Math Camp: Optimization and Comparative Statics in $\mathbb{R}^n$

Jacob Hukill

August 17, 2023

Functions in $\mathbb{R}^n$

We already covered single variables functions that take inputs from one dimensional space. Multivariate functions take inputs from more than one dimension. For example, the function $f : \mathbb{R}^2 \to [0, \infty)$ given by $f(x, y) = x^2 + y^2$ takes inputs from two dimensional space, i.e., $(x, y) \in \mathbb{R}^2$ and returns one dimensional scalars from $[0\infty)$. More specifically,

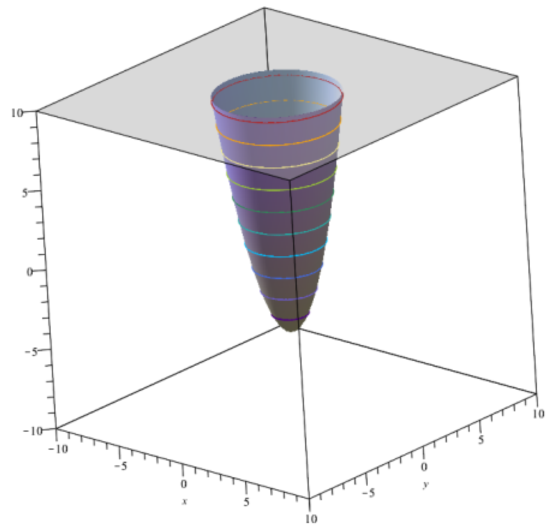
Definition 1. A function f of n variables $x_1, ..., x_n$ with domain $\mathcal{D}$ is a rule that assigns a specified number $f(x_1, ...x_n)$ to each $n-$vector $(x_1, ..., x_n)$ in $\mathcal{D}$.

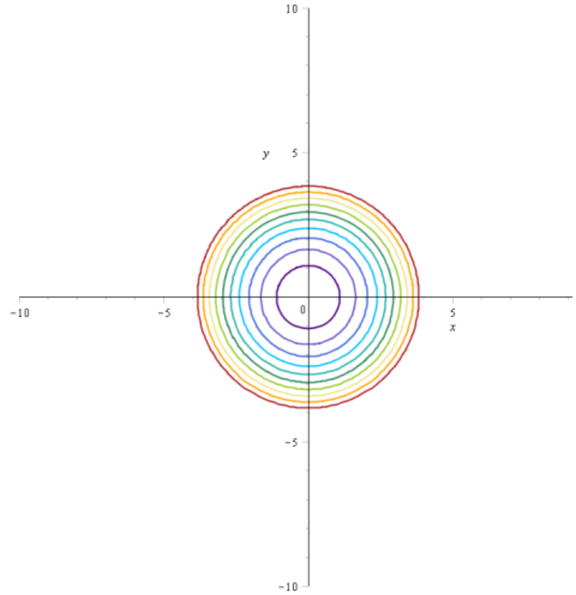
Given a function $f : \mathbb{R}^n \to: \mathbb{R}$, we are often interested in all the possible values of $x \in \mathbb{R}^n$ such that $f(x) = c$, where c is some given constant. We call the set of such points in the domain of a function f a *level set.* You may hear many names associated with this idea. For a function of two variables, it's called the level curve; for three, the level surface; in higher dimensions, the level hypersurface. You may also hear level sets be called implicit curves or countour sets. However, you are probably familiar with another term that they are known by - isoquants and indifference curves.

Definition 2. Given a function $f : \mathbb{R}^n \to: \mathbb{R}$, the level set is defined as

$$L_a(f) = \{x \in \mathbb{R}^n : f(x) = a\}$$

We can use this idea to get a geometric representation of some arbitrary function $z = f(x, y)$ and one of its level curves. We can visualize the graph of the function in a three dimensional space, with horizontal planes (parallel to the x-axis) cutting across the function. The intersections of these planes and the graph can be projected onto a two dimensional xy plane - when z=c (some constant), then the projection of this intersection is the level curve at c. This is shown for the function $f = x^2 + y^2$ below.

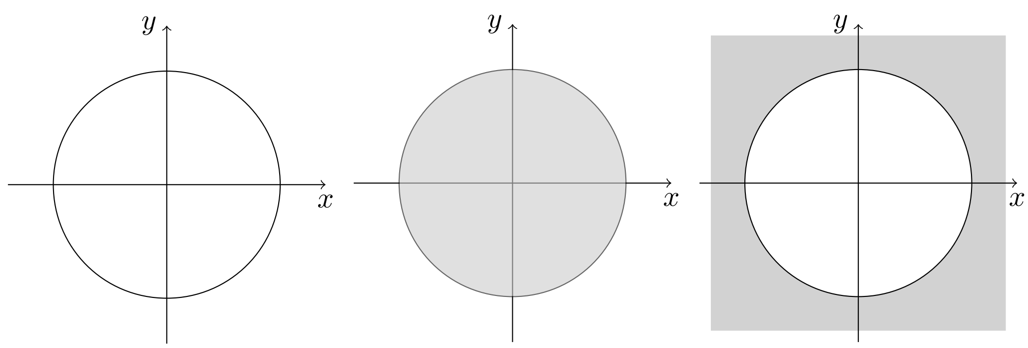




Sometimes we are also interested in all possible values of $x \in \mathbb{R}^n$ such that $f(x) \geq c$ (or $f(x) \leq c$), where c is some given constant. Such sets are called *upper level sets* (*lower level sets*.)

Definition 3. Given a function $f : \mathbb{R}^n \rightarrow: \mathbb{R}$, the upper and lower level sets are defined as $L_a^+(f) = \{x \in \mathbb{R}^n : f(x) \geq a\}$ and $L_a^-(f) = \{x \in \mathbb{R}^n : f(x) \leq a\}$, respectively.

Example 1. Assume that $f(x, y) = x^2 + y^2$, then $L(f)$ is a circle centered at the origin with a radius of $\sqrt{c}$. See the graph below illustrating $L(f), L^+(f)$, and $L^-(f)$.



$L_c(f)$ $\qquad$ $L_c^-(f)$ $\qquad$ $L_c^+(f)$

A more example pertinent for economics is the following: Suppose $u : \mathbb{R}_+^n \rightarrow: \mathbb{R}$ is a given utility function. Here, $x \in \mathbb{R}_+^n$ is a vector, called a bundle. Each component of this vector corresponds to an amount of a good consumed. The level set defined by $U(x) = \overline{u}$ is what we call the indifference curve - a curved that describes the set of bundles that leave the agent with the same level of utility. Below are some examples.

Example 2. Assume $u(x_1, x_2) = \alpha x_1 + \beta x_2$ and $\alpha, \beta > 0$. Then the indifference curves (ICs) are just straight lines in the (x_1, x_2) plane. To see this, assume that some level of utility $\overline{u}$ is given. Then $\alpha x_1 + \beta x_2 = \overline{u}$ defines a level curve - and as you might spot, this is an equation of a straight line. We could solve for x_2 to get:

$$x_2 = \frac{\overline{u}}{\beta} - \frac{\alpha}{\beta} x_1$$

This would correspond to the case when goods x_1, x_2 are perfect substitutes.

Example 3. Assume $u(x_1, x_2) = \min \alpha x_1, \beta x_2$ and $\alpha, \beta > 0$. Then the indifference curves (ICs) are L-shaped. To see this, suppose that first $\alpha x_1 \geq \beta x_2$. Then $u(x_1, x_2) = \beta x_2$ for all x_1 satisfying the inequality - the function does not change value as we increase x_1, so this part is given by a horizontal line in the (x_1, x_2) plane. Similarly, we could do the opposite (switch the inequality) to get a vertical line. Putting these together, we get L-shaped indifference curves. The set of points for which the curves have kinks is given by a line $\alpha x_1 = \beta x_2$. This would correspond to the case when goods x_1, x_2 are perfect complements, as goods are consumed in fixed proportions.

Example 4. Assume $u(x_1, x_2) = (\alpha x_1^\rho + \beta x_2^\rho)^{1\rho}$, and $\alpha, \beta > 0$, and $\rho < 1$. This is known as CES utility. Now suppose that we have some value of utility $\overline{u}$ given. We can get the equation of the IC by solving for one of the variables:

$$x_2 = (\frac{\overline{u}^\rho}{\beta} - \frac{\alpha}{\beta} x_1^\rho)^{1/\rho}$$

