Chapter 2. Constrained Optimization*

We in this chapter study the first order necessary conditions for an optimization problem with equality and/or inequality constraints. The former is often called the **Lagrange problem** and the latter is called the **Kuhn-Tucker problem**. We will not discuss the unconstrained optimization problem separately but treat it as a special case of the constrained problem because the unconstrained problem is rare in economics. Related materials of this chapter can be found in Chapter 17-19 of Simon and Blume (1994) and Chapter 4-6 of Sundaram (1996). Other useful references include Peressini et al. (1988), Bazaraa et al. (2006), Bertsekas (2016) and Luenberger and Ye (2016).

For an optimization problem, we first define what maximum/minimum and maximizer/minimizer mean.

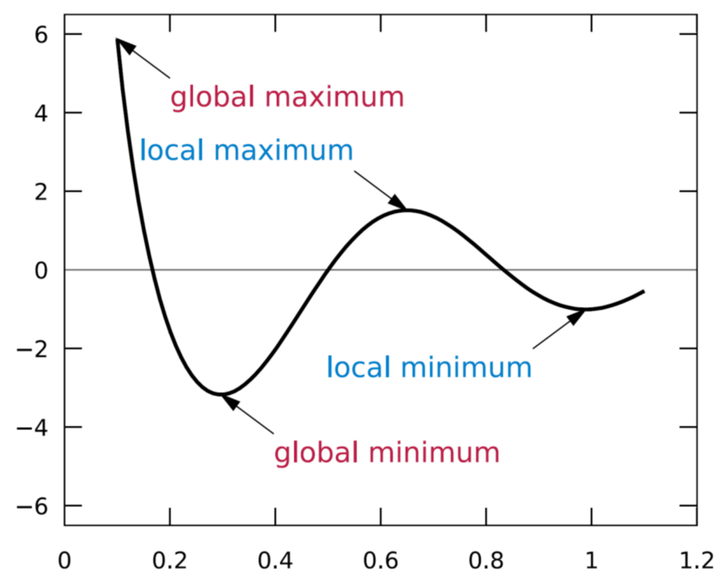


Figure 1: Local and Global Maxima and Minima for $\cos(3\pi x)/x$, $0.1 \leq x \leq 1.1$

Definition 1 *A function $f : X \to \mathbb{R}$ has a **global maximizer** at x^* if $f(x^*) \geq f(x)$ for all $x \in X$ and $x \neq x^*$. Similarly, the function has a **global minimizer** at x^* if $f(x^*) \leq f(x)$ for all $x \in X$ and $x \neq x^*$. If the domain X is a metric space, usually a subset of $\mathbb{R}^n$, then f is said*

*Email: pingyu@hku.hk

to have a ***local maximizer*** *at the point* x^* *if there exists* $r > 0$ *such that* $f(x^*) \geq f(x)$ *for all* $x \in B_r(x^*) \cap X \setminus \{x^*\}$. *Similarly, the function has a* ***local minimizer*** *at* x^* *if* $f(x^*) \leq f(x)$ *for all* $x \in B_r(x^*) \cap X \setminus \{x^*\}$.

Definition 2 *In both the global and local cases, the value of the function at a maximum point is called the* ***maximum (value)*** *of the function and the value of the function at a minimum point is called the* ***minimum (value)*** *of the function.*

Remark 1 *The maxima and minima (the respective plurals of maximum and minimum) are called* ***optima*** *(the plural of* ***optimum****), and the maximizer and minimizer are called the* ***optimizer****. Figure 1 shows local and global extrema. Usually, the optimizer and optimum without any qualifier means the global ones.*

Remark 2 *Note that a global optimizer is always a local optimizer but the converse is not correct.*

Remark 3 *In both the global and local cases, the concept of a* ***strict optimum*** *and a* ***strict optimizer*** *can be defined by replacing weak inequalities by strict inequalities. The* global *strict optimizer and optimum, if exist, are unique.*

The problem of maximization is usually stated as

$$\max_x f(x)$$
$$\text{s.t. } x \in X,$$

where "s.t." is a short for "subject to",[1] and X is called the **constraint set** or **feasible set**. The maximizer is denoted as

$$\arg\max \{f(x) | x \in X\} \text{ or } \arg\max_{x \in X} f(x),$$

where "arg" is a short for "arguments". The difference between the Lagrange problem and Kuhn-Tucker problem lies in the definition of X.

1 Equality-Constrained Optimization

1.1 Lagrange Multipliers

Consider the problem of a consumer who seeks to distribute her income across the purchase of the two goods that she consumes, subject to the constraint that she spends no more than her total income. Let us denote the amount of the first good that she buys x_1 and the amount of the second good x_2, the prices of the two goods p_1 and p_2, and the consumer's income y. The utility that the consumer obtains from consuming x_1 units of good 1 and x_2 of good two is denoted $u(x_1, x_2)$. Thus the consumer's problem is to maximize $u(x_1, x_2)$ subject to the constraint that $p_1x_1 + p_2x_2 \leq y$. (We shall soon write $p_1x_1 + p_2x_2 = y$, i.e., we shall assume that the consumer must spend all of her

[1] "s.t." is also a short for "such that" in some books.

