Application: Constrained Optimization

Single equality constraint

We consider the following problem:

$$\min f(\boldsymbol{x}) \qquad \text{subject to } g(\boldsymbol{x}) = 0 \tag{1}$$

where $f, g: \mathbb{R}^N \mapsto \mathbb{R}$.

Recall that the necessary condition involving first order derivatives is the following Lagrange multiplier theory. Define the Lagrange function:

$$L(\boldsymbol{x}, \lambda) := f(\boldsymbol{x}) - \lambda\, g(\boldsymbol{x}). \tag{2}$$

If $\boldsymbol{x}_0$ is a local minimizer for the equality constrained problem (1), then there is $\lambda_0 \in \mathbb{R}$ such that $(\boldsymbol{x}_0, \lambda_0)$ is a critical point of $L(\boldsymbol{x}, \lambda)$.

Exercise 1. Prove that $(\boldsymbol{x}_0, \lambda_0)$ is neither a local minimizer nor a local maximizer of L.

Clearly, if $(\boldsymbol{x}_0, \lambda_0)$ is a critical point of L, $\boldsymbol{x}_0$ may be neither local minimizer nor local maximizer of f.

Exercise 2. Give an example illustrating the above point.

Now we try to derive second order conditions that are sufficient or necessary for $\boldsymbol{x}_0$ to be a local minimizer.

Theorem 1. *Consider the constrained minimization problem (1). Let $(\boldsymbol{x}_0, \lambda_0)$ be a critical point of $L(\boldsymbol{x}, \lambda)$. Further assume $(\operatorname{grad} g)(\boldsymbol{x}_0) \neq \mathbf{0}$. Then $\boldsymbol{x}_0$ is a local minimizer if the following holds: $G^T H_L\, G$ is positive definite at $\boldsymbol{x}_0$, where*

$$H_L = \left(\frac{\partial^2 L}{\partial x_i \partial x_j} \right)_{i,j=1}^N, \qquad G = \frac{\partial(x_1, \ldots, x_{N-1}, X_N)}{\partial(x_1, \ldots, x_{N-1})} \tag{3}$$

with X_N the implicit function determined through $g(\boldsymbol{x}) = 0$ (assuming $\frac{\partial g}{\partial x_N} \neq 0$).

Proof. Since $(\operatorname{grad} g)(\boldsymbol{x}_0) \neq \mathbf{0}$, by Implicit Function Theorem we can represent on x_i as functions of other x_j's. Wlog assume $x_N = X_N(x_1, \ldots, x_{N-1})$.

Now define

$$F(x_1, \ldots, x_{N-1}) := f(x_1, \ldots, x_{N-1}, X_N(x_1, \ldots, x_{N-1})). \tag{4}$$

Observe that $\boldsymbol{x}_0 = \begin{pmatrix} x_{01} \\ \vdots \\ x_{0N} \end{pmatrix}$ is a local minimizer for (1) if and only if $\begin{pmatrix} x_{01} \\ \vdots \\ x_{0N-1} \end{pmatrix}$ is a local minimizer of F without any constraint.

The Lagrange multiplier theory dictates that $\begin{pmatrix} x_{01} \\ \vdots \\ x_{0N-1} \end{pmatrix}$ is a critical point of F. Also recall that from

$$\frac{\partial F}{\partial x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial x_N} \frac{\partial X_N}{\partial x_i}, \qquad \frac{\partial g}{\partial x_i} + \frac{\partial g}{\partial x_N} \frac{\partial X_N}{\partial x_i} = 0 \tag{5}$$

at $\boldsymbol{x}_0$, we have

$$\lambda_0 = \left(\frac{\partial g}{\partial x_N}(\boldsymbol{x}_0) \right)^{-1} \left(\frac{\partial f}{\partial x_N}(\boldsymbol{x}_0) \right). \tag{6}$$

We calculate the second derivatives of F.

