Math Camp 2020 - Optimization (Reference)*

Sarah Robinson

Department of Economics, UC Santa Barbara

1. **Extrema**

   (a) Definition. Let $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}^n$. $\overline{\mathbf{x}} \in D$ is a **local maximum** of f if and only if there exists an $\varepsilon > 0$ such that $f(\mathbf{x}) \leq f(\overline{\mathbf{x}})$ for all $\mathbf{x} \in B_\varepsilon(\overline{\mathbf{x}})$.

   If $f(\mathbf{x}) \leq f(\overline{\mathbf{x}})$ for all $\mathbf{x} \in D$, then $\overline{\mathbf{x}}$ is a **global maximum**.

   (b) Definition. Let $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}^n$. $\underline{\mathbf{x}} \in D$ is a **local minimum** of f if and only if there exists an $\varepsilon > 0$ such that $f(\mathbf{x}) \geq f(\underline{\mathbf{x}})$ for all $\mathbf{x} \in B_\varepsilon(\underline{\mathbf{x}})$.

   If $f(\mathbf{x}) \geq f(\underline{\mathbf{x}})$ for all $\mathbf{x} \in D$, then $\underline{\mathbf{x}}$ is a **global minimum**.

   (c) Theorem. Let $f(x)$ be a twice-continuously differentiable function of one variable. Then $f(x)$ reaches a local, interior

   - maximum at x^* iff $f'(x^*) = 0$ and $f''(x^*) \leq 0$
   - minimum at $\tilde{x}$ iff $f'(\tilde{x}) = 0$ and $f''(\tilde{x}) \geq 0$

   $f'(x) = 0$ is known as the **first-order condition**. The sign of $f''(x)$ is the **second order condition**.

   (d) Example. Consider the function $f(x) = \ln(x) - x$. This function is differentiable, so we can use derivatives to find the maximum:

$$\begin{aligned} f'(x) &= \frac{1}{x} - 1 && \text{(differentiating)} \\ 0 &= \frac{1}{x^*} - 1 && \text{(the first-order condition)} \\ x^* &= 1 && \text{(solving for } x\text{)} \end{aligned}$$

   Checking the second-order condition:

$$\begin{aligned} f''(x) &= -\frac{1}{x^2} && \text{(the 2nd derivative)} \\ f''(x^*) &= -\frac{1}{1^2} && \text{(plugging in for } x^*\text{)} \\ f''(x^*) &= -1 && \text{(simplifying)} \end{aligned}$$

   Thus, $x^* = 1$ is a (global) maximum.

*These lecture notes are drawn principally from the mathematical appendices from *Microeconomic Theory*, by Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green, and *Advanced Microeconomic Theory*, by Geoffrey A. Jehle and Philip J. Reny. The material posted on this website is for personal use only and is not intended for reproduction, distribution, or citation. James Banovetz created the first edition of these awesome notes and graciously shared them.

(e) Theorem. Let $f : D \to \mathbb{R}$ be a differentiable function, where $D \subseteq \mathbb{R}^n$. If $f(\mathbf{x})$ reaches a local extremum at $x^* \in \mathbb{R}^n$, then

$$\frac{\partial f(\mathbf{x})}{x_i} = 0 \quad \forall\, i = 1, \ldots, n$$

or, stating in terms of the gradient,

$$\nabla f(\mathbf{x}) = 0$$

(f) Example. Consider the function

$$f(x, y) = 8x^3 + 2xy - 3x^2 + y^2 + 1$$

The first order conditions are

$$\begin{aligned} f_x(x, y) &= 24x^2 + 2y - 6x = 0 && \text{(the FOC w.r.t. } x\text{)} \\ f_y(x, y) &= 2x + 2y = 0 && \text{(the FOC w.r.t. } y\text{)} \end{aligned}$$

We can manipulate these to find the extremum:

$$\begin{aligned} y &= -x && \text{(from the FOC w.r.t } y\text{)} \\ 0 &= 24x^2 + 2(-x) - 6x && \text{(plugging in for } y\text{)} \\ 0 &= 24x^2 - 8x && \text{(simplifying)} \\ 0 &= x(24x - 8) && \text{(factoring)} \end{aligned}$$

Thus, our two potential optima are

$$(0, 0) \quad \text{and} \quad \left(\frac{1}{3}, -\frac{1}{3}\right)$$

To figure out whether or not each point is a maximum or a minimum, however, we need to appeal to our multivariate second-order conditions.

(g) Theorem. Let $f : D \to \mathbb{R}$ be a twice-continuously differentiable function, where $D \subseteq \mathbb{R}^n$.

- If $f(\mathbf{x})$ reaches a local interior maximum at $\mathbf{x}^*$, then H is negative semidefinite at $\mathbf{x}^*$.
- If $f(\mathbf{x})$ reaches a local interior minimum at $\tilde{\mathbf{x}}$, then H is positive semidefinite at $\tilde{\mathbf{x}}$.

(h) Example Returning to our example above we can define both the gradient and the Hessian:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 24x^2 + 2y - 6x & 2x + 2y \end{bmatrix} \qquad \text{(the gradient)}$$

$$H(x, y) = \begin{bmatrix} 48x - 6 & 2 \\ 2 & 2 \end{bmatrix} \qquad \text{(the Hessian)}$$

At (0,0) and (1/3, −1/3) the Hessian is, respectively:

$$H(0, 0) = \begin{bmatrix} -6 & 2 \\ 2 & 2 \end{bmatrix} \qquad H\left(\frac{1}{3}, -\frac{1}{3}\right) = \begin{bmatrix} 10 & 2 \\ 2 & 2 \end{bmatrix}$$

$H(0, 0)$ has leading principle minors -6 and -16, so it is indefinite–$(0, 0)$ is a saddle point. $H(1/3, -1/3)$, however, has leading principle minors 10 and 16, so it is positive definite–$(1/3, -1/3)$ is a minimum.

(i) Aside. So far, we've found critical points and assigned them status as local maxima or minima. We're typically more interested, however, in finding global maxima and minima. The Weierstrauss Extreme Value theorem gave us a way to guarantee the existence of global extrema; we also have a theorem to find them via calculus criteria.

(j) Theorem. Let $f : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}^n$, be twice continuously differentiable and concave. Then the following statements are equivalent:

- $\nabla f(\mathbf{x}^*) = 0$
- f achieves a local maximum at $\mathbf{x}^*$
- f achieves a global maximum at $\mathbf{x}^*$

Further, if f is strictly concave, then x^* is the unique global maximizer, i.e., $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \in D$ such that $\mathbf{x} \neq \mathbf{x}^*$.

(k) Theorem. Let $f : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}^n$, be twice continuously differentiable and convex. Then the following statements are equivalent:

- $\nabla f(\mathbf{x}^*) = 0$
- f achieves a local minimum at $\mathbf{x}^*$
- f achieves a global minimum at $\mathbf{x}^*$

Further, if f is strictly convex, then x^* is the unique global minimizer, i.e., $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in D$.

