Review of Maximization Problem

This math review note is based on Mathematical Appendix of MWG, but mainly contains intuitions and applications. For rigorous proofs, please refer to other notes or textbooks on multivariable calculus. The lecture notes on the website `http://ocw.aca.ntu.edu.tw/ntu-ocw/index.php/ocw/cou/101S130` is a good reference.

M.J. Unconstrained Maximization (p.954)

Consider $f : \mathbb{R}^N \to \mathbb{R}$

Definition M.J.1. The vector $\overline{x} \in \mathbb{R}^N$ is a local maximizer of $f(\cdot)$ if there is an open neighborhood of $\overline{x}, A \subset \mathbb{R}^N$, s.t. $f(\overline{x}) \geq f(x)$ for every $x \in A$. If $f(\overline{x}) \geq f(x)$ for every $x \in \mathbb{R}^N$, then $\overline{x}$ is a global maximizer of $f(\cdot)$.

Theorem M.J.1. *Suppose that $f(\cdot)$ is differentiable and that $x \in \mathbb{R}^N$ is a local maximizer or local minimizer of $f(\cdot)$. Then $\frac{\partial f(\overline{x})}{\partial x_n} = 0$ for every n, or more concisely*

$$\nabla f(\overline{x}) = \begin{bmatrix} \frac{\partial f(\overline{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\overline{x})}{\partial x_N} \end{bmatrix} = 0.$$

Remark. $\nabla f(\overline{x}) = 0$ is only a necessary condition for local maximizer or local minimizer. Figure 1 below graphically illustrates this idea for the one-variable case.

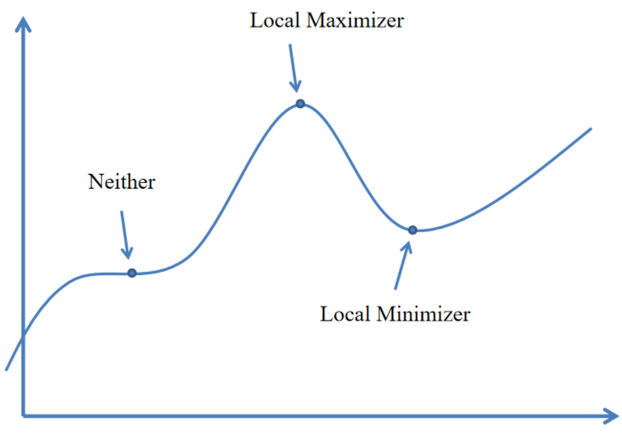

Figure 1: Local Maximizer and Local Minimizer

Theorem M.J.2. *Suppose that the function $f : \mathbb{R}^N \to \mathbb{R}$ is twice continuously differentiable (C^2) and that $\nabla f(\overline{x}) = 0$.*

(i) If $\overline{x} \in \mathbb{R}^N$ is a local maximizer, then the (symmetric) $N \times N$ matrix $D^2 f(\overline{x})$ is negative semidefinite.

(ii) If $D^2 f(\overline{x})$ is negative definite, then $\overline{x}$ is a local maximizer.

Remark. Replacing "negative" by "positive", the same is true for local minimizer.

Proof. Consider any arbitrary direction of displacement $\varepsilon z \in \mathbb{R}^N$, where $\varepsilon \geq 0$ is a scaler. By Taylor expansion,

$$\begin{aligned} f(\overline{x} + \varepsilon z) - f(\overline{x}) &= \varepsilon \nabla f(\overline{x}) \cdot z + \frac{1}{2}\varepsilon^2 z \cdot D^2 f(\overline{x}) z + \text{Remainder} \\ &= \frac{1}{2}\varepsilon^2 z \cdot D^2 f(\overline{x}) z + \text{Remainder} \end{aligned}$$

Note that ($\frac{1}{\varepsilon^2}$Remainder) is small when ε is small because Remainder contains $\varepsilon^3, \varepsilon^4, \ldots$terms.

(i) Suppose $\overline{x}$ is a local maximizer. Then for small enough ε, $[f(\overline{x} + \varepsilon z) - f(\overline{x})]/\varepsilon^2 \leq 0$ must hold. So taking limit,

$$z \cdot D^2 f(\overline{x}) z \leq 0$$

More explicitly,

$$\frac{[f(\overline{x} + \varepsilon z) - f(\overline{x})]}{\varepsilon^2} = \frac{1}{2} z \cdot D^2 f(\overline{x}) z + \frac{\textit{Remainder}}{\varepsilon^2} \tag{1}$$

$\lim\limits_{\varepsilon \to 0} \frac{\textit{Remainder}}{\varepsilon^2} = 0$ and $\lim\limits_{\varepsilon \to 0} \frac{[f(\overline{x}+\varepsilon z) - f(\overline{x})]}{\varepsilon^2} \leq 0$ ($\because \overline{x}$ is a local maximizer) implies

$$z \cdot D^2 f(\overline{x}) z \leq 0.$$

(ii) Suppose $z \cdot D^2 f(\overline{x}) z < 0$. Then RHS of (1) is negative for ε sufficiently small. Therefore, $\frac{[f(\overline{x}+\varepsilon z) - f(\overline{x})]}{\varepsilon^2} < 0$ for all sufficiently small ε. □

Remark. In the above proof, in part (ii), we rely on the assumption of $z \cdot D^2 f(\overline{x}) z < 0$. $z \cdot D^2 f(\overline{x}) z \leq 0$ is not enough to guarantee local maximization. To see this, consider the example, $f(x) = x^3$. $D^2 f(0)$ is negative semidefinite because $d^2 f(0)/dx^2 = 0$, but $\overline{x} = 0$ is neither a local maximizer nor a local minimizer.

