 1 + s/\sqrt{2}, t - 2). \end{aligned}$$

Note that $(s, t) = (0, 0)$ corresponds to the point $(1, 1, -2)$, where ∇F is zero.

To restrict F to the plane, we define $h(s, t) = F(\mathbf{r}(s, t))$. We know that h will have a critical point at $(0, 0)$. To see whether this is a local extremum, we use the second derivative test [**14**, p. 794]. First, compute the composition $F(\mathbf{r}(s, t))$ to obtain

$$h(s, t) = (\sqrt{2}s + s^2/2)t + 2.$$

The Hessian matrix of second partial derivatives,

$$\begin{bmatrix} \dfrac{\partial^2 h}{\partial s^2} & \dfrac{\partial^2 h}{\partial s\,\partial t} \\ \dfrac{\partial^2 h}{\partial s\,\partial t} & \dfrac{\partial^2 h}{\partial t^2} \end{bmatrix} = \begin{bmatrix} t & s + \sqrt{2} \\ s + \sqrt{2} & 0 \end{bmatrix},$$

has determinant -2 at $(s, t) = (0, 0)$. Since it is negative, we know that $h(s, t)$ has a saddle point at the critical point $(0, 0)$. This shows that $F(x, y, \lambda)$ cannot have a local maximum or minimum at $(1, 1, -2)$. In fact, $(1, 1, -2)$ and $(-1, -1, -2)$ are the only critical points of F, which therefore has *no* local maxima or minima.

This reveals the flaw in the transformation fallacy, which tells us to solve the constrained optimization problem by finding the unconstrained minimum of F. That is impossible for no minimum exists. Nevertheless, the constrained problem has a solution.

A theorem

What happened in the example always happens. For any objective function and constraint, the restriction of the Lagrangian function to a plane parallel to $\nabla g(x^*, y^*)$ and the λ axis has a saddle point at (x^*, y^*, λ^*), as shown in the following theorem. Although stated for functions of two variables, it extends naturally to functions of n variables.

First, we will establish a few notational conventions. Partial derivatives will be expressed using subscript notation, so f_x is the partial derivative of f with respect to x and f_{xy} is the second partial derivative with respect to x and y. We will abbreviate $\Phi(x^*, y^*)$ by Φ^* for any function Φ. In particular, $f^* = f(x^*, y^*)$, $\nabla f^* = \nabla f(x^*, y^*)$, and $f^*_{xx} = f_{xx}(x^*, y^*)$.

THEOREM. *Let f and g be functions of two variables, with continuous second derivatives. Let $F(x, y, \lambda) = f(x, y) + \lambda g(x, y)$. Then, if (x^*, y^*, λ^*) is a critical point of F at which ∇g^* is not the zero vector, F must have a saddle point at (x^*, y^*, λ^*).*

Proof. At any critical point (x^*, y^*, λ^*), all of the first partial derivatives of F vanish. This shows that $g^* = 0$ and $\nabla f^* + \lambda^* \nabla g^* = 0$. We also assume that ∇g^* is not zero.

Without loss of generality, we may assume $(x^*, y^*) = (0, 0)$ and ∇g^* points in the direction of the x axis: these conditions may be brought about by translating the x-y plane and rotating it about the λ-axis, neither of which alters the character of (x^*, y^*, λ^*) as a saddle point (or not). With these assumptions, $\nabla g^* = (g^*_x, g^*_y) = (a, 0)$ for some $a \neq 0$.

Next we will consider the restriction of F to the plane $y = 0$ in (x, y, λ) space. Let

$$h(x, \lambda) = F(x, 0, \lambda) = f(x, 0) + \lambda g(x, 0).$$

Then

$$\nabla h(x, \lambda) = (f_x(x, 0) + \lambda g_x(x, 0), g(x, 0)).$$

This vanishes at $(x, \lambda) = (0, \lambda^*)$, so $(0, \lambda^*)$ is a critical point of h.