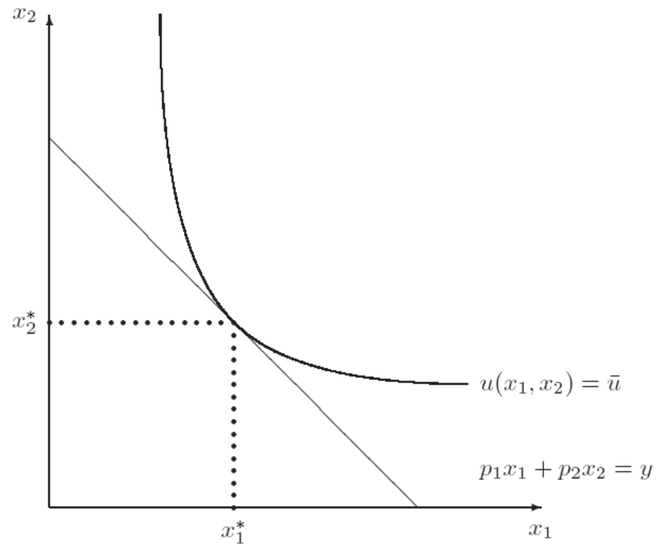


Figure 2: Utility Maximization Problem in Consumer Theory

income.) Before discussing the solution of this problem let us write it in a more "mathematical" way:

$$\begin{aligned} &\max_{x_1,x_2} u(x_1, x_2) \\ &\text{s.t. } p_1x_1 + p_2x_2 = y. \end{aligned} \tag{1}$$

We read this "Choose x_1 and x_2 to maximize $u(x_1, x_2)$ subject to the constraint that $p_1x_1+p_2x_2 = y$."

Let us assume, as usual, that the indifference curves (i.e., the sets of points (x_1, x_2) for which $u(x_1, x_2)$ is a constant) are convex to the origin. Let us also assume that the indifference curves are nice and smooth. Then the point (x_1^*, x_2^*) that solves the maximization problem (1) is the point at which the indifference curve is tangent to the budget line as given in Figure 2.

One thing we can say about the solution is that at the point (x_1^*, x_2^*) it must be true that the marginal utility with respect to good 1 divided by the price of good 1 must equal the marginal utility with respect to good 2 divided by the price of good 2. For if this were not true then the consumer could, by decreasing the consumption of the good for which this ratio was lower and increasing the consumption of the other good, increase her utility. Marginal utilities are, of course, just the partial derivatives of the utility function. Thus we have

$$\frac{\frac{\partial u}{\partial x_1}(x_1^*, x_2^*)}{p_1} = \frac{\frac{\partial u}{\partial x_2}(x_1^*, x_2^*)}{p_2}. \tag{2}$$

The argument we have just made seems very "economic." It is easy to give an alternate argument that does not explicitly refer to the economic intuition. Let x_2^u be the function that defines the indifference curve through the point (x_1^*, x_2^*), i.e.,

$$u(x_1, x_2^u(x_1)) \equiv \bar{u} \equiv u(x_1^*, x_2^*).$$

Now, totally differentiating this identity gives

$$\frac{\partial u}{\partial x_1}(x_1, x_2^u(x_1)) + \frac{\partial u}{\partial x_2}(x_1, x_2^u(x_1))\frac{dx_2^u}{dx_1}(x_1) = 0.$$

That is,

$$\frac{dx_2^u}{dx_1}(x_1) = -\frac{\frac{\partial u}{\partial x_1}(x_1, x_2^u(x_1))}{\frac{\partial u}{\partial x_2}(x_1, x_2^u(x_1))}.$$

Now $x_2^u(x_1^*) = x_2^*$. Thus the slope of the indifference curve at the point (x_1^*, x_2^*)

$$\frac{dx_2^u}{dx_1}(x_1^*) = -\frac{\frac{\partial u}{\partial x_1}(x_1^*, x_2^*)}{\frac{\partial u}{\partial x_2}(x_1^*, x_2^*)}.$$

Also, the slope of the budget line is $-\frac{p_1}{p_2}$. Combining these two results again gives result (2).

Since we also have another equation that (x_1^*, x_2^*) must satisfy, viz.,

$$p_1 x_1^* + p_2 x_2^* = y, \tag{3}$$

we have two equations in two unknowns and we can (if we know what the utility function is and what p_1, p_2, and y are) go happily away and solve the problem. (This isn't quite true but we shall not go into that at this point.) What we shall develop is a systemic and useful way to obtain the conditions (2) and (3). Let us first denote the common value of the ratios in (2) by λ. That is,

$$\frac{\frac{\partial u}{\partial x_1}(x_1^*, x_2^*)}{p_1} = \lambda = \frac{\frac{\partial u}{\partial x_2}(x_1^*, x_2^*)}{p_2}$$

and we can rewrite this and (3) as

$$\begin{aligned} &\tfrac{\partial u}{\partial x_1}(x_1^*, x_2^*) - \lambda p_1 = 0, \\ &\tfrac{\partial u}{\partial x_2}(x_1^*, x_2^*) - \lambda p_2 = 0, \\ &y - p_1 x_1^* - p_2 x_2^* = 0. \end{aligned} \tag{4}$$

Now we have three equations in x_1^*, x_2^*, and the new artificial or auxiliary variable λ. Again we can, perhaps, solve these equations for x_1^*, x_2^*, and λ. Consider the following function

$$\mathcal{L}(x_1, x_2, \lambda) = u(x_1, x_2) + \lambda(y - p_1 x_1 - p_2 x_2)$$

This function is known as the **Lagrangian**. Now, if we calculate $\frac{\partial \mathcal{L}}{\partial x_1}$, $\frac{\partial \mathcal{L}}{\partial x_2}$, and $\frac{\partial \mathcal{L}}{\partial \lambda}$, and set the results equal to zero we obtain exactly the equations given in (4). We now describe this technique in a somewhat more general way.

