Lagrangians[1]

Consider a constrained optimization problem:

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

We can write this as an analogous unconstrained problem:

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

$\mathcal{L}(x,y,\lambda)$ is the **Lagrangian** and λ is the **Lagrangian multiplier**. Now we can use our unconstrained maximization methods to solve!

In this case, with two choice variables and one constraint, our first-order conditions are:

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

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

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

And Lagrange's Theorem tells that any points (x^*, y^*, λ^*) that satisfy these FOCs are critical points of $f(x,y)$ along the constraint $c = g(x,y)$.

[1]Prepared by Sarah Robinson

More generally...

Let $f : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}^n$. If $m < n$, consider the optimization problem

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

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:

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

The Lagrange Theorem is extremely useful and will be used like this a lot during the first year. However, it has some limitations to keep in mind:

1. Notice the core structure of the theorem: If $\mathbf{x}^*$ is an interior optimum, then the Lagrangian FOCs will find it. We need to confirm whether each candidate found by the FOCs is a minimum, maximum, or neither.

2. The objective function and constraint need to be differentiable (can't use for $u(x, y) = \min\{x, y\}$).

3. It only finds interior optima (doesn't find corner solutions).

4. It only deals with equality constraints.

In essence, the Lagrangian FOCs are finding points of tangency between the objective function's level set and the constraint. It's up to us to confirm whether that tangency is a minimum or a maximum. And corner solutions won't have tangency, so can't be found using this method.

(We can consider inequality constraints and/or corner solutions with the Kuhn-Tucker conditions, which we'll discuss tomorrow).

Example: Consider the maximization problem

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

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

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

What are the FOCs?

$$\frac{\partial \mathcal{L}(\cdot)}{\partial x_1} =$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial x_2} =$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \lambda} =$$

Solving this three-equation, three-unknown system:

Our maximization problem:

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

We have one candidate point:

$$x_1^* = \frac{b}{a+b} \qquad x_2^* = \frac{a}{a+b} \qquad \lambda^* = -\frac{2ab}{a+b}$$

Notice that the objective function is concave in both x_1 and x_2 (think upside down parabolas). We could use this to show that (x_1^*, x_2^*) is a maximum.

What does λ^* tell us? It's the slope of the objective function at (x^*, y^*). This is also called the **shadow value**.

If we relaxed the constraint by one marginal unit (e.g., $2 = x_1 + x_2$), it tells us how the objective function at the solution (the value function) changes.

If $a = b = 1$, then $\lambda = -1$. If we increased the constraint by one *marginal* unit, and re-optimized, then

$$-a(x_1^*)^2 - b(x_2^*)^2$$

would decrease by one unit. (Remember that this is at the margin, so is a good approximation for small changes in the constraint. Also note that this interpretation holds because I put the constraint in as $\lambda[1 - x_1 - x_2]$.)

Illustration

Example: Solve this maximization problem using the Lagrangian.

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

We have found our solution from yesterday, plus the shadow value:

$$x^* = \frac{m}{2p_x} \qquad y^* = \frac{m}{2p_y} \qquad \lambda^* = \frac{m}{2p_x p_y}$$

Consider the solution for $p_x = 1$, $p_y = 2$, $m = 100$:

$$x^* = \frac{100}{2*1} = 50 \qquad y^* = \frac{100}{2*2} = 25 \qquad \lambda^* = \frac{100}{2*2*1} = 25$$

Our value function in this setting is:

$$V(p_x, p_y, m) = u\Big(x^*(p_x, p_y, m), y^*(p_x, p_y, m)\Big)$$

$$V(1, 2, 100) = x^* \; y^* = 50 * 25 = 1250$$

If we increased to $m' = 101$, then our solution would be:

$$x^* = \frac{101}{2*1} = 50.5 \qquad y^* = \frac{101}{2*2} = 25.25$$

$$V(1, 2, 101) = 50.5 * 25.25 = 1275.125$$

Said differently,

$$\lambda^* = \frac{\partial V(\cdot)}{\partial m}$$

The multiplier gives the marginal value of one more unit of money (relaxing the constraint by one marginal unit).

