Constrained optimization

Ever since Calculus I you have known how to solve a constrained optimization problem for an ordinary function of a single variable, namely, that of optimizing or extremizing (= maximizing or minimizing) a function f over an interval of the form $I = \{x|a \leq x \leq b\}$. First you observe that, because the interval is closed (= includes its endpoints a and b), the global—or absolute—optimum is bound to exist as long as f is continuous.* Now you find the critical points, thus identifying all possible candidates for global optimizer in the interior of the interval, that is, on (a, b). For the sake of simplicity, let's assume that our function is not only continuous but also smooth.† Then $x = x^*$ being a critical point means

$$f'(x^*) \;=\; 0. \tag{1}$$

Next you add the boundary points to the list of candidates. Finally, you hold an election, comparing all candidates x^* to see whose $f(x^*)$ is most extreme, and hence globally optimal. The upshot is that you proceed in two separate stages to solve a problem of the form

$$\text{Optimize} \quad f(x) \tag{2a}$$
$$\text{subject to} \quad x \in I \tag{2b}$$

where I is a closed subset of the real line (conceived of as the x-axis).

What changes if f is a function of two variables? We proceed by analogy. The two-dimensional analogue of the above optimization problem is to optimize $f(x, y)$ over a rectangle of the form $a_1 \leq x \leq b_1, a_2 \leq y \leq b_2$. First we observe that, because the rectangle is closed (= includes its boundary lines), the global optimum is bound to exist as long as f is continuous. Now we find the critical points, and hence all possible candidates for global optimizer in the interior of the rectangle, that is, on $(a_1, b_1) \times (a_2, b_2) = \{(x, y)|a_1 < x < b_1, a_2 < y < b_2\}$. For the sake of simplicity, let's again assume that our function is smooth. Then (x^*, y^*) being a critical point means

$$\nabla f(x^*, y^*) \;=\; 0. \tag{3}$$

We discard any of the critical points that happen to be saddle points, because a saddle point is not even a local optimizer, and hence cannot be a global optimizer. Now we add all possible candidates for global optimizer on the boundary of the rectangle to those in the interior.‡ Finally we hold an election, comparing all candidates (x^*, y^*) to see whose $f(x^*, y^*)$ is most extreme. Moreover, nothing fundamentally changes when we optimize over a closed (= includes its boundary curve) region D in the x-y plane that is not a rectangular region—we just replace "boundary of rectangle" by "boundary of D" in the above.

In the special case where D is indeed a rectangle, we can obtain the boundary candidates for local optimizer by separately optimizing along each of its four boundary lines $x = a_1$, $y = a_2$, $x = b_1$ and $y = b_2$. That is, we optimize $f(a_1, y)$ for $a_2 \leq y \leq b_2$, $f(x, a_2)$ for

*There may be many local optimizers, but either only one is a global optimizer; or else, if there is more than one global optimizer, then each is associated with the same global optimum.

†So there are no interior points at which, say, f' instantly changes sign from positive to negative to yield a local maximum, even though f' fails to be zero at the point in question because it is undefined there.

‡More on this step in a moment.

$a_1 \le x \le b_1$, $f(b_1, y)$ for $a_2 \le y \le b_2$ and $f(x, b_2)$ for $a_1 \le x \le b_1$. But each of these four problems is a standard Calculus-I problem that you already know how to solve—so no more need be said about it.

The upshot is that we proceed in two stages whenever we solve a problem of the form

$$\text{Optimize} \quad f(x, y) \tag{4a}$$
$$\text{subject to} \quad (x, y) \in D \tag{4b}$$

where D is some closed subset of the x-y plane in two separate stages. The first step is to find all the critical points of f, discarding any critical points that are saddle points or lie outside D. The remaining critical points are the interior candidates for global optimizer, corresponding to local optima or extrema (= maxima or minima, depending on whether you have a maximization or minimization problem) inside D. The second step is to find the boundary candidates for global maximizer. Finally we hold an election, comparing all candidates (x^*, y^*) for global optimizer to see whose $f(x^*, y^*)$ is indeed most extreme.

Often the statement that (x, y) lies on the boundary of D can be written in the form

$$g(x, y) \;=\; C \tag{5}$$

where g is a smooth function and C is a constant. Then the second step of the above procedure—the boundary optimization—requires us to know how to optimize $f(x, y)$ subject to the constraint defined by (5). But the chain rule enables us to reduce this problem to one we already know how to solve. Recall from Lecture 8 that

$$\frac{\partial f(x, y)}{\partial t} \;=\; f_x \frac{\partial x}{\partial t} + f_y \frac{\partial y}{\partial t} \tag{6}$$

where x and y depend on t (and possibly other variables). For the sake of simplicity, let's assume both that $x = t$, or $t = x$ (which of course is exactly the same thing); and that $y = \psi(t)$. Then $y = \psi(x)$, so that (6) implies

$$\frac{\partial f(x, \psi(x))}{\partial x} \;=\; f_x + f_y \frac{d\psi}{dx}. \tag{7}$$