Suppose that we have the following maximization problem

$$\begin{aligned} &\max_{x_1,\cdots,x_n} f(x_1,\cdots,x_n) \\ &\text{s.t. } g(x_1,\cdots,x_n) = c, \end{aligned} \tag{5}$$

and we let

$$\mathcal{L}(x_1,\ldots,x_n,\lambda) = f(x_1,\ldots,x_n) + \lambda(c - g(x_1,\ldots,x_n)),$$

then if $(x_1^*,\ldots,x_n^*)$ solves (5) there is a value of λ, say λ^* such that

$$\frac{\partial \mathcal{L}}{\partial x_i}(x_1^*,\ldots,x_n^*,\lambda^*) = 0,\ i = 1,\ldots,n, \tag{6}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda}(x_1^*,\ldots,x_n^*,\lambda^*) = 0. \tag{7}$$

Notice that the conditions (6) are precisely the first order conditions for choosing $x_1,\ldots,x_n$ to maximize $\mathcal{L}$, once λ^* has been chosen. This provides an intuition into this method of solving the constrained maximization problem. In the constrained problem we have told the decision maker that she must satisfy $g(x_1,\ldots,x_n) = c$ and that she should choose among all points that satisfy this constraint the point at which $f(x_1,\ldots,x_n)$ is greatest. We arrive at the same answer if we tell the decision maker to choose any point she wishes but that for each unit by which she violates the constraint $g(x_1,\ldots,x_n) = c$ we shall take away λ units from her payoff. Of course we must be careful to choose λ to be the correct value. If we choose λ too small the decision maker may choose to violate her constraint, e.g., if we made the penalty for spending more than the consumer's income very small the consumer would choose to consume more goods than she could afford and to pay the penalty in utility terms. On the other hand if we choose λ too large the decision maker may violate her constraint in the other direction, e.g., the consumer would choose not to spend any of her income and just receive λ units of utility for each unit of her income.

It is possible to give a more general statement of this technique, allowing for multiple constraints. Consider the problem

$$\begin{aligned} &\max_{x_1,\cdots,x_n} f(x_1,\cdots,x_n) \\ &\text{s.t. } g_1(x_1,\cdots,x_n) = c_1, \\ &\qquad \vdots \\ &\quad g_m(x_1,\cdots,x_n) = c_m, \end{aligned} \tag{8}$$

where $m \le n$, i.e., we have fewer constraints than we have variables. Again we construct the Lagrangian

$$\mathcal{L}(x_1,\ldots,x_n,\lambda_1,\cdots,\lambda_m) = f(x_1,\ldots,x_n) + \lambda_1(c_1 - g_1(x_1,\ldots,x_n)) + \cdots + \lambda_m(c_m - g_m(x_1,\ldots,x_n))$$

and again if $\mathbf{x}^* \equiv (x_1^*, \ldots, x_n^*)'$ solves (8) there are values of λ, say $\lambda_1^*, \ldots, \lambda_m^*$, such that

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x_i}(x_1^*, \ldots, x_n^*, \lambda_1^*, \cdots, \lambda_m^*) &= 0,\ i = 1, \ldots, n, \\
\frac{\partial \mathcal{L}}{\partial \lambda_j}(x_1^*, \ldots, x_n^*, \lambda_1^*, \cdots, \lambda_m^*) &= 0,\ j = 1, \cdots, m.
\end{aligned}$$

These conditions are often labeled as "first order conditions" or "**FOCs**" for the corresponding maximization problem.

1.2 Caveats and Extensions

Notice that we have been referring to the set of conditions which a solution to the maximization problem must satisfy. (We call such conditions **necessary conditions**, so the FOCs usually mean the first order necessary conditions) So far we have not even claimed that there necessarily is a solution to the maximization problem. There are many examples of maximization problems which have no solution. One example of an unconstrained problem with no solution is

$$\max_x \ 2x,$$

maximizing over the choice of x the function $2x$. Clearly the greater we make x the greater is $2x$, and so, since there is no upper bound on x there is no maximum. Thus we might want to restrict maximization problems to those in which we choose x from some bounded set. Again, this is not enough. Consider the problem

$$\max_{0 \leq x \leq 1} \ 1/x \ .$$

The smaller we make x the greater is $1/x$ and yet at zero $1/x$ is not even defined. We could define the function to take on some value at zero, say 7. But then the function would not be continuous. Or we could leave zero out of the feasible set for x, say $0 < x \leq 1$. Then the set of feasible x is not closed. Since there would obviously still be no solution to the maximization problem in these cases we shall want to restrict maximization problems to those in which we choose x to maximize some continuous function from some closed (and because of the previous example) bounded set. (Recall from the Heine-Borel Theorem that a set of numbers, or more generally a set of vectors, that is both closed and bounded is a **compact** set.) Is there anything else that could go wrong? No. The following result says that if the function to be maximized is continuous and the set over which we are choosing is compact, then there is a solution to the maximization problem.

Theorem 1 (The Weierstrass Theorem) *Let S be a compact set and $f : S \to \mathbb{R}$ be continuous. Then there is some x^* in S at which the function is maximized. More precisely, there is some x^* in S such that $f(x^*) \geq f(x)$ for any $x \in S$.*

Notice that in defining such compact sets we typically use inequalities, such as $x \geq 0$. However in Section 1 we did not consider such constraints, but rather considered only equality constraints.

However, even in the example of utility maximization at the beginning of Section 1.1, there were implicitly constraints on x_1 and x_2 of the form

$$x_1 \geq 0, \; x_2 \geq 0.$$

A truly satisfactory treatment would make such constraints explicit. It is possible to explicitly treat the maximization problem with inequality constraints, at the price of a little additional complexity. We shall return to this question later in this chapter.