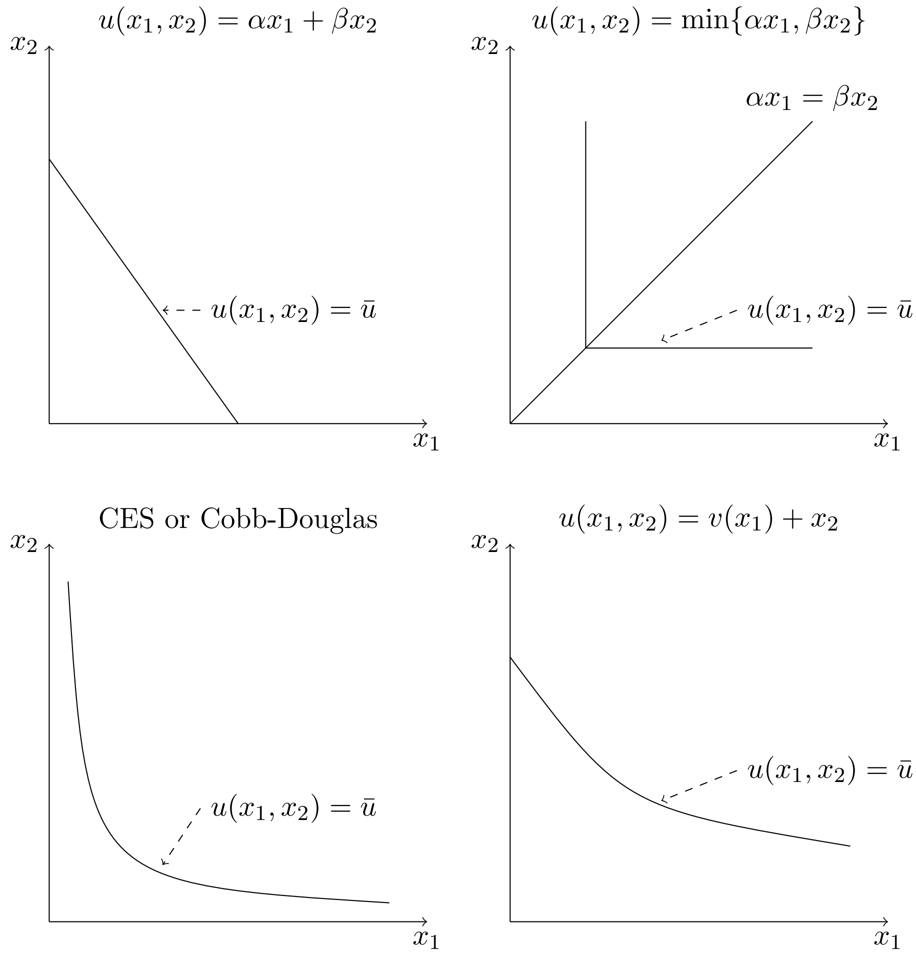
Another function form frequently used in eocnomics is $Cobb - Douglas$. This utiltiy would be given by $u(x_1, x_2) = x_1^\alpha x_2^\beta$ with IC given by $x_2 = \frac{\overline{u}^{1/\beta}}{x_1^{\alpha/\beta}}$. ICs now have a hyperbolic shape. Interestingly, each of the previous functional forms are just special cases of the CES function when ρ takes some particular values[1].

Example 5. Assume $u(x_1, x_2) = v(x_1) + x_2$, where $v(x_1)$ is some increasing function[2]. For example, if $v = \sqrt{x}$, then the ICs are given by $x_2 = \overline{u} - \sqrt{x_1}$.

These graphs below illustrate the level curves for each of the above discussed examples.

[1] If $\rho \to 1$, we have perfect substitutes; 0 and we have Cobb-Douglas; $-\infty$ and we have perfect complements.

[2] We actually need it to be increasing and concave, but we'll explain concavity of functions later.



Continuity and Differentiability

The concept of continuity for functions of one variable can be generalized to that of several variables - the idea being that a function of n variables is continuous if small changes in the independent variables give small changes in the value of the function (recall the epsilon-delta demonstration). We can apply the useful rule: Any function of n variables that can be constructed from continuous functions by combining the operations of addition, subtraction, multiplication, division, and function composition is continuous wherever it is defined. This is useful, because if our function is continuous, then the level set will have no "gaps' and will also be continuous.

We could use the general $(\varepsilon - \delta)$ definition in terms of metric spaces that we previously discussed, or we could use some sequential definition

Lemma 1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at $\boldsymbol{x}$, if and only if, for all sequences $x_n \to x$,* *we have that*

$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x).$$

This result states that the limit operator and the function can be interchanged if and only if the function is continuous at **x.**

In the multidimensional case, the derivative of a function conveys a similar idea to the univariate case with a slight twist when we hold other variables constant. We might for example want to know how a firm's profit might vary when we change only one input and hold the others constant. In terms of notation, we call the derivative with respect to the ith variable of f the partial derivative of f with respect to x_i, given by $\frac{\partial f}{\partial x_i}$. More formally, we can define it:

Definition 4. Suppose a function $f = f(x_1, ..., x_n)$. Then the partial derivative of f with respect to x_i is defined as

$$\frac{\partial f}{\partial x_i} = \lim_{h\to 0} \frac{f(x_1, ...x_i + h, ..., x_n) - f(x_1, ...x_i, ..., x_n)}{h}$$

It is usually pretty easy to find the partial derivatives of a function - we can simply think of a function as simply only a function of x_i and hold other variables constant, and just differentiate f as if it were only a function of x_i. You might see other notation as well: $\frac{\partial f}{\partial x_i} = f'_{x_i}(\cdot)$[3]$= f'_i(\cdot) = \frac{\partial f(\cdot)}{\partial x_i}$.

The generalization of the first order derivative in $\mathbb{R}^n$ is called the gradient of a function, which we denote by ∇f, or Df. It is simply a vector of first order partial derivatives:

$$\nabla f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}$$

Like in the univariate case, we are interested in how small changes to our variable changes the function at a point- but now we must consider changes to all of the variables, so we now have a vector of "changes". The gradient also has some really nice features that relate to what we've studied so far.

[3]It can be a little dangerous to use this notation, especially in the case of compound functions, in the sense of introducing ambiguity. For example, it may not be clear what $f_x(x^2y, 2y^3x^5)$ means. We will often instead use $f_1(x^2y, 2y^3x^5)$, which refers to the argument position, and taking the partial with respect to the first variable.

Properties of the Gradient

If $f(\mathbf{x}) = f(x_1, ..., x_n)$ is differentiable and $\nabla f(\mathbf{x})$ is nonzero, then:

- $\nabla f(\mathbf{x})$ is orthogonal to the level set $f(\mathbf{x})$=C.
- $\nabla f(\mathbf{x})$ points in the direction of maximal increase of $f(\mathbf{x})$.

You may already understand why this will be so useful in optimization. If the gradient points in the maximal increase of f, then to get to a local maximum, we would want to go in the direction of the gradient. Moreover, if the gradient is pointing in a direction, that means that the function is changing at that point - if the gradient is a zero vector at a point, then the point is called a *stationary point.* You may already see the role this will play in optimizing. We will come back to this point in a bit.

We can also define higher order derivatives as well:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \lim_{h \to 0} \frac{f_i(x_1, ...x_i + h, ..., x_n) - f_i(x_1, ...x_i, ..., x_n)}{h}$$

In general, many of the basic rules and properties of derivatives also translate to higher dimensions.

Example 6. Suppose $f(x, y) = e^{xy} + x^2 + y^2 + 4xy$, then

$$f_x(x, y) = ye^{xy} + 2x + 4y$$
$$f_y(x, y) = xe^{xy} + 2y + 4x$$

and the second order partials are:

$$f_{xx}(x, y) = y^2 e^{xy} + 2$$
$$f_{xy}(x, y) = e^{xy} + xye^{xy} + 4$$
$$f_{yx}(x, y) = e^{xy} + xye^{xy} + 4$$
$$f_{yy}(x, y) = x^2 e^{xy} + 2$$

Notice that $f_{xy}(x, y) = f_{yx}(x, y)$. This implies the order of differentiation does not matter. In most cases, the functions that we work with will have nice properties, so this above observation will be true. A result known as **Young's Theorem** makes a more general result, requiring a continuity condition.

Theorem 1 (Young's Theorem)**.** *Suppose that two mth order partial derivatives of the function $f(x_1, ..., x_n)$ involve the same number of differentiations with respect to each of the variables, and are both continuous in an open set S. Then the two partial derivatives are necessarily equal at all points in S.*

Almost all of the functions we will work with will have continuous partial derivatives everywhere in their domains. If a function f does have a continuous partial derivative of the first order in a domain A, then we call f **continuously differentiable** in A. In this case, f is called a C^1 **function** on A. If all partials up to order k exist and are continuous, f is called a C^k **function**.

We saw how second order partials were defined. In the case of a function of two variables, we would have four second-order partials. For a function of n variables, we would have n^2 second-order partials. The generalization of the second order derivative in $\mathbb{R}^n$ is given by something called the **Hessian**, or the **Hessian Matrix** of f. It is simply a nxn matrix of all second order derivatives.

$$H = \begin{pmatrix} f_{11}(\mathbf{x}) & f_{12}(\mathbf{x}) & \dots & f_{1n}(\mathbf{x}) \\ f_{21}(\mathbf{x}) & f_{22}(\mathbf{x}) & \dots & f_{2n}(\mathbf{x}) \\ \vdots & \vdots & \ddots \vdots & \\ f_{n1}(\mathbf{x}) & f_{n2}(\mathbf{x}) & \dots & f_{nn}(\mathbf{x}) \end{pmatrix}$$

As we mentioned before, usually $f_{12} = f_{21}$. Therefore, H will be a symmetric matrix.