Example: What if we use the Lagrangian for this problem?

$$\max_{x,y} u(x,y) = x^2 + y^2 \quad \text{s.t.} \quad 4x + 2y = 12$$

Envelope Theorem

Consider a maximization problem with K constraints::

$$\max_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\theta}) \quad s.t. \quad c_k = g_k(\mathbf{x}, \boldsymbol{\theta}) \;\; \forall k = 1, \ldots, K$$

Let $f(\cdot)$ and $g_k(\cdot)$ $\forall k$ be continuously differentiable with unique solution $\mathbf{x}^*(\cdot)$. Then:

$$\frac{\partial V(\boldsymbol{\theta})}{\partial \theta_i} = \frac{\partial \mathcal{L}}{\partial \theta_i}\bigg|_{optimum}$$

How the value function changes with respect to a parameter is equal to how the Lagrangian *at the optimum* changes with respect to that parameter.

We saw a simple case of this already when discussion the interpretation of the shadow value λ:

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

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

$$\frac{\partial \mathcal{L}}{\partial m}\bigg|_{optimum} = \lambda^* = \frac{\partial V(p_x, p_y, m)}{\partial m}$$

Kuhn-Tucker Conditions

We can use Kuhn-Tucker to solve optimization problems with inequality constraints and to find corner solutions. Let's start with the theorem.

Consider a constrained maximization problem with K inequality constraints:

$$\max_{\mathbf{x} \in D} f(\mathbf{x}) \quad s.t. \quad c_1 \geq g_1(\mathbf{x}) \ \ldots \ c_K \geq g_K(\mathbf{x})$$

If the following conditions hold:

- $f(\mathbf{x})$ and $g_k(\mathbf{x})$, $k = 1, \ldots, K$ are continuously differentiable over $D \subseteq \mathbb{R}^n$
- $\mathbf{x}^*$ is a solution to the maximization problem
- $\nabla g_i(\mathbf{x})$ are linearly independent for all constraints i that bind

Then there exist non-negative numbers $\lambda_1, \ldots, \lambda_K$ such that:

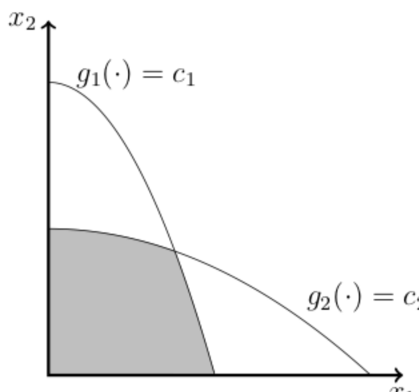
$$\nabla f(\mathbf{x}^*) = \sum_{k=1}^{K} \lambda_k \nabla g_k(\mathbf{x}^*)$$

$$\lambda_k [c_k - g_k(\mathbf{x}^*)] = 0$$

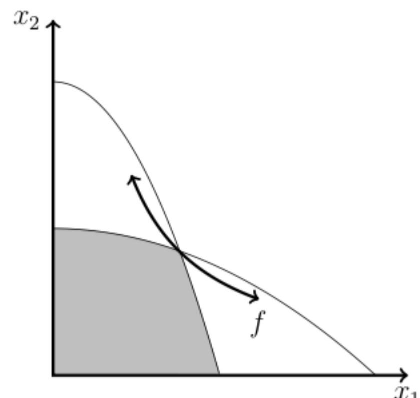
$$\nabla f(\mathbf{x}^*) = \sum_{k=1}^{K} \lambda_k \nabla g_k(\mathbf{x}^*) \quad \text{where} \quad \lambda_k \geq 0$$

These are our FOCs. The intuition is that, at the optimum, the gradient of the objective function is a linear combination of the gradient of the constraints.

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



The feasible set.



An optimal solution.

If one of the constraints does not bind, then that λ is equal to 0.

$$\lambda_k[c_k - g_k(\mathbf{x}^*)] = 0$$

These are complementary slackness conditions. The intuition is that either the constraint binds (so the right part is equal to 0). Or it does not bind, in which case the multiplier is 0 (we might as well have done the Lagrangian without it).