$$\frac{\partial F}{\partial x_i}(x_1,...,x_{N-1}) = \frac{\partial f}{\partial x_i}(x_1,...,x_{N-1},X_N) + \frac{\partial f}{\partial x_N}(x_1,...,x_{N-1},X_N)\,\frac{\partial X_N}{\partial x_i}(x_1,...,x_{N-1}). \tag{7}$$

Taking derivative again

$$\begin{aligned}\frac{\partial^2 F}{\partial x_i \partial x_j} &= \frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{\partial^2 f}{\partial x_i \partial x_N}\frac{\partial X_N}{\partial x_j} \\ &\quad + \left[\frac{\partial^2 f}{\partial x_j \partial x_N} + \frac{\partial^2 f}{\partial x_N^2}\frac{\partial X_N}{\partial x_j}\right]\frac{\partial X_N}{\partial x_i} \\ &\quad + \frac{\partial f}{\partial x_N}\frac{\partial^2 X_N}{\partial x_i \partial x_j}.\end{aligned} \tag{8}$$

Now using $\frac{\partial g}{\partial x_i} + \frac{\partial g}{\partial x_N}\frac{\partial X_N}{\partial x_i} = 0 \Longrightarrow \frac{\partial X_N}{\partial x_i} = -\left(\frac{\partial g}{\partial x_N}\right)^{-1}\left(\frac{\partial g}{\partial x_i}\right)$ the above becomes

$$\begin{aligned}\frac{\partial^2 F}{\partial x_i \partial x_j} &= \frac{\partial^2 f}{\partial x_i \partial x_j} - \left(\frac{\partial g}{\partial x_N}\right)^{-1}\left[\frac{\partial^2 f}{\partial x_i \partial x_N}\frac{\partial g}{\partial x_j} + \frac{\partial^2 f}{\partial x_j \partial x_N}\frac{\partial g}{\partial x_i}\right] \\ &\quad + \left(\frac{\partial g}{\partial x_N}\right)^{-2}\frac{\partial^2 f}{\partial x_N^2}\frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j} + \frac{\partial f}{\partial x_N}\frac{\partial^2 X_N}{\partial x_i \partial x_j}.\end{aligned} \tag{9}$$

Now differentiating $\frac{\partial g}{\partial x_i} + \frac{\partial g}{\partial x_N}\frac{\partial X_N}{\partial x_i} = 0$ we have

$$\begin{aligned}0 &= \frac{\partial^2 g}{\partial x_i \partial x_j} + \frac{\partial^2 g}{\partial x_i \partial x_N}\frac{\partial X_N}{\partial x_j} + \left[\frac{\partial^2 g}{\partial x_j \partial x_N} + \frac{\partial^2 g}{\partial x_N^2}\frac{\partial X_N}{\partial x_j}\right]\frac{\partial X_N}{\partial x_i} + \frac{\partial g}{\partial x_N}\frac{\partial^2 X_N}{\partial x_i \partial x_j} \\ &= \frac{\partial^2 g}{\partial x_i \partial x_j} - \left(\frac{\partial g}{\partial x_N}\right)^{-1}\left[\frac{\partial^2 g}{\partial x_i \partial x_N}\frac{\partial g}{\partial x_j} + \frac{\partial^2 g}{\partial x_j \partial x_N}\frac{\partial g}{\partial x_i}\right] \\ &\quad + \left(\frac{\partial g}{\partial x_N}\right)^{-2}\frac{\partial^2 g}{\partial x_N^2}\frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j} + \frac{\partial g}{\partial x_N}\frac{\partial^2 X_N}{\partial x_i \partial x_j}.\end{aligned} \tag{10}$$

which gives

$$\begin{aligned}\frac{\partial^2 X_N}{\partial x_i \partial x_j} &= -\left(\frac{\partial g}{\partial x_N}\right)^{-1}\left[\frac{\partial^2 g}{\partial x_i \partial x_j} - \left(\frac{\partial g}{\partial x_N}\right)^{-1}\left[\frac{\partial^2 g}{\partial x_i \partial x_N}\frac{\partial g}{\partial x_j} + \frac{\partial^2 g}{\partial x_j \partial x_N}\frac{\partial g}{\partial x_i}\right] + \right. \\ &\quad \left. \left(\frac{\partial g}{\partial x_N}\right)^{-2}\frac{\partial^2 g}{\partial x_N^2}\frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\right].\end{aligned} \tag{11}$$