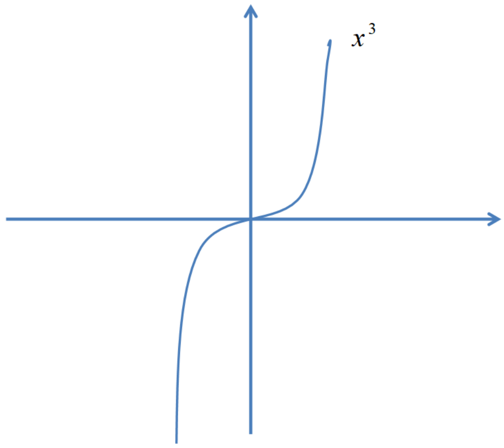

Figure 2: $f(x) = x^3$

Theorem M.J.3. *Any critical point* $\overline{x}$ *(i.e., any* $\overline{x}$ *satisfying* $\nabla f(\overline{x}) = 0$*) of a concave function* $f(\cdot)$ *is a global maximizer of* $f(\cdot)$.

Proof. Concavity implies $f(x) \leq f(\overline{x}) + \nabla f(\overline{x}) \cdot (x - \overline{x})$, $\forall x$. Since $\nabla f(\overline{x}) = 0$, we have $f(x) \leq f(\overline{x})$, $\forall x$. □

M.K. Constrained Maximization

Case I: Equality constraints We first study the maximization problem with M equality constraints, given by (C.M.P.1) below.

$$\begin{aligned} &\max_{x \in \mathbb{R}^N} f(x) \\ \text{s.t. } & g_1(x) = \overline{b}_1 \\ &\vdots \\ & g_M(x) = \overline{b}_M \end{aligned} \tag{C.M.P.1}$$

Assumption. $N \geq M$ *(Generically, solution doesn't exist if* $M > N$.*)*

The **Constraint Set** is

$$C = \{x \in \mathbb{R}^N : g_m(x) = \overline{b}_m \text{ for } m = 1, ..., M\}.$$

Theorem M.K.1. *Suppose that the objective and constraint functions of problem (C.M.P.1) are differentiable and that $\overline{x} \in C$ is a local constrained maximizer. Assume also the $M \times N$ matrix*

$$\begin{bmatrix} \nabla g_1(\overline{x})^T \\ \vdots \\ \nabla g_M(\overline{x})^T \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1(\overline{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\overline{x})}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_M(\overline{x})}{\partial x_1} & \cdots & \frac{\partial g_M(\overline{x})}{\partial x_N} \end{bmatrix}$$

*has rank M. (This is called **constraint qualification**: It says that the constraints are independent at $\overline{x}$.) Then, there are numbers $\lambda_m \in \mathbb{R}$ (Not $\mathbb{R}^+$), one for each constraint, such that*

$$\frac{\partial f(\overline{x})}{\partial x_n} = \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(\overline{x})}{\partial x_n} \text{ for every } n = 1, ..., N, \tag{M.K.2}$$

Or, equivalently,

$$\nabla f(\overline{x}) = \sum_{m=1}^{M} \lambda_m \nabla g_m(\overline{x}). \tag{M.K.3}$$

The numbers λ_m are referred to as **Lagrange multipliers**.

How to understand Theorem M.K.1? Note that we will not prove the theorem, but will explain the theorem using examples and graphs.

Two-variable, one-constraint Cases We first consider simple cases with two variables, and then extend the intuition to three and more variables.

Example M.K.1. Consider the following two-variable, one-constraint example.

$$\max_{(x_1,x_2)\in\mathbb{R}^2} x_1 + x_2$$

$$\text{s.t. } x_1^2 + x_2^2 = 1$$

Here, $f(x) = x_1 + x_2$, $g(x) = x_1^2 + x_2^2$. Graphically, the constraint set $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ is a circle with radius 1 centred at 0. And the level sets of $f(x_1, x_2)$, i.e., $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 = k\}$, are straight lines with slope -1. Figure 3 illustrates the idea.