We show that $(0, \lambda^*)$ is a saddle point of h by using the second derivative test. We need the Hessian matrix of second partial derivatives, which is defined by

$$H(x, \lambda) = \begin{bmatrix} h_{xx} & h_{x\lambda} \\ h_{x\lambda} & h_{\lambda\lambda} \end{bmatrix} = \begin{bmatrix} f_{xx}(x, 0) + \lambda g_{xx}(x, 0) & g_x(x, 0) \\ g_x(x, 0) & 0 \end{bmatrix}.$$

At the critical point, the determinant of the Hessian is $\det H(0, \lambda^*) = -(g^*_x)^2$, and this is negative because we know $g^*_x \neq 0$. Therefore, the second derivative test shows that h has a saddle point at $(0, \lambda^*)$. Hence F must have a saddle point at (x^*, y^*, λ^*), as asserted. ■

The theorem tells us that the Lagrangian function approach has *nothing* to do with finding a local maximum of F, and in practice we know that is true. Rather, *all* the critical points of F are found, and these are further examined to find the constrained maximum of f. Along the way, no one ever checks to see which critical points are local maxima of F. Otherwise, the transformation fallacy could not possibly persist. And indeed, specialists in optimization consider the preceding theorem common knowledge. I've consulted several economics faculty who specialize in mathematical methods, and they were all well aware of the result. But the result deserves to be better known among mathematicians, particularly the nonspecialists who teach Lagrange multipliers in calculus classes.

The theorem also crystalizes in a dramatic way what is needed in an intuitive justification of the Lagrangian approach. Because F essentially never has local extrema, a proper justification must explain why the critical points of F (and in fact, really why the *saddle* points of F) are significant. And it must do so without assuming that F is maximized somewhere. The transformation fallacy does not provide such a justification. Shortly we will see an intuitive argument that does. First, though, we take a brief look at the history of the Lagrange multiplier technique.

What did Lagrange say?

In the original development of the multiplier method by Lagrange, it is the Lagrangian function approach that appears, although without considering the multipliers to be independent variables of the function. What I have called the standard approach is a later development.

Interestingly, Lagrange did not initially formulate the multiplier method in the context of constrained optimization, but rather in the analysis of equilibria for systems of particles. He reported this application in *Mécanique Analytique*, published in 1788 and available now in English translation [**9**]. Using series expansions to analyze the effects of perturbation at the point of equilibrium, Lagrange derives conditions on the first order differentials for systems under very general assumptions. In the case that the particle motions are subject to external constraints, he points out that the constraints can be used to eliminate some of the variables that appear in his differential equations, before deriving the conditions for equilibrium. He goes on to observe that

> the same results will be obtained if the different equations of [constraint], each multiplied by an undetermined coefficient are simply added to [the general formula of equilibrium]. Then, if the sum of all the terms that are multiplied by the same differential are put equal to zero, as many particular equations as there are differentials will be obtained [**9**, p. 60].

Next, he explains how this procedure can be "treated as an ordinary equation of maxima and minima." However, this last statement must be understood in the context of Lagrange's earlier remarks, where he is careful to point out that "the equation of a differential set equal to zero does not always represent a maximum or minimum [**9**, p. 55]." Thus, Lagrange must have intended to convey that the conditions rigorously derived by his perturbation analysis could be found by formulating a Lagrangian function, and proceeding as if seeking a maximum or minimum. But there is no suggestion that the equilibrium point must actually correspond to a maximum or minimum in general, nor that the existence of such an extremum is necessary to establish the validity of the equilibrium conditions.

Having developed the multiplier technique in the analysis of equilibria, Lagrange proceeded to use it in other settings, notably in the calculus of variations [**4**]. Today's familiar application to constrained optimization problems was presented in two pages in his *Théorie des Fonctions Analytiques* of 1797, nearly a decade after the initial work with statics. As in the earlier work, Lagrange first establishes the conditions for a constrained extremum using series expansions. Then he points out that the conditions so described can be obtained according to a general principle [**10**, p. 198]. In translation, that principle runs as follows:

> When a function of several variables must be a maximum or a minimum, and there are one or more equations relating these variables, it will suffice to add to the proposed function the functions that must be equal to zero, multiplied each one by an indeterminate quantity, and next to find the maximum or minimum as if these variables were independent; the equations that one will find combined with the given equations will serve to determine all the unknowns.