Substituting into (9) we reach (denote $\lambda := \left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial f}{\partial x_N}$)

$$\begin{aligned}\frac{\partial^2 F}{\partial x_i \partial x_j} &= \frac{\partial^2 f}{\partial x_i \partial x_j} - \lambda\frac{\partial^2 g}{\partial x_i \partial x_j} \\ &\quad - \left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial g}{\partial x_j}\left(\frac{\partial^2 f}{\partial x_i \partial x_N} - \lambda\frac{\partial^2 g}{\partial x_i \partial x_N}\right) \\ &\quad - \left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial g}{\partial x_i}\left(\frac{\partial^2 f}{\partial x_j \partial x_N} - \lambda\frac{\partial^2 g}{\partial x_j \partial x_N}\right) \\ &\quad + \left(\frac{\partial g}{\partial x_N}\right)^{-2}\frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\left(\frac{\partial^2 f}{\partial x_N^2} - \lambda\frac{\partial^2 g}{\partial x_N^2}\right).\end{aligned} \tag{12}$$

Recalling the definition of the Lagrange function, we reach

$$\begin{aligned}\frac{\partial^2 F}{\partial x_i \partial x_j} &= \frac{\partial^2 L}{\partial x_i \partial x_j} - \left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial g}{\partial x_j}\frac{\partial^2 L}{\partial x_i \partial x_N} \\ &\quad - \left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial g}{\partial x_i}\frac{\partial^2 L}{\partial x_j \partial x_N} + \left(\frac{\partial g}{\partial x_N}\right)^{-2}\frac{\partial g}{\partial x_i}\frac{\partial g}{\partial x_j}\frac{\partial^2 L}{\partial x_N^2}.\end{aligned} \tag{13}$$

This leads to the following matrix relation

$$\left(\frac{\partial^2 F}{\partial x_i \partial x_j}\right) = \left(\frac{\partial g}{\partial x_N}\right)^{-2} G^T H_L G \tag{14}$$

where $H_L = \left(\frac{\partial^2 L}{\partial x_i \partial x_j}\right)_{i,j=1}^N$, and

$$G := \begin{pmatrix} 1 & 0 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & & 0 \\ 0 & 0 & \cdots & 0 & & 1 \\ -\left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial g}{\partial x_1} & \cdots & \cdots & -\left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial g}{\partial x_{N-2}} & & -\left(\frac{\partial g}{\partial x_N}\right)^{-1}\frac{\partial g}{\partial x_{N-1}} \end{pmatrix} = \frac{\partial(x_1, \ldots, x_{N-1}, X_N)}{\partial(x_1, \ldots, x_{N-1})}. \tag{15}$$

Thus ends the proof. □

Remark 2. Again, in fact $\boldsymbol{x}_0$ is a strict local minimizer.

Remark 3. The positive definiteness of $G^T H_L G$ is equivalent to

$$\boldsymbol{v}^T H_L \boldsymbol{v} > 0 \tag{16}$$

for every $\boldsymbol{v} \in \mathbb{R}^N$ that is a tangent vector of the surface $g(\boldsymbol{x}) = 0$.

Remark 4. Note that the following is not sufficient for $\boldsymbol{x}_0$ to be a local minimizer for the constrained optimization problem (1):

> $(\boldsymbol{x}_0, \lambda_0)$ is a critical point for $L(\boldsymbol{x}, \lambda)$, and for every $\boldsymbol{v} \in \mathbb{R}^N$ tangent to $g(\boldsymbol{x}) = 0$, $\boldsymbol{v}^T H \boldsymbol{v} > 0$ where $H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}_0)\right)$.