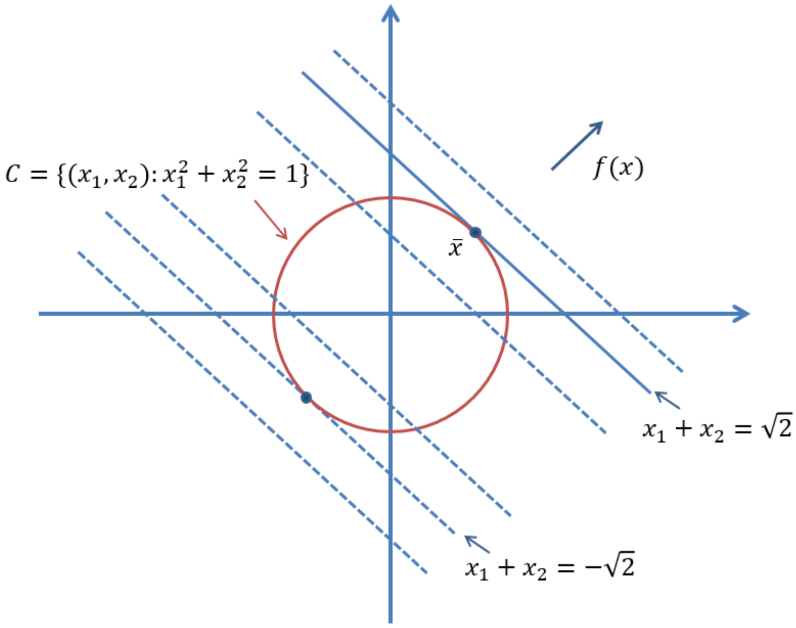

Figure 3: Two-variable example

From Figure 3, it is not hard to see that the objective function obtains its maximum when the level set $x_1+x_2=k$ is tangent to the circle $x_1^2+x_2^2=1$. Therefore, $(\overline{x}_1, \overline{x}_2) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ solves the problem. Figure 4 shows that $\nabla f(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ is parallel to $\nabla g(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

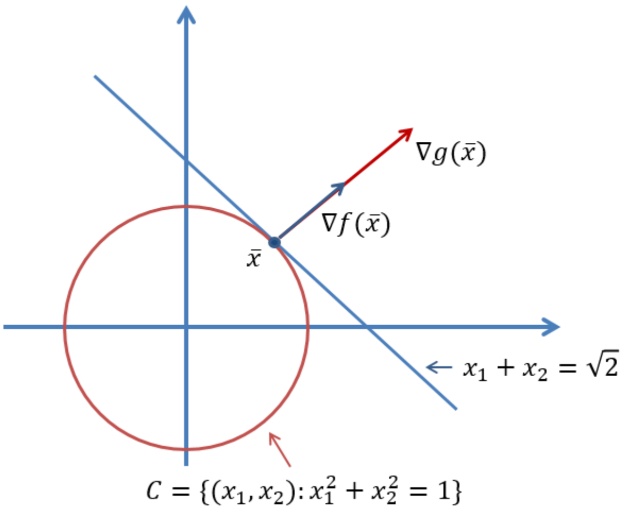

Figure 4: The gradients

Numerically, $\nabla f(x_1, x_2) = (1, 1)$ and $\nabla g(x_1, x_2) = (2x_1, 2x_2)$. At the solution $(\overline{x}_1, \overline{x}_2) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $\nabla f(\overline{x}_1, \overline{x}_2) = (1, 1)$ and $\nabla g(\overline{x}_1, \overline{x}_2) = (\sqrt{2}, \sqrt{2})$. Thus, we have $\nabla f(\overline{x}_1, \overline{x}_2) = \frac{1}{\sqrt{2}} \nabla g(\overline{x}_1, \overline{x}_2)$. That is, this simple example complies with Theorem M.K.1.

More generally, for two-variable, one-constraint cases, the maximum must be obtained where the level set of the objective function is tangent to the constraint set. See Figure 5 below.

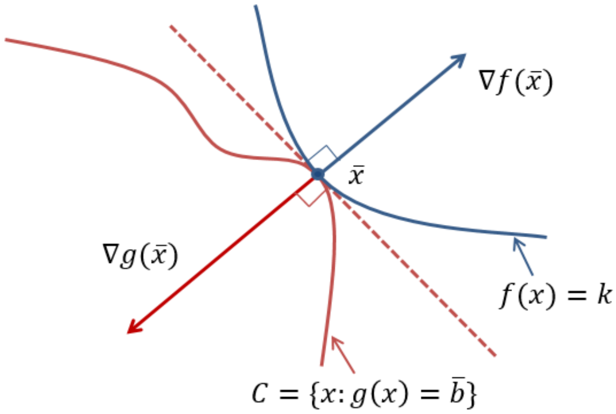

Figure 5: Two-variable case

Since $\nabla f(\overline{x})$ is orthogonal to the level set $\{x \in \mathbb{R}^2 : f(x) = k\}$ and $\nabla g(\overline{x})$ is orthogonal to the constraint set $C = \{x \in \mathbb{R}^2 : g(x) = \overline{b}\}$, and the two curves are tangent at $\overline{x}$, so $\nabla f(\overline{x})$ and $\nabla g(\overline{x})$ must lie on the same line. That is, $\exists \lambda$ such that $\nabla f(\overline{x}) = \lambda \nabla g(\overline{x})$. This gives rise to Theorem M.K.1 for the two-variable cases.

Three-variable, one-constraint cases Graphically, the constraint set

$$C = \{x \in \mathbb{R}^3 : g(x) = \overline{b}\}$$

is a surface. The level set of the objective function $\{x \in \mathbb{R}^3 : f(x) = k\}$ is also a surface. Similar to the two-variable cases, the maximum must be obtained where the level set of the objective function is tangent to the constraint set. See Figure 6 below.

Since $\nabla f(\overline{x})$ is orthogonal to the level set $\{x \in \mathbb{R}^3 : f(x) = k\}$ and $\nabla g(\overline{x})$ is orthogonal to the constraint set $C = \{x \in \mathbb{R}^3 : g(x) = \overline{b}\}$, and the two surfaces are tangent at $\overline{x}$, so $\nabla f(\overline{x})$ and $\nabla g(\overline{x})$ must lie on the same line. That is, $\exists \lambda$ such that $\nabla f(\overline{x}) = \lambda \nabla g(\overline{x})$. This gives rise to Theorem M.K.1 for three-variable, one-constraint cases.