We also have multivariate versions o Taylor's theorem. For example, if we have $f(x_1, x_2)$, then we can approximate f around $x^0 = (x_1^0, x_2^0)$ as

$$\begin{aligned} f(x_1, x_2) = f(x^0) + f_1(x^0)(x_1 - x_1^0) + f_2(x^0)(x_2 - x_2^0) + \frac{f_{11}(x^0)}{2!}(x_1 - x_1^0) + \\ 2\frac{f_{12}(x^0)}{2!}(x_1 - x_1^0)(x_2 - x_2^0) + \frac{f_{22}(x^0)}{2!}(x_2 - x_2^0) + ... \end{aligned}$$

Frequently, we write the second order approximation using matrix notation as

$$f(x) \approx f(x^0) + \nabla f(x^0)'(x - x^0) + \frac{1}{2!}(x - x^0)Hf(x^0)(x - x^0)$$

where

$$\nabla f(x^0) = \begin{pmatrix} f_1(x^0) \\ f_2(x^0) \end{pmatrix}, (x - x^0) = \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \end{pmatrix}, Hf(x^0) = \begin{pmatrix} f_{11}(x^0) & f_{12}(x^0) \\ f_{21}(x^0) & f_{22}(x^0) \end{pmatrix}$$

Many economic models involve composite functions. Functions of variables are themselves functions of other variables. What happens to the value of a composite function as its basic variables change?

Suppose that we have a function $z = F(x, y)$, where both $x = f(t), y = g(t)$ are functions of the variable t. Then we could substitute into z to get

$$z = F(f(t), g(t))$$

Now z is a function of t alone. How will t respond to small changes in t? This is given by the following rule:

Theorem 2 (Chain Rule)**.** *Let z= F(x,y), where x=f(t) and y=g(t). Then*

$$\frac{dz}{dt} = F_1'(x, y)\frac{dx}{dt} + F_2'(x, y)\frac{dy}{dt}$$

This derivative is normally called the **total derivative** of z with respect to t. Notice how z changes: it first depends on how the variable x changes with respect to time; it then depends on how the function z itself depends on how x changes. The same goes for y. There is both a direct change, and an indirect change.

We can generalize the chain rule to allow for our x, y to be functions of multiple variables. Now when we want to find the derivative, we must hold some of these other variables constant.

Theorem 3 (The general chain rule[4])**.** *Suppose that* $z = F(x_1, ..., x_n)$ *with* $x_1 = f_1(t_1, ..., t_m), ..., x_n = f_n(t_1, ..., t_m)$*. Then*

$$\frac{\partial z}{\partial t_j} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_j} + ... + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t_j}$$

Comparative Statics

Mostly in economics (and in other sciences), the model that describes something boils down to a key equation or system of equations that give the relationships between variables determined within the system (endogenous variables), and the variables determined outside of the system

[4]We must make some continuity and differentiability assumptions for this to hold.

(exogenous variables or parameters). We are often interested in how these endogenous variables change as we change exogenous variables:

- How might demand and supply respond to changes in parameters like price?
- How will consumption change as income changes?
- How will optimal savings decisions change as we increase interest rates?

We call this sort of exercise comparative statics.

Response of the maximizers

We will first focus on the response of the maximizers, or the endogenous choice variables. The first method that we can use is a simple one - if we have explicit parametric assumptions, we can simply differentiate with respect to the variable of interest.

While this can be useful, we often do not make explicit parametric assumptions.

Definition 5. An implicit solution to the equation of the form $f(x, \alpha) = 0$ is a function $x^* = x(\alpha)$, so that:

$$f(x(\alpha), \alpha) = 0$$

.

We now have the equation written as an identity that is satisfied for all values of α. We can then differentiate both sides with respect to α:

$$\frac{df(x(\alpha), \alpha)}{d\alpha} = 0$$

Now, look at the above derivative, what does it mean? How does f at $x(\alpha)$ change as we change α marginally? How can α affect f? In two ways - first, from the value of $x(\alpha)$ (called the indirect effect), and second, though direct dependence of f on α (this is called the direct effect). Therefore,

$$\frac{df(x(\alpha), \alpha)}{d\alpha} = \frac{\partial f(x(\alpha), \alpha)}{\partial x}\frac{dx(\alpha)}{d\alpha} + \frac{\partial f(x(\alpha), \alpha)}{\partial \alpha}.$$

If you are confused with the above sum, think about what this tells us. The total change is the sum of direct and indirect changes. This sum accounts for the only two possible ways that f can change along the solution of $x(\alpha)$. Then, if we rearrange (assuming that $\frac{\partial f(x(\alpha),\alpha)}{\partial x} \neq 0$

we get

$$\frac{dx(\alpha)}{d\alpha} = -\frac{\partial f/\partial \alpha}{\partial f/\partial x}|_*$$

where $|_*$ just means that the quantity is evaluated at the solution. This result is called the *implicit function theorem (IFT)*. Sometimes, we omit $|_*$ in order to not encumber notation. In terms o notation, we instead write

$$x'(\alpha) = -\frac{f_\alpha}{f_x}|_*.$$

The IFT has a particularly useful application in optimization theory. For example, suppose that we have a first order condition

$$\frac{df(x\alpha, \alpha)}{dx} = 0$$

Then we can apply the IFT to see how an optimized choice variable $x(\alpha)$ changes as we change α.

$$\frac{df(x\alpha, \alpha)}{dx} = -\frac{f_{x\alpha}}{f_{xx}}|_*.$$

Therefore, the sign of $dx(\alpha)/d\alpha$ is completely determined by the quantity the signs of the second derivatives - of which we usually have some assumptions on for there to be sufficient conditions for max.

Example 7. Suppose we want to maximize $f(x) = -2x^2 + \alpha^2 x + 5$. The FONC for this maximization problem is

$$-4x + \alpha^2 = 0$$

thus $x(\alpha) = \frac{\alpha^2}{4}$. Note that we can think of α as some parameter. Then for a SOSC, we would need that $f''(x) < 0$, which it is. From the explicit solution, we know that $dx(\alpha)/d\alpha = \alpha/2$, but let us verify this using the IFT. To do so, let $g(x, \alpha) = -4x + \alpha^2$ so that the FONC at the solution reads as $g(x(\alpha), \alpha) = 0$. Applying the IFT

$$\frac{dx(\alpha)}{d\alpha} = -\frac{\partial g/\partial \alpha}{\partial g/\partial x}|_*$$
$$\frac{-2\alpha}{f''(x)}|_*.$$

Since we know the sign (and actually the function itself) of f'', the sign is determined by α. Thus if $\alpha > 0$, then $x'(\alpha) > 0$.

Example 8. Suppose that the consumer's utility function is given by $u(x, y) = xy$. The problem that the agent faces is to choose the optimal amount of x and y that would maximize utility, subject to a budget constraint that total expenditure on two goods should equal to income: $p_x x + p_y y = I$. Thus, in the language of optimization theory:

$$\max_{x,y} u(x, y) \quad s.t. p_x x + p_y y = I.$$

We can convert the above problem into a one dimensional maximization problem by solving for (say) y form the budget constraint, and then substituting the resulting expression into the objective function:

$$u(x) \equiv u(x, \frac{I}{p_y} - \frac{\tilde{p_x}}{p_y} x) = x(\frac{I}{p_y} - \frac{p_x}{p_y} x).$$

Now we can apply the tools we have used so far. The FONC is given by

$$\tilde{u}'(x) = \frac{I}{p_y} - 2\frac{p_x}{p_y} x = 0$$
$$x^* = \frac{1}{2}\frac{I}{p_x}$$

Of course, y^* could be recovered by substituting back x^*.

$$y^* = \frac{1}{2}\frac{I}{p_y}$$

We need to check our SOSC, and so differentiating $\tilde{u}(x)$ once again gives

$$\tilde{u}''(x) = -2\frac{p_x}{p_y} < 0$$

and therefore the solution to the FONC is indeed a maximizer. Since it is globally satisfied, the a unique global maximizer.

Now we might be interested in asking some comparative statics questions. Notice that we have three exogenous variables here: prices and income. We are interested in how x^* changes as we change each of these exogenous variables. Of course, we could just take derivatives of our closed form solution above (it would be easy). For sake of illustration, let's use the IFT.

$$\begin{aligned}\frac{\partial x^*}{\partial p_y} &= -\frac{-I/p_y^2 + 2p_x/p_y^2 x}{\tilde{u}''(x)} \\ &= -\frac{-I/p_y^2 + 2p_x/p_y^2(\frac{1}{2}\frac{I}{p_x})}{\tilde{u}''(x)} \\ &= -\frac{-0}{\tilde{u}''(x)} = 0.\end{aligned}$$

so which is as expected, as our explicit solution for x does not depend on p_y at all.

Value Function and the Envelope Theorem

In various applications, we are interested in how does the maximized value of an objective function change whne we change some exogenous parameters.

Suppose we have some maximization problem

$$\max_{x \in D} f(x, \alpha)$$

where alpha is some exogenous parameter. Now, assume that we have found some solution, $x(\alpha)$ to the maximization problem such that we can define the value function as

$$V(\alpha) = f(x(\alpha), \alpha)$$

and if we want to know how does this maximized value of the objective change as we change alpha, then we just have to take the derivative of $V'(\alpha)$. Let's use a similar idea to the IFT.

$$V'(\alpha) = f_x(x(\alpha), \alpha)x'(\alpha) + f_\alpha(x(\alpha), \alpha).$$

However, if x is a maximizer, then it should satisfy the FONC, i.e. $f_x(x(\alpha), \alpha) = 0$. Then

$$V'(\alpha) = f_\alpha(x(\alpha), \alpha).$$

What does this tell us? The change in the optimized value only comes from the indirect effect. Intuitively, if we change alpha, then x also readjusts so that the change through $x(\alpha)$ is irrelevant. This result is known as the **envelope theorem.**

Example 9. Let's continue with the consumer problem from before. Suppose that we would want to know how maximized utility changes as we change the price of x. The value function is

$$V(p_x, p_y, I) = x^*(\frac{I}{p_y} - \frac{p_x}{p_y}x).$$

If we use the ET, then the answer to the above is:

$$\frac{\partial V}{\partial p_x} = x^*(-\frac{1}{p_y}x^*) < 0.$$

so that the consumer dislikes higher prices (of course), and maximum utility decreases.