(l) Example. Consider a profit function $f(K, L)$ given by

$$f(K, L) = p\big[\ln(K) + \ln(L)\big] - rK - wL$$

where p, r, and w are strictly positive parameters, while K and L are (strictly positive) choice variables. The first order conditions are

$$0 = \frac{p}{K^*} - r \qquad \text{(the FOC w.r.t. } K\text{)}$$

$$0 = \frac{p}{L^*} - w \qquad \text{(the FOC w.r.t. } L\text{)}$$

Solving these first order condition yields the extrema:

$$K^* = \frac{p}{r} \qquad\qquad L^* = \frac{p}{w}$$

To establish this as a maximum or a minimum, we can evaluate the definiteness of the Hessian:

$$H = \begin{bmatrix} -\frac{p}{K^2} & 0 \\ 0 & -\frac{p}{L^2} \end{bmatrix} \qquad \text{(the Hessian)}$$

$$|H_1| = -\frac{p}{K^2} \qquad \text{(the first LPM)}$$

$$|H_2| = \left(\frac{p}{KL}\right)^2 \qquad \text{(the second LPM)}$$

The leading principle minors follow the signing convention for negative definiteness, $|H_1| < 0$ and $|H_2| > 0$; thus, the function is strictly concave over it's domain, so (K^*, L^*) is the unique global maximum.

2. **Constrained Optimization**

(a) Aside. The real "work-horse" problem in economics is one of constrainted optimization–getting the most out of scarce resources. While much of the the following should be familiar to students, we will begin by defining terms in relatively simple cases; more general results will be stated later. Again, understanding how to work through constrained optimization problems is extremely important in the first-year classes.

(b) Definition. Let $f(x,y)$ be a real-valued function. Consider a relation between the elements (x,y) that must be satisified, denoted $g(x,y) = c$. Then the **constrained maximization** problem is written:

$$\max_{x,y} \; f(x,y) \quad \text{subject to} \quad g(x,y) = c$$

$f(x,y)$ is called the **objective function**, and $g(x,y) = c$ is the **constraint**.

(c) Aside. With a small number of variables and simple constraints, this problem can be solved easily via direct substitution. For example:

$$\max_{x_1,x_2} \; f(x_1,x_2) \quad \text{s.t.} \quad x_1 + x_2 = 1$$

can simply be written as a one-variable, unconstrained problem:

$$\max_{x_1} \; f(x_1, 1 - x_2)$$

This method is occasionally quite quick and easy to work though, and will be useful both micro and macro. We need a more general tool, however, to handle a broader class of problems.

(d) Definition. Consider a constrained optimization function of the form

$$\max_{x,y} \; f(x,y) \quad \text{s.t.} \quad g(x,y) = c$$

Then the 3-variable objective function

$$\mathcal{L}(x,y,\lambda) = f(x,y) - \lambda\big[c - g(x,y)\big]$$

is known as the **Lagrangian**.

(e) Aside. The importance of the Lagrangian in basic economics cannot be overstated. The key insight of Joseph-Louis Lagrange was that every constrained problem has an analagous *unconstrained* problem that we can solve with our usual methods.

(f) Theorem. Given a Lagrangian of the form

$$\mathcal{L}(x,y,\lambda) = f(x,y) - \lambda\big[c - g(x,y)\big]$$

The first-order conditions are

$$\frac{\partial \mathcal{L}(\cdot)}{\partial x} = \frac{\partial f(\cdot)}{\partial x} - \lambda \frac{\partial g(\cdot)}{\partial x} = 0$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial y} = \frac{\partial f(\cdot)}{\partial y} - \lambda \frac{\partial g(\cdot)}{\partial y} = 0$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \lambda} = c - g(\mathbf{x}) = 0$$

Any points (x^*, y^*, λ^*) that solve these three equations simultaneously must be critical points of $f(x,y)$ along the constraint $g(x,y) = c$.

(g) Example. Consider the maximization problem

$$\max_{x_1,x_2} \; (-ax_1^2 - bx_2^2) \quad \text{s.t.} \quad x_1 + x_2 = 1$$

We can employ the Lagrangian method to find potential extrema. The Lagrangian is given by:

$$\mathcal{L}(x_1, x_2, \lambda) = -ax_1^2 - bx_2^2 + \lambda(1 - x_1 - x_2)$$

The FOCs of the Lagrangian are:

$$\frac{\partial \mathcal{L}(\cdot)}{\partial x_1} = -2ax_1 - \lambda = 0 \qquad \text{(w.r.t. } x_1\text{)}$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial x_2} = -2bx_2 - \lambda = 0 \qquad \text{(w.r.t. } x_2$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \lambda} = 1 - x_1 - x_2 = 0 \qquad \text{(w.r.t. } \lambda\text{)}$$

Solving this three-equation, three-unknown system:

$$\begin{aligned}
2ax_1 &= 2bx_2 && \text{(combining the 1st and 2nd FOCs)} \\
x_2 &= \frac{a}{b}x_1 && \text{(solving for } x_2\text{)} \\
0 &= 1 - x_1 - \left(\frac{a}{b}x_1\right) && \text{(plugging in for } x_2 \text{ in the 3rd FOC)} \\
&\boxed{x_1^* = \frac{b}{a+b}} && \text{(solving for } x_1\text{)} \\
&\boxed{x_2^* = \frac{a}{a+b}} && \text{(solving for } x_2\text{)} \\
&\boxed{\lambda = -\frac{2ab}{a+b}} && \text{(solving for } \lambda\text{)}
\end{aligned}$$

(h) Aside. Depending on the textbook you read, λ is either considered to be of great importance, or of no importance whatsoever. In the context of economics problems, if we set up the Lagrangian a particular way, and if our typical assumptions are fulfilled, λ will end up having a very interesting interpretation.

While these formulations and examples have been with two variables and one constraint (a very common set up in 1st-year micro theory), Lagrange's theorem can handle a much wider class of problems.

(i) Theorem (JR THM A2.16). Let $f : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}^n$. If $m < n$, consider the optimization problem

$$\operatorname*{opt}_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad \begin{aligned} g_1(\mathbf{x}) &= c_1 \\ g_2(\mathbf{x}) &= c_2 \\ &\vdots \\ g_m(\mathbf{x}) &= c_m \end{aligned}$$

where $g_j(\mathbf{x})$ is real valued for all j. The associated Lagrangian is defined as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \left[c_j - g_j(\mathbf{x})\right]$$

If the following conditions hold:

- $f(\mathbf{x})$ and $g_j(\mathbf{x})$, $j = 1, \ldots, m$ are continuously differentiable over $D \subseteq \mathbb{R}^n$

- $\mathbf{x}^*$ is an interior optimum (maxima or minima) of $f(\mathbf{x})$ subject to the m constraints
- $\nabla g_i(\mathbf{x})$, $i = 1, \ldots, m$ are linearly independent

Then there exist m unique numbers λ_j^*, $j = 1 \ldots, m$ such that:

$$\frac{\mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial x_i} = \frac{\partial f(\mathbf{x}^*)}{\partial x_i} - \sum_{j=1}^{m} \lambda_j^* \frac{\partial g_j(\mathbf{x}^*)}{\partial x_i} = 0, \qquad i = 1, \ldots, n$$

(j) Aside. This is the more general statement of Lagrange's Theorem. Essentially, we need f and all the g's to be differentiable. Then, if we're at an interior optima, and our m constraints are actually different, then we will get m different values of λ (one for each constraint) that make all of our first order conditions equal to zero.

