Recitation 1 - Math Tools *

Nagisa Tadjfar

14.03/003 Fall 2025

Unconstrained Optimization: Single Variable

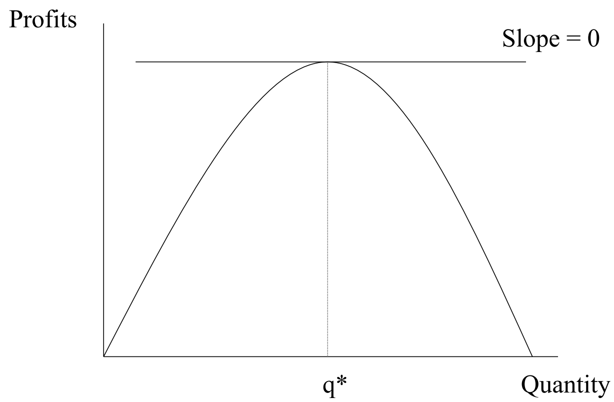
In many fields, but especially in economics, optimizing an objective function is a very important tool. For example, suppose a price-taking company has a profit function $\pi(q) = pq - q^2$ and wants to produce goods at quantity q^* that *maximizes* its profits. How do we find q^*?

Step 1: Find the 'critical point', or the value of q where $\frac{d\pi(q)}{dq} = p - 2q = 0$. Note that economists often refer to $\frac{d\pi(q)}{dq} = 0$ as a "first order condition", or FOC. Critical points are typically "candidates" for minimums or maximums of a function, but it is not guaranteed.

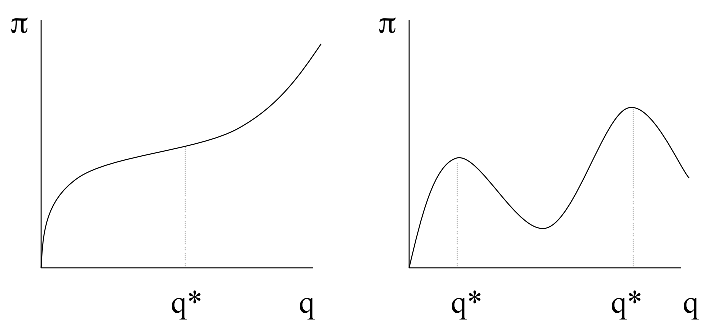
Step 2: In the single variable case, there is a simple way to check if if the critical point you found in step 1 is a local minimum or maximum. Check the sign of of the second derivative, $\frac{d^2\pi}{(dq)^2}\big|_{q^*}$, which can tell us whether we are at a local maximum or minimum. If $\frac{d^2\pi}{(dq)^2}\big|_{q^*} < 0$, the function is concave at $q*$, so we are at a local maximum. If $\frac{d^2\pi}{(dq)^2}\big|_{q^*} > 0$, the function is convex at q^*, so we are at a local minimum. This is often called a "second order condition", or SOC. In this example, $\frac{d^2\pi}{dq^2}\big|_{q^*} = -2$ and $-2 < 0$, which guarantees that q^* is a local maximum.

Because this profit function is *concave*, as shown in the figure below, it is intuitive that the sign of the second derivative, which captures the rate of change of the slope of the function, can tell you whether or not the critical point is a local minimum or maximum.

*Thanks to Professor Autor and Jon Cohen for sharing materials from previous years.



This method doesn't help when the function is not "well-behaved" (see figure below for a few examples). In 14.03/003, we will typically work with "well-behaved" functions that are continuous, differentiable, and concave.



Unconstrained Optimization: Multiple Variables

If we have a function of multiple variables, $y = f(x_1, x_2, ..., x_n)$, then the first step of finding critical points is similar to that of single variables. We take the partial derivatives of the objective function with respect to each variable and set them equal to zero. The partial derivative of a multivariate function such as f with respect to one variable (say x_1) is the derivative of f treating the other variables (in this case all the other $x_2 ... x_n$) as constants. Partial derivatives are denoted by the ∂ symbol or sometimes as f_1 (partial derivative of $f(x_1, x_2, ..., x_n)$ with respect to x_1).