Concavity and Convexity in $\mathbb{R}^n$

As in $\mathbb{R}$, the concavity and convexity of functions depend on the second order derivatives.

Theorem 4. *Let $f : X \to \mathbb{R}$ be a twice continuously differentiable function and $X \subseteq \mathbb{R}^n$ be an open and convex set. Then f is concave (convex) if and only if the Hessian matrix is negative (positive) semidefinite for any $x \in X$. Moreover, if the Hessian is negative (positive) definite, then f is strictly concave (convex).*

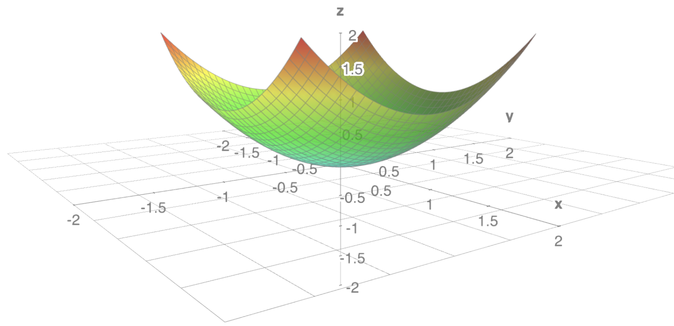
Notice that the second part of the theorem is not an if and only if statement, but just gives the sufficient condition for strict concavity of f. Recall that negative definiteness or semidefiniteness of matrices can be checked using appropriate principal minors of the Hessian. Thus, for strict concavity of a function f, we have to check the following sign pattern:

$$(-1)^r \begin{vmatrix} f_{11} & f_{12} & \dots & f_{1r} \\ f_{21} & f_{22} & \dots & f_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ f_{r1} & f_{r2} & \dots & f_{rr} \end{vmatrix} > 0 \quad for \quad r = 1, 2, ...n$$

i.e., we would need the determinants of of the leading principal minors to alternate in sign, with the first minor $|f_{11}| < 0$. For strict convexity, we can check for the following sign pattern:

$$\begin{vmatrix} f_{11} & f_{12} & \dots & f_{1r} \\ f_{21} & f_{22} & \dots & f_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ f_{r1} & f_{r2} & \dots & f_{rr} \end{vmatrix} > 0 \quad for \quad r = 1, 2, ...n$$

As we discussed earlier, in the case of $\mathbb{R}$, the second order derivative can be used to check whether some function looks like x^2 or $-x^2$ - the idea here is the same. The above theorem just checks whether the function looks like $x^2 + y^2$ or $-x^2 - y^2$. Lastly, I want to emphasize



that a strictly concave function does not have to follow the sign pattern above - however, if we find that the function does follow the sign pattern, then we know it is strictly concave.

Example 10. Let $f(x_1, x_2) = x_1^2 + x_2^2$. Then

$$f_1(x_1, x_2) = 2x_1, f_2(x_1, x_2) = 2x_2$$

and

$$f_{11}(x_1, x_2) = 2, f_{12}(x_1, x_2) = 0 = f_{21}(x_1, x_2), f_{22}(x_1, x_2) = 2.$$

The Hessian is given by

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Then the first LPM, M_1 is just $f_{11}(x_1, x_2) = 2 > 0$. The second LPM, M_2, is $f_{11}f_{22} - f_{12}^2 = 2^2 - 0 = 4 > 0$. Therefore, the Hessian is positive definite, and f is strictly convex. In fact, it is also globally strictly convex, since this sign pattern holds for all values of x_1, x_2. Here is the graph of this function.

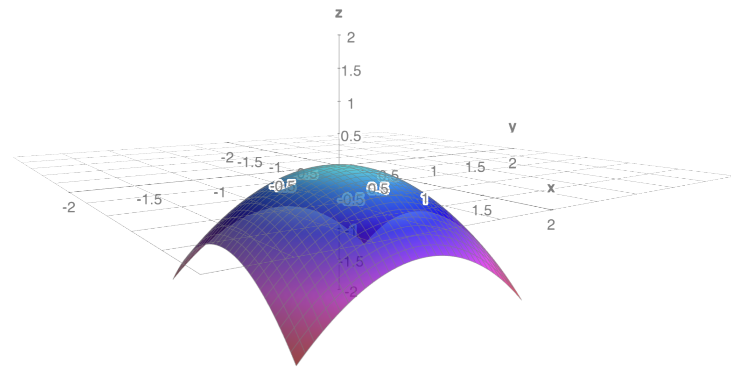
Example 11. If $f(x_1, x_2) = -x_1^2 - x_2^2$, then the Hessian is

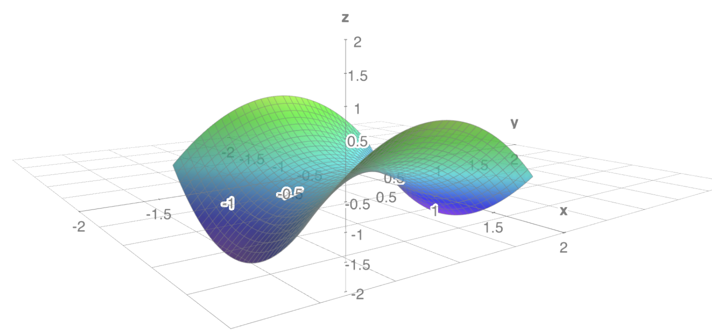
$$\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

and LPMs are

$$M_1 = -2 < 0, M_2 = 4 > 0$$

and thus f is globally strictly concave, since this holds for all x. Here is the graph of this





function.

Example 12. If $f(x_1, x_2) = x_1^2 - x_2^2$, then the Hessian is

$$\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

and LPMs are

$$M_1 = 2 > 0, M_2 = -4 < 0$$

and thus H is indefinite. Therefore, f is neither concave, nor convex. Here is the graph of this function.

Notice that among one dimension, it looks like x^2, but along the other dimension it looks like $-x^2$.

We can also equivalently characterize concave (convex) functions in terms of first derivatives, as we have discussed in the case of functions of a single variable.

Theorem 5. *Let $f : X\mathbb{R}$ be a differentiable function and $X \subseteq \mathbb{R}^n$ be an open and convex set, then f is concave if and only if for all $x, y \in X$,*

$$f(y) - f(x) \leq Df(x)(y - x),$$

that is,

$$f(y) - f(x) \leq \frac{\partial f(x)}{\partial x_1}(y_1 - x_1) + ... + \frac{\partial f(x)}{\partial x_n}(y_n - x_n).$$

Similarly, f is convex if and only if the above inequality is flipped.

The geometric interpretation is similar to one we discussed in the case of univariate functions. For example, if we rewrite the condition for concavity as $f(y) \leq f(x) + Df(x)(y - x)$, then the righthand side corresponds to a tangent hyperplane and the inequality just says that the graph of the function is below the tangent hyperplane.

Example 13. We already saw that the function $f = x_1^2 + x_2^2$ is strictly convex since the Hessian passed the sign test. Let us verify that the same conclusion holds when we use the above theorem. We need to verify that:

$$\begin{aligned} y_1^2 + y_2^2 - x_1^2 - x_2^2 \geq & \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \cdot \begin{pmatrix} y_1 - x_1 \\ y_2 - x_2 \end{pmatrix} \\ &= 2x_1(y_1 - x_1) + 2x_2(y_2 - x_2) \\ &= 2x_1y_1 - 2x_1^2 + 2x_2y_2 - 2x_2^2. \end{aligned}$$

Now rearrange to get

$$y_1^2 + y_2^2 + x_1^2 + x_2^2 - 2x_1y_1 - 2x_2y_2 \geq 0.$$

We can complete the squares to get

$$(y_1 - x_1)^2 + (y_2 - x_2)^2 \geq 0$$

which is always true.

We can add one last definition for concavity.

Definition 6. f is concave $\iff M_f = \{(x, y) : x \in S \text{ and } y \leq f(x)\}$ is convex.
f is convex $\iff N_f = \{(x, y) : x \in S \text{ and } y \geq f(x)\}$ is convex.

Finally, it is useful to remember that the sum of concave (convex) functions is concave (convex). This is especially useful in the multivariate case.

Example 14. Suppose that $f(x) = \sum_{-=1}^{n} ln(x_i)$. We want to check whether f is concave or convex. A simple argument would be to note that $ln(x)$ is a (strictly) concave function, so f is also (strictly) concave (since it is the sum of (strictly) concave functions).

Here is a crazy approach that will work in this case, but would generally not recommend. Let us start with 2:

$$H_2 f(x) = \begin{pmatrix} \frac{-1}{x_1^2} & 0 \\ 0 & \frac{-1}{x_2^2} \end{pmatrix}$$

Then the LPMs are

$$M_1 = \frac{-1}{x_1^2} < 0, M_2 = \frac{1}{x_1^2 x_2^2} > 0.$$

Then for n=2, f is concave.
Now consider n=3,

$$H_2 f(x) = \begin{pmatrix} \frac{-1}{x_1^2} & 0 & 0 \\ 0 & \frac{-1}{x_2^2} & 0 \\ 0 & 0 & \frac{-1}{x_3^2} \end{pmatrix}$$

Then the LPMs are

$$M_1 = \frac{-1}{x_1^2} < 0, M_2 = \frac{-1}{x_1^2 x_2^2 x_3^2} < 0.$$

so for n=3, f is concave. A pattern emerges, and f will be concave for any n.

