Simon Fraser University, Department of Economics
Econ 798 – Introduction to Mathematical Economics
Prof. Alex Karaivanov
Lecture Notes 4

1 Unconstrained optimization

In this section we address the problem of maximizing (minimizing) a function in the case when there are *no constraints* on its arguments. This is not a very interesting case for economics, which typically deals with problems where resources are constrained, but represents a natural starting point to solving the more economically relevant constrained optimization problems.

1.1 Univariate case

Let $f : U \subseteq \mathbb{R} \to R$ be C^2. We are interested in finding maxima (or minima) of this function. We need to start with defining what do we mean by these concepts.

- **Definition (local maximum)** – done before.

  A point $x_0 \in U$ is a local maximum for the function f if $\exists \varepsilon > 0$, such that $f(x_0) \geq f(x)$, $\forall x \in U \cap N_\varepsilon(x_0)$, where $N_\varepsilon(x_0)$ denotes an ε-ball around x_0. If $f(x_0) > f(x)$, $\forall x \in U \cap N_\varepsilon(x_0)$ with $x \neq x_0$ we say that the local maximum is *strict*.

  Clearly a function can have many or no local maxima in its domain.

- **Definition (global maximum)**

  A point $x_0 \in U$ is a global maximum for the function f if $f(x_0) \geq f(x)$, $\forall x \in U$.

So how do we go about finding local (global) maxima? Most of the time we use differentiation and set the first derivative to zero but, in general, a zero first derivative is *neither necessary* (e.g., corner maximum; kink maximum), *nor sufficient* (minimum, inflection point) condition for maximum. Thus, some care is needed to ensure that what one finds by setting $f' = 0$ is indeed what one is looking for. Let us call both local maximum and local minimum local extremum. The following theorem is the basic result used for univariate unconstrained optimization problems.

- **Theorem 19 (sufficient conditions for local extrema)**

  Let $f'(x_0) = 0$. If:

  (i) $f''(x_0) < 0$ then x_0 is a local maximum of f.

  (ii) $f''(x_0) > 0$ then x_0 is a local minimum of f.

  (iii) $f''(x_0) = 0$ then we cannot conclude whether x_0 is a local extremum of f.

The following result is about existence of a maximum for a *continuous* function on a *compact* set:

- **Theorem 20 (Existence of global extrema)**

  A continuous function f with domain the closed interval $[a, b] \in R$ attains a global maximum and global minimum in the interval.

1.2 The multivariate case

Now consider more general functions of the type $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ (multivariate).

- **Theorem 21 (First-order necessary conditions for local extremum)**

  Let $f : U \subseteq R^n \rightarrow R$ be a C^1 function (continuously differentiable). If x^0 is a local extremum of f in the interior of U then:

$$\frac{\partial f(x^0)}{\partial x_i} = 0, \ i = 1, ..n$$

The above theorem states that at an interior local extremum of a C^1 function **all first partial derivatives must be equal to zero**, i.e. we can solve the system of n equations defined by the condition above and look for interior extrema only among its solutions. Note also that the above can be written equivalently as

$$\nabla f(x^0) = \mathbf{0}_{n\times 1}$$

i.e., **at interior local extremum the gradient of f is zero**. Remembering that the gradient was a vector pointing in the direction in which the function changes fastest, we see that the above condition implies that at the extremum there's no such best direction, i.e. if we go in any direction we will reach a lower functional value (if we are talking about a maximum).

The first-order condition $\nabla f(x^0) = \mathbf{0}_{n\times 1}$ is **only necessary**. Also, the theorem does not apply for kink maxima or corners (think why!).To obtain sufficient conditions, as in the univariate case (Thm 19) we need to know something about the *second derivatives* of f. In order to be able to do so, we need some useful concepts from linear algebra.

- **Definition (Principal minor)**

  Let A be an $n \times n$ matrix. A principal minor of A of order k is the *determinant* of the matrix formed by deleting some $n - k$ rows and their corresponding $n - k$ columns of A where $k = 0, ..., n - 1$.

- **Definition (Leading principal minor)**

The k-th order *leading principal minor* (LPM) of A is the determinant of the matrix formed by deleting the *last* $n-k$ columns and rows of A.

For a *symmetric* $n \times n$ matrix A define the following concepts.

- **Definition (positive/negative definite symmetric matrix)** (*Note: we saw alternative definitions earlier, using quadratic forms*)

  (a) The matrix A is positive definite (p.d.) iff all its n LPMs are positive.

  (b) The matrix A is negative definite (n.d.) iff all its LPMs are not zero and alternate in sign, that is $\det(A_1) < 0$, $\det(A_2) > 0$, etc.

- **Definition (positive/negative semidefinite symmetric matrix)**

  (a) The matrix A is positive semi-definite (p.s.d.) iff all its principal minors are non-negative.

  (b) The matrix A is negative semi-definite (n.s.d.) iff all its odd-order principal minors are non-positive and all its even-order principal minors are non-negative.

Note: the above must be true for all principal minors, not just the leading ones.
Finally we are ready to state the sufficient conditions for local extrema.

