Lecture 5: Constrained Optimization & Envelope Theorem

Originally written by Mauricio Caceres Bravo
Revised by Cole Davis
Webpage: cj-davis99.github.io

August 19, 2024

Contents

Notation

- $\forall$ translates to "for all"
- $\exists$ translates to "there exists"
- $\mathbb{N} = \{1, 2, \ldots\}$ is the set of natural numbers
- $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ is the set of integers
- $\mathbb{Q} = \{p/q : p \in \mathbb{Z} \text{ and } q \in \mathbb{Z}\backslash\{0\}\}$ is the set of rational numbers
- $\mathbb{R}$ is the set of real numbers
- If S is a set and $n \in \mathbb{N}$, then S^n is the n^{th} order Cartesian product of S. E.g., $S^2 = S \times S$
- For any $\varepsilon > 0$, $B_\varepsilon(x)$ is the Euclidean ball around x with radius ε
- Unless otherwise specified, $d(x, y)$ is a metric on the contextual set x, y belong to
- The origin is always denoted as 0 regardless of the dimension of the space considered
- If v is a vector, then both v^T and v' can represent the vector transpose. Preference is usually given to the v^T notation.

1 Constrained Optimization

The general problem of constrained optimization can be written as

$$\max_{x \in \mathbb{R}^N} f(x) \quad g(x) \leq b \quad h(x) = c \tag{1}$$

$f : \mathbb{R}^N \to \mathbb{R}$ is a real-valued function, $g : \mathbb{R}^N \to \mathbb{R}^K$, $h : \mathbb{R}^N \to \mathbb{R}^M$ are vector-vaued functions, and $b \in \mathbb{R}^K$ and $c \in \mathbb{R}^M$ are constants. Recall

$$g(x) \leq b \equiv g_k(x) \leq b_k \ \ \forall k = 1, \dots, K \quad \text{and} \quad h(x) = c \equiv h_m(x) = c_m \ \ \forall m = 1, \dots, M$$

The function f is called the ***objective***, the functions that comprise g are called the ***inequality constraints***, and the functions that comprise h are called ***equality constraints***.

Example 1. What is the maximum of $f(x) = x^2$ on $[0, 1]$? We can write this as

$$\max_{x \in \mathbb{R}} f(x) \quad g(x) \leq b$$

where $g : \mathbb{R} \to \mathbb{R}^2$ is given by $g(x) = (-x, x)$ and $b = (0, 1)$. A more common way to write this, however, is

$$\max_{x \in \mathbb{R}} x^2 \quad \text{s.t.} \quad 0 \leq x \leq 1$$

In this case, the maximum occurs at $x = 1$ (note $f(x)$ is strictly increasing).

1.1 One Equality Constraint

Suppose that we only have a single equality constraint:

$$\max_{x \in \mathbb{R}^N} f(x) \quad h(x) = c$$

with $c \in \mathbb{R}$ some scalar. With unconstrained optimization, we look for candidate extrama by solving for x using the FOC. In this case, however, solving for the FOC doesn't immediately help because the candidate points might not be in the constraint. What to do?

- Suppose that we could solve for x_i for some i explicitly as a function of c and the other variables. WLOG let this be x_1; that is, given $h(x) = c$ we find φ s.t.

  $$x_1 = \varphi(x_{-1}) \quad \text{and} \quad h(x) = c$$

  Then we can plug in x_1 and we have

  $$\max_{x \in \mathbb{R}^N} f(\varphi(x_{-1}), x_{-1})$$

The FOC for this transformed problem is

$$0 = Df(\varphi(x^*_{-1}), x^*_{-1}) = D_{x_1} f(\varphi(x^*_{-1}), x^*_{-1}) D_{x_{-1}} \varphi(x^*_{-1}) + D_{x_{-1}} f(\varphi(x^*_{-1}), x^*_{-1}) \tag{2}$$

and any x^*_{-1} satisfying (2) for the unconstrained problem gives a candidate point of the constrained problem: $(\varphi(x^*_{-1}), x^*_{-1})$ with $\varphi(x^*_{-1}) = x^*_1$.