The following are some useful properties of convex functions

Theorem 6. *Let f and g be functions defined over a convex set S in $\mathbb{R}^n$. Then*

- *f and g concave and $a \geq 0, b \geq 0 \implies af + bg$ concave.*
- *f and g convex and $a \geq 0, b \geq 0 \implies af + bg$ convex.*
- *$f(x)$ concave and F(u) concave and increasing $\implies$ U(x)=F(f(x)) concave.*
- *$f(x)$ convex and F(u) convex and increasing $\implies$ U(x)=F(f(x)) convex.*
- *f and g concave $\implies h(x) = min\{f(x), g(x)\}$ is concave.*
- *f and g convex $\implies h(x) = max\{f(x), g(x)\}$ is convex.*

Quasi-concavity and Quasi-convexity

The generalization of concavity (convexity) is quasi-concavity (quasi-convexity). The concepts are extremely helpful in optimization theory, especially in the presence of constraints as you will soon see.

Definition 7. The function f, defined over a convex set $S \subset \mathbb{R}^n$, is quasi-concave if the upper level set $L_a^+ = \{x \in S : f(x) \geq a\}$ is convex for each number a. That is, if

$$f(x) \geq a, \text{ and } f(x') \geq a \implies f(\alpha x + (1-\alpha)x') \geq a$$

for any $t \in \mathbb{R}, x, x' \in S, \text{ and } \alpha \in [0,1]$. If the inequality is strict for $x \neq x'$ and $\alpha \in (0,1)$, then we say that f is strictly quasi-concave.

We will say that f is quasi-convex if $-f$ is quasi-concave. Therefore,

Definition 8. The function f, defined over a convex set $S \subset \mathbb{R}^n$, is quasi-convex if the lower level set $L_a^- = \{x \in S : f(x) \leq a\}$ is convex for each number a.

The following theorem gives the connection between concave (convex) and quasi-concave functions.

Theorem 7. *If a function $f(x)$ is concave (convex), then it is quasi-concave (quasi-convex).*

The converse of the above is not true - there are quasi-concave functions that are not concave. Some useful properties of quasi-concave functions are these:

Theorem 8.

1. *A sum of quasi-concave (quasi-convex) functions is not necessarily quasi-concave (quasi-convex).*
2. *If $f(x)$ is quasi-concave (quasi-convex) and F is strictly increasing, then $F(f(x))$ is quasi-concave (quasi-convex).*
3. *If f(x) is quasi-concave (quasi-convex) and F is strictly decreasing, the F(f(x)) is quasi-convex (quasi-concave).*

Example 15. Suppose a function $f(x,y) = x^\alpha y^\beta$, with $\alpha, \beta > 0$ is defined for all $x, y > 0$.. Then the function is quasi-concave, since $ln(f(x)) = \alpha lnx + \beta lny$ is concave as a sum of concave functions. We could take the strictly increasing transformation $e^{ln(f(x))}$ of a quasi-concave function, and thus the original function $f(x)$ is quasi-concave. Notice that if we were to compute the Hessian of f, we would have

$$H = \begin{bmatrix} \alpha(\alpha-1)x^{\alpha-2}y^{\beta} & \alpha\beta x^{\alpha-1}y^{\beta-1} \\ \alpha\beta x^{\alpha-1}y^{\beta-1} & \beta(\beta-1)x^{\alpha}y^{\beta-1} \end{bmatrix}$$
$$= x^{\alpha-1}y^{\beta-1}\begin{bmatrix} \alpha(\alpha-1)\frac{y}{x} & \alpha\beta \\ \alpha\beta & \beta(\beta-1)\frac{x}{y} \end{bmatrix}$$

Then $f(x,y)$ is concave iff H is negative definite. Then the sign pattern of LPMs should be $-, +$, thus we need

$$M_1 = \alpha(\alpha-1)\frac{y}{x} < 0$$
$$M_2 = \alpha(\alpha-1)\frac{y}{x}\beta(\beta-1)\frac{x}{y} - (\alpha\beta)^2 > 0$$

The first inequality implies that $\alpha<1$. The second simplifies to

$$\alpha\beta[1-\alpha-\beta>0]$$

implying that $\alpha+\beta<1$. Thus, if $\alpha,\beta\in(0,1)$ and $\alpha+\beta<1$, then we know that the function is strictly concave[5]. We can tell pretty easily that if the sum is greater than 1, then the second LPM will be negative: in which case, the function is neither convex nor concave. However, the function is quasi-concave for all $\alpha,\beta>0$. Recall that we showed the function was quasi-concave in the beginning without putting any requirements on the parameters.

We could also define quasi-concavity (quasi-convexity) in a couple more ways.

Theorem 9. *Let f be a function of n variables defined over a convex set $S\in\mathbb{R}^n$. Then f is quasi-concave iff, for all x, x_0, and all $\lambda\in[0,1]$,*

$$f(x)\geq f(x_0) \implies f((1-\lambda)x+\lambda x_0)\geq f(x_0)$$

Theorem 10. *Let a function f of n variables defined over an open, convex set $S\in\mathbb{R}^n$. Then f is quasi-concave iff*

$$f(\alpha x+(1-\alpha)x')\geq Min\{f(x), f(x)'\}$$

for all $x, x'\in A$, and $\alpha\in[0,1]$.

The last point that I want to make here, is that quasi-concavity is weaker in a sense than concavity. While concavity is a cardinal property, which may not be preserved under increas-

[5]Though this is only the sufficient condition, it is true here. Moreover, since the Hessian is negative-semidefinite at $\alpha+\beta=1$, then it is concave here.

ing transformations, quasi-concavity is an ordered property: so it will be preserved.

This section will end by giving a criterion for checking the quasi-concavity of a function by examining the signs of certain determinants, in something called the Bordered Hessian. We call it the "bordered" Hessian, as the normal Hessian is bordered by an extra row and column of the first order partials of the function.

Theorem 11 (Bordered Hessian)**.** *Let f be a C^2 function defined on an open, convex set $S \in \mathbb{R}^n$. Define the bordered Hessian determinants D_r, r=1,...,n by*

$$D_r(x) = \begin{vmatrix} 0 & f_1'(x) & \dots f_r'(x) & \\ f_1'(x) & f_{11}''(x) & \dots f_{1r}''(x) & \\ \vdots & \vdots & \ddots & \vdots \\ f_r'(x) & f_{r1}''(x) & \dots & f_{rr}''(x) \end{vmatrix}$$

1. *A necessary condition for f to be quasi-concave is that $(-1)^r D_r(x) \geq$ for r=1,...,n, and all $x \in S$.*
2. *A sufficient condition for f to be quasi-concave is that $(-1)^r D_r(x) >$ for r=1,...,n, and all $x \in S$.*

Unconstrained Optimization

The FONCs for unconstrained optimization problems in $\mathbb{R}^n$ are simple genearlizations of the FONCs that we saw with one variable. Recall that the derivative in $\mathbb{R}$ is genealized by a vector of partial derivatives called the gradient of a function. Then FONCs are a system of equations that basically says that no deviation along any dimension should change the objective function.

Theorem 12 (Necessary First-Order Conditions)**.** *Let f be defined in a set S in $\mathbb{R}^n$ and let $c = (c_1, ..., c_n)$ be an interior point in S at which f is differentiable. A ncessary condition for c to be a maximum or minimum point for f is that c is a stationary point of f - that is*

$$f_i'(c) = 0 \quad (i = 1, ..., n)$$

If we are maximizing $f(x_1, ..., x_n)$ then FONCs are

$$\frac{\partial f(x)}{\partial x_1} = 0$$
$$\vdots \frac{\partial f(x)}{\partial x_n} = 0$$

or in vector notation,

$$\nabla_x f(x) = 0.$$

This will give us n equations and n unknowns to solve for. Geometrically, FONCs now mean that at the peak (or valley) where a maximum (or minimum) occurs, the tangent hyperplane should be horizontal.
Similarly, SOSCs convey the same idea as in the case of $\mathbb{R}$. We need our objective function to look like a hill or a valley.

Theorem 13 (Second Order Sufficent Conditions for Local Optima). *Let $f : D \to \mathbb{R}$ be a C^2 function on D, and open and convex subset of $\mathbb{R}$. Let $x^* \in D$ be a point such that $\nabla_x f(x^*) = 0$. If the Hessian matrix of f at X^* is negative (positive) definite, then f achieves a strict local maximum (minimum) at x^*.*

Just like in $\mathbb{R}$, we could define second order necessary conditions, that would be "weaker" than the sufficient conditions. We could get rid of the strict result and replace the negative (positive) definite conditions in the sufficient conditions above, with negative (positive) semi-definite conditions.

So far, these are only conditions for local optima. If we allow these results to hold for all x, then we can get conditions for global maxima and minima.