(k) Example. Geometrically, we can interpret the Lagrangian as finding the tangency between our objective function $f(x_1, x_2)$ and our constraint $g(x_1, x_2) = c$. Consider a level set of our objective function at y_0:

$$y_0 = f(x_1, x_2) \qquad \text{(a level set at } y_0\text{)}$$

We can think about the total differential:

$$0 = f_1(x_1, x_2)dx_1 + f_2(x_1, x_2)dx_2 \qquad \text{(the total differential)}$$

$$\frac{dx_2}{dx_1} = -\frac{f_1(\cdot)}{f_2(\cdot)} \qquad \text{(solving for } dx_2/dx_1\text{)}$$

This gives the slope of the level curve $L(y_0)$. Similarly, we can manipulate the constraint:

$$c = g(x_1, x_2) \qquad \text{(the constraint)}$$

$$0 = g_1(x_1, x_2)dx_1 + g_2(x_1, x_2)dx_2 \qquad \text{(the total differential)}$$

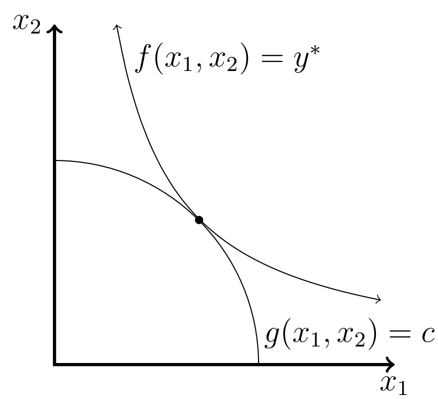
$$\frac{dx_2}{dx_1} = -\frac{g_1(\cdot)}{g_2(\cdot)} \qquad \text{(solving for } dx_1/dx_2\text{)}$$

This gives the slope of of the constraint. From the FOCs of the Lagrangian associated with this problem, we obtain the conditions:

$$\frac{f_1(x_1^*, x_2^*)}{f_2(x_1^*, x_2^*)} = \frac{g_1(x_1^*, x_2^*)}{g_2(x_1^*, x_2^*)} \qquad \text{(from the FOCs w.r.t. } x_1 \text{ and } x_2\text{)}$$

$$c = g(x_1^*, x_2^*) \qquad \text{(from the FOC w.r.t. } \lambda\text{)}$$

Thus, the slopes are equal at the critical points:



(l) Aside. This is the graph for a "well-behaved" maximum, but we're not always guaranteed these shapes. We're only guaranteed that the two curves are tangent, not that we've found a constrained maximum. For that, we need to appeal to the second-order conditions. Our second-order conditions are related to the quasiconcavity of a function; but we actually have a slightly less strict condition for Lagrangians. Indeed, the function only needs to be "well-behaved" along our constraint.

(m) Definition. The matrix associated with the Lagrangian, denoted $\bar{H}$, is defined as

$$\bar{H} = \begin{bmatrix} 0 & g_1 & g_2 \\ g_1 & \mathcal{L}_{11} & \mathcal{L}_{12} \\ g_2 & \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix}$$

This matrix is the **bordered Hessian** of the Lagrangian function.

(n) Aside. We can find the first and second derivatives to help with the intuition of this matrix. Consider the system of first derivatives, arranged to be:

$$\begin{aligned} \mathcal{L}_\lambda &= c - g(x_1, x_2) && \text{(derivative of } \mathcal{L} \text{ w.r.t. } \lambda) \\ \mathcal{L}_1 &= f_1(x_1, x_2) + \lambda g_1(x_1, x_2) && \text{(derivative of } \mathcal{L} \text{ w.r.t. } x_1) \\ \mathcal{L}_2 &= f_2(x_1, x_2) + \lambda g_2(x_1, x_2) && \text{(derivative of } \mathcal{L} \text{ w.r.t. } x_2) \end{aligned}$$

If we interpret this as the gradient, we can take another set of partial derivatives to obtain the "Hessian" of $\mathcal{L}$:

$$\tilde{H} = \begin{bmatrix} 0 & -g_1(\cdot) & -g_2(\cdot) \\ -g_1(\cdot) & f_{11}(\cdot) + \lambda g_{11}(\cdot) & f_{12}(\cdot) + \lambda g_{12}(\cdot) \\ -g_2(\cdot) & f_{21}(\cdot) + \lambda g_{21}(\cdot) & f_{22}(\cdot) + \lambda g_{22}(\cdot) \end{bmatrix} \qquad \text{(the "Hessian" of } \mathcal{L})$$

It turns out that we're mostly interested in the leading principle minors of this matrix. Recall that we can scale rows of a matrix, and the determinant will scale correspondingly. Thus, if we scale two rows by -1, the determinant will remain unchanged. Further, we can substitute in $\mathcal{L}_{ij}$ to obtain:

$$\bar{H} = \begin{bmatrix} 0 & g_1 & g_2 \\ g_1 & \mathcal{L}_{11} & \mathcal{L}_{12} \\ g_2 & \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix} \qquad \text{(the bordered Hessian)}$$

Note that this is *very* similar to our previous bordered Hessian, except that our border is formed by the first derivative of our constraint, rather than the first derivative of the function.

(o) Theorem (JR THM A2.17). Given a Lagrangian of the form

$$\mathcal{L}(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda\big[c - g(x_1, x_2)\big]$$

If (x_1^*, x_2^*, λ) solves the first order conditions and $|\bar{H}| > 0$ ($|\bar{H}| < 0$) at that point, then (x_1^*, x_2^*) is a local maximimum (minimum) of $f(x_1, x_2)$ subject to $g(x_1, x_2)$. This is the *second order* condition.

(p) Example. Recall our previous example:

$$\mathcal{L}(x_1, x_2, \lambda) = -ax_1^2 - bx_2^2 + \lambda\big[1 - x_1 - x_2\big]$$

This yielded the critical point

$$(x_1^*, x_2^*, \lambda^*) = \left(\frac{b}{a+b}, \frac{a}{a+b}, -\frac{2ab}{a+b}\right)$$

Writing out the Hessian to evaluate our second-order conditions:

$$\bar{H} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -2a & 0 \\ 1 & 0 & -2b \end{bmatrix} \qquad \text{(the bordered Hessian)}$$

$$|\bar{H}_1| = 0 \qquad \text{(the first LPM)}$$

$$|\bar{H}_2| = -1 \qquad \text{(the second LPM)}$$

$$|\bar{H}_3| = 2(a+b) \qquad \text{(the third LPM)}$$

Note that our first LPM should always be zero; the second should always be non-positive; the third here is postive, meaning that we're at a local maximum.