- **Theorem 22 (Second-order (sufficient) conditions for local extrema)**

  Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a C^2 function. Let also $x^0 \in U$ satisfy $\nabla f(x^0) = \mathbf{0}_{n\times 1}$ and $H(x^0)$ be the Hessian of f at x^0. Then:

  (i) If $H(x^0)$ is negative definite, then x^0 is a strict local maximum of f.

  (ii) If $H(x^0)$ is positive definite, then x^0 is a strict local minimum of f.

If the Hessian is only p.s.d. (n.s.d.) the extrema may be not strict.

- **Theorem 23 (Second-order necessary conditions)**

  Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be a C^2 function. Let also $x^0 \in intU$ (the interior of U, i.e. not a boundary point). If x^0 is a local maximum (minimum) of f then $\nabla f(x^0) = \mathbf{0}_{n\times 1}$ and $H(x^0)$ is n.s.d. (p.s.d.).

The following examples illustrate how the theory from above is applied.

- **Example 1 (Multi-product firm)**

Suppose we have a firm producing two goods in quantities q_1 and q_2 and with prices p_1 and p_2. Let the cost of producing q_1 units of good 1 and q_2 units of good 2 is given by $C(q_1, q_2) = 2q_1^2 + q_1 q_2 + 2q_2^2$. The firm maximizes profits, i.e., it solves:

$$\max_{q_1, q_2} \quad \pi = p_1 q_1 + p_2 q_2 - (2q_1^2 + q_1 q_2 + 2q_2^2) = p^T q - q^T A q$$

where $p = (p_1, p_2)^T$, $q = (q_1, q_2)^T$ and $A = \begin{bmatrix} 2 & .5 \\ .5 & 2 \end{bmatrix}$.

How to solve for the optimal quantities the firm will choose? Take first the partial derivatives of π with respect to q_1 and q_2 and set them to zero, to find $q_i^* = \dfrac{4p_i - p_j}{15}$, $i, j = 1, 2$. We also need to verify that this is a maximum indeed. The Hessian of the objective is $H = \begin{bmatrix} -4 & -1 \\ -1 & -4 \end{bmatrix}$. Let's check if the leading principal minors alternate in sign, we have $H_1 = \det[-4] = -4 < 0$ and $H_2 = \det(H) = 15 > 0$, i.e., the candidate solution is a maximum indeed.

- **Example 2 (OLS)**

Think of some variable y which depends on $x_1, x_2, ..x_k$ and assume we have a dataset of n observations (i.e. n vectors $X_i = (x_{1i}, x_{2i}, ..x_{ki})$, $i = 1..n$). Assume that x_1 is a vector of ones. We are looking for the "best fit" between a linear function of the observations, $X\beta$ and our dependent variable y. (Note that X is $n \times k$ and β is $k \times 1$ vector of coefficients). Thus we can write:

$$y_i = \beta_1 x_{1i} + ...\beta_k x_{ki} + \varepsilon_i, \;\; i = 1..n$$

where ε_i are the 'residuals' (errors) between the fitted line $X\beta$ and y. The above can be written more compactly in matrix form as:

$$y = X\beta + \varepsilon$$

Remember, we want to find the best fit, i.e. the coefficients β which minimize the $\varepsilon's$ in some sense. One possible criterion (used by the OLS method) is to choose β to minimize $\sum_{i=1}^n \varepsilon_i^2$, i.e. we want to solve the problem:

$$\begin{aligned} \min_\beta S(\beta) &= \sum_i (y_i - \beta_1 x_{1i} - ...\beta_k x_{ki})^2 = (y - X\beta)^T (y - X\beta) = \\ &= y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta \end{aligned}$$

The first order condition for the above minimization problem is (using the matrix differentiation rules – differentiate wrt each β_i and stack):

$$\frac{\partial S(\beta)}{\partial \beta} = -2X^T y + 2X^T X\beta = \overrightarrow{0}$$

from which we find $\beta^* = (X^T X)^{-1} X^T y$ – a candidate minimum. So is β^* indeed a minimum of $S(\beta)$? We need to check if the Hessian is positive semi-definite, i.e., whether $H(\beta^*) = \dfrac{\partial^2 S(\beta^*)}{\partial \beta^2} = 2X^T X$, a k-by-k matrix is p.s.d. (Exercise: prove that the Hessian is p.s.d. using one of the given definitions).

1.3 Constrained optimization

1.3.1 Introduction

In this section we look at problems of the following general form:

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & f(x) \\ s.t. \quad & g(x) \leq b \\ & h(x) = c \end{aligned} \tag{NLP}$$

We call the above problem, a Non-Linear Optimization Problem (NLP). In it, $f(x)$ is called the *objective function*, $g(x) \geq b$ are *inequality constraints*, and $h(x) = c$ are *equality constraints*. Note that any optimization problem can be written in the above canonical form. For example if we want to minimize a function $h(x)$, we can do this by maximizing $-h(x)$.

It turns out that it is easier not to solve (NLP) directly, but instead solve another, related problem (Lagrange's or Kuhn-Tucker's) for x^* and then verify that x^* solves the original NLP as well. We will also be interested in whether we are obtaining *all solutions* to the NLP in this way, i.e., whether it is true that if x^* solves the NLP it solves the related problem as well. Thus we would like to see when the Lagrange's or Kuhn-Tucker's methods are both *necessary and sufficient* for obtaining solutions to the original NLP.