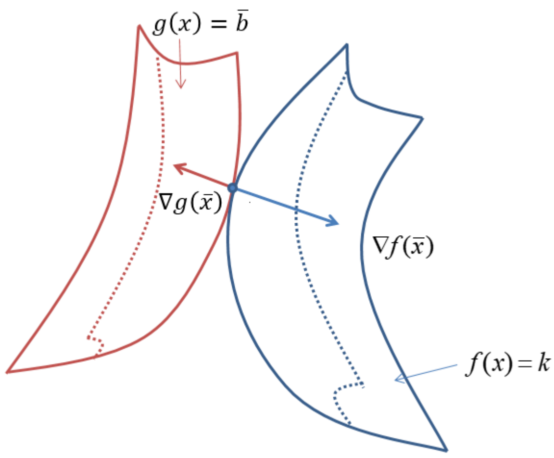

Figure 6: Three-variable, one-constraint case

Three-variable, two-constraint cases Graphically,the two constraints are surfaces. The constraint set

$$C = \{x \in \mathbb{R}^3 : g_1(x) = \bar{b}_1 \text{ and } g_2(x) = \bar{b}_2\}$$

is the intersection of the two surfaces, forming a curve. See Figure 7 below.

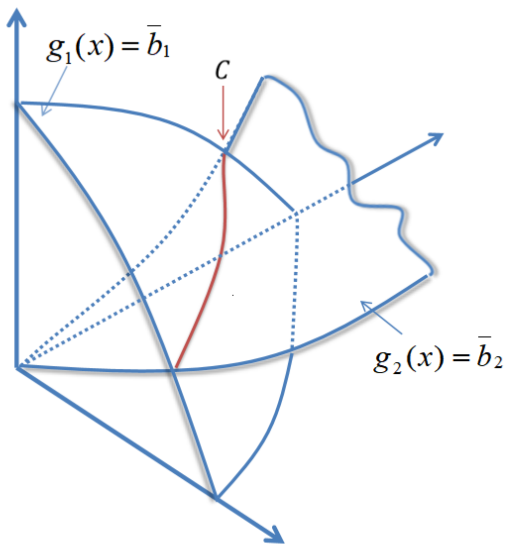

Figure 7: The constraints

Similar to previous cases, the maximum must occur when the level set of the objective function (which is a surface in this case) is tangent to the constraint set. See Figure 8 below.

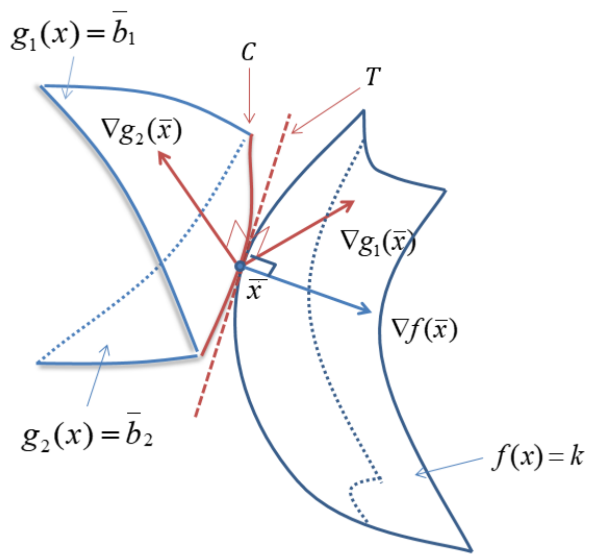

Figure 8: Three-variable, two-constraint case

In Figure 8, the level set of the objective function (the surface $\{x \in \mathbb{R}^3 : f(x) = k\}$) and the constraint set (Curve C) are tangent at point $\overline{x}$. Line T is the tangent line to Curve C at $\overline{x}$. So, Line T is also tangent to the surface $\{x \in \mathbb{R}^3 : f(x) = k\}$.

We will show below that $\nabla f(\overline{x})$, $\nabla g_1(\overline{x})$ and $\nabla g_2(\overline{x})$ lie on the same plane.

(i) Since $\nabla f(\overline{x})$ is orthogonal to the surface $\{x \in \mathbb{R}^3 : f(x) = k\}$, $\nabla f(\overline{x})$ is orthogonal to Line T.

(ii) Since Line T is tangent to Curve C at $\overline{x}$, and Curve C is in the surface formed by the first constraint $\{x \in \mathbb{R}^3 : g_1(x) = \overline{b}_1\}$, so Line T is tangent to the surface $\{x \in \mathbb{R}^3 : g_1(x) = \overline{b}_1\}$ at $\overline{x}$. Therefore, given that $\nabla g_1(\overline{x})$ is orthogonal to the surface $\{x \in \mathbb{R}^3 : g_1(x) = \overline{b}_1\}$, $\nabla g_1(\overline{x})$ is orthogonal to Line T.

(iii) Following the same logic as in Part (ii), $\nabla g_2(\overline{x})$ is also orthogonal to Line T.