The critical points of this function are the points where:

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = ... = \frac{\partial f}{\partial x_n} = 0$$

and these n equations are the "first order conditions". Again, these are points where the slope of the function is zero in every direction, and they are "candidates" for minimums and maximums of $f(x_1, x_2, ..., x_n)$. To know whether these critical points are global maximums or minimums, we again need to consider the second order conditions. To know whether these critical points are global maximums or minimums, we again need to consider the second order conditions. For multivariable functions the sufficient conditions for a critical point to be a minimum or a maximum turn out to be whether the Hessian matrix is positive definite or negative definite (all negative or all positive eigenvalues). Fortunately, we won't worry about this much in 14.03/003 and when we do it'll be in a case with two variables in which case there's a shortcut. The Second Order Condition (SOC) for a maximum when there is more than one variable has the following form:

$$\begin{aligned} f_{11} &< 0 \\ f_{11}f_{22} - (f_{12})^2 &> 0 \end{aligned}$$

Note that $f_{22} < 0$ is implied by the above, since if f_{11} and f_{22} have opposite signs there is no way to get $f_{11}f_{22} - (f_{12})^2 > 0$. Also note that it is not sufficient to have just $f_{11} < 0$ and $f_{22} < 0$. For a minimum, the SOC takes the form

$$\begin{aligned} f_{11} &> 0 \\ f_{11}f_{22} - (f_{12})^2 &> 0 \end{aligned}$$

If $f_{11}f_{22} - (f_{12})^2 < 0$, then the point in question is neither a minimum nor a maximum. To simplify this we will often work with concave functions because they have the nice property that the FOC is sufficient for a global max.

A *concave function* is a function that always lies below any hyperplane that is tangent to it. More formally, the definition of concave functions is that for any two points a, b and $\alpha \in (0, 1)$:

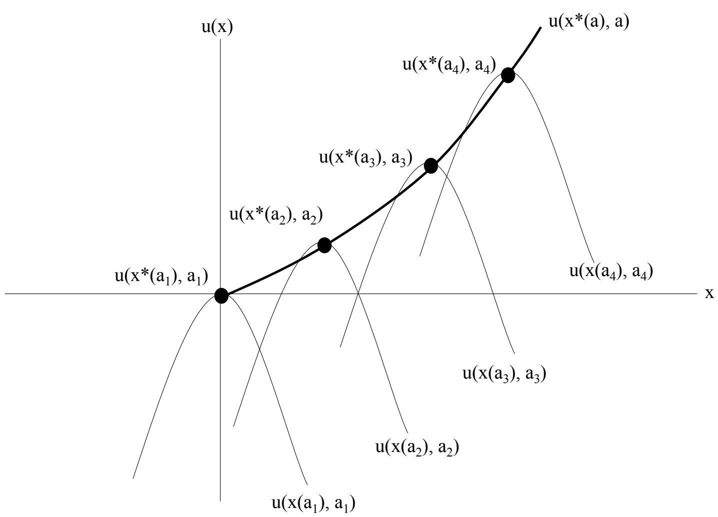
$$f(\alpha a + (1-\alpha)b) \geq \alpha f(a) + (1-\alpha)f(b)$$

Implicit Function Theorem and Envelope Theorem

So far, we've discussed how to find values (e.g. q^*) that maximize an objective function (e.g. profits $\pi(q)$) along with the maximized values of the objective (π^*). But what if we want to know how a *maximized* value changes if some parameters are altered? We will first discuss an example and walk through some math “tricks" that allow us to more easily find these “changes" and deal with maximized functions, which are themselves functions of certain parameters but distinct from the objective functions. Suppose x is the amount of ice cream you buy, u is the utility you get, and a is the “ambient temperature" that day (out of your control when you are maximizing your utility). Your utility function is as follows:

$$u(x;a) = -x^2 + ax$$

On any day, at a given ambient temperature a, you optimize maximizing your utility to buy $x^*(a)$ ice creams to get maximal utility $u^* = u(x^*(a);a)$. If we fix the temperature at some $a = a_i$, the utility function becomes a univariate function (a downward facing parabola). The figure below plots $u(x;a)$ for four different values of a_i. The peak of each parabola is the maximized utility $u^* = u(x^*(a);a)$. The function $u^* = u(x^*(a);a)$ is traced out by the peaks of each $u^* = u(x^*(a);a)$ (the “upper envelope" - shown by the bold line on the graph).