Theorem 14. *Let $F : U \to \mathbb{R}$ be a C^2 function whose domain is a convex open subset $U \in \mathbb{R}^n$.*

- *The following conditions are equivalent*
  1. *F is a concave function on U; and*
  2. *$F(y) - F(x) \leq DF(x)(y - x) \forall x, y \in U$; and*
  3. *$D^2F(x)$ is negative semi-definite for all $x \in U$.*
- *The following conditions are equivalent*
  1. *F is a convex function on U; and*

2. *$F(y) - F(x) \geq DF(x)(y-x) \forall x, y \in U$; and*

3. *$D^2F(x)$ is positive semi-definite for all $x \in U$.*

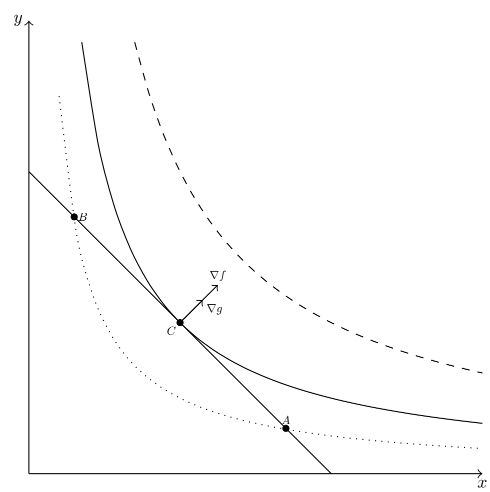
- *If F is a concave function on U and $DF(x^*)=0$ for some $x^* \in U$, then x^* is a global max in U.*

- *If F is a convex function on U and $DF(x^*)=0$ for some $x^* \in U$, then x^* is a global min in U.*

Dealing With Equality Constraints

Let's start with a simple example. Suppose we want to maximize $f(x, y) = xy$ subject to the constraint set $g(x, y) = 1$, where $g(x, y) = x + y$. What does this mean? We should find values of x, y such that at the same time that $x + y = 1$, we try and make xy as high as possible. Let's start with some geometric intuition.

First, notice that the level curves of f are just hyperbolas, since $xy = c \iff y = c/x$. The constraint is just a straight line joining points (0,1) and (1,0). Below is the figure illustrating this.



Now, if we take the dotted level curve, then it crosses the constraint at two points, A and B. These solutions are feasible, but not optimal since we could move from point A along the constraint and achieve a higher level of f (similar for B). If the constraint passes through the level curve, it can't be a maximizer - thus, the level curve must be tangent to our constraint. Continuing with this logic, we arrive at point C. Note that at C, we cannot increase the objective function without violating the constraint, and therefore, point C should be the maximizer. In terms of optimally, going higher to the dashed level curve would give us a higher value of the objective function, but would not be feasible since it would violate the constraint.

Now thinking about this point C, it is characterized by two conditions. The first is that it is feasible, i.e. it lies within the constraint set. The second, is that the gradient of the objective function, ∇f and the gradient of the constraint are co-linear. The slope of the tangent to the level curve, is equal to the slope of the tangent to the constraint at this point. So we can write our two conditions that characterize C:

$$
\begin{aligned}
g(x, y) &= 1 \\
\nabla f(x, y) &= \lambda \nabla g(x, y)
\end{aligned}
$$

where λ is some scalar we have to find. Because the gradient gives some magnitude as well as direction, we have to account for the magnitude being able to scale. We know that the two gradients are colinear, so there must be some λ that could scale the gradient of the constraint such that it equals the gradient of the function. However, notice that for the above idea to work, we need $\nabla g(x^*, y^*) \neq 0$, or the above does not work! This is called a constraint qualification - we will need to account for this later. For now, let's solve the problem using the equations we just wrote.

$$
\begin{aligned}
f_x(x, y) &= \lambda g_x(x, y) \\
f_y(x, y) &= \lambda g_y(x, y) \\
g(x, y) &= 1
\end{aligned}
$$

using the functional form from the problem above,

$$
\begin{aligned}
y &= \lambda \\
x &= \lambda \\
x + y &= 1
\end{aligned}
$$

so $x = \lambda = y$, substituting $x = y$ into the constraint, we get that

$$2y = 1.$$

Therefore, $y^* = 1/2, x^* = 1/2, \lambda^* = 1/2$. This is the geometric idea behind Lagrange's method of solving constrained optimization problems.

In particular, for a general maximization problem,

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

we write a new function called the Lagrangian as

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

the auxiliary variable λ is called the Lagrange multiplier, and is some constant. Then, we can take FONCs with respect to x, y, λ to get

$$\begin{aligned}
&\langle x\rangle : f_x(x,y)-\lambda g_x(x,y)=0\\
&\langle y\rangle : f_y(x,y)-\lambda g_y(x,y)=0\\
&\langle \lambda\rangle : -[g(x,y-c]=0
\end{aligned}$$

Notice that the first order conditions for x, y just reconstruct the condition $\nabla f(x,y) = \lambda\nabla g(x,y)$, and the first order condition for λ just reconstructs the constraint $g(x,y) = c$. So, what the Lagrangian does is converts a constrained optimization problem into an unconstrained one. Intuitively, Lagrange's multiplier attaches a cost of λ per unit of violation of the constraint.

Theorem 15. *Let f and h be c^1functions of two variables. Suppose that $x^* = (x_1^*, x_2^*)$ is a solution of the problem*

$$maximize \quad f(x_1,x_2)$$

$$subject\ to \quad h(x_1,x_2)=c$$

Suppose further that (x_1^*, x_2^*) is not a critical point of h. Then there is a real number λ^*such that $(x_1^*, x_2^*, \lambda^*)$ is a critical point of the Lagrangian function

$$\mathcal{L}\left(x_1,x_2,\lambda\right)=f\left(x_1,x_2\right)-\lambda\left[g\left(x_1,x_2\right)-c\right].$$

In other words, at $(x_1^*, x_2^*, \lambda^*)$,

$$\frac{\partial L}{\partial x_1}=0, \quad \frac{\partial L}{\partial x_2}=0, \quad \frac{\partial L}{\partial \lambda}=0$$

You may notice that there is a small qualification that the point is not a critical point of the constraint. This is the constraint qualification - meaning that the gradient of the constraint is not 0. We need to consider these points in addition.

We can use the following steps to apply the theorem:

- First, check the constraint qualification by calculating the critical points of the constraint function h
- If none of these lie in the constraint, we can proceed. If they do, then we add these points to potential candidates for optima, along with the critical points of the Lagrangian
- Write out the Lagrangian, set its partial derivatives equal to zero, and solve the resulting system.
- Evaluate optima to find the maximizing solution

If we add multiple equality constraints, we can generally use the same process, but include a different $\lambda_j$for each constraint h_j. However, we need some method to calculate the constraint qualification of multiple constraints. With $j > 1$ constraints, this generalization involves the Jacobian derivative of the constraint functions:

$$Dh(x^*) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x^*) & \cdots & \frac{\partial h_1}{\partial x_n}(x^*) \\ \frac{\partial h_2}{\partial x_1}(x^*) & \cdots & \frac{\partial h_2}{\partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_j}{\partial x_1}(x^*) & \cdots & \frac{\partial h_j}{\partial x_n}(x^*) \end{bmatrix}$$

In general, a point $x^*$is called a critical point of $h = (h_1, ..., h_j)$if the rank of the matrix $Dh(x^*)$is $<$ j. Formally, we say that the **nondegenerate constraint qualification (NDCQ)** at x^* is satisfied if the rank of the Jacobian matrix at $x^*$is j.

What about second order conditions? Recall the second order conditions for unconstrained maximization

- At a maximum f(x^*), D$f(x^*)$ must be zero, and $D^2f(x^*)$must be negative semidefinite (necessary conditions)
- To guarantee that a point $x^*$is a local maximizer, we need $Df(x^*) = 0$ and $D^2f(x^*)$ be negative definite (sufficient conditions).

Our main focus will be on the sufficient conditions here, which will be similar. They will once again be related to the definiteness of a quadratic form, but restricted to a linear subspace.

Intuitively, the second order condition for a constrained maximization problem:

- should involve the negative definiteness of some Hessian matrix, but

- should only be concerned with directions along the constraint set

Theorem 16. *Let f and h be C^2 functions on R^2. Consider the problem of maximizing f on the constraint set $C_h = \{(x,y) : h(x,y) = c\}$. Form the Lagrangian*

$L(x,y,\lambda)= f(x,y) - \lambda(h(x,y) - c)..$

Suppose that (x^*,y^*,λ^*)satisfies at (x^*,y^*,λ^*):

- $\frac{\partial L}{\partial x} = 0$
- $\frac{\partial L}{\partial y} = 0$
- $\frac{\partial L}{\partial \lambda} = 0$

and at (x^*,y^*,λ^*)

- *det* $\begin{bmatrix} 0 & \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial^2 L}{\partial x^2} & \frac{\partial L}{\partial x \partial y} \\ \frac{\partial h}{\partial y} & \frac{\partial L}{\partial y \partial x} & \frac{\partial L}{\partial y^2} \end{bmatrix} > 0$

Then, (x^*,y^*) is a local max of f on C_h.