- Given the above, if there was a general way of finding φ we'd be good. This is not possible, but we don't really need to solve for φ *explicitly*. Looking at the FOC, all the expressions are computable except for $D_{x_{-1}} \varphi(x^*_{-1})$. Looking back at the constraint, if we plug in φ, we can write any x^* satisfies

$$h(\varphi(x^*_{-1}), x^*_{-1}) - c = 0$$

and now you should be reminded of the IFT!

- In other words, even when we cannot find an explicit expression for φ, we know implicitly such a function exists. Following the theorem, we can compute

$$D_{x_{-1}} \varphi(x^*_{-1}) = -(D_{x_1} h(x^*))^{-1} (D_{x_{-1}} h(x^*))$$

Plugging this into (2), we find

$$\begin{aligned} 0 &= -D_{x_1} f(x^*)(D_{x_1} h(x^*))^{-1}(D_{x_{-1}} h(x^*)) + D_{x_{-1}} f(x^*) \\ D_{x_{-1}} f(x^*) &= \left[D_{x_1} f(x^*)(D_{x_1} h(x^*))^{-1}\right] D_{x_{-1}} h(x^*) \end{aligned}$$

Let $\lambda_1 \equiv D_{x_1} f(x^*)(D_{x_1} h(x^*))^{-1}$ and we have any optimum must satisfy

$$D_{x_{-1}} f(x^*) = \lambda_1 D_{x_{-1}} h(x^*)$$

This should be a very familiar expression!

- Indeed, what we just found is almost exactly the method of Lagrangians you know and love. What's missing? Well, remember 1 was chosen arbitrarily, so we can write down the above expression $2, \dots, N$ and find the corresponding expression with $\lambda_2, \dots, \lambda_N$. However, the gradients at x^* will not change, which implies that $\lambda_i = \lambda_j$. Hence we can simply write

$$D_x f(x^*) = \lambda D_x h(x^*)$$

for any local optimum, which is exactly the method of Lagrange multipliers.

Example 2. Consider $f(x,y) = \sqrt{x \cdot y}$, $h(x,y) = x^2 + y^2$, and $c = 1$. Solve

$$\max_{x,y} (xy)^{1/2} \quad \text{s.t.} \quad x^2 + y^2 = 1$$

We know that $(Df)^T = \lambda(Dh)^T$ for some λ, or

$$\begin{aligned}
\begin{bmatrix} \frac{1}{2}x^{-1/2}y^{1/2} \\ \frac{1}{2}x^{1/2}y^{-1/2} \end{bmatrix} &= \lambda \begin{bmatrix} 2x \\ 2y \end{bmatrix} \\
\begin{bmatrix} x^{-1/2}y^{1/2} \\ x^{1/2}y^{-1/2} \end{bmatrix} &= 4\lambda \begin{bmatrix} x \\ y \end{bmatrix} \\
\frac{x^{-1/2}y^{1/2}}{x^{1/2}y^{-1/2}} &= \frac{x}{y} \\
x^2 &= y^2
\end{aligned}$$

The last step is to use the constraint, which must hold at the optimum as well:

$$x^2 + x^2 = 1 \implies x = y = \pm\sqrt{1/2}$$

Note that we can't have $x = \sqrt{1/2}$ and $y = -\sqrt{1/2}$ because then $f(x, y) \notin \mathbb{R}$.

1.2 Multiple Equality Constraints

Theorem 1 (Lagrange Multipliers)**.** *Let $f : \mathbb{R}^N \to \mathbb{R}$ and $h : \mathbb{R}^N \to \mathbb{R}^M$ be continuously differentiable and x^* be a solution to*

$$\max_{x\in\mathbb{R}^N} f(x) \quad s.t. \quad h(x) = c \tag{3}$$

*for a given $c \in \mathbb{R}^M$. If $\operatorname{rank}(Dh(x^*)) = M \leq N$ (this rank condition is called a **non-degenerate constraint qualification**; why is a NDCQ required?*[1]*) then $\exists\lambda^* \in \mathbb{R}^M$ s.t.*

$$\begin{aligned}
Df(x^*) &= (\lambda^*)^T Dh(x^*) \\
\frac{\partial f(x^*)}{\partial x_k} &= \sum_{i=1}^{M} \frac{\partial h_i(x^*)}{\partial x_k}\lambda_i^*
\end{aligned}$$