It might be interesting to know how your *maximized utility*, u^*, changes with temperature. Because for any given value of a, you optimize x to purchase $x^*(a)$ ice creams, what we are really after is $\frac{d}{da}u(x^*(a);a)$. Another object of interest may be how this optimized bundle of ice cream, $x^*(a)$, changes as a changes (i.e. $\frac{dx^*(a)}{da}$). One way to do this, of course, is the explicit way using the tools we have already learned. First, we could solve for $x^*(a)$ by taking the first order conditions, then find the expression for $u(x^*(a);a)$ by substituting $x^*(a)$ into the functional form of u and then taking the derivative with respect to a for both.

$$\frac{du(x;a)}{dx} = -2x + a = 0 \rightarrow x^* = \frac{a}{2}$$
$$\frac{dx^*(a)}{da} = \frac{d}{da}\frac{a}{2} = \frac{1}{2}$$
$$u(x^*(a);a) = -\left(\frac{a}{2}\right)^2 + a\left(\frac{a}{2}\right) = \frac{a^2}{4}$$
$$\frac{du(x^*(a);a)}{da} = \frac{d}{da}\frac{a^2}{4} = \frac{a}{2}$$

It turns out that with a few mathematical tricks, however, there is a shortcut to find both of these terms more easily. Consider a multivariate function $f(x,y)$. Recall that when taking the partial derivative of a multivariate function such as f with respect to one variable (say x), we treat the other variable (in this case y) as a constant. Partial derivatives are denoted by the ∂ symbol, or sometimes as f_x (partial derivative of $f(x,y)$ with respect to x). Next, we introduce the notion of a total derivatives. In general, it is not the case that the partial derivative of a multivariate function is equal to the total derivative. That is, in general $\frac{\partial f}{\partial x} \neq \frac{df}{dx}$. Why? If y is also a function of x, then the total derivative of the function $f(x,y(x))$ is a combination of a *direct* effect of varying x on the value of f but also the *indirect* effect of how varying x affects y, which in turn also affects f.

The chain rule for total derivatives is as follows:

$$\underbrace{\frac{df}{dx}}_{\text{Total derivative}} = \underbrace{\frac{\partial f}{\partial x}}_{\text{Direct Effect}} + \underbrace{\frac{\partial f}{\partial y}\frac{dy}{dx}}_{\text{Indirect Effect}}$$

Now, let's go back to our ice cream example. At the optimized level of ice cream, x^*, the first order condition is $\frac{\partial u(x;a)}{\partial x} = -2x^* + a = 0$. This first order condition is actually a *level curve*, which is often mathematically denoted as $G(x^*, a) = 0$. We can now take the total derivative of G with respect to a on both sides of this level curve. Notice that a nice feature of level curves is that the total derivative is equal to zero (by definition), which gives us a solvable equation that allows us to simply “rearrange" and obtain an expression for $\frac{dx^*}{da}$:

$$G(x^*, a) = -2x^* + a = 0$$
$$\frac{dG}{da} = \frac{\partial G}{\partial a} + \frac{\partial G}{\partial x^*}\frac{dx^*}{da} = 1 + (-2)\frac{dx^*}{da}$$
$$1 + (-2)\frac{dx^*}{da} = 0 \rightarrow \frac{dx^*}{da} = \frac{1}{2}$$

which is the same answer as we previously obtained when solving explicitly for $x^*(a)$. In taking this shortcut, in addition to using the expression for total derivatives, we relied on the *implicit function theorem*, which guarantees that under some mild mathematical conditions (that will almost always apply in 14.03/003 problems) that $x^*(a)$ exists and that we can therefore use the expression for the total derivatives shown above to solve for $\frac{dx^*(a)}{da}$. In economic applications, we often end up with implicit functions where exogenous variables (e.g. a in this example) and endogenous variables (e.g. $x^*(a)$ in this example) are all mixed together but we still want to know how the change in one variable (say a) affects another say (x). While in this simple example, it's straightforward to solve for $x^*(a)$ explicitly, sometimes with more complicated functional forms we can't solve explicitly for $x^*(a)$. However, the derivative $\frac{dx^*(a)}{da}$ may still exist and the implicit function theorem gives us the green light to solve for it. Here is a formal statement of the implicit function theorem:

Theorem 1 *(Implicit Function Theorem). Assume that F is a scalar function of class C^1 defined for all (x, y) in an open set $U \subset \mathbb{R}^2$. If $F(a, b) = 0$ and $\partial_y F(a, b) \neq 0$, then the equation $F(x, y) = 0$ implicitly determines y as a C^1 function of x, i.e. $y = f(x)$, for x near a. Moreover, the function f is of class C^1, and its derivatives may be determined by differentiating the identity $F(x, f(x)) = 0$ and solving to find the partial derivatives of f.*

We now use the expression of the total derivative to obtain an expression for $\frac{d}{da}u(x^*(a); a)$:

$$\frac{d}{da}u(x^*(a); a) = \frac{\partial u(x^*(a); a)}{\partial a} + \frac{\partial u(x^*(a); a)}{\partial x}\frac{dx^*(a)}{da}$$