Since $\nabla f(\overline{x})$, $\nabla g_1(\overline{x})$ and $\nabla g_2(\overline{x})$ are all orthogonal to Line T, they lie on the same plane. Therefore, $\exists \lambda_1, \lambda_2$ such that $\nabla f(\overline{x}) = \lambda_1 \nabla g_1(\overline{x}) + \lambda_2 \nabla g_2(\overline{x})$, **if $\nabla g_1(\overline{x})$ and $\nabla g_2(\overline{x})$ are linearly independent ("Constraint Qualification")**. This gives rise to Theorem M.K.1 for the three-variable, two-constraint cases.

More variables Theorem M.K.1 says that $\nabla f(x)$ lies on the hyperplane spanned by $\nabla g_m(x)$ for $m = 1, ..., M$, **if the constraints $g_m(x)$ are linearly independent ("Constraint Qualification")**. The same intuition from previous simple cases apply:

1. The constraint set $C = \{x \in \mathbb{R}^N : g_m(x) = \bar{b}_m \text{ for } m = 1, ..., M\}$ and the level set $\{x \in \mathbb{R}^n : f(x) = k\}$ are tangent at the maximum $\overline{x}$.

2. $\nabla g_m(\overline{x}), m = 1, ..., M$ is orthogonal to the constraint set.

3. $\nabla f(\overline{x})$ is orthogonal to the level set $\{x \in \mathbb{R}^N : f(x) = k\}$.

Therefore, $\nabla f(\overline{x})$ and $\nabla g_m(\overline{x}), m = 1, ..., M$ lie one the same hyperplane. $\nabla f(\overline{x}) = \sum_{m=1}^{M} \lambda_m \nabla g_m(\overline{x})$, if the matrix

$$\begin{bmatrix} \nabla g_1(\overline{x})^T \\ \vdots \\ \nabla g_M(\overline{x})^T \end{bmatrix}$$

has Rank M.

What happens when *Constraint Qualification* fails?

Example. Consider the case $\nabla g_1(\overline{x}) = -\alpha \nabla g_2(\overline{x})$. See Figure 9 for graphical illustration.

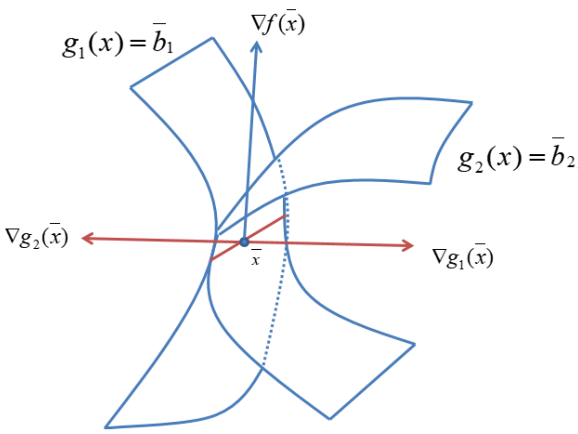

Figure 9: Failure of *Constraint Qualification*

Here, although $\overline{x}$ is a local maximizer, $\nabla f(\overline{x})$ cannot be written as $\lambda_1 \nabla g_1(\overline{x}) + \lambda_2 \nabla g_2(\overline{x})$.

How to use Theorem M.K.1? We will introduce an alternative presentation of Theorem M.K.1.

Define *Lagrangian function*:

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{m=1}^{M} \lambda_m (g_m(x) - \bar{b}_m)$$

The constrained maximization problem can be rewritten as the following *unconstrained maximization problem:*

$$\max_{x\in\mathbb{R}^N,\lambda\in\mathbb{R}^M} \mathcal{L}(x,\lambda).$$

First Order Conditions (F.O.C.) give:

$$\frac{\partial f(x)}{\partial x_n} - \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(x)}{\partial x_n} = 0, \text{ for } n = 1, ..., N;$$
$$g_m(x) - \bar{b}_m = 0, \text{ for } m = 1, ..., M.$$

Remark. In practice, failure of *Constraint Qualification* is rarely a problem. However, you should be alerted and check *Constraint Qualification* if you find the above standard methods problematic. If you find no solution, it may be that the maximization problem itself has no solution, or *Constraint Qualification* may fail so that F.O.C is not applicable.

Example. Let's get back to Example M.K.1 and apply Theorem M.K.1 to solve it.

Solution. Lagrange function:

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

F.O.C. gives

$$\frac{\partial \mathcal{L}}{\partial x_1} = 1 - 2\lambda x_1 = 0$$
$$\frac{\partial \mathcal{L}}{\partial x_2} = 1 - 2\lambda x_2 = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - x_1^2 - x_2^2 = 0.$$

We obtain two solutions

$$(x_1^*, x_2^*, \lambda^*) = (1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$$
$$\text{and } (x_1^*, x_2^*, \lambda^*) = (-1/\sqrt{2}, -1/\sqrt{2}, -1/\sqrt{2}).$$

The values are $f(1/\sqrt{2}, 1/\sqrt{2}) = \sqrt{2}$ and $f(-1/\sqrt{2}, -1/\sqrt{2}) = -\sqrt{2}$.

Similar as shown in Theorem M.J.1, F.O.C. is only a necessary condition for local maximum. We also need to check Second Order Condition.

Second Order Condition To ensure that the solution is a maximum, we need similar conditions as in Theorem M.J.2: If $\overline{x}$ is a local maximizer, then

$$D_x^2\mathcal{L}(\overline{x},\lambda) = D^2 f(\overline{x}) - \sum_{m=1}^{M} \lambda_m D^2 g_m(\overline{x})$$

is negative semidefinite on the subspace

$$\{z \in \mathbb{R}^N : \nabla g_m(\overline{x}) \cdot z = 0 \text{ for all } m\}.$$

The other direction also applies, i.e., negative definiteness on the subspace implies local maximization.