Now, in the vicinity of any potential optimizer, we can solve (5) for y in terms of x to yield, say, $y = \psi(x)$. Thus $y = \psi(x)$ is *defined* by

$$g(x, \psi(x)) \;=\; C. \tag{8}$$

Here an ordinary function of a single variable is equal to a constant; if we differentiate, then we are bound to get zero. So, from (7) with $f = g$:

$$\frac{dg}{dx} \;=\; \frac{\partial g(x, \psi(x))}{\partial x} \;=\; g_x + g_y \frac{d\psi}{dx} \;=\; 0. \tag{9}$$

But $y = \psi(x)$ also makes f a function of a single variable, whose only candidates for optimizer are its critical points. So, from (7), we require

$$\frac{df}{dx} \;=\; \frac{\partial f(x, \psi(x))}{\partial x} \;=\; f_x + f_y \frac{d\psi}{dx} \;=\; 0. \tag{10}$$

Note that (9) holds everywhere, whereas (10) holds only at a critical point; however, if we are going to optimize f, then at a critical point is precisely where we need to be. Thus, from (9)-(10), we have

$$f_x + f_y \frac{d\psi}{dx} \;=\; g_x + g_y \frac{d\psi}{dx} \;=\; 0 \tag{11}$$

for any potential optimizer. A little algebra now reveals that (11) implies

$$\frac{f_x}{g_x} \;=\; \frac{f_y}{g_y}, \tag{12}$$

where both sides are evaluated at the critical point. Let λ be the common value of the left- and right-hand sides of (12). Then we have $f_x = \lambda\, g_x$ and $f_y = \lambda\, g_y$ or, which is the very same thing (at least in two dimensions)

$$\nabla f \;=\; \lambda \nabla g. \tag{13}$$

So at any potential optimizer, say (x^*, y^*), the gradient of f must be parallel to the gradient of g. The constant of proportionality λ is called a *Lagrange multiplier*.[§]

Note that (5) and (13) are three equations for three unknowns, namely, x^*, y^* and λ. Note also that although the parameter λ is a "constant," it is different for different candidates; or if you prefer, although λ is independent of x and y it still depends on x^* and y^* (and so you will often see λ written as $\lambda(x^*, y^*)$ in texts on optimization).

Now suppose that f is a function of three variables instead of two. What then? The problem of optimizing $f(x, y)$ subject to $g(x, y) = C$ becomes the problem of optimizing $f(x, y, z)$ subject to

$$g(x, y, z) \;=\; C \tag{14}$$

where g is a smooth function. The chain rule enables us to reduce this 3-D problem to a 2-D problem. For recall from Lecture 6 that, if t and u are independent, implying

$$\frac{\partial u}{\partial t} \;=\; 0 \;=\; \frac{\partial t}{\partial u}, \tag{15}$$

then

$$\frac{\partial f(x, y, z)}{\partial t} \;=\; f_x \frac{\partial x}{\partial t} + f_y \frac{\partial y}{\partial t} + f_z \frac{\partial z}{\partial t} \tag{16a}$$

$$\frac{\partial f(x, y, z)}{\partial u} \;=\; f_x \frac{\partial x}{\partial u} + f_y \frac{\partial y}{\partial u} + f_z \frac{\partial z}{\partial u} \tag{16b}$$

where x and y depend on t and u (and possibly other variables). For the sake of simplicity, let's assume that $x = t$ and $y = u$ or (which of course is exactly the same thing) $t = x$, $u = y$; then $\frac{\partial y}{\partial t} = 0 = \frac{\partial x}{\partial u}$ and $\frac{\partial x}{\partial t} = 1 = \frac{\partial y}{\partial u}$. So (16) becomes

$$\frac{\partial f(x, y, z)}{\partial x} \;=\; f_x + f_z \frac{\partial z}{\partial x} \tag{17a}$$

$$\frac{\partial f(x, y, z)}{\partial y} \;=\; f_y + f_z \frac{\partial z}{\partial y}. \tag{17b}$$

[§]We assume $\nabla g \neq \mathbf{0}$. So the method of Lagrange multipliers turns out to be unsuitable for optimization on D when D is a rectangle, because although the boundary can be written as a single constraint of the form $g(x, y) = (x - a_1)(b_1 - x)(y - a_2)(b_2 - y) = 0$, $\nabla g = \mathbf{0}$ at any corner, where there may be an optimum.

Now, in the vicinity of any potential optimizer, we can solve (14) for z in terms of x and y to yield, say, $z = \psi(x, y)$. Thus $z = \psi(x, y)$ is *defined* by

$$g(x, y, \psi(x, y)) \;=\; C. \tag{18}$$

Here a function of two variables is equal to a constant; and so, if we differentiate it with respect to either x or y, then we are bound to get zero in either case. That is, from (17) with $f = g$ and $z = \psi$:

$$\frac{\partial g(x, y, \psi(x, y))}{\partial x} \;=\; g_x + g_z\,\frac{\partial \psi}{\partial x} \;=\; 0 \tag{19a}$$

$$\frac{\partial g(x, y, \psi(x, y))}{\partial y} \;=\; g_y + g_z\,\frac{\partial \psi}{\partial y} \;=\; 0. \tag{19b}$$