How to Use Kuhn-Tucker

$$\max_{\mathbf{x}\in D} f(\mathbf{x}) \quad s.t. \quad c_1 \geq g_1(\mathbf{x}) \ \ldots \ c_K \geq g_K(\mathbf{x})$$

Step 1: Set up the Lagrangian, with one multiplier for each constraint. Make sure that the constraints are set up such that violating the constraint incurs a "penalty" (breaking the constraint decreases $\mathcal{L}$)

$$\mathcal{L} = f(\mathbf{x}) + \lambda_1[c_1 - g_1(\mathbf{x})] + \cdots + \lambda_K[c_K - g_K(\mathbf{x})]$$

Step 2: Write out all of your conditions that a solution candidate must meet. They are:

(i) The FOC for each choice variable:

$$\frac{\partial \mathcal{L}}{\partial x_1} = 0 \qquad \ldots \qquad \frac{\partial \mathcal{L}}{\partial x_n} = 0$$

(ii) The constraints: $c_k \geq g_k(\mathbf{x}) \ \forall k$

(iii) The complementary slackness conditions for each constraint:

$$\lambda_k[c_k - g_k(\mathbf{x})] = 0 \ \forall k$$

(iv) Non-negative multipliers: $\lambda_k \geq 0 \ \forall k$

Step 3: Find all of the solution candidates $\mathbf{x}$ that meet all of the conditions, by investigating every potential combination of binding constraints (e.g., from no constraints bind to all constraints bind). Make sure to solve for all of the multipliers to confirm they are non-negative.

Step 4: If you have multiple solution candidates, plug them into $f(\cdot)$ to see which one (or ones) return the highest value.

Example: Use Kuhn-Tucker to solve, where $p_x, p_y, m > 0$:

$$\max_{x,y} \; xy \quad s.t. \quad m \geq p_x x + p_y y \quad x \geq 0 \quad y \geq 0$$

Lagrangian:

(i) FOCs:

(ii) Constraints:

(iii) Complementary slackness conditions:

(iv) Non-negative multipliers:

1. No constraints bind
2. The budget constraint binds
3. $x \geq 0$ binds
4. $y \geq 0$ binds
5. Budget constraint and $x \geq 0$ bind
6. Budget constraint and $y \geq 0$ bind
7. $x \geq 0$ and $y \geq 0$ bind
8. All constraints bind

Example: Use Kuhn-Tucker to solve, where $p_x, p_y, m > 0$:

$$\max_{x,y}\ x + y \quad s.t. \quad m \geq p_x x + p_y y \quad x \geq 0 \quad y \geq 0$$

Non-Negativity Constraints

Consider $\max_{\mathbf{x}\in\mathbb{R}^n} u(\mathbf{x}) \quad s.t. \quad w \geq p_1x_2 + \cdots + p_nx_n \quad x_i \geq 0 \; \forall i$

$$\mathcal{L} = u(\mathbf{x}) + \lambda[w - p_1x_1 - \cdots - p_nx_n]) + \mu_1[x_1 - 0] + \cdots + \mu_n[x_n - 0]$$

Our FOCs are $\dfrac{\partial u}{\partial x_i} = \lambda p_i - \mu_i \; \forall i$

Our complementary slackness conditions are $\mu_i x_i = 0 \; \forall i$

Because we know that $\mu_i \geq 0 \; \forall i$, we can rewrite the FOCs as:

$$\frac{\partial u}{\partial x_i} \leq \lambda p_i \text{ with equality if } x_i > 0$$

You will often see this formulation of a FOC with an inequality, that says "with equality if $x_i = 0$". This is exactly the conditions we get out of Kuhn-Tucker, just condensed to make it easier to write.

Kuhn-Tucker Special Case

Let us be in a situation where we can use Kuhn-Tucker (differentiable functions, linearly independent constraints that bind, etc.).

If also:

- $f(\cdot)$ is quasiconcave
- $g_k(\cdot)$ $\forall k$ are quasiconvex (or, the constraint set D is convex)

Then any point $\mathbf{x}^*$ that satisfies the Kuhn-Tucker conditions is a solution to the constrained optimization problem.

This is an extremely convenient theorem, because it means that if we have a quasiconcave objective function and a convex choice set, then any point that meets the Kuhn-Tucker conditions (FOCs, constraints, complementary slackness, non-negative multipliers) is in fact a solution.