*$\lambda^* = (\lambda_1^*, \ldots, \lambda_M^*)$ are called **Lagrange multipliers**. (**Note**: Here Dh is $M \times N$.)*

Note the above is equivalent to stating that if x^* solves (3) then $\exists\lambda^*$ s.t. (x^*, λ^*) satisfy the FOC for

$$\mathcal{L}(\lambda, x) = f(x) - \lambda^T(h(x) - c) \tag{4}$$

In the above equation $\mathcal{L}$ is called the ***Lagrangian***, so the result from Theorem 1 that $Df(x^*) = (\lambda^*)^T Dh(x^*)$ is referred to as the ***FOC*** for the Lagrangian.

[1]There cannot be more non-redundant constraints than there are variables, lest the problem have no solution!

Example 3. Let $f(x, y, z) = xyz$, $h_1(x, y, z) = x^2 + y^2$, $h_2(x, y, z) = x + z$, and $c_1 = c_2 = 1$. Solve

$$\max_{x,y,z} f(x, y, z) \quad \text{s.t.} \quad h(x, y, z) = c$$
$$\max_{x,y,z} xyz \quad \text{s.t.} \quad \begin{bmatrix} x^2 + y^2 \\ x + z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Now we can apply the theorem:

$$\begin{aligned}
[Df(x)]^T &= [Dh(x)]^T \lambda \\
\begin{bmatrix} yz \\ xz \\ xy \end{bmatrix} &= \begin{bmatrix} 2x & 1 \\ 2y & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \\
xy &= \lambda_2 \\
\frac{xz}{2y} &= \lambda_1 \\
yz &= 2x\lambda_1 + \lambda_2 = \frac{x^2 z}{y} + xy \\
\implies y^2 z &= x^2 z + xy^2
\end{aligned}$$

To finish we need to leverage the constraints: $z = 1 - x$ and $y^2 = 1 - x^2$, so

$$\begin{aligned}
(1 - x^2)(1 - x) &= x^2(1 - x) + x(1 - x^2) \\
0 &= (1 - x - 3x^2)(1 - x)
\end{aligned}$$

The solutions are given by $x = 1, \dfrac{-1 \pm \sqrt{13}}{6}$. Note that $x = 1$ cannot be a solution because it implies $y = z = 0$ which yields an objective of 0, which we can readily check this is not a max, but it is also an issue for how we solved the problem, where $y \neq 0$ is required. Our possible maximizing points are given by

$$x = \frac{-1 \pm \sqrt{13}}{6} \quad z = 1 - x \quad y = \pm\sqrt{1 - x^2}$$

That is, we have 4 candidate points (because of the "$\pm$" in the x and y equations). Now, we could manually go through and check each point, or we could be clever about how we go about systematically checking for a max: Noting that

- $f(x, y, z) > 0$ as long as all arguments are not zero and an even number of arguments are negative
- $y = \sqrt{1 - x^2} > 0$ and $y = -\sqrt{1 - x^2} < 0$
- $x = \left(-1 + \sqrt{13}\right)/6 > 0$ and $x = \left(-1 - \sqrt{13}\right)/6 < 0$
- $z > 0$ no matter the value of x

it must be that

$$(x_0^*, y_0^*, z_0^*) = \left(\frac{-1 + \sqrt{13}}{6}, \sqrt{1 - (x_0^*)^2}, 1 - x_0^*\right) \quad \text{or} \quad (x_1^*, y_1^*, z_1^*) = \left(\frac{-1 - \sqrt{13}}{6}, -\sqrt{1 - (x_1^*)^2}, 1 - x_1^*\right)$$

are our optimal arguments. Finally, a little bit of algebra shows us that (x_1^*, y_1^*, z_1^*) yields the maximum (if this felt a little tedious, that's because it was! We'll discuss some shortcuts in Section 2).

1.3 Inequality Constraints

The trick is to realize we can think of inequality constraints as either

1. ***Binding***, that is, the optimum occurs along the boundary of the constraint. In this case we can treat them as equality constraints.
2. ***Non-binding***, in which case we can de facto ignore them.