(q) Theorem (CW 12.3). Consider a constrained maximization problem with $\mathbf{x} \in \mathbb{R}^n$ and $m < n$ equality constraints:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \left[c_j - g_j(\mathbf{x})\right]$$

Then the bordered Hessian is given by

$$\bar{H} = \begin{bmatrix} 0 & \cdots & 0 & \frac{\partial g_1(\cdot)}{\partial x_1} & \cdots & \frac{\partial g_1(\cdot)}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{\partial g_m(\cdot)}{\partial x_1} & \cdots & \frac{\partial g_m(\cdot)}{\partial x_n} \\ \frac{\partial g_1(\cdot)}{\partial x_1} & \cdots & \frac{\partial g_m(\cdot)}{\partial x_1} & \frac{\partial^2 \mathcal{L}(\cdot)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 \mathcal{L}(\cdot)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1(\cdot)}{\partial x_n} & \cdots & \frac{\partial g_m(\cdot)}{\partial x_n} & \frac{\partial^2 \mathcal{L}(\cdot)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 \mathcal{L}(\cdot)}{\partial x_n \partial x_n} \end{bmatrix}$$

Let $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ solve the first-order conditions. Then:

- $\mathbf{x}^*$ is a local maximum of $f(\mathbf{x})$ if $|\bar{H}_{2m}|$ has the same sign as $(-1)^m$ and the leading principle minors alternate in sign thereafter.
- $\mathbf{x}^*$ is a local minimum if $|\bar{H}_{2m}|$ and all subsequent leading principle minors have the same sign as $(-1)^m$.

(r) Aside. Note a few things about the signing convention:

- The first m leading principle minors are zero matrices (every element of $|\bar{H}_1|$ through $|\bar{H}_m|$ is zero).
- The next $m-1$ leading principle minors will have a determinant of zero ($|\bar{H}|_{m+1}$ through $|\bar{H}_{2m-1}|$). For example, with two constraints, consider $|\bar{H}_3|$:

$$|\bar{H}_3| = \begin{vmatrix} 0 & 0 & x_1 \\ 0 & 0 & y_1 \\ x_1 & y_1 & z_1 \end{vmatrix} = 0$$

- The $2m$th leading principle minor will be signed deterministically, and thus contains no relevant information. For Example, with two constraints, consider $|\bar{H}_4|$:

$$|\bar{H}_4| = \begin{vmatrix} 0 & 0 & x_1 & x_2 \\ 0 & 0 & y_1 & y_2 \\ x_1 & y_1 & z_1 & z_2 \\ x_2 & y_2 & z_2 & w \end{vmatrix} = (x_1 y_2 - y_1 x_2)^2$$

An alternative way to remember signing conventions is to think about about flipping the conventions once for each border we add. For example:

- 0 Borders: $-, +, -, + \ldots$ (max) or $+, +, +, +, \ldots$ (min)
- 1 Border: $+, -, +, -, \ldots$ (max) or $-, -, -, -, \ldots$ (min)
- 2 Borders: $-, +, -, +, \ldots$ (max) or $+, +, +, +, \ldots$ (min)

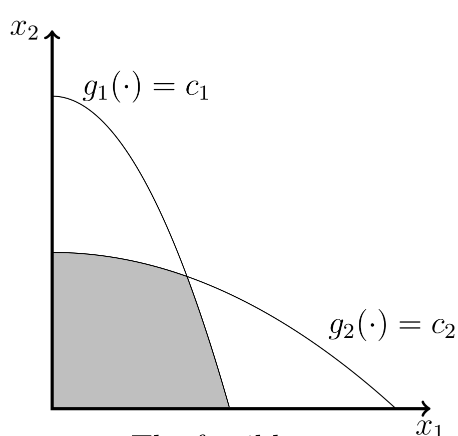
This may be more convenient to remember, but we also need to recall the fact that our first $2m - 1$ LPMs will be zero.

(s) Aside. This is a good fact to know, but it will virtually never come up in the first year courses with more than two constraints. In terms of testing material, neither the students nor the professors learn much by taking determinants of large-scale matrices.

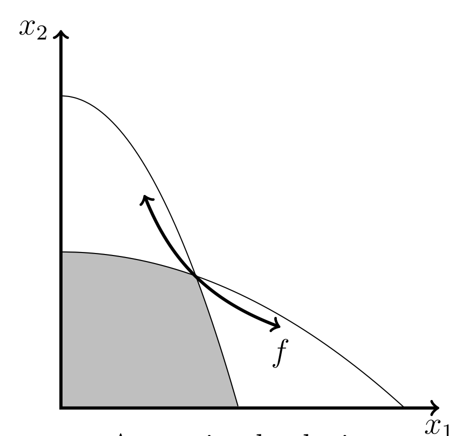
3. **Kuhn-Tucker Conditions**

(a) Aside. In the preceding discussions, we limited ourselves to about equality constraints, e.g., the condition $g(x_1, x_2) = c$ has to hold with *strict* equality. We can relax this to *inequality* constraints, to handle a wider array of problems. This leads us tothe Kuhn-Tucker conditions.

(b) Example. Consider a case with two constraint sets, $g_1(x_1, x_2) \leq c_1$ and $g_2(x_2, x_2) = c_2$, where WLOG, g_1 is the steeper of the constraints:



The feasible set.



An optimal solution.

Note that the graphical interpretation can be represented in terms of slopes of the objective function and constraints, i.e.,

$$-\frac{\partial g_1(\cdot)/\partial x_1}{\partial g_1(\cdot)/\partial x_2} \leq -\frac{\partial f(\cdot)/\partial x_1}{\partial f(\cdot)/\partial x_2} \leq -\frac{\partial g_2(\cdot)/\partial x_1}{\partial g_2(\cdot)/\partial x_2}$$

This says that the slope of the function at the optimum must be no flatter than $g_2(\cdot)$ and no steeper than $g_1(\cdot)$. In other words, for well-behaved functions, there are three possible solutions:

- $f(\cdot)$ is tangent to $g_2(\cdot)$
- $f(\cdot)$ is tangent to $g_1(\cdot)$
- $f(\cdot)$ is optimized at the intersection of $g_2(\cdot) = g_1(\cdot)$

More formally, we still satisfy our typical first-order conditions, e.g.,

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_i} - \sum_{j=1}^{2} \lambda_j \frac{\partial g_j(\mathbf{x}^*)}{\partial x_i}, \quad \text{for } i = 1, 2$$

where $\lambda_1^* \geq 0$ and $\lambda_2^* \geq 0$. Then the slope of the level curve is a linear combination of the slopes of the two constraints! Both constraints are binding in the picture above; this is not necessary, and we can formalize all of this with the Kuhn-Tucker theorem to help us find solutions in these types of cases.

(c) Theorem (JR A2.20). Let $f(\mathbf{x})$ and $g_j(\mathbf{x})$, $j = 1, \ldots, m$ be continuously differentiable real-valued functions defined over $D \subseteq \mathbb{R}^n$.

- Let $\mathbf{x}^*$ be an interior maxima of $f(\cdot)$ subject to the constraints $g_j(\cdot)$, $j = 1, \ldots, m$.
- Let $\nabla g_j(\mathbf{x}^*)$ be linearly independent for every binding constraint

Then there exists a unique vector $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ satisfies the **Kuhn-Tucker** conditions:

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*)}{\partial x_i} = \frac{\partial f(\mathbf{x}^*)}{\partial x_i} + \sum_{j=1}^{m} \lambda_j^* \frac{\partial g_j(\mathbf{x}^*)}{\partial x_i} = 0, \quad \text{for } i = 1, \ldots, n$$

with $\lambda_j \geq 0$ for all $j = 1, \ldots m$. Further, $\lambda_j^* \mathcal{L}_{\lambda_j} = 0$ for all $j = 1, \ldots, m$. This is known as **complementary slackness**.