1.3.2 Equality constraints

Start simple, assuming that the problem we deal with has only equality constraints, i.e.,

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & f(x) \\ s.t. \quad & h(x) = c, \ c \in \mathbb{R}^m \end{aligned}$$

The equality constraints restrict the domain over which we maximize. Notice that if the number of the constraints is equal to the number of variables ($m = n$) and if we assume that the constraints are linearly independent, potentially we can solve for x from the constraints and there will be nothing to be optimized. Thus a well-defined problem will typically have $m < n$ (less constraints than choice variables).

(a) The Lagrange multipliers method

The method for solving problems of the above type is called the *Lagrange Multipliers Method* (LMM). What it does is convert the NLP into a related problem (call it the LMM problem) with a **new objective function and no constraints**, so that we can then use the usual unconstrained optimization techniques.

What is the price we have to pay for this simplification? During the conversion to LMM we end up with m **more variables** to optimize over. We next verify what is the connection between the solutions to the LMM and the original NLP and most importantly, what conditions are needed for the solutions to the LMM to be solutions to our NLP with equality constraints.

Let us describe the Lagrange method works. First we form the new objective function, called *the Lagrangean*:

$$\Lambda(x, \lambda) \equiv f(x) + \lambda^T (c - h(x))$$

Notice that we added m new variables, λ_j, $j = 1, ..., m$ – one for each constraint. These are called *Lagrange multipliers*. Note that they multiply zeros, so in fact the functional value of the objective does not change.

The LMM problem is:

$$\max_{x,\lambda} \Lambda(x, \lambda) \tag{LMM}$$

Suppose we have set all partial derivatives to zero and arrived at a candidate-solution (x^*, λ^*). We need to check if it is indeed a maximum, i.e. a second-order condition must be verified as well.

Let's now go through the above steps in more detail. First, write down the first-order (necessary) conditions for local extremum in the LMM problem:

$$\begin{aligned} \frac{\partial \Lambda}{\partial x_i} &= -\lambda_1 \frac{\partial h_1}{\partial x_1} - ... - \lambda_m \frac{\partial h_m}{\partial x_m} + \frac{\partial f}{\partial x_i} = 0, \ i = 1..n \\ \frac{\partial \Lambda}{\partial \lambda_j} &= c_j - h_j(x) = 0, \ j = 1..m \end{aligned}$$

Note we have $m + n$ equations in the same number of unknowns.

(b) Second-order conditions of LMM and the bordered Hessian

Suppose the above system of first-order conditions has a solution (x^*, λ^*). We need to check if it is a maximizer indeed. The standard way in unconstrained problems was to see if the Hessian is n.s.d. Here, we form so-called *bordered Hessian*, defined as:

$$\hat{H}_{(m+n)\times(m+n)}(x^*, \lambda^*) \equiv \begin{bmatrix} 0 & ... & 0 & -\frac{\partial h_1}{\partial x_1} & ... & -\frac{\partial h_1}{\partial x_n} \\ ... & ... & ... & ... & ... & ... \\ 0 & ... & 0 & -\frac{\partial h_m}{\partial x_1} & ... & -\frac{\partial h_m}{\partial x_n} \\ -\frac{\partial h_1}{\partial x_1} & ... & -\frac{\partial h_m}{\partial x_1} & \frac{\partial^2 \Lambda}{\partial x_1^2} & ... & \frac{\partial^2 \Lambda}{\partial x_1 \partial x_n} \\ ... & ... & ... & ... & ... & ... \\ -\frac{\partial h_1}{\partial x_n} & ... & -\frac{\partial h_m}{\partial x_n} & \frac{\partial^2 \Lambda}{\partial x_n \partial x_1} & ... & \frac{\partial^2 \Lambda}{\partial x_n^2} \end{bmatrix}$$

where all derivatives are evaluated at (x^*, λ^*). This is nothing but our usual Hessian for the Lagrangean (from the unconstrained optimization method) but notice that we have ordered the matrix of second partials in a particular way – first taking all second partials with respect to the $\lambda's$ and then with respect to the $x's$. The bordered Hessian, $\hat{H}$ can be written in a more compact way as:

$$\hat{H} = \begin{bmatrix} \mathbf{0}_{m\times m} & -\mathbf{J}^h_{m\times n} \\ -(\mathbf{J}^h)^T_{n\times m} & H(\Lambda(x))_{n\times n} \end{bmatrix}$$

where $\mathbf{J}^h$ is the Jacobian of $h(x)$ and $H(\Lambda(x))$ is the "Hessian" of $\Lambda(X)$ (the matrix of second partials of Λ taken only with respect to the x_i).