In other words, we can solve a problem with inequality constraints in almost the exact way we would a problem with equality constraints. Consider $\max_x f(x)$ s.t. $g(x) \leq b$; one way to approach this:

- Solve for $x^* : Df(x^*) = Dg(x^*)\lambda$ like in a constrained problem assuming $g(x^*) = b$.
- Solve for $x^* : Df(x^*) = 0$ like in an unconstrained problem and check $g(x^*) < b$.

Once you have all the candidate points, check which is largest to determine the max.

Example 4. Consider maximizing $f(x, y) = \sqrt{xy}$ subject to $x + y \leq 1$, $x, y \geq 0$.

$$Df(x,y) = \begin{bmatrix} \frac{1}{2}x^{-1/2}y^{1/2} \\ \frac{1}{2}x^{1/2}y^{-1/2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \lambda = Dg(x,y)\lambda$$
$$\implies x = y$$

assuming $x + y = 1$ at the optimum (note that if x or y are 0 then $f(x, y) = 0$, which we can readily check is not a max). Here the candidate point is $(x, y) = (1/2, 1/2)$; now for the unconstrained problem, we note that f is strictly increasing if $x, y > 0$, which means that any such point cannot be an unconstrained max (we can increase either x or y slightly and make the objective bigger). To recap:

- Solving assuming binding constraints we find $x = 0$ or $y = 0$ cannot be max, and the Lagrange multiplier method gives $(x, y) = (1/2, 1/2)$ as a possible maximum.
- Solving assuming non-binding constraints yields no solution.

Hence the only candidate point is $(x, y) = (1/2, 1/2)$.

2 Karush-Kuhn-Tucker (KKT) Conditions

Theorem 2 (Karush-Kuhn-Tucker)**.** *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a real-valued function, $g : \mathbb{R}^N \to \mathbb{R}^K$, $h : \mathbb{R}^N \to \mathbb{R}^M$ vector-valued functions, and $b \in \mathbb{R}^K$ and $c \in \mathbb{R}^M$ constants. Suppose that x^* is such that the NDCQ is satisfied (i.e., the number of binding constrains is less than or equal to N) and that it solves*

$$\max_{x \in \mathbb{R}^N} f(x) \quad s.t. \quad g(x) \leq b \quad h(x) = c$$

Then there exist $\mu \in \mathbb{R}_+^K$ (non-negative Lagrange multipliers for the inequality constants) and $\lambda \in \mathbb{R}^M$ s.t.

1. *$Df(x^*) = \mu^T Dg(x^*) + \lambda^T Dh(x^*)$. This is still the first order condition of the Lagrangian; in this case,*

$$\mathcal{L}(\mu, \lambda, x) = f(x) - \mu^T(g(x) - b) - \lambda^T(h(x) - c)$$

2. *$(g_j(x^*) - b_j)\mu_j = 0$ for each $j = 1, \ldots, K$; this is called the **complementary slackness** condition.*

Complementary slackness states that either the constraint binds with equality or that we can ignore it for the purposes of the FOC (i.e. either $g_j(x^*) = b$ or the corresponding $\mu_j = 0$). Note as we did when solving equality constraints, we typically have to leverage the fact that $h(x^*) = c$ and $g(x^*) \leq b$ to arrive at a solution.

Remark 1. The result of the above theorem is often known as the Kuhn-Tucker conditions, named after two mathematicians who showed the theorem in in the 1950s. However, Karush had already proven this result more than a decade earlier in his (apparently unpublished) masters' thesis.

2.1 Second Order Conditions

Consider the general optimization problem outlined in (1) and suppose the functions are twice continuously differentiable, so the Lagrangian

$$\mathcal{L}(\mu, \lambda, x) = f(x) - \mu^T(g(x) - b) - \lambda^T(h(x) - c)$$

is twice continuously differentiable as well. Suppose (μ^*, λ^*, x^*) is s.t. x^* is a solution to the problem and the KKT conditions are satisfied. Further suppose the constraints $g_1, \ldots, g_L$ are binding and the other $K - L$ constrains are non-binding. Then the Hessian of $\mathcal{L}$ with respect to x is negative definite on the set