On casual inspection, this appears to be a statement of the transformation fallacy. However, two points are significant. First, in saying to find the extremum as if the variables were independent, Lagrange clearly does not consider these variables to include the multipliers. This is reflected not only by the context in which he uses the phrase *these variables,* but also by his inclusion of the phrase *combined with the given equations.* If the multipliers were considered as variables, there would be no need to separately mention the constraint equations, which would appear as partial derivatives with respect to the multipliers. Second, consistent with my earlier remarks, it seems evident that Lagrange only advised proceeding *as if* seeking a maximum or minimum, and that the key point is the assertion of the final clause: the variables can be found by solving the given set of equations. Remember, too, that the validity of that assertion, established in a separate argument, did not depend on the existence of an unconstrained extremum at the solution point.

Apparently the idea of considering the multipliers as new variables for the optimization problem originated elsewhere, probably when someone recognized that the partial derivatives with respect to the multipliers are exactly the quantities that the constraints hold at zero. To a reader with that observation in mind however, it is easy to imagine that the quoted passage above, taken out of context, might encourage a belief in the transformation fallacy.

Lagrangian leveling We come at last to my proposed rationale for the Lagrangian function approach. The transformation fallacy says that the Lagrangian function is a device by which a constrained problem becomes an unconstrained problem. An equally memorable idea, and one that is correct, is that the Lagrangian is an approach that *levels the playing field.*

To make this idea clear, let us consider once again the fundamental problem, to maximize $f(x, y)$ subject to $g(x, y) = 0$. Suppose that the solution point occurs at (x^*, y^*). If we are very lucky, this will actually be a local maximum of f disregarding the constraint. In this case the graph of f is a surface G with a high point at $(x^*, y^*, f(x^*, y^*))$, so the tangent plane is horizontal there. That is why we can find such points by setting the partial derivatives equal to 0.

Usually, though, this is not what occurs in a constrained problem. As portrayed in FIGURE 2, we can only consider points (x, y) on a curve C in the domain. The graph of the restricted function $f|_C$ is a curve G_C on surface G. The constrained maximum occurs at a high point of G_C but it is not a high point of G, and the tangent plane of G is not horizontal there. Traveling along the plane in the direction of G_C we would experience a slope of 0, because the curve has a high point, but traveling off of the curve we can go even higher. In particular, traveling on the tangent plane perpendicular to G_C the slope is not 0.

In order to rectify this problem, we would like to flatten out the graph of f, so that the tangent plane becomes horizontal. In the process, we do not want to disturb G_C, because that might change the location of the constrained maximum. But it should be perfectly alright to alter the values of the function at points that are not on the constraint curve. In fact, near the high point of G_C, we can imagine pivoting the graph of f around the curve's horizontal tangent line. If we pivot by just the right amount, the

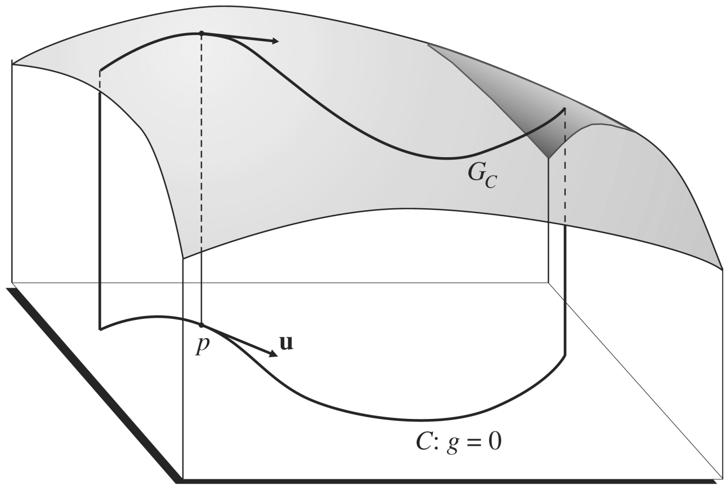

Figure 2 Constraint curve C in the domain of f, the graph G of f, and the graph C_G of $f|_C$

tangent plane will become horizontal, and so detectable by setting partial derivatives to zero. That is the image of leveling the playing field.