The last part is a Bordered Hessian. For constrained optimization problems, we can write the bordered Hessian that characterizes the concavity or convexity of the Lagrangian[6].

$$\tilde{H} = \begin{pmatrix} 0 & g_x & g_y \\ g_x & f_{xx} & f_{xy} \\ g_y & f_{yx} & f_{yy} \end{pmatrix}$$

or written more generally in blocks:

$$\tilde{H} = \begin{pmatrix} 0 & \nabla g' \\ \nabla g & Hf \end{pmatrix}$$

This method generalizes to arbitrary number of variables and multiple constraints. If we want to maximize $f(x)$ subject to $g^j(x) = c_j, j = 1,...,m$ (here x=$(x_1,...,x_n)$ so we have n variables to choose from, and m constraints to satisfy), then we construct the Lagrangian:

$$\S, \lambda = f(x) - \sum_{j=1}^{m} \lambda_j [g^j(x) - c_j]$$

[6]Now that we are dealing with constraints, what we really need is for the objective function to be concave on the constrained set.

and the FONCs will be

$$\langle x_i \rangle : \frac{\partial f(x)}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial g^i}{\partial x_i} = 0$$
$$\langle \lambda \rangle : -[g^j(x) - c_j] = 0$$

Example 16. Suppose we want to maximize $f(x_1, x_2) = -(x_1)^2 + (x_2)^2$ subject to $g(x_1, x_2) = x_1 + x_2 = c$. Start with the Lagrangian:

$$\mathcal{L} = (x_1)^2 + (x_2)^2 - \lambda(x_1 + x_2 - c)$$

Then the FONCs are

$$\langle x_1 \rangle : 2x_1 - \lambda = 0$$
$$\langle x_2 \rangle : 2x_2 - \lambda = 0$$
$$\langle \lambda \rangle : -[x_1 + x_2 - c] = 0$$

From the first two equations, we have that $2x_1 = \lambda = 2x_2$, with $x_1 = x_2$, and from the constraint we have that $x_1^* = c/2 = x_2^*$. So there is a unique solution to the FONCs. Now we need to check whether this solution really is a minimizer. The bordered Hessian is

$$\tilde{H} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

For a candidate to be a solution, we need the full matrix to have a negative determinant. Since $|\tilde{H}| = -4 < 0$ everywhere, then x^* is indeed a maximizer, and a global maximizer.

Comparative Statics, the Value Function, and the Envelope Theorem

Comparative statics for constrained optimization problems are not that different from unconstrained ones. The idea is to apply the IFT to FONCS of the Lagrangian. The only caveat in this case, is that both x, λ are endogenous variables. For example, in the case of maximizing $f(x_1, x_2, \alpha)$ subject to $g(x_1, x_2, \alpha) = c$, we have the following set of FONCs at the solution

$(x_1(\alpha), (x_2(\alpha), \lambda(\alpha)) is$

$$\begin{aligned}
\langle\lambda\rangle : -[g(x_1(\alpha), x_2(\alpha), \alpha) - c] = 0 \\
\langle x_1\rangle : f_1(x_1(\alpha), x_2(\alpha), \alpha) - \lambda(\alpha) g_1(x_1(\alpha), x_2(\alpha), \alpha) = 0 \\
\langle x_2\rangle : f_2(x_1(\alpha), x_2(\alpha), \alpha) - \lambda(\alpha) g_2(x_1(\alpha), x_2(\alpha), \alpha) = 0
\end{aligned}$$

then, if we differentiate all equations with respect to α we will get

$$\begin{pmatrix} 0 & -g_1 & -g_2 \\ -g_1 & f_{11} & f_{12} \\ -g_2 & f_{21} & f_{22} \end{pmatrix} \begin{pmatrix} \lambda'(\alpha) \\ x_1'(\alpha) \\ x_2'(\alpha) \end{pmatrix} = - \begin{pmatrix} -g_\alpha \\ f_{1\alpha} - \lambda g_{1\alpha} \\ f_{2\alpha} - \lambda g_{2\alpha} \end{pmatrix}$$

Notice that the matrix the pre-multiplies the vector of comparative statics quantities is almost the bordered Hessian, however the determinant of this matrix, and the bordered Hessian matrix is exactly the same. So what do we do with this system? Previously, we use the IFT to determine how the maximizer changes in response to exogenous variables. In multi-dimensions, the IFT can be hard to work with. However, we can solve the above system using Cramer's rule. For example,

$$x_1'(\alpha) = |\tilde{H}|^{-1} \begin{vmatrix} 0 & g_\alpha & -g_2 \\ -g_1 & -f_{1\alpha} + \lambda g_{1\alpha} & f_{12} \\ -g_2 & -f_{2\alpha} + \lambda g_{2\alpha} & f_{22} \end{vmatrix}$$

We are mostly interest in the sign of derivatives, therefore we do not need to compute $|\tilde{H}|^{-1}$ explicitly, we just need to sign it. For maximization (minimization) $|\tilde{H}|^{-1}$ should be positive (negative), and therefore the sign of comparative statics would just depend on the quantity

$$\begin{vmatrix} 0 & g_\alpha & -g_2 \\ -g_1 & -f_{1\alpha} + \lambda g_{1\alpha} & f_{12} \\ -g_2 & -f_{2\alpha} + \lambda g_{2\alpha} & f_{22} \end{vmatrix} = -g_\alpha(-g_1 f_{22} + g_2 f_{12}) - g_2(-g_1[-f_{2\alpha} + \lambda g_{2\alpha}] + g_2[-f_{1\alpha + \lambda g_{1\alpha}}]).$$

Often, we will have have assumptions that characterize our functions, even if we do not have a functional form for them. For example, we might require that $f_2 2 < 0$. Then using these assumption, we are often able to give a sign to $x_1'(\alpha)$, how x_1 changes with respect to α. As an example, we might be interested in how on of the *endogenous variables*, a commodity good x, changes with respect to one of the *exogenous variables,* price.

I do want to caution you, however. If you are tasked to find how $\frac{\partial x}{\partial y}$, you do not necessarily need to totally differentiate the system and use Cramer's rule for the answer. If you have solved for a function form of $x^*(\cdot)$, then you can just differentiate x^* in terms of the variable of interest.

Now, we turn to the value function and the envelope theorem. Notice that the way we constructed the Lagrangian function, at the optimum, it exactly equals to the value of the objective function:

$$\begin{aligned}\mathcal{L}(x^*, y^*, \lambda^*) &= f(x^*, y^*) - \lambda^*[g(x^*, y^*) - c] \\ &= f(x^*, y^*)\end{aligned}$$

since at the solution, the constraint should be satisfied, $g(x^*, y^*) - c = 0$.
The envelope theorem that we discussed earlier also applies to constrained optimization problems. For example, if we want to know how does the maximized value of the object function change when we increase c (this is what we call a relaxation of the constraint), then we would apply the envelope theorem.
First, note that since $\mathcal{L}^* = f^*$, we will have

$$\frac{\partial f^*}{\partial c} = \frac{\partial \mathcal{L}(x^*, y^*, \lambda^*)}{\partial c}.$$

According to this result, the total effect on the value of $f^*(\cdot)$ of a small change in c is found by simply differentiating the Lagrangian partially with respect to c, treating the $x's, y's, \lambda's$ as constants. This is the (general) **envelope theorem.**

Note that in our above example, since we chose to look at the change in c, we have that

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial c} &= \frac{\partial}{\partial c}(f(x^*, y^*) - \lambda^*[g(x^*, y^*) - c]) \\ &= -\lambda^*(\frac{-\partial}{\partial c}(c)) \\ &= \lambda\end{aligned}$$

and therefore,

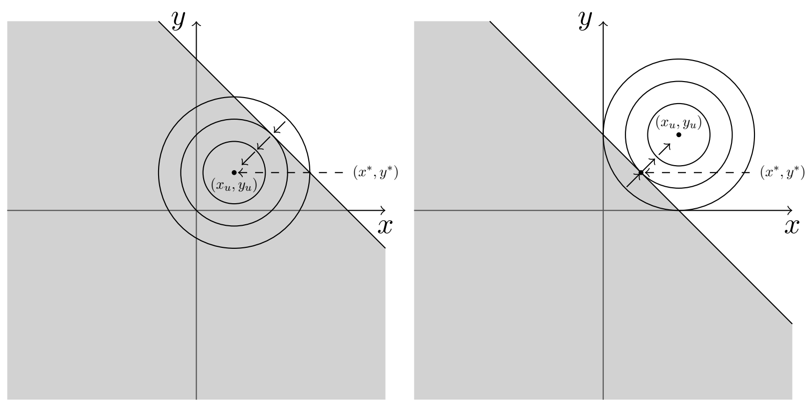
$$\frac{\partial f^*}{\partial c} = \frac{\partial \mathcal{L}^*}{\partial c} = \lambda^*.$$

So what does this say? It gives an important interpretation of λ^*, as it measures the change in the objective function as we relax the constraint. It measures the change in value as the constraint relaxes by a unit - economists call λ the shadow price.

Introducing Inequalities

Most of the problems in economics are written with inequalities, so it is important to know how to deal with optimization problems that use them.

Suppose we have some $f(x, y)$ subject to a constraint $g(x, y) \leq c$. Assume that the function attains its unconstrained (if we ignored the constraint) maximum at (x_0, y_0). Now, does the constraint matter if $g(x_0, y_0) < c$? Since f attains its unconstrained maximum at a point where the constraint is satisfied with strict inequality, then the constraint does not really matter, as we could just max f and ignore the constraint, but get the same answer. In such cases, we say the constraint is slack, or does not bind. Below are figures that illustrate such a case.



The first case is on the left, while the second is on the right. The shaded area is feasible, and the arrows indicate which way f is increasing.