Because of all the zeros, turns out we need only the last $n-m$ leading principal minors of $\hat{H}$ to determine if it is n.s.d. Let $\det(\hat{H}_{m+1})$ be the LPM of the matrix with bottom-right element $\frac{\partial^2\Lambda}{\partial x_1^2}$, $\det(\hat{H}_{m+2})$ be the LPM of the matrix with bottom-right element $\frac{\partial^2\Lambda}{\partial x_2^2}$, etc. Then we have the following result:

- **Theorem:**

  If $sign(\det(\hat{H}_{m+l})) = (-1)^l$, $l = 1, ..n-m$, *then the bordered Hessian is n.s.d. and the candidate solution is a maximum of the LMM.*

- **Example: A Consumer's problem**

A consumer has income y and wants to choose the quantities of n goods $q_1, ..q_n$ to buy to maximize his strictly concave utility $U(q_1, ..q_n)$, taking as given the prices of the goods $p_1, ..p_n$. Her problem can be written as (using vector notation):

$$\max_q U(q)$$
$$s.t. \quad p^T q = y$$

Set up the Lagrangean:

$$\Lambda(q, \lambda) = U(q) + \lambda(y - p^T q)$$

The FOCs are:

$$\begin{aligned} y - p^T q &= 0 \\ \frac{\partial U}{\partial q_i} - \lambda p_i &= 0 \end{aligned}$$

which can be solved for (q^*, λ^*). Check that the SOC (the fact that the bordered Hessian is n.s.d.) is satisfied as an exercise (Hint: use the concavity of U).

(c) The constraint qualification

Notice that the above proposition *does not say anything about whether x^* obtained as the solution to LMM will solve the original NLP problem.* In general, this is not true since the FOCs of the Lagrangean may be neither necessary, nor sufficient for a maximum and thus additional conditions are needed. One possible necessary condition so that the solutions of the LMM be solutions of the NLP as well is the so-called *constraint qualification (CQ)*:

$$\mathbf{J}(h(x^*)) = \begin{bmatrix} \frac{\partial h_1(x^*)}{\partial x_1} & ... & \frac{\partial h_1(x^*)}{\partial x_n} \\ ... & ... & ... \\ \frac{\partial h_m(x^*)}{\partial x_1} & ... & \frac{\partial h_m(x^*)}{\partial x_n} \end{bmatrix} \textbf{ is rank } m$$

If there is only one constraint the CQ is equivalent to the **gradient of h being not a vector of zeros** at x^*. The CQ is only a *necessary condition*, so if the Jacobian is singular at some $\hat{x}$ we should treat it as candidate maximum and we would need to check (separately) whether it solves the NLP.

- **Theorem 26 (Equality constraints)**

  Consider the problem:

$$\max_{x\in\mathbb{R}^n} \; f(x)$$
$$s.t. \quad h(x) = c \in \mathbb{R}^m$$

  Let $C = \{x \in \mathbb{R}^n : h_1(x) = c_1, ...h_m(x) = c_m\}$, i.e., the set of feasible points. Let x^* be local maximum in C and suppose it satisfies the constraint qualification. $rank(\mathbf{J}(h(x^*)) = m$. Then $\exists\lambda^* \in \mathbb{R}^m$ such that (x^*, λ^*) is a critical point of the Lagrangean $\Lambda(x, \lambda) = f(x) + \lambda^T(c - h(x))$.

The above theorem implies that if the constraint qualification holds, we can use the FOCs of the LMM to find candidate maxima and then verify which of them is solves the original NLP problem.

- **Example:** solve $\max_{x_1,x_2} \; x_1^2x_2$ subject to $2x_1^2 + x_2^2 = 3$.

Solution: First check for any points which may violate the CQ. We have $J^h = [4x_1 \; 2x_2]$ i.e. we do not want both x_1 and x_2 to be zero. But it is obvious (it violates the constraint) that $(0, 0)$ is not a solution to our problem, so we can safely ignore this point as canidate maximum.

The Lagrangean is $L(x_1, x_2, \mu) = x_1^2x_2 - \mu(2x_1^2 + x_2^2 - 3)$; the FOCs are $\frac{\partial L}{\partial x_1} = 2x_1(x_2 - 2\mu) = 0$ and $\frac{\partial L}{\partial x_2} = x_1^2 - 2\mu x_2 = 0$ and $2x_1^2 + x_2^2 = 3$. The first equation yields $x_1 = 0$ or $x_2 = 2\mu$. Case 1: if $x_1 = 0$ then $x_2 = \pm\sqrt{3}$ from the 3rd FOC and $\mu = 0$ from the 2nd. Two candidate maxima.

Case 2: If $x_2 = 2\mu$ we get $x_1^2 = x_2^2$ from the 2nd FOC. Plug into the 3rd to get $x_1 = \pm 1$. So, we also have $x_2 = \pm 1$ and then $\mu = 0.5$ if $x_2 = 1$ and $\mu = -0.5$ when $x_2 = -1$. Four more candidate maxima. Since the CQ holds at any point $\neq (0, 0)$, we know that the FOCs are necessary for maximum (see Theorem 26), so just check which of the six candidates delivers it. Answer – there are 2 solutions to the original problem: $(1, 1)$ and $(-1, 1)$.