$$V = \{v : D\widetilde{g}(x^*)v = 0 \quad \text{and} \quad Dh(x^*)v = 0\}$$

with $\widetilde{g} = (g_1, \ldots, g_L)$. That is,

$$v^T D_x^2 \mathcal{L}(\mu^*, \lambda^*, x^*)v < 0$$

for any $v \in V$. To check this condition, consider the ***bordered Hessian***:

$$H = \begin{bmatrix} 0_{L\times L} & 0_{L\times M} & \underbrace{D\widetilde{g}(x^*)}_{L\times N} \\ 0_{M\times L} & 0_{M\times M} & \underbrace{Dh(x^*)}_{M\times N} \\ D\widetilde{g}(x^*)^T & Dh(x^*)^T & \underbrace{D_x^2\mathcal{L}(\mu^*, \lambda^*, x^*)}_{N\times N} \end{bmatrix}$$

You will note the bordered Hessian is basically the Hessian of the Lagrangian, but it is so-called because it looks like we are "bordering" $D_x^2\mathcal{L}$ with $D\widetilde{g}, Dh$. To check if a candidate point (derived using the FOC of the Lagrangian) is a max or min, it is sufficient to check the last $N - L - M$ leading principal minors of H:

- For a max, the last $N-L-M$ leading principle minors alternate signs, with the sign of the largest principal minor (i.e., the determinant of the bordered Hessian) equal to $(-1)^N$. Equivalently, the sign of the $(2(L+M)+1)^{\text{th}}$ leading principle minor (i.e., the first leading principle minor considered) has a determinant equal to $(-1)^{L+M+1}$.
- For a min, the signs of the last $N-L-M$ leading principle minors of H are equal to $(-1)^{L+M}$.

You will note we are discarding the first $2(L+M)$ leading principal minors. Why?[2]

2.2 Application to Economics: Cobb-Douglas Utility Maximization

Consider a consumer with wealth $w>0$ maximizing Cobb-Douglas utility $U(x,y)=x^\alpha y^{1-\alpha}$ for goods $x,y\in\mathbb{R}$ with prices p_x,p_y such that $x,y\geq 0$, $p_x,p_y>0$.

$$\max_{x,y} x^\alpha y^{1-\alpha} \quad \text{s.t.} \quad x,y\geq 0 \quad \text{and} \quad p_x x + p_y y \leq w$$

Note that $f(x,y)=x^\alpha y^{1-\alpha}$, $g(x,y)=(-x,-y,p_x x+p_y y)$, and $b=(0,0,w)$. We know from Theorem 2 that any maximum satisfies

$$(Df(x,y))^T = \left(D\begin{bmatrix}\tilde{g}(x,y)\\ h(x,y)\end{bmatrix}\right)^T \mu$$
$$\begin{bmatrix}\alpha x^{\alpha-1}y^{1-\alpha}\\ (1-\alpha)x^\alpha y^{-\alpha}\end{bmatrix} = \begin{bmatrix}-1 & 0 & p_x\\ 0 & -1 & p_y\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\end{bmatrix}$$

and $\mu_1 x = \mu_2 y = (p_x x + p_y y - w)\mu_3 = 0$. Let us consider cases: If $x,y\neq 0$ then $\mu_1=\mu_2=0$ and from the FOC we find

$$\frac{\alpha x^{\alpha-1}y^{1-\alpha}}{(1-\alpha)x^\alpha y^{-\alpha}} = \frac{p_x}{p_y}$$
$$\alpha p_y y = (1-\alpha)p_x x$$

If $p_x x + p_y y < w$ then since f is strictly increasing in either argument, we can find some ε s.t. $p_x(x+\varepsilon)+p_y(y+\varepsilon)<w$ but $f(x+\varepsilon,y+\varepsilon)>f(x,y)$. Therefore the constraint is binding, and

$$w = \frac{\alpha}{1-\alpha}p_y y + p_y y$$
$$y = \frac{(1-\alpha)w}{p_y}$$
$$x = \frac{\alpha w}{p_x}$$

[2] The bordered Hessian cannot be positive or negative definite due to the matrix of zeros in the top left corner (second derivatives with respect to the multipliers are a 0). That discards checking the first $L+M$ leading principal minors. What about the next? Intuitively, we can think that binding conditions impose restrictions of parameters. Remember we motivated our sketch of the proof of the Lagrange multiplier method by invoking the IFT. Why did we do this? Because if for each constraint we express a variable in terms of the others, we can solve the modified unconstrained problem. In other words, each constraint removes a variable from the problem, and there goes the other $L+M$ degrees of freedom.