Example. Apply Second Order Condition to the solutions of Example M.K.1.

Solution. We first calculate $D_x^2\mathcal{L}(x_1, x_2, \lambda)$:

$$D_x^2\mathcal{L}(x_1, x_2, \lambda) = D^2 f(x_1, x_2) - \lambda D^2 g(x_1, x_2) = \begin{bmatrix} -2\lambda & 0 \\ 0 & -2\lambda \end{bmatrix}$$

At $(x_1^*, x_2^*, \lambda^*) = (1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$,

$$D_x^2\mathcal{L}(1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2}) = \begin{bmatrix} -\sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{bmatrix}.$$

Then, we obtain the subspace.

$$\nabla g(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}.$$

At $(x_1^*, x_2^*, \lambda^*) = (1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$,

$$\nabla g(x_1, x_2) = \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}.$$

So, the subspace is given by

$$\{z \in \mathbb{R}^2 : \nabla g(x_1^*, x_2^*, \lambda^*) \cdot z = 0\} = \{z \in \mathbb{R}^2 : \sqrt{2}z_1 + i\sqrt{2}z_2 = 0\} = \{z \in \mathbb{R}^2 : z_2 = -z_1\}.$$

Thus, what we need to check is the negative semi-definiteness of $D_x^2\mathcal{L}(1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$ on the subspace $\{z \in \mathbb{R}^2 : z_2 = -z_1\}$. Since

$$z \cdot \begin{bmatrix} -\sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{bmatrix} z = \begin{bmatrix} -\sqrt{2}z_1 & \sqrt{2}z_1 \end{bmatrix} \begin{bmatrix} z_1 \\ -z_1 \end{bmatrix} = -2\sqrt{2}z_1^2 \leq 0,$$

$(x_1^*, x_2^*) = (1/\sqrt{2}, 1/\sqrt{2})$ is a maximizer.

Repeating the above process for the other critical point $(-1/\sqrt{2}, -1/\sqrt{2}, -1/\sqrt{2})$ reveals that $(x_1^*, x_2^*) = (-1/\sqrt{2}, -1/\sqrt{2})$ is a local minimizer. (You should check it by yourself.)

Remark. You could also use *bordered Hessian matrix* to check Second Order Condition.

What does λ_m measure?

Claim. λ_m measures the sensitivity of $f(x^*)$ to a small increase in $\bar{b}_m$, i.e., $\lambda_m = \frac{\partial f(x^*(\bar{b}))}{\partial \bar{b}_m}$.

To see this, we first generate a family of maximization problems with different values of $\bar{b}^T = \begin{bmatrix} \bar{b}_1 & \bar{b}_2 & \ldots & \bar{b}_M \end{bmatrix}$: $\max f(x)$ subject to $g(x) = \bar{b}$.

Let $x^*(\bar{b})$ be a solution, and suppose that the constraint qualification holds at all $\bar{b} \in \mathbb{R}^M$. Then, $\exists \lambda^*(\bar{b}) \in \mathbb{R}^M$ such that

$$\nabla f(x^*(\bar{b})) - \sum_{m=1}^{M} \lambda_m^*(\bar{b}) \nabla g_m(x^*(\bar{b})) = 0 \tag{FOC}$$

$$g_m(x^*(\bar{b})) - \bar{b}_m = 0 \;\; m = 1, ..., M. \tag{Constraints}$$

By chain rule,

$$\nabla_{\bar{b}} f(x^*(\bar{b}) = \nabla f(x^*(\bar{b}) \nabla x^*(\bar{b}) \underbrace{=}_{\text{(FOC)}} \sum_{m=1}^{M} \lambda_m^*(\bar{b}) \nabla g_m(x^*(\bar{b})) \nabla x^*(\bar{b}) \tag{2}$$

Differentiating both sides of (Constraints) by $\bar{b}$ gives

$$\nabla g_m(x^*(\bar{b})) \nabla x^*(\bar{b}) - e_m = 0, \;\; m = 1, ..., M, \tag{3}$$

where e_m is a vector in $\mathbb{R}^M$ that has 1 in the m^{th} place and 0s elsewhere.

Plugging (3) into (2) gives

$$\nabla_{\bar{b}} f(x^*(\bar{b}) = \sum_{m=1}^{M} \lambda_m^* e_m = \lambda^*(\bar{b}).$$

That is, $\frac{\partial f(x^*(\bar{b}))}{\partial \bar{b}_m} = \lambda_m^*(\bar{b})$.

Case II: Inequality Constraints

$$\max_{x\in\mathbb{R}^N} f(x) \tag{C.M.P.2}$$
$$\text{s.t. } g_1(x) \leq \bar{b}_1$$
$$\vdots$$
$$g_M(x) \leq \bar{b}_M$$

Remark. Problem (C.M.P.2) is a simplified version of Problem (M.K.4) in MWG. Here, the coexistence of equality constraints is ignored.

The **Constraint Set** is

$$C = \{x \in \mathbb{R}^N : g_m(x) \leq \bar{b}_m \text{ for } m = 1, ..., M\}.$$

Similar to Theorem M.K.1, we require **Constraint Qualification**: $\nabla g_m(\overline{x})$ with the binding constraints, i.e., for those $m = 1, ..., M$ such that $g_m(\overline{x}) = \bar{b}_m$ at the optimum, are linearly independent.