(d) Aside. We need to be a litte more careful when constructing Kuhn-Tucker FOCs, to make sure that all of our λs are positive to match the theorem directly; in our contexts however, we frequently end up simplifying things a big.

(e) Example. Consider a typical utility-maximization problem with an income-inequality constraint and non-negativity constraints on consumption:

$$\max_{x,y} U(x, y) \quad \text{s.t.} \quad p_x x + p_y y \leq M \;\text{ and }\; x \geq 0, y \geq 0$$

Writing out the Lagrangian to conform to our Kuhn-Tucker formulation requires the use of three explicit constraints:

$$\mathcal{L}(x, y, \lambda_M, \lambda_x, \lambda_y) = U(x, y) + \lambda_m[M - p_x x - p_y y] + \lambda_x[x - 0] + \lambda_y[y - 0]$$

Then our Kuhn-Tucker first-order conditions are:

$$\begin{aligned}
\mathcal{L}_x &= U_x(x, y) + \lambda_M[-p_x] + \lambda_x[1] = 0 && \text{(w.r.t. } x) \\
\mathcal{L}_y &= U_y(x, y) + \lambda_M[-p_y] + \lambda_y[1] = 0 && \text{(w.r.t. } y) \\
\mathcal{L}_{\lambda_m} &= M - p_x x - p_y y \geq 0 && \text{(w.r.t. } \lambda_m) \\
\mathcal{L}_{\lambda_x} &= x - 0 \geq 0 && \text{(w.r.t. } \lambda_x) \\
\mathcal{L}_{\lambda_y} &= y - 0 \geq 0 && \text{(w.r.t. } \lambda_y)
\end{aligned}$$

with $\lambda_m \mathcal{L}_{\lambda_m} = 0$, $\lambda_x \mathcal{L}_{\lambda_x} = 0$, and $\lambda_y \mathcal{L}_{\lambda_y} = 0$. These complementary slackness conditions can be interpreted as follows:

- If our income constraint binds, then $M - p_x x - p_y y = 0$ and there is positive value to relaxing the constraint, so $\lambda_m > 0$. If our income constraint *does not bind*, then $M - p_x x - p_y y \geq 0$ and there is no value in relaxing the constraint, so $\lambda_m = 0$.
- If our non-negativity of x binds, then $x = 0$ and there is positive value in relaxing the condition, so $\lambda_x > 0$. If $x > 0$, the constraint *does not bind*, so there is no value in relaxing the constraint and $\lambda_x = 0$.
- If $y = 0$, there is value in relaxing the constraint, so $\lambda_y > 0$. If $y > 0$, there is no value in relaxing the constraint, so $\lambda_y = 0$.

This gives us a algorithmic framework with which to tackle more complicated problems.

(f) Example. Consider a quasi-linear utility maximization problem:

$$\max_{x,y} \ \ln(x) + y \quad \text{s.t.} \quad M \geq p_x x + p_y y \ \text{ and } \ x \geq 0, y \geq 0$$

where $p_x > 0$, $p_y > 0$, and $M > 0$. Then our associated Lagrangian is

$$\mathcal{L}(x, y, \lambda, \pi_x, \pi_y) = \ln(x) + y + \lambda[M - p_x x - p_y y] + \pi_x[x - 0] + \pi_y[y - 0]$$

Note that if either of our latter two constraints bind, it implies a corner solution. The Kuhn-Tucker formulation helps us spot potential corner solution in a systematic way. Our KTFOCs:

$$\begin{aligned}
\mathcal{L}_x &= \frac{1}{x} + \lambda[-p_x] + \pi_x[1] = 0 && \text{(w.r.t. } x\text{)} \\
\mathcal{L}_y &= 1 + \lambda[-p_y] + \pi_y[1] = 0 && \text{(w.r.t. } y\text{)} \\
\mathcal{L}_\lambda &= M - p_x x - p_y y \geq 0 && \text{(w.r.t. } \lambda\text{)} \\
\mathcal{L}_{\pi_x} &= x - 0 \geq 0 && \text{(w.r.t. } \lambda_x\text{)} \\
\mathcal{L}_{\pi_y} &= y - 0 \geq 0 && \text{(w.r.t. } \lambda_y\text{)}
\end{aligned}$$

With our KTFOCs, we can systematically consider the various combinations of binding constraints to find solutions:

- All three are binding ($x = 0$, $y = 0$, and $M = p_x x + p_y y$). This clearly is inconsistent, so we can discard this as a potential solution.
- $\mathcal{L}_{\pi_x}$ and $\mathcal{L}_\lambda$ are binding. Then our FOCs become:

$$\begin{aligned}
\mathcal{L}_x &= \frac{1}{x} - \lambda p_x + \pi_x = 0 && \text{(w.r.t. } x\text{)} \\
\mathcal{L}_y &= 1 - \lambda p_y = 0 && \text{(w.r.t. } y\text{)} \\
\mathcal{L}_\lambda &= M - p_x x - p_y y = 0 && \text{(w.r.t. } \lambda\text{)} \\
\mathcal{L}_{\pi_x} &= x = 0 && \text{(w.r.t. } \pi_x\text{)}
\end{aligned}$$

  This is also inconsistent, as $1/0$ is undefined, so we can discard this as a potential solution.
- $\mathcal{L}_{\pi_y}$ and $\mathcal{L}_\lambda$ are binding. Then our FOCs become:

$$\begin{aligned}
\mathcal{L}_x &= \frac{1}{x} - \lambda p_x = 0 && \text{(w.r.t. } x\text{)} \\
\mathcal{L}_y &= 1 - \lambda p_y + \pi_y = 0 && \text{(w.r.t. } y\text{)} \\
\mathcal{L}_\lambda &= M - p_x x - p_y y = 0 && \text{(w.r.t. } \lambda\text{)} \\
\mathcal{L}_{\pi_y} &= y = 0 && \text{(w.r.t. } \pi_x\text{)}
\end{aligned}$$

  From our third FOC (w.r.t. λ) and our fourth FOC (w.r.t. π_y):

$$\begin{aligned}
&\boxed{x^* = \frac{M}{p_x}} && \text{(solving for } x\text{)} \\
&\boxed{y^* = 0} && \text{(by assumption)} \\
&\boxed{\lambda^* = \frac{1}{M}} && \text{(solving for } \lambda\text{)}
\end{aligned}$$

These are usually the pieces we're after, but for completeness we can examine the final two Lagrange multipliers:

$$\boxed{\pi_x^* = 0} \qquad \text{(by assumption)}$$

$$\boxed{\pi_y^* = \frac{p_y}{M} - 1} \qquad \text{(solving for } \pi_y\text{)}$$

Thus, this is a solution to the system, so long as $\frac{M}{p_y} < 1$.