1.3.3 Inequality constraints

(a) The Lagrange multipliers method again

The Lagrange method can be also used to solve problems involving inequality constraints in addition to the equality ones. Consider the following problem:

$$\max_{x\in\mathbb{R}^n} f(x)$$
$$s.t. \quad g(x) \leq b, \;\; b \in \mathbb{R}^k$$

Note that the inequality constraints can be either *binding*, i.e. they **hold as equality** at the solution x^* or *non-binding* if they **hold as strict inequality**. Again the Lagrange method calls for setting up the Lagrangean:

$$\Lambda(x,\lambda) = f(x) + \lambda(b - g(x))$$

It is very tempting to use the previously described method, take derivatives with respect to the x_i and set them equal to zero. The problem with that approach is what we are going to do if we have a non-binding constraint? It clearly enters the FOCs and so may potentially change the value of the objective – if we proceed in this way, we will end up maximizing a different function.

What to do? If a constraint binds, we can proceed as before, but if it does not, **we must set** $\lambda_j = 0$. This is intuitively clear – if a constraint is not binding at the solution it **need not be included in the maximization**. Thus, together with the FOCs, we need to impose the additional conditions that:

$$\lambda_j[b_j - g_j(x)] = 0, \;\; j = 1..k$$

It can be demonstrated (but I will not do it here) that the Lagrange multipliers in the case of inequality constraints *cannot be negative*. Intuitively, this is true since $\nabla f(.) = \lambda \nabla g(x)$ and the fact that the gradients of f and g must point in the same direction (towards outside the constraint set, i.e., where the function increases).

Note: if you have also equality constraints in the problem, there is no need to impose the $\lambda_j[b_j - g_j(x)] = 0$ *condition for them (why?). But there is no harm if you do – it will be automatically satisfied (think why)*

- **First-order conditions**

In total the first-order conditions for the LMM in this case are:

$$\begin{aligned} \frac{\partial \Lambda(x,\lambda)}{\partial x_i} &= 0 \\ \lambda_j(b_j - g_j(x)) &= 0 \\ g_j(x) &\leq b_j \\ \lambda_j &\geq 0 \end{aligned} \tag{FOC}$$

- **Second-order conditions**

Suppose that, by solving the above FOCs we have found a candidate solution (x^*, λ^*) and wish to check if it solves the NLP as well. Again, we need to set up the bordered Hessian, but we must include only the elements of λ^* that are non-zero (i.e., **include only the binding constraints**, $g_1, ..g_{k_0}$). For example, let us assume that the first k_0 constraints are binding. Then the bordered Hessian is given by:

$$\hat{H}_{(n+k_0)\times(n+k_0)}(x^*,\lambda^*) = \begin{bmatrix} 0 & ... & 0 & -\frac{\partial g_1}{\partial x_1} & ... & -\frac{\partial g_1}{\partial x_n} \\ ... & ... & ... & ... & ... & ... \\ 0 & ... & 0 & -\frac{\partial g_{k_0}}{\partial x_1} & ... & -\frac{\partial g_{k_0}}{\partial x_n} \\ -\frac{\partial g_1}{\partial x_1} & ... & -\frac{\partial g_1}{\partial x_n} & \frac{\partial^2 \Lambda}{\partial x_1^2} & ... & \frac{\partial^2 \Lambda}{\partial x_1 \partial x_n} \\ ... & ... & ... & ... & ... & ... \\ -\frac{\partial g_{k_0}}{\partial x_1} & ... & -\frac{\partial g_{k_0}}{\partial x_n} & \frac{\partial^2 \Lambda}{\partial x_n \partial x_1} & ... & \frac{\partial^2 \Lambda}{\partial x_n^2} \end{bmatrix}$$

where all derivatives are evaluated at (x^*, λ^*). If $\hat{H}$ is n.s.d. we indeed have a maximum in the LMM problem.

Once again a possible **constraint qualification** (to ensure the solution to LMM solves NLP) is that the Jacobian matrix of the k_0 binding constraints be rank k_0.

- **Theorem 27 (Inequality constraints)**

Consider the problem:

$$\max_{x\in\mathbb{R}^n} \quad f(x)$$
$$s.t. \quad g(x) \leq b \in \mathbb{R}^k$$

Let $C = \{x \in \mathbb{R}^n : g(x) \leq b\}$, i.e. the set of feasible points. Let x^* be local maximum in C that satisfies the constraint qualification (CQ) for the k_0 binding constraints $rank(\mathbf{J}^g(x^*)) = k_0$. Then $\exists\lambda^* \in \mathbb{R}^{k_0}$ such that (x^*, λ^*) is a critical point of the Lagrangean $\Lambda(x,\lambda) = f(x) + \sum_{j=1}^{k_0} \lambda_j(b_j - g_j(x))$.

Again, what this meas is that if the CQ holds, we can use the Lagrange method to obtain the solution to the NLP.

Sometimes it is hard to check if the constraint qualification holds. So what to do instead? We can impose some conditions on the shapes of the objective function and the constraints. For example an alternative to the above full-rank-of-the-Jacobian condition is given by the following:

- **Theorem 24 (Slater's constraint qualification)**

Consider the problem $\max_x f(x)$ *s.t.* $g(x) \geq 0$ *(NLP). Let* $f(.)$ *and* $g(.)$ *be concave functions on* $\mathbb{R}^n$*. Then:*

(i) If the FOCs of the Lagrangean are satisfied at x^* *then* x^* *solves the NLP exhibited above.*

(ii) (the Slater condition) A candidate maximum satisfies the FOCs of the Lagrangean if $\exists\bar{x}$*, such that* $g_j(\bar{x}) > 0$*,* $j = 1..m$*.*

- **Example:** Solve using the Lagrange method the problem $\max_{x,y}\ x - y^2\ s.t.\ x^2 + y^2 = 4$, $x \geq 0,\ y \geq 0$.