Note if x or y are 0 then the objective is 0, which is lower than the objective when both $x, y > 0$ (as is the case above). Hence the only candidate point is given by

$$(x, y) = \left(\frac{(1-\alpha)w}{p_y}, \frac{\alpha w}{p_y}\right)$$

In this case, the Lagrangian was

$$\mathcal{L}(\mu, x) = x^\alpha y^{1-\alpha} - x\mu_1 - y\mu_2 - (p_x x + p_y y - w)\mu_3$$

Note here only the budget constraint is binding, so the bordered Hessian is

$$\begin{bmatrix} 0 & p_x & p_y \\ p_x & \alpha(\alpha-1)x^{\alpha-2}y^{1-\alpha} & \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} \\ p_y & \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} & -\alpha(1-\alpha)x^\alpha y^{-\alpha-1} \end{bmatrix}$$

Hence we have two variables and one binding constraint, so we check the last $2 - 1 = 1$ leading principal minors, which is the third leading principal minor, or just the matrix itself. The determinant is

$$\begin{aligned}
&0 \cdot \det \begin{bmatrix} \alpha(\alpha-1)x^{\alpha-2}y^{1-\alpha} & \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} \\ \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} & -\alpha(1-\alpha)x^\alpha y^{-\alpha-1} \end{bmatrix} \\
&- p_x \cdot \det \begin{bmatrix} p_x & \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} \\ p_y & -\alpha(1-\alpha)x^\alpha y^{-\alpha-1} \end{bmatrix} \\
&+ p_y \cdot \det \begin{bmatrix} p_x & \alpha(\alpha-1)x^{\alpha-2}y^{1-\alpha} \\ p_y & \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} \end{bmatrix} \\
&= -p_x \cdot \left[-p_x\alpha(1-\alpha)x^\alpha y^{-\alpha-1} - p_y\alpha(1-\alpha)x^{\alpha-1}y^{-\alpha}\right] \\
&+ p_y \cdot \left[p_x\alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} + p_y\alpha(1-\alpha)x^{\alpha-2}y^{1-\alpha}\right] \\
&> 0
\end{aligned}$$

Finally, $L + M + 1 = 1 + 1 = 2$, so $(-1)^2 > 0$ means the 3rd leading principal minor must be positive for a max, and that is exactly what we found.

3 Envelope Theorem

Consider the function

$$f(x; \theta) = x^\theta - x$$

for $\theta \in (0, 1)$ and suppose we are interested in maximizing $f(x; \theta)$. From the FOC, $\theta x^{\theta-1} = 1$, and we can parametrize the maximum for any value of θ, $x^*(\theta) = \theta^{1/(1-\theta)}$.

$$f(x^*(\theta); \theta) = (x^*(\theta))^\theta - x^*(\theta) = \theta^{\theta/(1-\theta)} - \theta^{1/(1-\theta)}$$

If we are interested in how f changes with respect to θ, we can take this derivative. This derivative, while possible to derive, is challenging and needlessly time consuming in an exam setting. Instead we can do some wrangling. Note