How do we put this into effect analytically? The simplest way to alter the values of the function f is to add some perturbation. But we want the amount we add to be 0 along the constraint curve, so as not to alter f for those points. On the other hand, we know that g is 0 along the constraint curve. So, adding g to f is the sort of perturbation we need. It leaves the values of f unchanged along the constraint curve, but modifies the values away from that curve. More generally, we can add any multiple of g to f and achieve the same effect. That is the motivation behind defining for each λ a perturbed function

$$F(x, y) = f(x, y) + \lambda g(x, y).$$

In this conception, we do not think of F as a function of three variables. Rather, we have in mind an entire family of two variable functions, one for each value of the parameter λ. A representation of this situation is shown in FIGURE 3. The surface marked $\lambda = 0$ is the unperturbed graph of f. The other surfaces are graphs of different members of the family of F functions. All of the surfaces intersect in the graph of f over the constraint curve, so the constrained maximum is the high point on the intersection curve.

The figure shows that choosing different values of λ imposes just the sort of pivoting action described earlier. Intuition suggests (and a little further analysis proves) that there must actually be a choice of λ that makes the tangent plane horizontal. That is, if (x^*, y^*) is the location of the maximum value of f subject to the constraint $g = 0$, then for some λ it will be the case that $\nabla F(x^*, y^*) = 0$. In that case, the following equations must hold:

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

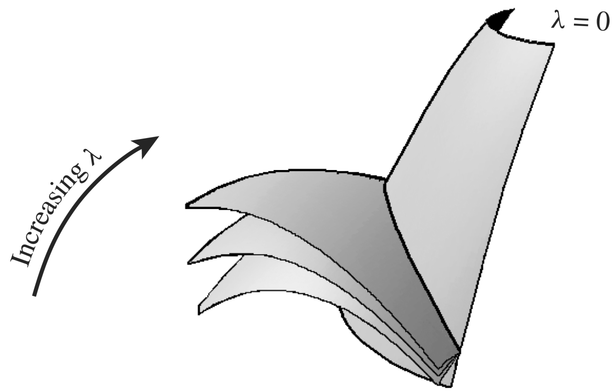

Figure 3 Graphs of several members of the family of F functions

By finding every possible triple (x, y, λ) for which these equations hold, we obtain a candidate set that must contain the solution to the constrained optimization problem. That is what the Lagrange multipliers theorem says.

With the concept of *leveling*, we can also rederive an interpretation of λ at the solution point, popular in the optimization literature for applications in economics [**2**, pp. 376–377], [**6**, p. 611]. At the maximum point over the constraint curve, we know that the directional derivative $D_{\mathbf{u}} f$ in the direction of the curve is zero. And we are assuming that the derivative $D_{\mathbf{u}} f$ in the direction normal to the constraint curve is not 0.

To level the tangent plane, we need to choose λ so that F will have zero directional derivative normal to the curve. That is, for $\mathbf{u}$ normal to the curve, we want $D_{\mathbf{u}}(f + \lambda g) = 0$. Clearly, we need to choose $\lambda = -D_{\mathbf{u}} f / D_{\mathbf{u}} g$. This shows, by the way, when λ exists. We may take for $\mathbf{u}$ the unit vector in the direction of ∇g. Then $D_{\mathbf{u}} g = \nabla g \cdot \mathbf{u} = |\nabla g|$. In particular, $D_{\mathbf{u}} g$ can only be 0 if $\nabla g = 0$. Otherwise, for $\lambda = -D_{\mathbf{u}} f / D_{\mathbf{u}} g$, the modified function F will have a horizontal tangent plane.

But there is also a meaningful interpretation of this choice for λ. The ratio $D_{\mathbf{u}} f / D_{\mathbf{u}} g$ is the rate of change of the objective function f relative to a change in the constraint function g. It shows, for a given perturbation of the point (x^*, y^*) orthogonally away from the constraint curve, how the change in f compares to the change in g. The economists interpret this as the marginal change in the objective function relative to the constraint. It indicates to first order, how the maximum value will change under relaxation or tightening of the constraint.