Exercise 3. Give an example justifying the above remark. (Hint: Consider $g(x, y) = y - x^2$).

Exercise 4. Prove that if g is linear, then the claim

> $(\boldsymbol{x}_0, \lambda_0)$ is a critical point for $L(\boldsymbol{x}, \lambda)$, and for every $\boldsymbol{v} \in \mathbb{R}^N$ tangent to $g(\boldsymbol{x}) = 0$, $\boldsymbol{v}^T H \boldsymbol{v} > 0$ where $H = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}_0)\right)$.

is indeed true.

Question 5. *Derive the theory for general equality constrained problem:*

$$\min f(\boldsymbol{x}) \qquad \textit{subject to } \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}. \tag{17}$$

Question 6. *Prove the following result from [H. Hancock, Theory of Maxima and Minima, Dover, New York, 1960]: A a matrix A satisfies $\boldsymbol{v}^T A \boldsymbol{v} > 0 (\geqslant 0)$ for every $\boldsymbol{v}$ satisfying $G \boldsymbol{v} = \boldsymbol{0}$ if and only if all solutions to*

$$\det \begin{pmatrix} A - z I & G^T \\ G & 0 \end{pmatrix} = 0 \tag{18}$$

are positive (non-negative). Here $G \in \mathbb{R}^{M\times N}$. Discuss how this result can be applied to checking optimality of critical points. Note that (18) is an algebraic equation in z of order $N-M$.

Exercise 5. (S. S. Rao, *Engineering Optimization: Theory and Practice, 2009*) Apply the above result to solve

$$\max f(x,y) = \pi x^2 y \qquad \text{subject to } 2\pi x^2 + 2\pi x y = 24\pi. \tag{19}$$

(Solution: $(2,4)$.)

Single inequality constraint and KKT conditions

Now we consider the problem

$$\min f(\boldsymbol{x}) \qquad \text{subject to } g(\boldsymbol{x}) \geqslant 0. \tag{20}$$

Then if $\boldsymbol{x}_0$ is a local minimizer, we have to discuss two cases:

1. $g(\boldsymbol{x}_0) > 0$ (the constraint is "not active");

2. $g(\boldsymbol{x}_0) = 0$ (the constraint is "active");

We discuss the two cases. The discussion in this section will not be fully rigorous.

- $g(\boldsymbol{x}_0) > 0$. In this case there is $r > 0$ such that $B(\boldsymbol{x}_0.r) \subseteq \{\boldsymbol{x} | \, g(\boldsymbol{x}) \geqslant 0\}$ and therefore the condition is the same as unconstrained minimization:

  $\boldsymbol{x}_0$ is a local minimizer if

  1. $\boldsymbol{x}_0$ is a critical point for f: $(\text{grad } f)(\boldsymbol{x}_0) = \mathbf{0}$;

  2. The Hessian matrix of f, $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}_0)\right)$ is positive definite.

  On the other hand, if $\boldsymbol{x}_0$ is a local minimizer, then

  1. $\boldsymbol{x}_0$ is a critical point for f: $(\text{grad } f)(\boldsymbol{x}_0) = \mathbf{0}$;

  2. The Hessian matrix of f is positive semi-definite.

- $g(\boldsymbol{x}_0) = 0$. In this case the situation is more complicated. To obtain sufficient conditions, we realize that

  1. $\boldsymbol{x}_0$ must be a local minimizer for the equality constrained problem:

  $$\min f(\boldsymbol{x}) \qquad \text{subject to } g(\boldsymbol{x}) = 0. \tag{21}$$