Also, notice that had we wished to solve a minimization problem we could have transformed the problem into a maximization problem by simply multiplying the objective function by -1. That is, if we wish to minimize $f(x)$ we could do so by maximizing $-f(x)$. From the following exercise, we know that if x_1^*, x_2^*, and λ^* satisfy the FOCs for $\max_{x_1,x_2} u(x_1, x_2)$ s.t. $p_1x_1 + p_2x_2 = y$, then x_1^*, x_2^*, and $-\lambda^*$ satisfy the FOCs for $\min_{x_1,x_2} u(x_1, x_2)$ s.t. $p_1x_1 + p_2x_2 = y$. Thus we cannot tell whether there is a maximum at (x_1^*, x_2^*) or a minimum. This corresponds to the fact that in the case of a function of a single variable over an unconstrained domain at a maximum we require the first derivative to be zero, but that to know for sure that we have a maximum we must look at the second derivative. We shall not develop the analogous conditions for the constrained problem with many variables here. However, again, we shall return to it in the next chapter.

Finally, the unconstrained problem is a special case of the constrained problem where λ^* is set at zero. In other words, since no constraints exist, no penalty is imposed on constraints.

Exercise 1 *Write out the FOCs for* $\min_{x_1,x_2} u(x_1, x_2)$ *s.t.* $p_1x_1 + p_2x_2 = y$*. Show that if* x_1^**,* x_2^**, and* λ^* *satisfy (3), then* x_1^**,* x_2^**, and* $-\lambda^*$ *satisfy the new FOCs.*

2 Inequality-Constrained Optimization

Up until now we have been concerned with solving optimization problems where the variable could be any real number and where the constraints were of the form that some function should take on a particular value. However, for many optimization problems, and particularly in economics, the more natural formulation of the problem has some inequality restriction on the variables, such as the requirement that the amount consumed of any good should be non-negative. And it is often the case that the constraints on the problem are more naturally thought of as inequality constraints. For example, rather than thinking of the budget constraint as requiring that a consumer consume a bundle exactly equal to what she can afford the requirement perhaps should be that she consume no more than she can afford. The theory of such optimization problems is called **nonlinear programming**. We give here an introduction to the most basic elements of this theory.

2.1 Kuhn-Tucker Conditions

We start by considering one of the very simplest optimization problems, namely the maximization of a continuously differentiable function of a single real variable:

$$\max_x f(x). \tag{9}$$

We read this as "Choose x to maximize $f(x)$." As mentioned at the end of Section 1.2, the first order necessary condition for x^* to be a solution to this maximization problem is that

$$\frac{df}{dx}(x^*) = 0.$$

Suppose now we add the constraint that $x \geq 0$. How does this changes the problem and its solution? We can write the problem as follows:

$$\begin{aligned} &\max_x f(x) \\ &\text{s.t. } x \geq 0. \end{aligned} \tag{10}$$

We read this as "Choose x to maximize $f(x)$ subject to the constraint that $x \geq 0$."

How does it change the conditions for x^* to be a solution? Well, if there is a solution to the maximization problem, there are two possibilities. Either the solution could occur when $x^* > 0$ or it could occur when $x^* = 0$. In the first case then x^* will also be (at least a local) maximum of the unconstrained problem and a necessary condition for this is that $\frac{df}{dx}(x^*) = 0$. In the second case it is not necessary that $\frac{df}{dx}(x^*) = 0$ since we are in any case not permitted to decrease x^*. (It's already 0 which is as low as it can go.) However, it should not be the case that we increase the value of $f(x)$ when we *increase* x from x^*. A necessary condition for this is that

$$\frac{df}{dx}(x^*) \leq 0.$$

Figure 3 illustrates the intuition why $\frac{df}{dx}(x^*) \leq 0$ when $x^* = 0$. The solution $x^* = 0$ is called a **corner solution**.

So, our FOCs for a solution to this maximization problem are that $\frac{df}{dx}(x^*) \leq 0$, that $x^* \geq 0$ and that either $x^* = 0$ or that $\frac{df}{dx}(x^*) = 0$. We can express this last by the equivalent requirement that $x^* \frac{df}{dx}(x^*) = 0$, that is that the product of the two is equal to zero. A pair of inequalities, not both of which can be strict (or slack) (i.e., at least one of them is effective), is said to show **complementary slackness**. We state this as a (very small and trivial) theorem.

Theorem 2 *Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable (or a C^1 function). Then, if x^* maximizes $f(x)$ over all $x \geq 0$, x^* satisfies*

$$\begin{aligned} &\tfrac{df}{dx}(x^*) \leq 0, \\ &x^* \tfrac{df}{dx}(x^*) = 0, \\ &x^* \geq 0. \end{aligned} \tag{11}$$

Exercise 2 *Give a version of Theorem 2 for minimization problems.*

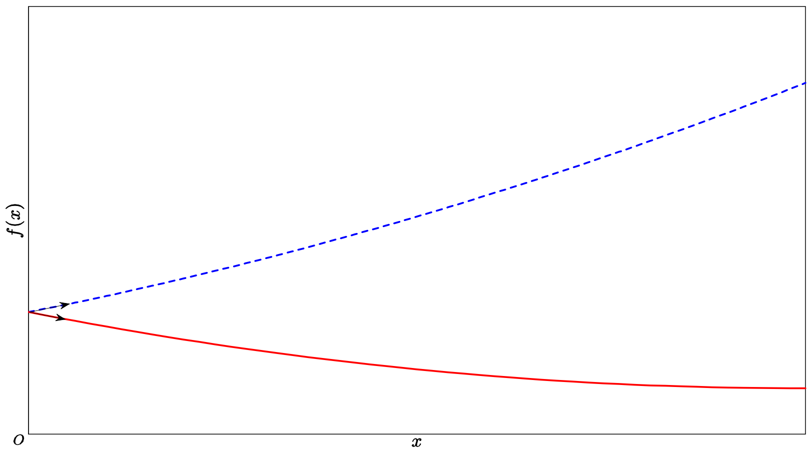


Figure 3: Illustration of Why $\frac{df}{dx}(x^*) \leq 0$ When $x^* = 0$

How to understand these FOCs? If we form the Lagrangian

$$\mathcal{L}(x, \lambda) = f(x) + \lambda x,$$

then we can express these FOCs as

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) &= \frac{df}{dx}(x^*) + \lambda^* = 0, \\
\frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) &= x^* \geq 0, \\
\lambda^* &\geq 0, \lambda^* \frac{\partial \mathcal{L}}{\partial \lambda}(x^*, \lambda^*) = \lambda^* x^* = 0.
\end{aligned}$$