Solution: Verify that the CQ first considering all cases (note that the first constraint is binding). Form the Lagrangean $L = x - y^2 - \mu(x^2 + y^2 - 4) + \lambda_1 x + \lambda_2 y$. The nine first-order conditions for maximum are:

$$\begin{aligned}
\frac{\partial L}{\partial x} &= 1 - 2\mu x + \lambda_1 = 0 \ (1)\\
\frac{\partial L}{\partial y} &= -2y - 2\mu y + \lambda_2 = 0 \ (2)\\
\frac{\partial L}{\partial \mu} &= x^2 + y^2 - 4 = 0 \ (3)\\
\lambda_1 x &= 0 \ (4) \text{ and } \lambda_2 y = 0 \ (5)\\
\lambda_1 &\geq 0 \ (6) \text{ and } \lambda_2 \geq 0 \ (7)\\
x &\geq 0 \ (8) \text{ and } y \geq 0 \ (9)
\end{aligned}$$

From (1) we have $1 + \lambda_1 = 2\mu x$. Since $\lambda_1 \geq 0$, then $1 + \lambda_1 > 0$ and so it must be $x > 0$ and $\mu > 0$ (think why!). But then from (4) we have $\lambda_1 = 0$. From (2) we have $2y(1+\mu) = \lambda_2$. Since $1 + \mu > 0$ either both y and λ_2 are zero or both are > 0. But by (5) both cannot be positive. So, $\lambda_2 = y = 0$. Now $x = 2$ from (3), $\lambda_1 = 0$ by (4) and $\mu = 1/4$ by (1) . This leads to the candidate solution $(x, y, \mu, \lambda_1, \lambda_2) = (2, 0, 1, 1/4, 0, 0)$.Verify that it is a maximum.

(b) The Kuhn-Tucker method

An alternative to the Lagrangean method is the so-called Kuhn-Tucker method. Consider the following canonical non-linear optimization problem:

$$\begin{aligned}
&\max_{x\in\mathbb{R}^n}\ f(x)\\
&s.t.\quad g(x) \leq b, \quad b \in \mathbb{R}^k\\
&\qquad\quad x_i \geq 0, \quad i = 1, ..n
\end{aligned} \tag{KT}$$

- **Remarks**

(1) the above problem features so-called *non-negativity constraints* on the choice variables. Because most of the objects (prices, quantities, etc.) that economics deals with are non-negative, the above method is highly applicable in economic settings.

(2) note that the formulation above is without loss of generality and does allow for equality constraints – if we have an equality constraint $h(x) = c$ we write it as the two inequality constraints: $h(x) \leq c$ and $-h(x) \leq -c$.

Let $\Lambda(.)$ denotes the Lagrangean of the above problem, **excluding the non-negativity constraints**, i.e.

$$\Lambda(x, \lambda) = f(x) + \lambda^T(b - g(x))$$

and $\hat{\Lambda}(.)$ denote the Lagrangean, including them, i.e.,

$$\hat{\Lambda}(x,\lambda) = f(x) + \lambda^T(b - g(x)) + \mu^T x$$

Notice that if we have $x_i = 0$ at the solution, then $\frac{\partial \Lambda}{\partial x_i} = \frac{\partial \hat{\Lambda}}{\partial x_i} - \mu_i \leq 0$ as $\frac{\partial \hat{\Lambda}}{\partial x_i} = 0$ at the solution. If, instead, $x_i > 0$ then the non-negativity constraint is non-binding and can be excluded, thus it must be that $\frac{\partial \Lambda}{\partial x_i} = 0$ at the solution. Combining these two observations, the following conditions must hold:

$$x_i \frac{\partial \Lambda}{\partial x_i} = 0,\ x_i \geq 0,\ \frac{\partial \Lambda}{\partial x_i} \leq 0$$

Thus, the first-order conditions for the KT method are:

$$\begin{aligned} x_i \frac{\partial \Lambda}{\partial x_i} &= 0 \\ x_i &\geq 0 \\ \frac{\partial \Lambda}{\partial x_i} &\leq 0 \\ \lambda_j(b_j - g_j(x)) &= 0 \\ g_j(x) &\leq b_j \\ \lambda_j &\geq 0 \end{aligned} \tag{KT FOCs}$$

- **Remarks**

1. The *second order conditions* involve the bordered Hessian of the binding constraints as before.
2. A possible *constraint qualification* is that the Jacobian of the partial derivatives of the **binding constraints** with respect to the non-zero $x_i's$ to have maximum possible rank when evaluated at (x^*, λ^*).

Advantages over the Lagrange method: less equations and less unknowns to solve for.