We are interested in the derivative of the *maximized* utility function $u(x^*(a); a)$, and we know that at the optimum x^*, the term $\frac{\partial u(x^*(a);a)}{\partial x} = 0$. This means that $\frac{d}{da}u(x^*(a); a)$ reduces to the partial derivative, i.e. $\frac{\partial u(x^*(a);a)}{\partial x}$, which is simply $x^* = \frac{a}{2}$. That's the same answer that we got before! This equivalence between the total derivative and the partial derivative at the optimum is the *envelope theorem*. In other words, the envelope theorem states that because the slope of $u(x; a)$ when we are close to $x^*(a)$ is zero, if we move a around, $x^*(a)$ will change, but $u(x^*(a); a)$ won't change much and so the indirect effects don't do much and only the direct effect remains. Here is a formal statement of the envelope theorem:

Theorem 2 *(Envelope Theorem for the unconstrained case). Let $f(x, a)$ be a C^1 function of $x \in \mathbb{R}^n$ and the scalar a. For each a consider the unconstrained maximization:*

$$\max_x \ f(x; a)$$

Let $x^*\ (a)$ be a solution of this problem. Suppose that $x^*(a)$ is a C^1 function of a. Then,

$$\frac{d}{da}f(x^*(a),a) = \frac{\partial}{\partial a}f(x^*(a),a)$$

Proof.

$$\begin{aligned}\frac{d}{da}f(x^*(a),a) &= \underbrace{\sum \frac{\partial f}{\partial x_i}(x^*(a),a)\frac{\partial x_i^*(a)}{\partial a}}_{=0} + \frac{\partial f}{\partial a}(x^*(a),a) = \\ &= \frac{\partial f}{\partial a}(x^*(a),a)\end{aligned}$$

where the first term of the derivative is zero because of the FOCs of the maximization problem to obtain x^*:

$$\frac{\partial f}{\partial x_i}(x^*(a),a) = 0 \ \ \forall i = 1,2,...,n$$

■

Constrained Optimization

Most of the optimization problems we deal with in economics are subject to constraints (e.g. maximizing utility subject to a budget). A technique to solve constrained optimization problems is the method of *Lagrange multipliers*, which is a special case of the Karush-Kuhn-Tucker (KKT) conditions used more broadly in optimization. Suppose we want to maximize an objective function $f(x_1,x_2,...,x_n)$ subject to the constraint that $g(x_1,x_2,...,x_n) \leq 0$. We typically assume that this inequality is an equality, $g(x_1,x_2,...,x_n) = 0$ (i.e. "the constraint binds"). Why? Because we are usually trying to maximize something like utility subject to a budget constraint. Utility is (usually) everywhere increasing in the goods you're consuming, so you'll want to exhaust your budget constraint (i.e. spend all your money - leave nothing on the table!) However, if you want to be very careful you can also find the critical points of the unconstrained function. If any of those satisfy the constraint, plug them in along with the solution(s) from the constrained optimization. That will ensure you find the maximum and/or minimum, even if the constraint does not bind. Assuming the constraint *binds*, the problem setup is then:

$$\begin{aligned}\max\ y &= f(x_1,x_2,...,x_n) \\ s.t.\ g(x_1,x_2,...,x_n) &= 0\end{aligned}$$

The Lagrangian function function for this problem is

$$\mathcal{L} = f(x_1, x_2, ..., x_n) + \lambda g(x_1, x_2, ..., x_n)$$

where λ is called the *Lagrange multiplier*. We then treat the Lagrangian function like an unconstrained optimization problem. We have $n+1$ first-order conditions to take (one for each x plus one for the newly introduced variable λ), which allows us to solve the system of equations for the optimized values x_1, x_2,...,x_n. These first order conditions are:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x_1} &= \frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} = 0 \\
\frac{\partial \mathcal{L}}{\partial x_2} &= \frac{\partial f}{\partial x_2} - \lambda \frac{\partial g}{\partial x_2} = 0 \\
&\ldots \\
\frac{\partial \mathcal{L}}{\partial x_n} &= \frac{\partial f}{\partial x_n} - \lambda \frac{\partial g}{\partial x_n} = 0 \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= -g(x_1, x_2, ..., x_n) = 0
\end{aligned}$$

We then solve simultaneously for x_1, ..., x_n, λ to find x^*.