But $z = \psi(x, y)$ also makes f a function of two variables, whose only candidates for optimizers are its critical points. Therefore, from (17) with $z = \psi$, we require

$$\frac{\partial f(x, y, \psi(x, y))}{\partial x} \;=\; f_x + f_z\,\frac{\partial \psi}{\partial x} \;=\; 0 \tag{20a}$$

$$\frac{\partial f(x, y, \psi(x, y))}{\partial y} \;=\; f_y + f_z\,\frac{\partial \psi}{\partial y} \;=\; 0. \tag{20b}$$

Note that (19) holds anywhere, whereas (20) holds only at a critical point; however, if we are going to optimize f, then at a critical point is precisely where we need to be. Thus, from (19)-(20), we have

$$f_x + f_z\,\frac{\partial \psi}{\partial x} \;=\; g_x + g_z\,\frac{\partial \psi}{\partial x} \;=\; 0 \tag{21a}$$

$$f_y + f_z\,\frac{\partial \psi}{\partial y} \;=\; g_y + g_z\,\frac{\partial \psi}{\partial y} \;=\; 0 \tag{21b}$$

at any potential optimum. A little algebra now reveals that (21) implies

$$\frac{f_x}{g_x} \;=\; \frac{f_y}{g_y} \;=\; \frac{f_z}{g_z}, \tag{22}$$

where all three ratios are evaluated at the critical point; and if λ denotes the common value of all three ratios in (22), then $f_x = \lambda\, g_x$, $f_y = \lambda\, g_y$ and $f_z = \lambda\, g_z$ or (which is exactly the same thing, at least in three dimensions)

$$\nabla f \;=\; \lambda \nabla g. \tag{23}$$

Thus (13) holds in three dimensions as well as in two. That is, at any potential optimizer, say (x^*, y^*, z^*), the gradient of f must still be parallel to the gradient of g (and the constant of proportionality λ is still called a Lagrange multiplier): geometrically, the level surfaces of f and g through (x^*, y^*, z^*) must have a common normal, and hence a common tangent plane. Furthermore, in (14) and (23) we have four simultaneous nonlinear equations to determine x^*, y^*, z^* and $\lambda = \lambda(x^*, y^*, z^*)$.

A related problem to that of optimizing $f(x, y, z)$ subject to a single constraint, as in (14), is that of optimizing $f(x, y, z)$ subject to *two* constraints, namely,

$$g(x, y, z) = C_1 \tag{24a}$$
$$h(x, y, z) = C_2 \tag{24b}$$

where g and h are smooth functions. Each of (24a) and (24b) defines a surface, and we assume that these surfaces intersect in a curve, which we denote by Γ.[¶] Let Γ be parameterized by

$$\boldsymbol{\rho}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}. \tag{25}$$

Because Γ lies in both surfaces, it follows from (24) that

$$g(x(t), y(t), z(t)) = C_1 \tag{26a}$$
$$h(x(t), y(t), z(t)) = C_2. \tag{26b}$$

Any candidate for global optimizer of f subject to (24) must clearly lie on Γ, and hence must correspond to a particular value of the parameter t, say t^*; and because this value must optimize f on Γ, we must have $\omega'(t^*) = 0$ where

$$\omega(t) = f(x(t), y(t), z(t)). \tag{27}$$

By the chain rule, however, differentiation of (27) with respect to t yields

$$\omega'(t) = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial f}{\partial z} \cdot \frac{dz}{dt} = \nabla f \cdot \frac{d\boldsymbol{\rho}}{dt} \tag{28}$$

while differentiation of (26a)–(26b) with respect to t implies

$$\frac{\partial g}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial g}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial g}{\partial z} \cdot \frac{dz}{dt} = \frac{C_1}{dt} = 0 \tag{29a}$$
$$\frac{\partial h}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial h}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial h}{\partial z} \cdot \frac{dz}{dt} = \frac{C_2}{dt} = 0 \tag{29b}$$

for all t. Hence for $t = t^*$ we must have

$$\nabla f \cdot \frac{d\boldsymbol{\rho}}{dt} = \nabla g \cdot \frac{d\boldsymbol{\rho}}{dt} = \nabla h \cdot \frac{d\boldsymbol{\rho}}{dt} = 0, \tag{30}$$

implying that ∇f, ∇g and ∇h are all normal to $\frac{d\boldsymbol{\rho}}{dt}$ at any point $(x(t^*), y(t^*), z(t^*))$ that is a candidate for constrained global optimizer of f. It follows that ∇f, ∇g and ∇h must all lie in the same plane at the point in question, and hence that λ, μ exist such that[‖]

$$\nabla f = \lambda \nabla g + \mu \nabla h. \tag{31}$$

In other words, there are now *two* Lagrange multipliers—one for each constraint—and in (24) and (31) we have five simultaneous nonlinear equations to determine x^*, y^*, z^*, $\lambda = \lambda(x^*, y^*, z^*)$ and $\mu = \mu(x^*, y^*, z^*)$.

[¶] Needless to say, unless the surfaces intersect, the problem has no solution.

[‖] We assume that none of ∇f, ∇g or ∇h either equals $\mathbf{0}$ or is parallel to one of the other two.