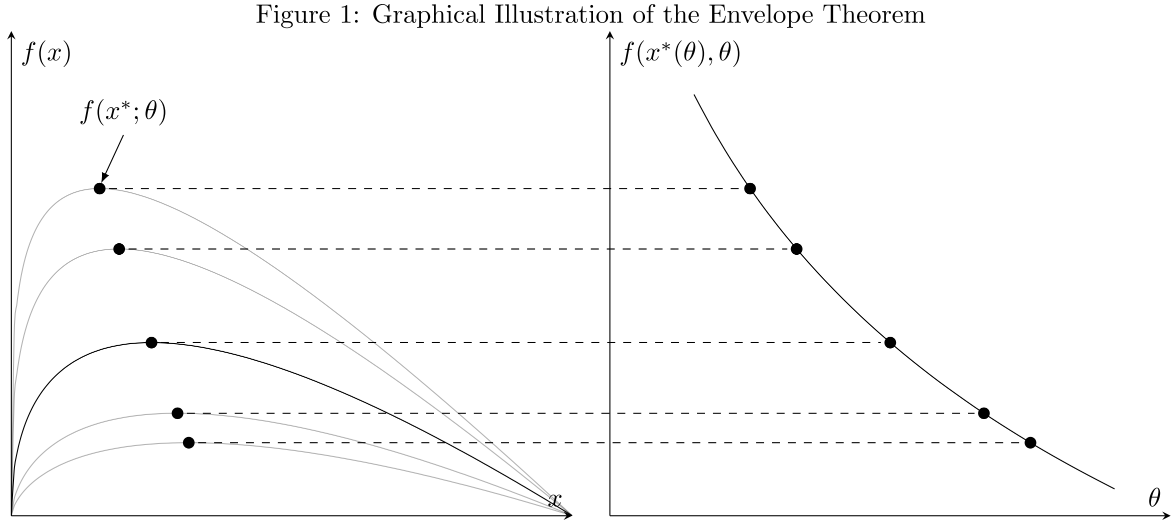
$$\frac{d}{d\theta}f(x^*(\theta);\theta) = \left(\frac{\partial}{\partial x}f(x^*(\theta);\theta)\right)\frac{dx^*(\theta)}{d\theta} + \frac{\partial}{\partial\theta}f(x;\theta)\bigg|_{x=x^*(\theta)}$$

From the FOC, however, we know $\partial_x f(x^*(\theta);\theta) = 0$ for each θ, meaning (writing just x^* for ease of notation),

$$\frac{d}{d\theta}f(x^*;\theta) = \frac{\partial}{\partial\theta}f(x;\theta)\bigg|_{x=x^*} = \left[\frac{\partial}{\partial\theta}\Big(\exp(\theta\log(x)) - x\Big)\right]_{x=x^*} = (x^*)^\theta\log(x^*)$$

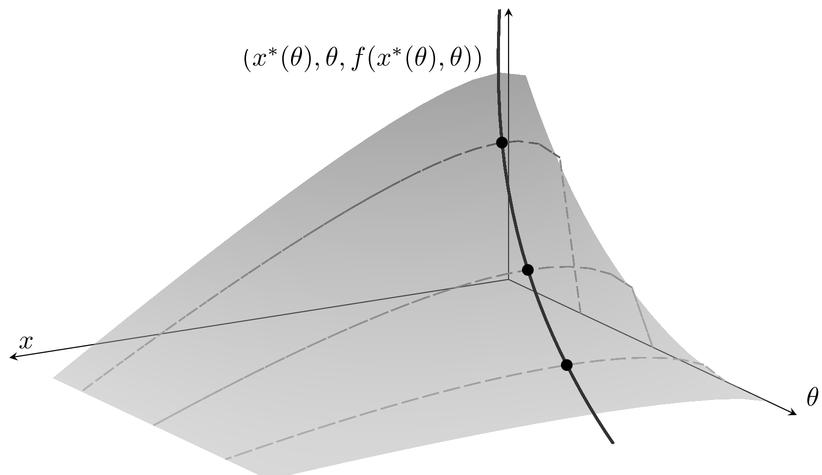
This trick of leveraging the FOC to obtain the derivative of the optimized objective with respect to a parameter is called the ***envelope theorem***. Graphically, we can see that

Figure 1: Graphical Illustration of the Envelope Theorem



In x–θ space, we can see that $f(x^*(\theta),\theta)$ ***envelops*** $f(x,\theta)$ at each $(x^*(\theta),\theta)$ (hence the name).

Figure 2: Graphical Illustration of the Envelope Theorem Ctd.