- $\mathcal{L}_\lambda$ is binding. Then our FOCs become

$$\mathcal{L}_x = \frac{1}{x} - \lambda p_x = 0 \qquad \text{(w.r.t. } x\text{)}$$

$$\mathcal{L}_y = 1 - \lambda p_y = 0 \qquad \text{(w.r.t. } y\text{)}$$

$$\mathcal{L}_\lambda = M - p_x x - p_y y = 0 \qquad \text{(w.r.t. } \lambda\text{)}$$

Solving our first two FOCs for λ and setting them equal:

$$\frac{1}{p_x x} = \frac{1}{p_y} \qquad \text{(from } \mathcal{L}_x \text{ and } \mathcal{L}_y\text{)}$$

$$\boxed{x^* = \frac{p_y}{p_x}} \qquad \text{(solving for } x\text{)}$$

$$M = p_x\left(\frac{p_y}{p_x}\right) + p_y y \qquad \text{(plugging } x \text{ into } \mathcal{L}_\lambda\text{)}$$

$$\boxed{y^* = \frac{M}{p_y} - 1} \qquad \text{(solving for } y\text{)}$$

$$\boxed{\lambda^* = \frac{1}{p_y}} \qquad \text{(solving for } \lambda\text{)}$$

$$\boxed{\pi_x^* = 0} \qquad \text{(by assumption)}$$

$$\boxed{\pi_y^* = 0} \qquad \text{(by assumption)}$$

- No constraints bind. This is clearly inconsistent, since our utility function is strictly increasing, so we can discard this as a potential solution.

(g) Aside. Hopefully, we've seen these types of problems before. We should become familiar enough with optimization problems to recognize when we can skip the entire Kuhn-Tucker formulation. That said, for more complicated utility functions, e.g.,

$$U(x_1, x_2, x_3) = x_1 + \ln(x_2 + \ln(1 + x_3))$$

where it's not as easy to spot potential corner solutions, Kuhn-Tucker helps us be systematic in solving the problems.

(h) Aside. Note that we will frequently see an alternative formulation for maximization problems with inequality constraints. If we have the maximization problem:

$$\max_{x,y} U(x, y) \quad \text{s.t.} \quad g(x, y) \leq c \text{ and } x \geq 0, y \geq 0$$

We will frequently write the Lagrangian with just one constraint:

$$\mathcal{L}(x, y, \lambda) = U(x, y) + \lambda\big[c - g(x, y)\big]$$

Then the Kuhn-Tucker First Order Conditions are:

$$\begin{aligned}
&\mathcal{L}_x = U_x(x, y) - \lambda g_x(x, y) \leq 0, && x\mathcal{L}_x = 0 && \text{(w.r.t. } x\text{)} \\
&\mathcal{L}_y = U_y(x, y) - \lambda g_y(x, y) \leq 0, && y\mathcal{L}_y = 0 && \text{(w.r.t. } y\text{)} \\
&\mathcal{L}_\lambda = c - g(x, y) \geq 0, && \lambda\mathcal{L}_\lambda = 0 && \text{(w.r.t. } \lambda\text{)}
\end{aligned}$$

This is equivalent to our KTFOCs above, since π_x and π_y were non-negative values. Again, the statement that the variable multiplied by the condition equals zero is complementary slackness.

(i) <u>Example</u>. Consider a situation with multiple linear constraints: a consumer faces the typical budget constraint, but also has a coupon constraint (sometimes used as a rationing mechanism). Suppose we have $U = xy^2$, $M = 100$, $p_x = p_y = 1$, $C = 120$, $c_x = 2$, $c_y = 1$ and $x, y \geq 0$. Then the Lagrangian can be written:

$$\mathcal{L} = xy^2 + \lambda_1[100 - x - y] + \lambda_2[120 - 2x - y]$$

The KTFOCs can be written:

$$\begin{aligned}
\mathcal{L}_x &= y^2 - \lambda_1 - 2\lambda_2 \leq 0 && x \geq 0 && x\mathcal{L}_x = 0 \\
\mathcal{L}_y &= 2xy - \lambda_1 - \lambda_2 \leq 0 && y \geq 0 && y\mathcal{L}_y = 0 \\
\mathcal{L}_{\lambda_1} &= 100 - x - y \geq 0 && \lambda_1 \geq 0 && \lambda_1\mathcal{L}_{\lambda_1} = 0 \\
\mathcal{L}_{\lambda_2} &= 120 - 2x - y \geq 0 && \lambda_2 \geq 0 && \lambda_2\mathcal{L}_{\lambda_2} = 0
\end{aligned}$$

Note that we can dispense with thinking about corner solution here: given the utility function, it should be clear that positive amounts of both goods need to be consumed. Now, check the situations for each of the two budget constraints:

- Let the first budget constraint bind:

$$\begin{aligned}
y^2 - \lambda_1 &= 0 && \text{(w.r.t. } x\text{)} \\
2xy - \lambda_1 &= 0 && \text{(w.r.t. } y\text{)} \\
100 - x - y &= 0 && \text{(w.r.t. } \lambda_1\text{)}
\end{aligned}$$

From the first two constraints, we can obtain a relationship for x and y:

$$\begin{aligned}
y^2 &= 2xy && \text{(from the 1st two FOCs)} \\
y &= 2x && \text{(solving for } y\text{)} \\
0 &= 100 - x - 2x && \text{(plugging into the constraint)} \\
&\boxed{\tilde{x} = \frac{100}{3}} && \text{(solving for } x\text{)} \\
&\boxed{\tilde{y} = \frac{200}{3}} && \text{(solving for } y\text{)} \\
&\boxed{\tilde{\lambda}_1 = \frac{4000}{9}} && \text{(solving for } \lambda_1\text{)}
\end{aligned}$$

But, we need to check if our second constraint is indeed slack:

$$0 \not\leq 120 - 2\left[\frac{100}{3}\right] - \frac{200}{3} \qquad \text{(testing the constraint)}$$

This violates our assumption that the second constraint is non-binding, so this is not a solution.

- Let the second budget constraint bind:

$$\begin{aligned}
y^2 - 2\lambda_2 &= 0 && \text{(w.r.t. } x\text{)}\\
2xy - \lambda_2 &= 0 && \text{(w.r.t. } y\text{)}\\
120 - 2x - y &= 0 && \text{(w.r.t. } \lambda_2\text{)}
\end{aligned}$$

From the first two constraints, we can obtain a relationship for x and y:

$$\begin{aligned}
\frac{y^2}{2} &= 2xy && \text{(from the 1st two FOCs)}\\
y &= 4x && \text{(solving for } y\text{)}\\
0 &= 100 - x - 4x && \text{(plugging into the constraint)}\\
&\boxed{x^* = 20} && \text{(solving for } x\text{)}\\
&\boxed{y^* = 80} && \text{(solving for } y\text{)}\\
&\boxed{\lambda_2^* = 3200} && \text{(solving for } \lambda_2\text{)}
\end{aligned}$$

Again, we need to check if our first constraint is indeed slack:

$$0 \leq 100 - 20 - 80 \qquad \text{(testing the constraint)}$$

This is an odd case, where the constraint "binds," but only "just so." That is, it is technically mathematically binding, but not economically. That is, the second constraint holds with equality, but there is no value in relaxing the constraint! Thus, this is both the solution when only the second constraint binds, *and* the solution when both constraints bind!