- **Theorem 28 (Necessity of the Kuhn-Tucker conditions)**

Consider the problem of optimization with inequality and non-negativity constraints:

$$\begin{aligned} &\max_{x \in \mathbb{R}^n} \ f(x) \\ &s.t. \quad g(x) \leq b \in \mathbb{R}^k \\ &x \geq 0 \end{aligned}$$

Let C be the feasible set; x^* be a local maximum in C and suppose x^* satisfies the CQ formed of *the k_0 binding constraints with partials taken with respect to the j_0 non-zero x_j*. Then $\exists \lambda^* \geq 0$

such that (x^*, λ^*) satisfies the system of KT first-order conditions:

$$\begin{aligned} x_i \frac{\partial \Lambda(.)}{\partial x_i} &= 0,\ x_i \geq 0,\ \frac{\partial \Lambda}{\partial x_i} \leq 0,\ i = 1..n \\ \lambda_j [b_j - g_j(x)] &= 0,\ \lambda_j \geq 0,\ g_j(x) \leq b_j,\ j = 1..k \end{aligned}$$

where $\Lambda(x, \lambda) = f(x) + \lambda^T (b - g(x))$.

The interpretation is that if x^* solves the NLP and the (modified) CQ holds, then x^* solves the KT equations, i.e., the latter are necessary for x^* to be maximum if the CQ holds.

- **Theorem 29 (Kuhn-Tucker – sufficient conditions for maximum)**

  Consider the problem from the previous theorem. If the following conditions are satisfied:

  (i) $f(x)$ is differentiable and concave on $\mathbb{R}^n_+$.

  (ii) $g_j(x)$ are differentiable and convex on $\mathbb{R}^n_+$ for all $j = 1, ..k$.

  (iii) x^* satisfies the KT first-order conditions (KT FOCs).

  Then x^* is the global maximum of $f(x)$ subject to the constraints $g(x) \leq b$.

Thus, with concavity of the objective function and convexity of the constraints the KT method is sufficient to find a maximum. Many economic problems satisfy these criteria, which makes the KT method the economists' "weapon of choice".

- **Example:** *Solve using the KT method the problem* $\max_{x,y}\ x - y^2$ *subject to* $x^2 + y^2 \leq 4$, $x \geq 0$, $y \geq 0$.

Solution: Verify the CQ as before. Form the KT Lagrangean $L = x - y^2 - \mu(x^2 + y^2 - 4)$. The KT first-order conditions for maximum are:

$$\frac{\partial L}{\partial x} = 1 - 2\mu x \leq 0 \ (1)$$

$$\frac{\partial L}{\partial y} = -2y - 2\mu y \leq 0 \ (2)$$

$$x^2 + y^2 - 4 \leq 0 \ (3)$$

$$(1 - 2\mu x)x = 0 \ (4) \text{ and } (-2y - 2\mu y)y = 0 \ (5)$$

$$\mu(x^2 + y^2 - 4) = 0 \quad (6)$$

$$x, y, \mu \geq 0 \ (7)$$

Brute force method: check all eight cases about whether x, y, μ are 0 or > 0 or be smart and eliminate some cases upfront.

1.3.4 The Envelope theorem

This section considers the effect of changes in some parameters on which the objective function and/or the constraints depend on the outcome of an optimization problem.

- **Envelope theorem 1 (unconstrained problem)**

Suppose we have the problem:

$$\max_x \; f(x, a)$$

where $x \in \mathbb{R}^n$ (vector of unknowns) and $a \in \mathbb{R}$ (a parameter). Suppose f is C^1 and suppose $x^*(a)$ solves the above problem. Then:

$$\frac{d}{da} f(x^*(a), a) = \frac{\partial f}{\partial a} f(x^*(a), a)$$

Note that we have the total derivative on the l.h.s. and the partial on the r.h.s.

Proof: Use the chain rule

$$\begin{aligned} \frac{d}{da} f(x^*(a), a) &= \sum_{i=1}^{n} \frac{\partial f(x^*(a), a)}{\partial x_i} \frac{\partial x_i^*(a)}{\partial a} + \frac{\partial f(x^*(a), a)}{\partial a} = \\ &= \frac{\partial f(x^*(a), a)}{\partial a} \end{aligned}$$

since $\frac{\partial f(x^*(a),a)}{\partial x_i} = 0$ for $i = 1, .. n$ by the first-order conditions for maximum.

- **Interpretation:** if we are interested in the (total) change of the maximized functional value with respect to the parameter a we can simply take the partial derivative of f with respect to a and evaluate at $x^*(a)$ – no need to worry that a affects f both directly and indirectly through x^*. This saves you a lot of work when doing comparative statics.