Now that we have a little geometric intuition, how would we approach such a problem? Recall our discussion of the envelope theorem, where we found that $\frac{\partial f}{\partial c} = \lambda$: in the first case, the λ is just zero, right? Since the constraint does not bind, then if we relax the constraint a bit ($\frac{\partial f}{\partial c}$), the value of our function does not change at all. When the constraint is binding, f^* will respond to changes in c.

The bottom line is that it will either be the case that $\lambda = 0$ and we are in the world of unconstrained optimization, or $\lambda > 0$ and the constraint is binding such that $g(x, y) = c$. A convenient way of putting the above into a tractable framework is by writing down the following system with **Kuhn-Tucker** complementary slackness conditions. We write the Lagrangian with the inequality of the form $g(x, y) \leq c$:

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

FONCs are given by

$$\begin{aligned}
\langle x \rangle : f_x(x,y) - \lambda g_x(x,y) = 0 \\
\langle y \rangle : f_y(x,y) - \lambda g_y(x,y) = 0 \\
\langle PF \rangle : g_x(x,y) \leq c \\
\langle DF \rangle : \lambda \geq 0 \\
\langle CS\lambda[g(x,y) - c] = 0.
\end{aligned}$$

The first two conditions are the same: our first order derivatives and our constraint. The last three (called primal feasibility, dual feasibility, and complementary slackness) effectively describe the intuition that we developed above. Importantly though, we added one final condition for λ in the fourth equation - it will either be 0 or positive, but never negative. The reason for this is simple - it now matters which way the gradient of the constraint is pointing. Recall that the gradient points to the way that it is increasing most - meaning that ∇g points to the set $g(x,y) \geq c$, not $g(x,y) \leq c$. Moreover, it has to be the case that ∇f also points outside the set. Otherwise, we could increase f, and still be inside the constraint set! Therefore, it must be that the gradients of f, g point in the same direction, and thus $\lambda \geq 0$.

Theorem 17. *Suppose that f and g are C^1functions on R^2 and that (x^*, y^*) maximizes f on the constraint set $g(x,y) \leq b$. If $g(x^*, y^*) = b$,suppose that:*

$$\frac{\partial g}{\partial x}(x^*, y^*) \neq 0 \quad or \quad \frac{\partial g}{\partial y}(x^*, y^*) \neq 0$$

In any case, form the Lagrangian function

$$L(x, y, \lambda) = f(x,y) - \lambda[g(x,y) - b].$$

Then there is a multiplier λ^*such that:

- $\frac{\partial L}{\partial x}(x^*, y^*, \lambda^*) = 0$
- $\frac{\partial L}{\partial y}(x^*, y^*, \lambda^*) = 0$
- $\lambda^*[g(x^*, y^*) - b] = 0$
- $\lambda^* \geq 0$
- $g(x^*, y^*) \leq b.$

Just like equality constraints, we can generalize our theorem to several inequality constraints.

- Similar to before, you need to check the Jacobian matrix, and ensure that its rank is as large as possible
- However, you form the matrix using only the binding constraints! If you have k total constraints, with k_0 binding, and $k - k_0$ that don't bind, then you want the rank to be of size k_0.
- Nonbinding constraints play no role in the first order conditions (they fall out of the Lagrangian, since the multiplier is set equal to zero)

Our discussion has focused on maximization thus far. Minimization requires a slightly different treatment

- If we wanted to minimze rather than maximize over the same constraint set, we could require $\lambda \leq 0$ instead.
- More commonly, we could set our constraints to be $g(x) \geq b$.
- You may see f replaced by $-f$
- Or the multipliers in the Lagrangian with plus instead of minus signs

Now, let's generalize the above thoughts.

Theorem 18. *Let f, $g_1, ..., g_m, h_1, ..., h_k$be c^2functions on R^n.Consider the problem of maximizing f on the constraint set:*

$$C_{g,h} = \{x : g_1(x) \leq b_1, ..., g_m(x) \leq b_m, h_1(x) = c_1, ..., h_k(x) = c_k\}.$$

Form the Lagrangian

$$L(x_1, ..., x_n, \lambda_1, ..., \lambda_m, \mu_1, ..., \mu_k) = f(x) - \lambda_1(g_1(x) - b_1) - ... - \mu_k(h_k(x) - c_k)$$

1. Suppose there exist $\lambda_1^*, ..., \lambda_m^*, \mu_1^*, ..., \mu_k^*$ such that the first order conditions are satisfied. That is,

$$\frac{\partial L}{\partial x_1} = 0, ..., \frac{\partial L}{\partial x_n} = 0, \quad at\ (x^*, \lambda^*, \mu^*),$$
$$\lambda_1^* \geq 0, ..., \lambda_m^* \geq 0,$$

$$\lambda_1^*(g_1(x^*) - b_1) = 0, ..., \lambda_m^*(g_m(x^*) - b_m) = 0,$$

$$h_1(x^*) = c_1, ..., h_k(x^*) = c_k.$$

2. For notation's sake, suppose that $g_1, ..., g_e$ are binding at x^* and $g_{e+1}, ..., g_m$ are not binding. Write $(g_1, ..., g_e)$ as g_E. Suppose that the Hessian of L with respect to x at (x^*, λ^*, μ^*) is negative definite on the constraint set

$$\{v : Dg_E(x^*)v = 0 \quad and \quad Dh(x^*)v = 0\}$$

Then x^* is a strict local constrained max of f on $C_{g,h}$.

To check this second condition, form the bordered Hessian:

$$\begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{\partial g_e}{\partial x_1} & \cdots & \frac{\partial g_e}{\partial x_n} \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{\partial h_k}{\partial x_1} & \cdots & \frac{\partial h_k}{\partial x_n} \\ \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_e}{\partial x_1} & \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_k}{\partial x_1} & \frac{\partial^2 L}{\partial x_1^2} & \cdots & \frac{\partial^2 L}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_n} & \cdots & \frac{\partial g_e}{\partial x_n} & \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_k}{\partial x_n} & \frac{\partial^2 L}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 L}{\partial x_n^2} \end{bmatrix}$$

- If the last n-(e+k) leading principal minors alternate in sign, with the sign of the determinant of the largest matrix the same as the sign of (-1)^n, then the condition holds.
  - This may seem unwieldy, but lets apply it to the simple theorem we had a few slides ago. There was only one constraint, which was an equality constraint, and two x variables. So n-(e+k)=2-(0+1)=1. Therefore we only had to check one principal minor. Furthermore, this principal minor had to be the same sign of (-1)$^n = (-1)^2 > 0$. Which is exactly what we saw in the theorem.
- To state the minimization problem instead,
  - Change the word maximizing to minimizing
  - write the inequality constraints as $g_i(x) \geq b_i$,
  - change "negative definite" to "positive definite"
  - change "max" to "min" in the concluding sentence.

Finally, I want to make some remarks about nonlinear programming problems.

Theorem 19 (Kuhn-Tucker Sufficient Conditions)**.** *Suppose the problem :*

$$max f(x) \quad subject\ to \quad g_j(x) \leq c_j, j = 1, ..., m$$

Suppose f concave and $g_j, ...g_m$ all convex, and continuously differentiable. If there exist multipliers for each constraint λ_j. and a feasible vector x_0 that satisifes the first order conditions and complementary slackness conditions, then x_0 solves the problem.

Example 17. Suppose we want to maximize $f(x_1, x_2) = -(x_1 - 1)^2 - (x_2 - 1)^2$ subject to $g(x_1, x_2) = x_1 + x_2 \leq 4$. Start with the Lagrangian:

$$\mathcal{L} = -(x_1 - 1)^2 - (x_2 - 1)^2 - \lambda(x_1 + x_2 - 4)$$

Then the FONCs are

$$\begin{aligned} \langle x_1 \rangle &: -2(x_1 - 1) - \lambda = 0 \\ \langle x_2 \rangle &: -2(x_2 - 1) - \lambda = 0 \\ \langle PF \rangle &: x_1 + x_2 \leq 4 \\ \langle DF \rangle &: \lambda \geq 0 \\ \langle CS\lambda(x_1 + x_2 - 4) &= 0. \end{aligned}$$

Now we need to consider two cases: when $\lambda = 0$, and when $\lambda > 0, and x_1$. If $\lambda = 0$, then $x_1 = x_2 = 1$. Since all FONCs are satisfied, we have a candidate solution. Since f is globally concave (check the Hessian), the candidate solution is indeed a maximizer.

Now assume that $\lambda > 0$, then $x_1 + x_2 = 4$. From the first two FONCs, we have that $x_1 = x_2 = 2$. Now we need to find a lambda for this to be true, but there does not exist one from the first condition (lambda would be negative).

Concavity is very restrictive, so we can edit just slightly these conditions.

Theorem 20 (Kuhn-Tucker Sufficient Conditions for Quasi-Concave Programming)**.** *Suppose the problem :*

$$max f(x) \quad subject\ to \quad g_j(x) \leq c_j, j = 1, ..., m$$

Suppose f concave and $g_j, ...g_m$ all convex, and continuously differentiable. If there exist multipliers for each constraint λ_j. and a feasible vector x_0 such that:

- *x_0 is feasible and the FONCs and complementary slackness conditions hold.*
- $(f_1'(x_0)...f_n'(x_0)) \neq 0$

- *If f(x) is quasi-concave, and* $\lambda_j g_j(x)$ *is quasi-convex*

then x_0 *solves the problem.*