(j) <u>Aside</u>. Note that we don't need to think about the constraint qualification (the linear independence of the gradient of binding constraints) all that frequently. We can think about an example, however, to demonstrate when they might fail. Consider the maximization problem:

$$\max_{x_1, x_2} x_1 \quad \text{s.t.} \quad x_2 - (1 - x_1)^3 \leq 0 \text{ and } x_1, x_2 \geq 0$$

The Lagrangian is then

$$\mathcal{L}(x_1, x_2, \lambda, \pi_1, \pi_2) = x_1 + \lambda\big[0 - x_2 + (1 - x_1)^3\big] + \pi_1[x - 0] + \pi_2[y - 0]$$

The KTFOC are:

$$\begin{aligned}
\mathcal{L}_1 &= 1 + \lambda\big[-3(1 - x_1)^2\big] + \pi_1 = 0 && \text{(w.r.t. } x_1\\
\mathcal{L}_2 &= -\lambda + \pi_2 = 0 && \text{(w.r.t. } x_2\text{)}\\
\mathcal{L}_\lambda &= 0 - x_2 + (1 - x_1)^3 \geq 0 && \text{(w.r.t. } \lambda\text{)}\\
\mathcal{L}_{\pi_1} &= x_1 - 0 \geq 0 && \text{(w.r.t. } \pi_1\text{)}\\
\mathcal{L}_{\pi_2} &= x_2 - 0 \geq 0 && \text{(w.r.t. } \pi_2\text{)}
\end{aligned}$$

Where we have complementary slackness. The optimal solution here is $(1,0)$, but note then the inconsistency in the 1st FOC:

$$\begin{aligned}\mathcal{L}_1 &= 1 + \lambda[0] + \pi_1 = 0 && \text{(the first FOC)}\\ \pi_1 &= -1 && \text{(solving for } \pi_1\text{)}\end{aligned}$$

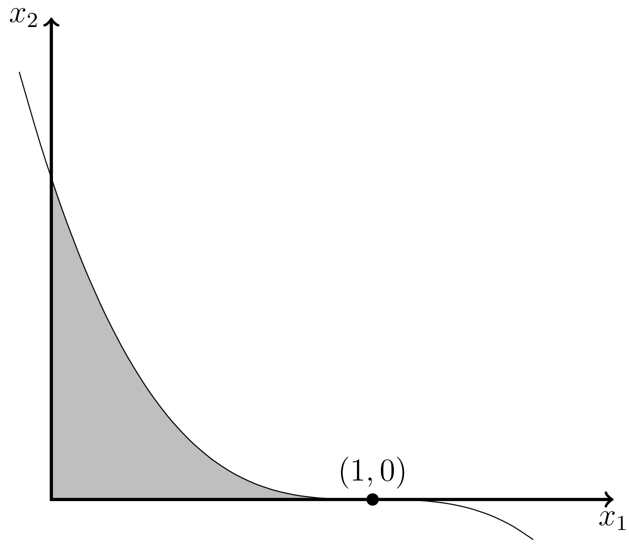
This is a result of a failure of our constraint qualification. The gradients of our constraints are:

$$\nabla[x_1] = \begin{bmatrix}1 & 0\end{bmatrix} \qquad \nabla[x_2] = \begin{bmatrix}0 & 1\end{bmatrix} \qquad \nabla[x_2 - (1-x_1)^3] = \begin{bmatrix}3(1-x_1)^2 & 1\end{bmatrix}$$

At the optimum, however, the gradient of the two *binding* constraints are:

$$\nabla[x_2^*] = \begin{bmatrix}0 & 1\end{bmatrix} \qquad \nabla[x_2^* - (1-x_1^*)^3] = \begin{bmatrix}0 & 1\end{bmatrix}$$

Graphically, these constraints produce a somewhat odd constraint set:



As we can see, the slope of our constraint $x_2 = 0$ is the same as the slope of our constraint $x_2 - (1-x_1)^3 = 0$ at the optimal point (1,0)–one interpretation is that we don't actually have two different constraints *at the optimum.* While this type of problem won't come up much in economics, it is a possbility in more complicated non-linear programming problems. Luckily, we have a a few useful theorem that avoids a lot of the technical details in constrained optimization.

(k) Theorem (CW 13.2). Consider the constrained optimization problem

$$\operatorname*{opt}_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g_1(\mathbf{x}) \leq c_1, \ldots, g_m(\mathbf{x}) \leq c_m$$

If $g_j(\mathbf{x})$ is a linear function for $j = 1, \ldots m$ and the feasible set is convex, then the constraint qualification will be met and the Kuhn-Tucker conditions will hold at an optimal solution.

(l) Aside. Recall the example where we were looking for corner solutions in a very mechanical way (e.g., without using our economic intuition). This theorem guaranteed that we would find solutions, since our constraint set was the budget-set "triangle." Further, we were guaranteed that our solutions were optimal (without checking any second-order conditions) by the following theorem.

(m) Theorem (MWG THM M.K.4). Consider the constrained maximization problem

$$\max_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g_1(\mathbf{x}) = c_1, \ldots, g_m(\mathbf{x}) = c_m$$

Suppose the feasible set formed by $g_j(\mathbf{x}) \leq c_j$, $j = 1, \ldots, m$ is a convex set and that $f(\mathbf{x})$ is strictly quasiconcave. Then there is a unique global constrained maximizer.

(n) Aside. This ends up going a *long* way in helping us with problems. Most functions we use in the first-year sequences are strictly quasiconcave, so we can rely on this without checking all of our more technical definitions to ensure we have a solution.

4. **The Envelope Theorem**

(a) Definition. Consider the maximization problem

$$\max_{\mathbf{x}} f(\mathbf{x}, \mathbf{a}) \quad \text{s.t.} \quad \tilde{g}_j(\mathbf{x}, \mathbf{a}) \leq 0,\ j = 1, \ldots, m$$

Where $\mathbf{x} \in \mathbb{R}^n$ is the vector of choice variables and $\mathbf{a} \in \mathbb{R}^k$ is a vector of parameters that enter the objective function, constraint, or both. Then with the following assumptions:

- Let A be the set of parameters under consideration. For every $\mathbf{a} \in A$, there exists a $\mathbf{x}$ such that the constraints are satsified.
- For each $\mathbf{a} \in A$ there is a solution to the optimization problem.
- For a vector $\mathbf{a}$, the maximized value of the objective function is $f(\mathbf{x}(\mathbf{a}), \mathbf{a})$.

Then the **value function** $V(\mathbf{a})$ is defined as

$$V(\mathbf{a}) = \max_{\mathbf{x}} f(\mathbf{x}, \mathbf{a}) \quad \text{s.t.} \quad \tilde{g}_j(\mathbf{x}, \mathbf{a}) \leq 0,\ j = 1, \ldots, m$$

(b) Example. Recall a typical utility maximization problem, e.g.,

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

If we set up and solve this Lagrangian, we obtain the demand functions:

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

Further, our Lagrange multiplier is:

$$\lambda^*(M, p_x, p_y) = \left(\frac{\alpha}{p_x}\right)^\alpha \left(\frac{1-\alpha}{p_y}\right)^{1-\alpha}$$

We can plug the demands back into the utility function to obtain the indirect utility function:

$$u(x^*, y^*) = M\left(\frac{\alpha}{p_x}\right)^\alpha \left(\frac{1-\alpha}{p_y}\right)^{1-\alpha}$$

This is precisely the value function $V(M, p_x, p_y)$ described in our definition. The interpretation of this function is "plug in any parameter values, the function gives the optimal utility level for those parameters."