Theorem M.K.2 (Kuhn-Tucker Conditions)**.** *Suppose that $\overline{x} \in C$ is a local maximizer of problem (C.M.P.2). Assume also that the constraint qualification is satisfied. Then, there are multipliers $\lambda_m \in \mathbb{R}_+$ (Not $\mathbb{R}$), one for each inequality constraint, such that*

(i) For every $n = 1, ..., N$,

$$\frac{\partial f(\overline{x})}{\partial x_n} = \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(\overline{x})}{\partial x_n}$$

or

$$\nabla f(\overline{x}) = \sum_{m=1}^{M} \lambda_m \nabla g_m(\overline{x})$$

(No change compared to case of equality constraints, except $\lambda_m \in \mathbb{R}_+$.)

(ii) For every $m = 1, ..., M$,

$$\lambda_m(g_m(\overline{x}) - \bar{b}_m) = 0$$

i.e., $\lambda_m = 0$ for any constraint k that doesn't hold with equality.

Condition (ii) is called "*complementary slackness*" condition. It refers to the fact that one of the two inequalities $\lambda_m \geq 0$ and $g_m(\overline{x}) \leq \bar{b}_m$ must be binding.

Explanation of (i) Here, the constraint set is the entire space on or below the surfaces. Recall, in the case of **equality constraints**, $\nabla f(\overline{x})$ can be any linear combination of $\nabla g_1(\overline{x}), .., \nabla g_M(\overline{x})$. With equality constraints, we need to ensure that movement locally along the constraint set does not change $f(x)$.

Here, with **inequality constraints**, λ_m must be *non-negative* to make sure that any local movement from $\overline{x}$ into the constraint space does not have a component pointing towards $\nabla f(\overline{x})$ and thus leading to an increase in $f(\cdot)$.

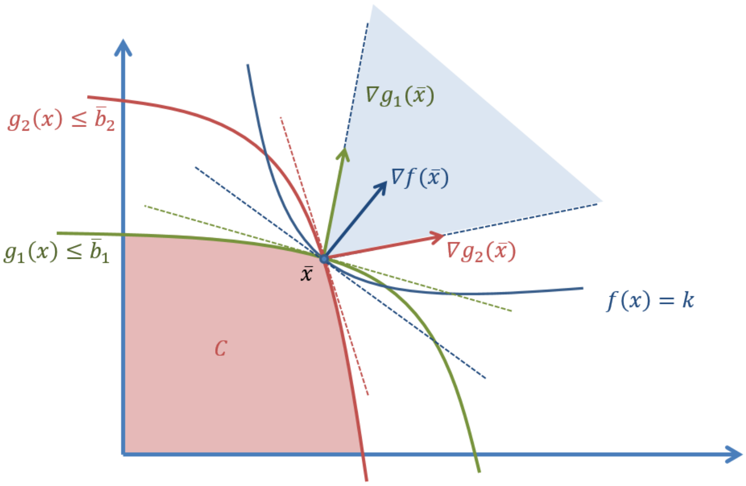

Figure 10: Kuhn-Tucker Condition (i)

Explanation of (ii) When $g_m(\overline{x}) < \overline{b}_m$, the constraint is not binding. So it doesn't affect the F.O.C locally $\implies \lambda_m = 0$.

How to use Theorem M.K.2? Define *Lagrangian function*:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{m=1}^{M} \lambda_m (g_m(x) - \overline{b}_m)$$

Kuhn-Tucker conditions give:

$$\frac{\partial f(x)}{\partial x_n} - \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(x)}{\partial x_n} = 0, \text{ for } n = 1, ..., N \qquad \text{(FOC for } x_n\text{)}$$

$$g_m(x) - \overline{b}_m \leq 0, \text{ for } m = 1, ..., M \qquad \text{(Constraints)}$$

$$\lambda_m \geq 0, \text{ for } m = 1, ..., M \qquad \text{(Non-negativity of } \lambda\text{)}$$

$$\lambda_m (g_m(x) - \overline{b}_m) = 0, \text{ for } m = 1, ..., M \qquad \text{(Complementary slackness)}$$

Second Order Condition Second order conditions for inequality problems (C.M.P.2) is exactly the same as those for equality problems (C.M.P.1). The only adjustment is that the constraints that count are those that bind, that is, those that hold with equality at the point $\overline{x}$ under consideration.[1]

Example M.K.2. Use Theorem M.K.2 to solve the following problem:

$$\max_{(x_1,x_2)\in\mathbb{R}^2} x_1^2 - x_2$$
$$\text{s.t. } x_1^2 + x_2^2 \leq 1$$

Solution. Here, $f(x) = x_1^2 - x_2, g(x) = x_1^2 + x_2^2$.

Lagrange function:

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

Kuhn-Tucker conditions:

$$\frac{\partial \mathcal{L}}{\partial x_1} = 2x_1 - 2\lambda x_1 = 0 \quad \text{(FOC for } x_1\text{)}$$
$$\frac{\partial \mathcal{L}}{\partial x_2} = -1 - 2\lambda x_2 = 0 \quad \text{(FOC for } x_2\text{)}$$
$$x_1^2 + x_2^2 \leq 1 \quad \text{(Constraint)}$$
$$\lambda \geq 0 \quad \text{(Non-negativity of } \lambda\text{)}$$
$$\lambda(x_1^2 + x_2^2 - 1) = 0 \quad \text{(C-S)}$$

From (FOC for x_1), we have $\lambda = 1$ or $x_1 = 0$.