For the general inequality-constrained problem,

$$\begin{aligned}
&\max_{x_1, \cdots, x_n} f(x_1, \cdots, x_n) \\
\text{s.t.} \quad &g_j(x_1, \cdots, x_n) \geq 0, \; j = 1, \cdots, J,
\end{aligned}$$

or more compactly,

$$\begin{gathered}
\max_{\mathbf{x}} f(\mathbf{x}) \\
\text{s.t. } \mathbf{g}(\mathbf{x}) \geq \mathbf{0},
\end{gathered}$$

we can form the Lagrangian

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda} \cdot \mathbf{g}(\mathbf{x}).$$

and express the FOCs as

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathbf{x}}(\mathbf{x}^*,\boldsymbol{\lambda}^*) &= \frac{\partial f}{\partial\mathbf{x}}(\mathbf{x}^*)+\frac{\partial\mathbf{g}(\mathbf{x}^*)'}{\partial\mathbf{x}}\boldsymbol{\lambda}^*=\mathbf{0},\\ \frac{\partial\mathcal{L}}{\partial\boldsymbol{\lambda}}(\mathbf{x}^*,\boldsymbol{\lambda}^*) &= \mathbf{g}(\mathbf{x}^*)\geq\mathbf{0},\\ \boldsymbol{\lambda}^* &\geq \mathbf{0},\boldsymbol{\lambda}^*\odot\frac{\partial\mathcal{L}}{\partial\boldsymbol{\lambda}}(\mathbf{x}^*,\boldsymbol{\lambda}^*)=\boldsymbol{\lambda}^*\odot\mathbf{g}(\mathbf{x}^*)=\mathbf{0},\end{aligned}$$

where $\odot$ is the element-by-element product. These FOCs are called the **Kuhn-Tucker conditions** due to Kuhn and Tucker[2] (1951).[3] For more intuitions on the Kuhn-Tucker conditions when $n=2$, $J=1$, see Section 18.3 of Simon and Blume (1994), and when $n=2$, $J=2$, see the Appendix of Dixit (1990).

We summarize the discussions on the equality-constrained and inequality-constrained problem in the following theorem; a rigorous proof can be found in Section 19.6 of Simon and Blume (1994). First, define the **mixed constrained problem** as follows,

$$\begin{aligned}&\max_{x_1,\cdots,x_n} f(x_1,\cdots,x_n)\\ \text{s.t. } &g_j(x_1,\cdots,x_n)\geq 0,\ j=1,\cdots,J,\\ &h_k(x_1,\cdots,x_n)=0,\ k=1,\cdots,K,\end{aligned}$$

or more compactly,

$$\begin{aligned}&\max_{\mathbf{x}} f(\mathbf{x})\\ \text{s.t. } &\mathbf{g}(\mathbf{x})\geq\mathbf{0},\\ &\mathbf{h}(\mathbf{x})=\mathbf{0},\end{aligned}$$

where $K\leq n$. The term "mixed constrained problem" is only for convenience because any equality constraint can be transformed to two inequality constraints, e.g., $h_k(\mathbf{x})=0$ is equivalent to $h_k(\mathbf{x})\geq 0$ and $h_k(\mathbf{x})\leq 0$. Form the Lagrangian

$$\mathcal{L}(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu})=f(\mathbf{x})+\boldsymbol{\lambda}\cdot\mathbf{g}(\mathbf{x})+\boldsymbol{\mu}\cdot\mathbf{h}(\mathbf{x}).$$

Theorem 3 (Theorem of Kuhn-Tucker) *Suppose that $f:\mathbb{R}^n\to\mathbb{R}$, $\mathbf{g}:\mathbb{R}^n\to\mathbb{R}^J$ and $\mathbf{h}:\mathbb{R}^n\to\mathbb{R}^K$ are C^1 functions. Then, if $\mathbf{x}^*$ maximizes $f(\mathbf{x})$ over all $\mathbf{x}$ satisfying the constraints $\mathbf{g}(\mathbf{x})\geq\mathbf{0}$ and $\mathbf{h}(\mathbf{x})=\mathbf{0}$, and if $\mathbf{x}^*$ satisfies the **nondegenerate constraint qualification** (NDCQ) as will be specified in the next subsection, then there exists a vector $(\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$ such that $(\mathbf{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$ satisfies*

[2] Albert W. Tucker (1905-1995) is the supervisor of John Nash, the Nobel Prize winner in Economics in 1994, and Lloyd Shapley, the Nobel Prize winner in Economics in 2012.