Concluding remarks There is a vast literature on optimization. It covers in great depth ideas that have been barely touched upon here, and far more. A good general reference most closely related to the topics discussed in this paper is Hestenes [**5**]. This provides a general discussion of Lagrange multipliers, including the transformation of constrained to unconstrained problems through a process called *augmentation*, as well as a detailed account of penalty functions. Although Hestenes does not use the terminology of *leveling*, this idea is implicit in his treatment of *augmentability*. For a general account of multiplier rules, I highly recommend Pourciau [**13**] (winner of a Ford award).

An expanded version of the present paper appears in [**8**]. Among other things, it gives examples of textbooks that encourage a belief in the transformation fallacy, implicitly or explicitly, as well as several geometric arguments in support of the standard approach to Lagrange multipliers.

The image of leveling the graph of f using the Lagrangian function F offers an attractive way to think about Lagrange multipliers. In this formulation, we find a clear and convincing intuitive justification of the Lagrangian approach. It also makes the existence of the necessary λ completely transparent, while leading naturally to the interpretation of λ as the marginal rate of change of the objective function relative to the constraint. This has the potential of enriching the insight of students, particularly those interested in the applications of mathematics to economics. And it is certainly superior to perpetuating the transformation fallacy.

Acknowledgment. Bruce Pourciau contributed many ideas that improved this paper; Ed Barbeau shared a wealth of material about Lagrange multipliers; Jennifer Kalman Beal generously provided a translation of Lagrange's description of his multiplier method [**10**].

REFERENCES

1. G. C. Archibald and Richard G. Lipsey, *An Introduction to Mathematical Economics: Methods and Applications,* Harper and Row, New York, 1976.
2. Alpha C. Chiang, *Fundamental Methods of Mathematical Economics,* 3rd ed., McGraw-Hill, New York, 1984.
3. Frank A. Farris, private communication, July 2008.
4. C. G. Fraser, Isoperimetric problems in the variational calculus of Euler and Lagrange, *Historia Mathematica* **19**(1) (1992) 4–23.
5. Magnus R. Hestenes, *Optimization Theory the Finite Dimensional Case,* Wiley, New York, 1975.
6. Michael Hoy, John Livernois, Chris McKenna, Ray Rees, and Thanasis Stengos, *Mathematics for Economics,* 2nd ed., MIT Press, Cambridge, MA, 2001.
7. Deborah Hughes-Hallett, Andrew M. Gleason, William G. McCallum, et al., *Calculus Single and Multivariable,* 2nd ed., Wiley, New York, 1998.
8. Dan Kalman, *Uncommon Mathematical Excursions: Polynomia and Related Realms,* MAA, Washington, DC, 2009.
9. J. L. Lagrange, *Analytical Mechanics,* translated from the Mécanique analytique, nouvelle édition of 1811 by Auguste Boissonnade and Victor N. Vagliente, Kluwer, Dordrecht, 1997.
10. J. L. Lagrange, *Théorie des fonctions analytiques contenant les principes du calcul différentiel, dégagés de toute considération d'infiniment petits ou d'évanouissans, de limites ou de fluxions, et réduits à l'analyse algébrique des quantités finies*, De l'imprimerie de la République, Paris, 1797.
11. Ronald E. Larson, Robert P. Hostetler, and Bruce H. Edwards, *Calculus with Analytic Geometry,* 6th ed., Houghton Mifflin, Boston, 1998.
12. E. J. McShane, The Lagrange multiplier rule, *Amer. Math. Monthly*, **80**(8) (1973) 922–925.
13. B. H. Pourciau, Modern multiplier rules, *Amer. Math. Monthly*, **87**(6) (1980) 433–452.
14. James Stewart, *Calculus Early Transcendentals,* 3rd ed., Brooks/Cole, Pacific Grove, CA, 1995.
15. George B. Thomas, Jr., Maurice D. Weir, Joel D. Hass, and Frank R. Giordano, *Thomas' Calculus Early Transcendentals,* 10th ed., Addison-Wesley, Boston, 2001.

Editor's Note: Unfortunately, an error crept into our April issue (**82**:2, p. 116). In the Proof Without Words: Ordering Arithmetic, Geometric, and Harmonic Means, the vertical line segment labeled $H(a, b)$ should stop at the hypotenuse of the triangle, and not go all the way up to the circle. A corrected image is shown here. We regret the error.