Theorem 3 (Envelope Theorem). *Let $f : \mathbb{R}^N \times \Theta \to \mathbb{R}$ and $h : \mathbb{R}^N \times \Theta \to \mathbb{R}^M$ be continuously differentiable and $(x^*(\theta), \lambda^*(\theta))$ satisfy the FOC of the Lagrangian for the optimization problem at a given $\theta \in \Theta$:*

$$\max_x f(x;\theta) \quad s.t. \quad h(x;\theta) = c$$

*(**Note**: θ may be a vector of parameters, not just a single parameter.) The **value function** is the maximized value of the objective at each θ, given by*

$$V(\theta) \equiv f(x^*(\theta);\theta)$$

If $V(\theta)$ is continuously differentiable at θ then

$$\begin{aligned} D_{\theta'} V(\theta) &= \Big[D_{\theta'} f(x;\theta) - \lambda^T D_{\theta'} h(x;\theta) \Big]_{(x,\lambda)=(x^*(\theta),\lambda^*(\theta))} \\ \frac{\partial V(\theta)}{\partial \theta_k} &= \frac{\partial f(x^*(\theta),\theta)}{\partial \theta_k} - \sum_{m=1}^{M} \lambda_m^*(\theta) \frac{\partial h_m(x^*(\theta),\theta)}{\partial \theta_k} \\ &= \frac{\partial f(x,\theta)}{\partial \theta}\bigg|_{x=x^*(\theta)} - \sum_{m=1}^{M} \lambda_m^*(\theta) \frac{\partial h_m(x^*(\theta),\theta)}{\partial \theta_k} \end{aligned}$$

where $\lambda_m^*(\theta)$ is the mth entry of $\lambda^*(\theta)$ and similarly for $h_m(\cdot)$.

Note that in our example, there is no constraint so the second term in the theorem is just 0. (Alternatively, if θ does not appear in the constraint, the second term drops out of the expression as well.)

Proof. As we showed the envelope theorem for unconstrained optimization by leveraging the FOC, we can do the same for constrained optimization. Note, getting rid of the the stars, we have

$$D_{\theta'} V(\theta) = D_{x'} f(x(\theta);\theta)(D_{\theta'} x(\theta)) + D_{\theta'} f(x(\theta);\theta)$$

Now recall the FOC:

$$D_{x'}f(x(\theta);\theta) = \lambda(\theta)^T D_{x'}h(x(\theta);\theta)$$

Plugging this in,

$$D_{\theta'}V(\theta) = \lambda(\theta)^T D_{x'}h(x(\theta);\theta)(D_\theta x(\theta)) + D_\theta f(x(\theta);\theta)$$

To finish, differentiate the constraint $h(x(\theta);\theta) = c$ with respect to θ' to get

$$-D_{\theta'}h(x(\theta);\theta) = D_{x'}h(x(\theta);\theta)(D_{\theta'}x(\theta))$$

by the chain rule. The right-hand appears in our expression for $D_{\theta'}V(\theta)$; plug in $-D_{\theta'}h(x(\theta);\theta)$ and you get the envelope theorem. □

3.1 Application to Economics: Cost Minimization

Consider a consumer minimizing expenditures subject to a minimum utility level u, with utility function $U(x,y) = x^\alpha y^{1-\alpha}$ for goods $x, y \in \mathbb{R}$ with prices p_x, p_y such that $x, y \geq 0$, $p_x, p_y > 0$.

$$\min_{x,y} p_x x + p_y y \quad \text{s.t.} \quad x^\alpha y^{1-\alpha} \geq u$$

If the constraint is not binding at (x,y), that is, $x^\alpha y^{1-\alpha} > u$, then $\exists \varepsilon > 0$ s.t. $(x-\varepsilon)^\alpha (y-\varepsilon)^{1-\alpha} > u$. Since $p_x(x-\varepsilon) + p_y(y-\varepsilon) < p_x x + p_y y$, this point cannot be a minimum. Thus the constraint is binding, and we can apply the envelope theorem. The value function of this problem is called the ***expenditure function***, given by

$$e(p_x, p_y, u) = p_x x(p_x, p_y, u) + p_y y(p_x, p_y, u)$$

with $x(\cdot), y(\cdot)$ the so-called Hicksian demand functions. We can see that to find the gradient of the expenditure function with respect to price we can apply the envelope theorem! Noting p_x, p_y are not in the constraint,

$$D_{(p_x,p_y)}e(p_x,p_y,u) = \begin{bmatrix} x(p_x,p_y,u) & y(p_x,p_y,u) \end{bmatrix}$$

Index