[3] The Kuhn-Tucker conditions were originally named after Harold W. Kuhn, and Albert W. Tucker. Later scholars discovered that the necessary conditions for this problem had been stated by William Karush in his master's thesis in 1939, so this group of conditions is also labeled as the Karush-Kuhn-Tucker (KKT) conditions.

the Kuhn-Tucker conditions given as follows:

$$\begin{array}{l}\frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0},\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{h}(\mathbf{x}^*) = \mathbf{0}, \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{g}(\mathbf{x}^*) \geq \mathbf{0},\ \boldsymbol{\lambda}^* \geq \mathbf{0}, \\ \boldsymbol{\lambda}^* \odot \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \boldsymbol{\lambda}^* \odot \mathbf{g}(\mathbf{x}^*) = \mathbf{0}.\end{array} \tag{12}$$

Remark 4 *Rigorously speaking, the Kuhn-Tucker conditions are necessary conditions for "local" optima, and of course are also necessary conditions for global optima.*

The $\mathbf{x}^*$'s that satisfy the Kuhn-Tucker conditions are called the **critical points** of $\mathcal{L}$. Usually, critical points mean the points that satisfy the FOCs; the Kuhn-Tucker conditions are a special group of FOCs. Parallel to Lagrange multipliers in the Lagrange problem, $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ are called **Kuhn-Tucker multipliers**.

Exercise 3 *Write out the Kuhn-Tucker conditions given in (12) in the long form without using vector and matrix notation.*

Exercise 4 *Give a version of the Kuhn-Tucker conditions such as in (12) for a constrained minimization problem.*

Exercise 5 *Show that the Kuhn-Tucker conditions for the maximization problem,* $\max_{\mathbf{x}} f(\mathbf{x})$ *s.t.* $\mathbf{x} \geq \mathbf{0}$ *and* $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$*, can be expressed as*

$$\begin{array}{rcl}\frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}^*, \boldsymbol{\lambda}^*) & \leq & 0, \\ \mathbf{x}^* \odot \frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}^*, \boldsymbol{\lambda}^*) & = & 0, \\ \mathbf{x}^* & \geq & 0, \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}(\mathbf{x}^*, \boldsymbol{\lambda}^*) & \geq & 0, \\ \boldsymbol{\lambda}^* \odot \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}(\mathbf{x}^*, \boldsymbol{\lambda}^*) & = & 0, \\ \boldsymbol{\lambda}^* & \geq & 0,\end{array}$$

where the Lagrangian is defined as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda} \cdot \mathbf{g}(\mathbf{x}).$$

2.2 The Constraint Qualification

We in this subsection describe the NDCQ in the Theorem of Kuhn-Tucker. A constraint $g_j(\mathbf{x}) \geq 0$ is **binding** (or **effective**, or **active**, or **tight**) at $\mathbf{x}^*$ if $g_j(\mathbf{x}^*) = 0$. Suppose the first J_0 inequality constraints are binding at $\mathbf{x}^*$; then the NDCQ states that the rank at $\mathbf{x}^*$ of the Jacobian matrix of the equality constraints and the binding inequality constraints

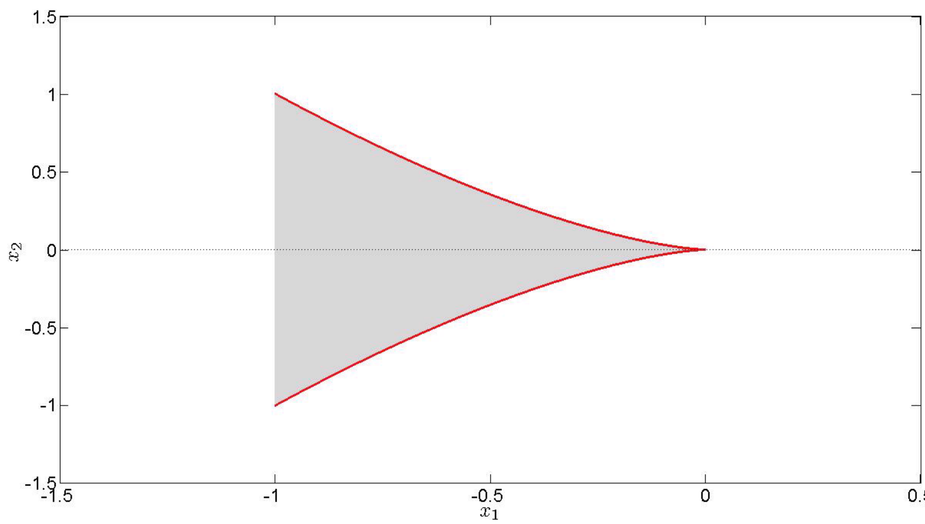
$$\mathbf{J} \equiv \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_{J_0}}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial g_{J_0}}{\partial x_n}(\mathbf{x}^*) \\ \frac{\partial h_1}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_K}{\partial x_1}(\mathbf{x}^*) & \cdots & \frac{\partial h_K}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}$$

is $J_0 + K$ - as large as it can be. When for some $\mathbf{x}$'s the NDCQ does not hold, compare the values of $f(\cdot)$ at critical points and also these $\mathbf{x}$'s to determine the ultimate maximizer.

Exercise 6 *Write out the Kuhn-Tucker conditions and the NDCQ in the unconstrained problem and the equality-constrained problem.*

Exercise 7 *Show that NDCQ in Theorem 2 holds.*

The following example provides a case where the constraint qualification fails.



The Constraint Set $\left\{(x_1, x_2) | x_1^3 + x_2^2 \leq 0\right\}$

Example 1 *This example follows Example 19.9 of Simon and Blume (1994). We want to maximize* $f(x_1, x_2) = x_1$ *s.t.* $g(x_1, x_2) = x_1^3 + x_2^2 \leq 0$. *From Figure 2.2, the constraint set is a cusp and it is easy to see that* $(x_1^*, x_2^*) = (0, 0)$. *However, at* (x_1^*, x_2^*), *there is no* λ^* *satisfying the Kuhn-Tucker conditions. To see why, set the Lagrangian*

$$\mathcal{L}(\mathbf{x}, \lambda) = x_1 - \lambda \left(x_1^3 + x_2^2\right),$$

and then the Kuhn-Tucker conditions are

$$
\begin{aligned}
1-3\lambda x_1^2 &= 0, 2\lambda x_2 = 0, \\
x_1^3 + x_2^2 &\leq 0, \lambda \geq 0, \lambda\left(x_1^3+x_2^2\right) = 0.
\end{aligned}
$$