Example: Optimal fence dimensions

Suppose we want to maximize the area inside a rectangular fence, subject to a fixed amount of fencing p. Formally, the problem setup is:

$$\begin{aligned}
&\max\ xy \\
s.t.\ 2x + 2y &= p
\end{aligned}$$

The Lagrangian for this problem is:

$$\mathcal{L} = xy + \lambda(p - 2x - 2y)$$

And the three FOCs for this problem are:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x} &= y - 2\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial x} &= x - 2\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= p - 2x - 2y = 0\end{aligned}$$

Solving the system of equations yields the optimized width and length of the rectangle (x^* and y^*), as well as the Lagrange multiplier λ:

$$\begin{aligned}x^* &= y^* = \frac{p}{4} \\ \lambda &= \frac{p}{8}\end{aligned}$$

We say λ is the "shadow price of the constraint." In other words, λ represents how much the objective function changes if we relax the constraint by one unit. To see this, suppose $p = 80$, so $x = y = 20$ and $\lambda = 10$. Then the objective function $xy = 400$. Now suppose we relax the constraint by 1, so $p = 81$. Then $x = y = 20.25$. The objective function $xy = 410.0625$. The objective changed by $\sim$10, which is the same as λ .

Envelope Theorem for Constrained Problems

Let $x^*(a)$ denote the solution to the following problem:

$$\begin{aligned}\max\ y = & \ f(x) \\ s.t.\ g(x;a) &= 0\end{aligned}$$

Let λ be the Lagrange multiplier for the constraint in this problem.
Then:

$$\underbrace{\frac{d}{da} f(x^*(a))}_{\substack{Total\ derivative\ of \\ the\ original\ function\ f}} = \underbrace{\lambda \frac{\partial g(x;a)}{\partial a} = \frac{\partial}{\partial a} \mathcal{L}(x^*(a), \lambda(a), a)}_{\substack{Partial\ derivative \\ of\ Lagrangian}}$$

Why is this true? First use the chain rule:

$$\underbrace{\frac{d}{da} f(x^*(a))} = \sum_i \frac{\partial f(x^*(a), a)}{\partial x_i} \frac{dx_i^*}{da}$$

The FOC for maximizing the Lagrangean $\mathcal{L}(x, \lambda, a) = f(x) + \lambda g(x; a)$ are

$$\frac{\partial \mathcal{L}(x, \lambda, a)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} + \lambda^* \frac{\partial g}{\partial x_i} = 0$$

So

$$\frac{d}{da} f(x^*(a)) = -\lambda \sum_i \frac{\partial g}{\partial x_i} \frac{dx_i^*}{da}$$

But by taking the derivative of $g(x; a) = 0$ with respect to a we get

$$\sum_i \frac{\partial g}{\partial x_i} \frac{dx_i^*}{da} + \frac{\partial g}{\partial a} = 0$$

Duality

Now that we have discussed constrained optimization, we want to briefly introduce the notion of duality. Broadly, duality is the notion that optimization problems can be viewed from either of two perspectives:

1. Primal problem

2. Dual problem

In constrained convex optimization problems under certain conditions (basically all problems we will consider in 14.03/003), the solutions to the dual problem and the primal problem are equivalent. In economics, this fact is useful because the two perspectives may have different economic interpretations (while have the same solutions). For instance, cost minimization (holding constant a target profit level) is the *dual problem* of profit maximization (subject to a budget constraint) and utility maximization (subject to a budget constraint) is the *dual problem* of expenditure minimization (holding constant a certain utility level). Below is an example to illustrate duality. Consider the *primal* problem:

$$\begin{aligned} \max\ z &= x^{\frac{1}{2}} y^{\frac{1}{2}} \\ s.t.\ x + y &= 4 \\ \mathcal{L} &= x^{\frac{1}{2}} y^{\frac{1}{2}} + \lambda(4 - x - y) \\ x^* &= y^* = 2,\ \ \lambda^* = \frac{1}{2},\ z^* = 2 \end{aligned}$$

The *dual* problem is then:

$$\begin{aligned}
\min\ k &= x+y \\
s.t.\, 2 &= x^{\frac{1}{2}}y^{\frac{1}{2}} \\
\mathcal{L}^D &= x+y+\lambda^D(2-x^{\frac{1}{2}}y^{\frac{1}{2}}) \\
x_D^* &= y_D^* = 2,\ \ \lambda_D^* = 2,\ z^* = 2,\ k = 4
\end{aligned}$$

Notice that $\lambda_D \neq \lambda_P$, where P stands for primal and D stands for dual. Note that in this case the choice of setting the maximization as the primal (and hence the minimization as the dual) is arbitrary and it could have been set up the other way around with the minimization as the primal and the maximization as the dual.