The constrained case

Suppose we have the problem:

$$\begin{aligned} &\max_{x \in \mathbb{R}^n} \; f(x, \alpha) \\ &\text{s.t.} \quad g(x, \alpha) \leq b \in \mathbb{R}^k \end{aligned}$$

Let λ_j be the multipliers on the inequality constraints and suppose $\alpha \in \mathbb{R}^m$ are parameters (below I just use α for one of them). Suppose the solution $x^*(\alpha)$ to the above problem exists and is such that **all** k **constraints are binding at** $x^*(a)$. Suppose also that the FOCs are necessary and sufficient for a maximum. At x^*, λ^* the Lagrangean is

$$\Lambda(x^*(\alpha), \alpha, \lambda^*(\lambda)) = f(x^*(\alpha), \alpha) + \lambda^*(\alpha)^T (b - g(x^*(\alpha), \alpha))$$

and (x^*, λ^*) satisfy the LMM FOCs:

$$\frac{\partial \Lambda}{\partial x_i} = \frac{\partial f(x^*(\alpha),\alpha)}{\partial x_i} - \sum_{j=1}^{k} \lambda_j^*(\alpha) \frac{\partial g_j(x^*(\alpha),\alpha)}{\partial x_i} = 0 \quad i = 1,..n$$
$$b_j - g_j(x^*(\alpha),\alpha) = 0, \quad j = 1..k \text{ (all constraints bind, as assumed)}$$

Notice that if we plug in $x^*(\alpha)$ and $\lambda^*(\alpha)$ into Λ it becomes just a function of α. Let us write this more explicitly. Define $V(\alpha) \equiv f(x^*(\alpha),\alpha)$ called the *value function*, i.e., this is the objective function evaluated at the solution $x^*(\alpha)$. Define also $\Psi(\alpha) \equiv \Lambda(x^*(\alpha),\alpha,\lambda^*(\alpha))$, i.e., the value function of the Lagrangean at the solution.

Consider now a small change in α. We have, for the first derivative of Ψ (reflecting the *total* change of the Lagrangean maximized value with respect to α):

$$\begin{aligned}\frac{d\Psi(\alpha)}{d\alpha} &= \sum_{i=1}^{n} \frac{\partial \Lambda}{\partial x_i}\frac{dx_i^*}{d\alpha} + \sum_{j=1}^{k} \frac{\partial \Lambda}{\partial \lambda_j}\frac{\partial \lambda_j^*(\alpha)}{\partial \alpha} + \frac{\partial \Lambda}{\partial \alpha} = \\ &= \sum_{i=1}^{n} \frac{\partial \Lambda}{\partial x_i}\frac{dx_i^*}{d\alpha} + \sum_{j=1}^{k} [b_j - g_j(x^*(\alpha),\alpha)]\frac{\partial \lambda_j^*(\alpha)}{\partial \alpha} + \frac{\partial \Lambda}{\partial \alpha}\end{aligned}$$

All above derivatives are evaluated at $x^*(a), \lambda^*(a)$. At (x^*, λ^*) the FOCs hold, i.e. $\frac{\partial \Lambda}{\partial x_i} = 0$, and also we assumed that all constraints bind, so $b_j = g_j(x^*(a), a)$ for all $j = 1, ...k$. Thus, the first two terms in $\frac{d\Psi(\alpha)}{d\alpha}$ above equal zero, and therefore,

$$\frac{d\Psi(\alpha)}{d\alpha} = \frac{\partial \Lambda(x^*(\alpha),\lambda^*(\alpha),\alpha)}{\partial \alpha}$$

Now do this for $V(\alpha)$ – the total change in the maximized value of f:

$$\frac{dV(a)}{d\alpha} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i^*(\alpha)}{d\alpha} + \frac{\partial f}{\partial \alpha}$$

Again, all derivatives are evaluated at $x^*(\alpha)$, arguments are omitted to save on notation. Using the Lagrangean FOCs again, we have

$$\frac{\partial f(x^*(\alpha),\alpha)}{\partial x_i} = \sum_{j=1}^{k} \lambda_j^*(\alpha)\frac{\partial g_j(x^*(\alpha),\alpha)}{\partial x_i}$$

Using also that $g_j(x^*(\alpha),\alpha) = b_j$ in a neighborhood of α, i.e., differentiating both sides,

$$\sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i}\frac{dx_i^*}{d\alpha} + \frac{\partial g_j}{\partial \alpha} = 0$$

, we obtain (changing the order of summation with respect to i and j):

$$\begin{aligned}
\frac{dV(\alpha)}{d\alpha} &= \sum_{j=1}^{k} \lambda_j^* \sum_{i=1}^{n} \frac{\partial g_j(x^*(\alpha), \alpha)}{\partial x_i} \frac{dx_i^*(\alpha)}{d\alpha} + \frac{\partial f}{\partial \alpha} \\
&= -\sum_{j=1}^{k} \lambda_j^* \frac{\partial g_j}{\partial \alpha} + \frac{\partial f}{\partial \alpha} = \frac{\partial \Lambda}{\partial \alpha}
\end{aligned}$$

The last line follows from the definition of the Lagrangean – take a partial derivative with respect to the α argument alone. Finally combining (1) and (2) we have that:

$$\frac{d\Psi(\alpha)}{d\alpha} = \frac{\partial \Lambda(x^*(\alpha), \alpha, \lambda^*(\alpha))}{\partial \alpha} = \frac{dV(\alpha)}{d\alpha}$$

- **What is the use of this theorem?**

Suppose you are maximizing a function f and you have found x^* that maximizes it. Suppose also you would like to do comparative statics, i.e., see how x^* and $f(x^*)$ change as you vary some parameter α_i. From the above theorem, the effect of changing α_i on the value function can be calculated **without solving the whole problem again**, but simply by taking the **partial derivative** of the Lagrangean with respect to α_i alone (i.e., keeping the other x and the other α's constant while differentiating).