(c) Theorem (JR THM A2.22). Consider the constrained maximization problem:

$$\max_{\mathbf{x}} f(\mathbf{x}, \mathbf{a}) \quad \text{s.t.} \quad \tilde{g}(\mathbf{a}) \leq 0$$

Assume that:

- f and g are differentiable
- $\mathbf{x}(\mathbf{a})$ is the unique solution

- The constraint g binds for all values of $\mathbf{a}$
- $\mathcal{L}(\mathbf{x}, \mathbf{a}, \lambda)$ is the associated Lagrangian and $V(\mathbf{a})$ is the associated value function
- $(\mathbf{x}(\mathbf{a}), \lambda(\mathbf{a})$ satisfy the Kuhn-Tucker conditions

Then the **Envelope Theorem** states that for every $\mathbf{a}$:

$$\frac{\partial V(\mathbf{a})}{\partial a_j} = \frac{\partial \mathcal{L}}{\partial a_j}\bigg|_{\mathbf{x}(\mathbf{a}),\lambda(\mathbf{a})}$$

(d) Example. We can "prove" that this is true by taking derivatives and manipulating our conditions with a little bit of algebra. Note that the Lagrangian is

$$\mathcal{L} = f(\mathbf{x}, \mathbf{a}) - \lambda[\tilde{g}(\mathbf{x}, \mathbf{a})]$$

The first order conditions from this, evaluated at the optimum, are:

$$0 = \frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial x_i} - \lambda(\mathbf{a})\frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial x_i} \qquad \text{for } i = 1, \dots, n \qquad (*)$$

$$0 = g(\mathbf{x}(\mathbf{a}), \mathbf{a}) \qquad \text{(the constraint)}$$

We can also differentiate both sides of the constraint, using the chain rule, for reasons that will become apparent shortly:

$$0 = \sum_{i=1}^{n} \left[\frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial x_i} \cdot \frac{\partial x_i(\mathbf{a})}{\partial a_j}\right] + \frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} \qquad (**)$$

Note that the derivative of $\mathcal{L}$ with respect to a_j is simply:

$$\frac{\partial \mathcal{L}}{\partial a_j} = \frac{\partial f(\mathbf{x}, \mathbf{a})}{\partial a_j} - \lambda\frac{\partial g(\mathbf{x}, \mathbf{a})}{\partial a_j} \qquad \text{(differentiating w.r.t. } a_j\text{)}$$

$$\frac{\partial \mathcal{L}}{\partial a_j}\bigg|_{\mathbf{x}(\mathbf{a},\mathbf{a}} = \frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} - \lambda\frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} \qquad \text{(evaluating at } \mathbf{x}(\mathbf{a})\text{)}$$

Further, consider the value function:

$$V(\mathbf{a}) = f(\mathbf{x}(\mathbf{a}), \mathbf{a}) \qquad \text{(the value function)}$$

$$\frac{\partial V(\mathbf{a})}{\partial a_j} = \sum_{i=1}^{n} \left[\frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial x_i} \cdot \frac{\partial x_i(\mathbf{a})}{\partial a_j}\right] + \frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} \qquad \text{(by the chain-rule)}$$

Consider the preceding sum. We have n terms in the brackets, but these correspond to the n conditions in our 1st "general" FOC (*):

$$\frac{\partial V(\mathbf{a})}{\partial a_j} = \sum_{i=1}^{n} \left[\lambda(\mathbf{a})\frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial x_i} \cdot \frac{\partial x_i(\mathbf{a})}{\partial a_j}\right] + \frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} \qquad \text{(substituting)}$$

$$= \lambda(\mathbf{a})\sum_{i=1}^{n} \left[\frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial x_i} \cdot \frac{\partial x_i(\mathbf{a})}{\partial a_j}\right] + \frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} \qquad \text{(pulling out the constant)}$$

Finally, note that the sum can be put in terms of our differentiated constraint (**):

$$\frac{\partial V(\mathbf{a})}{\partial a_j} = \lambda(\mathbf{a})\left[-\frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j}\right] + \frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} \qquad \text{(substituting)}$$

$$= \frac{\partial f(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} - \lambda(\mathbf{a})\frac{\partial g(\mathbf{x}(\mathbf{a}), \mathbf{a})}{\partial a_j} \qquad \text{(rearranging)}$$

Which is precisely what we got above.

(e) Example. Consider a constrainted utility-maximization problem:

$$\max_{x,y}\ u(x,y) \quad \text{s.t.} \quad M \geq p_x x + p_y y, \ \text{ and } \ x \geq 0, y \geq 0$$

The associated Lagrangian is

$$\mathcal{L}(x,y,\lambda) = u(x,y) + \lambda\big[M - p_x x - p_y y\big]$$

Assuming we have an increasing, strictly quasi-concave utility function, then the envelope theorem gives us several key results:

$$\begin{aligned}
\frac{\partial u(x^*,y^*)}{\partial M} &= \frac{\partial \mathcal{L}(x,y,\lambda)}{\partial M}\bigg|_{x^*,y^*,\lambda^*} && \text{(differentiating w.r.t. } M\text{)} \\
&= \lambda\,\Big|_{x^*,y^*,\lambda^*} && \text{(taking the derivative of } \mathcal{L}\text{)} \\
\frac{\partial u(x^*,y^*)}{\partial M} &= \lambda^* && \text{(evaluating at the optimum)}
\end{aligned}$$

Thus, λ^* can be interpreted as the *marginal utility of income*; it measures how utility changes as income changes. Somteims, λ is referred to as the "shadow value" of income. From our Cobb-Douglas utility example above, we can verify that this holds:

$$\begin{aligned}
\lambda^*(M,p_x,p_y) &= \left(\frac{\alpha}{p_x}\right)^{\alpha}\left(\frac{1-\alpha}{p_y}\right)^{1-\alpha} && (\lambda^* \text{ from above}) \\
u(x^*,y^*) &= M\left(\frac{\alpha}{p_x}\right)^{\alpha}\left(\frac{1-\alpha}{p_y}\right)^{1-\alpha} && \text{(indirect utility from above)} \\
\frac{\partial u(\cdot)}{\partial M} &= \left(\frac{\alpha}{p_x}\right)^{\alpha}\left(\frac{1-\alpha}{p_y}\right)^{1-\alpha} && \text{(differentiating w.r.t. } M\text{)}
\end{aligned}$$

(f) Aside. Note that our envelope ideas hold for unconstrained maximization problems as well as constrained. Minimization problems, furthermore, will produce envlelope theorem results as well. Concepts such as "Roy's Identity," "Shephard's Lemma," and "Hotelling's Lemma," which we will encounter during our first-year sequences, are fundamentally envelope theorem results.