1. $\lambda = 1$. Then from (FOC for x_2), we have $x_2 = -\frac{1}{2}$. Then from (C-S), we have $x_1 = \pm\frac{\sqrt{3}}{2}$. In this case, $f(\frac{\sqrt{3}}{2}, -\frac{1}{2}) = f(-\frac{\sqrt{3}}{2}, -\frac{1}{2}) = \frac{5}{4}$.

2. $x_1 = 0$. Then from (FOC for x_2) and (C-S), we have $\begin{cases} x_2 = -1 \\ \lambda = \frac{1}{2} \end{cases}$ or $\begin{cases} x_2 = 1 \\ \lambda = -\frac{1}{2}. \end{cases}$

   The second solution is rejected since (Non-negativity of λ) requires $\lambda \geq 0$. In this case $f(0, -1) = 1$.

[1] More accurately, we should keep the binding constraints *with strictly positive corresponding Lagrange multipliers*.

We need to apply Second Order Condition to the three candidate solutions. (The procedure is similar to Example M.K.1 and thus is omitted here.)

In the end, the solutions are $(x_1, x_2) = (\frac{\sqrt{3}}{2}, -\frac{1}{2})$ and $(x_1, x_2) = (-\frac{\sqrt{3}}{2}, -\frac{1}{2})$.

The problem in Example M.K.2 is simple and we could easily graph it.

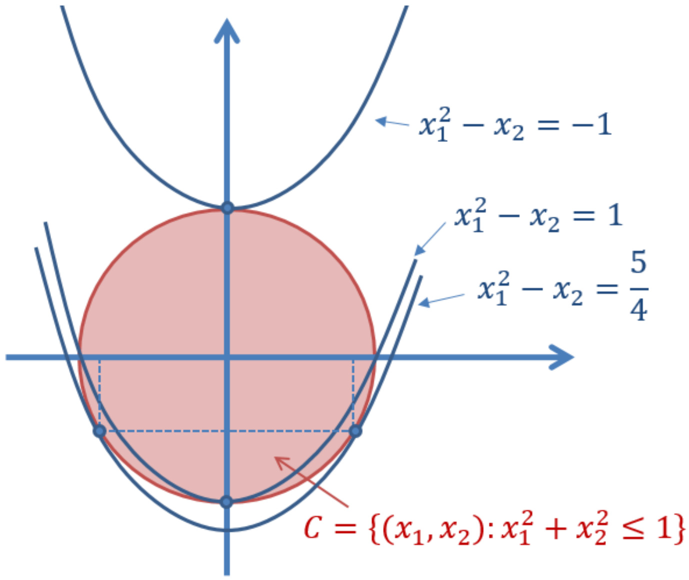

Figure 11: Example M.K.2

From Figure 11, we have the 4 critical points. And it is not hard to see that the solutions to the maximization problem are the two points that lie on the indifference curve that gives value $\frac{5}{4}$.

Adding Non-negativity Constraints If we add non-negativity constraints $x_n \geq 0, n = 1, ..., N$ to the maximization problem (C.M.P.2):

$$\begin{aligned} &\max_{x \in \mathbb{R}^N} f(x) \\ \text{s.t. } &g_1(x) \leq \bar{b}_1 \\ &\vdots \\ &g_M(x) \leq \bar{b}_M \\ &x_1 \geq 0 \\ &\vdots \\ &x_N \geq 0 \end{aligned}$$

We only need to modify Part (i) of Theorem M.K.2 to

$$\frac{\partial f(\overline{x})}{\partial x_n} \leq \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(\bar{x})}{\partial x_n}, \text{with equality if } \overline{x}_n > 0.$$

Explanation Suppose we explicitly add the non-negativity constraints, $-x_n \leq 0, n = 1, ..., N$.

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{m=1}^{M} \lambda_m (g_m(x) - \overline{b}_m) + \sum_{n=1}^{N} \lambda_{M+n} x_n \quad (\text{or} - \sum_{n=1}^{N} \lambda_{M+n}(-x_n - 0))$$

F.O.C for x_n :

$$\frac{\partial f(\overline{x})}{\partial x_n} = \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(\overline{x})}{\partial x_n} - \lambda_{M+n}, \text{ where } \lambda_{M+n} \geq 0.$$

We also need to add another corresponding complementary slackness condition

$$-\lambda_{M+n} x_n = 0.$$

(i) If $x_n = 0$, then $\lambda_{M+n} \geq 0$ & $\frac{\partial f(\overline{x})}{\partial x_n} \leq \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(\overline{x})}{\partial x_n}$;

(ii) If $x_n > 0$, then $\lambda_{M+n} = 0$ & $\frac{\partial f(\overline{x})}{\partial x_n} = \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(\overline{x})}{\partial x_n}$.

Combining the two cases, we have

$$\frac{\partial f(\overline{x})}{\partial x_n} \leq \sum_{m=1}^{M} \lambda_m \frac{\partial g_m(\overline{x})}{\partial x_n}, \text{with equality if } \overline{x}_n > 0.$$