It is not hard to see that there is no λ^* *satisfying these conditions when* $(x_1^*, x_2^*) = (0,0)$*.*

What can we learn from this example? Note that $g(x_1, x_2)$ *is binding at* $(0,0)$*, while* $(0,0)$ *is the critical point of* $g(x_1, x_2)$ *(i.e.,* $\frac{\partial g_1}{\partial x_1}(0,0) = \frac{\partial g_1}{\partial x_2}(0,0) = 0$*), so the constraint qualification fails. If we compare* $f(\cdot)$ *at the critical values of* $\mathcal{L}$ *(which is empty) and* $(0,0)$*, we indeed get the correct maximizer* $(0,0)$*.* $\square$

We provide some intuition on why the NDCQ is required. This heuristic discussion is based on the Appendix of Dixit (1990). Suppose $\mathbf{x}^*$ maximizes $f(\mathbf{x})$ over all $\mathbf{x}$ satisfying the constraints $\mathbf{g}(\mathbf{x}) \geq \mathbf{0}$, where the first J_0 constraints are binding at $\mathbf{x}^*$, and $K=0$. Then there is no neighboring $\mathbf{x}$ such that

$$g_j\left(\mathbf{x}\right) \geq g_j\left(\mathbf{x}^*\right) = 0, j = 1, \cdots, J_0, \tag{13}$$

and

$$f(\mathbf{x}) > f(\mathbf{x}^*).$$

Note that we need not consider the other $J - J_0$ constraints because they hold as strict inequalities at $\mathbf{x}^*$, by continuity they will go on holding for $\mathbf{x}$ sufficiently near $\mathbf{x}^*$. (13) implies

$$Dg_j\left(\mathbf{x}^*\right) d\mathbf{x} \geq 0, j = 1, \cdots, J_0, \tag{14}$$

and

$$Df(\mathbf{x}^*)d\mathbf{x} > 0.$$

This is valid provided the $J_0 \times n$ submatrix of $\mathbf{J}$, say $\mathbf{J}_0$, has rank J_0. If it has a smaller rank, too many vectors $d\mathbf{x}$ yield zero when multiplied by this matrix.[4] Then many more $d\mathbf{x}$ satisfy the linear approximation (14) than do $\mathbf{x}^* + d\mathbf{x}$ satisfy the true constraints (13). As a result, the FOCs can fail even though $\mathbf{x}^*$ is optimum. In the above example, the linear approximation (14) becomes

$$(0,0)\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} \geq 0,$$

which is the whole plane. However, only the set $\{(x_1, x_2)\,|x_1 \leq 0, x_2 = 0\}$, i.e., the left half of the x_1 axis, is a linear approximation to the feasible set. So essentially, the NDCQ guarantees that local to $\mathbf{x}^*$, the binding constraints and their first order approximations are equivalent.

More straightforwardly, let $J = J_0 = 1$ and $K = 0$, then that the NDCQ fails implies $Dg\left(\mathbf{x}^*\right) = \mathbf{0}$ or $Df(\mathbf{x}^*) + \lambda^* Dg\left(\mathbf{x}^*\right) = Df(\mathbf{x}^*) = \mathbf{0}$, which is the FOCs for the unconstrained problem. In other words, the binding constraint $g\left(\mathbf{x}^*\right) = 0$ does not play any role in the Kuhn-Tucker conditions,

[4] Strictly speaking, the dimension of such $d\mathbf{x}$ is $n-\text{rank}(\mathbf{J}_0)$.

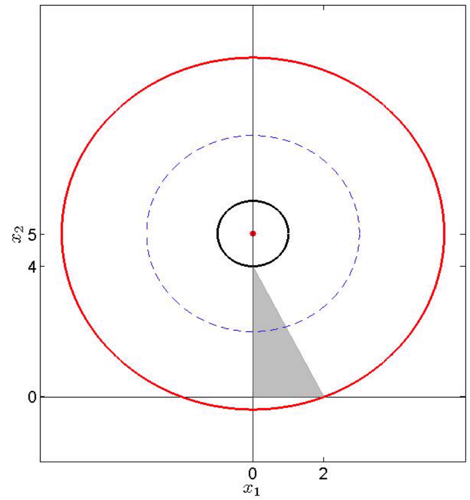


Figure 4: Intuitive Illustration of Example

which is of course absurd!

We use an example to illustrate how to find the maximizer in practice. A caution here is that "Never blindly apply the Kuhn-Tucker conditions".

Example 2 max $x_1^2 + (x_2 - 5)^2$ *s.t.* $x_1 \geq 0$, $x_2 \geq 0$, *and* $2x_1 + x_2 \leq 4$.

Solution: First, since the objective function is continuous and the constraint set is compact (why?), by the Weierstrass theorem, the maximizer exists. We then check the NDCQ. $g_1(x) = x_1$, $g_2(x) = x_2$ and $g_3(x) = 4 - 2x_1 - x_2$, so the Jacobian of the constraint functions is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -2 & -1 \end{pmatrix},$$

whose any one or two rows are linearly independent. Since at most two of the three constraints can be binding at any one time, the NDCQ holds at any solution candidate.

The Lagrangian is

$$\mathcal{L}(x, \lambda, \mu) = x_1^2 + (x_2 - 5)^2 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3(4 - 2x_1 - x_2),$$

and the Kuhn-Tucker conditions are

$$\begin{aligned} 2x_1 + \lambda_1 - 2\lambda_3 &= 0,\ 2(x_2 - 5) + \lambda_2 - \lambda_3 = 0, \\ x_1 &\geq 0,\ x_2 \geq 0,\ 4 - 2x_1 - x_2 \geq 0, \lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0, \\ \lambda_1 x_1 &= 0,\ \lambda_2 x_2 = 0,\ \lambda_3(4 - 2x_1 - x_2) = 0. \end{aligned}$$

Totally eight possibilities depending whether $\lambda_j = 0$ or not, $j = 1, 2, 3$.

(i) $\lambda_1 > 0, \lambda_2 > 0$ and $\lambda_3 > 0 \Longrightarrow x_1 = 0, x_2 = 0$, and $2x_1 + x_2 = 4$. Impossible.

(ii) $\lambda_1 = 0, \lambda_2 > 0$ and $\lambda_3 > 0 \Longrightarrow x_1 \geq 0, x_2 = 0$, and $2x_1 + x_2 = 4$. So $(x_1, x_2) = (2, 0)$. From $4 - 2\lambda_3 = 0$ and $-10 + \lambda_2 - \lambda_3 = 0$, we have $(\lambda_1, \lambda_2, \lambda_3) = (0, 12, 2)$.

(iii) $\lambda_1 > 0, \lambda_2 = 0$ and $\lambda_3 > 0 \Longrightarrow x_1 = 0, x_2 \geq 0$, and $2x_1 + x_2 = 4$. So $(x_1, x_2) = (0, 4)$. From $\lambda_1 - 2\lambda_3 = 0$ and $-2 - \lambda_3 = 0$, we have $(\lambda_1, \lambda_2, \lambda_3) = (-4, 0, -2)$. Impossible.

(iv) $\lambda_1 = \lambda_2 = 0$ and $\lambda_3 > 0 \Longrightarrow x_1 \geq 0, x_2 \geq 0$, and $2x_1 + x_2 = 4$. So from $2x_1 - 2\lambda_3 = 0, 2(x_2 - 5) - \lambda_3 = 0$, and $2x_1 + x_2 = 4$, we have $(x_1, x_2) = (-2/5, 24/5)$. Impossible.

(v) $\lambda_1 > 0, \lambda_2 > 0$ and $\lambda_3 = 0 \Longrightarrow x_1 = 0, x_2 = 0$, and $2x_1 + x_2 \leq 4$. So $(x_1, x_2) = (0, 0)$. From $\lambda_1 = 0$ and $-10 + \lambda_2 = 0$, we have $(\lambda_1, \lambda_2, \lambda_3) = (0, 10, 0)$. Impossible.

(vi) $\lambda_1 = \lambda_3 = 0, \lambda_2 > 0 \Longrightarrow x_1 \geq 0, x_2 = 0$, and $2x_1 + x_2 \leq 4$. From $2x_1 = 0$ and $-10 + \lambda_2 = 0$, we have $(x_1, x_2) = (0, 0)$ and $(\lambda_1, \lambda_2, \lambda_3) = (0, 10, 0)$.

(vii) $\lambda_1 > 0, \lambda_2 = \lambda_3 = 0 \Longrightarrow x_1 = 0, x_2 \geq 0$, and $2x_1 + x_2 \leq 4$. So from $\lambda_1 = 0$ and $2(x_2 - 5) = 0$, we have $(x_1, x_2) = (0, 5)$ and $(\lambda_1, \lambda_2, \lambda_3) = (0, 0, 0)$. Impossible.

(viii) $\lambda_1 = \lambda_2 = \lambda_3 = 0 \Longrightarrow x_1 \geq 0, x_2 \geq 0$, and $2x_1 + x_2 \leq 4$. So from $2x_1 = 0$ and $2(x_2 - 5) = 0$, we have $(x_1, x_2) = (0, 5)$. Impossible.

Candidate maximizers are $(2, 0)$ and $(0, 0)$. The objective function values at these two candidates are 29 and 25, so $(2, 0)$ is the maximizer and the associated Lagrange multipliers are $(0, 12, 2)$. □

From this above example and discussions in this chapter, we summarize a "cookbook" procedure for a constrained optimization problem. First, define the **feasible set** of the general mixed constrained maximization problem as

$$G = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}.$$

Step 1: Apply the Weierstrass theorem to show that the maximum exists. If the feasible set G is compact, this is usually straightforward; if G is not compact, truncate G to a compact set, say G_o, such that there is a point $x_o \in G_o$ and $f(x_o) > f(x)$ for all $x \in G \backslash G_o$.

Step 2: Check whether the constraint qualification is satisfied. If not, denote the set of possible violation points as Q.

Step 3: Set up the Lagrangian and find the critical points. Denote the set of critical points as R.

Step 4: Check the value of f on $Q \cup R$ to determine the maximizer or maximizers.

It is quite often for practitioners to apply Step 3 directly to find the maximizer. Although this may work in most cases, it is possible to fail in some cases which do not seem bizarre at all. First, the Lagrangian may fail to have any critical points due to nonexistence of maximizers or failure of constraint qualification. Second, even if the Lagrangian does have one or more critical points, this set of critical points need not contain the solution still due to these two reasons. Let us repeat our caveat, "Never blindly apply the Kuhn-Tucker conditions"!

This cookbook procedure works well in most cases, especially when the set $Q \cup R$ is small, e.g., $Q \cup R$ includes only a few points. If this set is large, it is better to employ more necessary conditions (e.g., the second order conditions (SOCs)) to screen the points in $Q \cup R$. Another solution is to employ sufficient conditions, i.e., as long as $\mathbf{x}^*$ satisfies these conditions, it must be the maximizer. Sufficient conditions are very powerful especially combined with the uniqueness result because as long as we find one solution, it is the solution and we can stop. These topics are the main theme of the next chapter.