  This can be guaranteed by requiring

  a. There is $\lambda_0 \in \mathbb{R}$ such that $(\text{grad } f)(\boldsymbol{x}_0) = \lambda_0 \, (\text{grad } g)(\boldsymbol{x}_0)$;

  b. For every $\boldsymbol{v}$ tangent to $g(\boldsymbol{x}) = 0$ at $\boldsymbol{x}_0$, that is for every $\boldsymbol{v} \perp (\text{grad } g)(\boldsymbol{x}_0)$, we have

  $$\boldsymbol{v}^T \left(\frac{\partial f}{\partial x_i \partial x_j} - \lambda_0 \frac{\partial g}{\partial x_i \partial x_j} \right)_{i,j=1}^{N} \boldsymbol{v} > 0. \tag{22}$$

2. There is $r > 0$ such that for all $\boldsymbol{x} \in B(\boldsymbol{x}_0, r) \cap \{\boldsymbol{x} | \; g(\boldsymbol{x}) > 0\}$, $f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}_0)$. This can be guaranteed by requiring

$$\frac{\partial f}{\partial \boldsymbol{v}} > 0 \tag{23}$$

for every $\boldsymbol{v}$ "pointing into" $\{\boldsymbol{x} | \; g(\boldsymbol{x}) > 0\}$. Such $\boldsymbol{v}$ can be characterized by

$$\boldsymbol{v} \cdot (\operatorname{grad} g)(\boldsymbol{x}_0) > 0. \tag{24}$$

Recalling $(\operatorname{grad} f)(\boldsymbol{x}_0) = \lambda_0 \, (\operatorname{grad} g)(\boldsymbol{x}_0)$, we see that this is equivalent to $\lambda_0 > 0$.

Exercise 6. Prove that if

$$\frac{\partial f}{\partial \boldsymbol{v}} > 0 \tag{25}$$

for every $\boldsymbol{v}$ satisfying

$$\boldsymbol{v} \cdot (\operatorname{grad} g)(\boldsymbol{x}_0) > 0. \tag{26}$$

then for all $\boldsymbol{x} \in B(\boldsymbol{x}_0, r) \cap \{\boldsymbol{x} | \; g(\boldsymbol{x}) > 0\}$, $f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}_0)$.

One way to summarize the above is as follows. $\boldsymbol{x}_0$ is a local minimizer for

$$\min f(\boldsymbol{x}) \qquad \text{subject to } g(\boldsymbol{x}) \geqslant 0. \tag{27}$$

if the following are satisfied: There exists $\lambda_0 \in \mathbb{R}$ such that

i. $(\boldsymbol{x}_0, \lambda_0)$ is a critical point of the Lagrange function $L(\boldsymbol{x}, \lambda) := f(\boldsymbol{x}) - \lambda \, g(\boldsymbol{x})$;

ii. $\lambda_0 \geqslant 0$;

iii. $g(\boldsymbol{x}_0) \geqslant 0$;

iv. $\lambda_0 \, g(\boldsymbol{x}_0) = 0$; $\lambda_0, g(\boldsymbol{x}_0)$ not both 0.

v. The Hessian matrix of f at $\boldsymbol{x}_0$ is positive definite if $\lambda_0 = 0$; The matrix $\left(\frac{\partial^2 L}{\partial x_i \partial x_j} \right)_{i,j=1}^{N}$ satisfies

$$\boldsymbol{v}^T \left(\frac{\partial^2 L}{\partial x_i \partial x_j} \right) \boldsymbol{v} > 0 \tag{28}$$

for all $\boldsymbol{v}$ satisfying $\boldsymbol{v} \cdot (\operatorname{grad} g)(\boldsymbol{x}_0) = 0$.

Problem 1. (S. S. Rao, *Engineering Optimization: Theory and Practice*, 2009) Solve

$$\max f(x, y) = 2\,x + y + 10 \qquad \text{subject to } x + 2\,y^2 = 3. \tag{29}$$

Discuss the effect of changing the right hand side of the constraint to the optimum value of f.

General KKT conditions

The analysis in the previous section can be readily generalized to the following general constrained optimization:

$$\min f(\boldsymbol{x}) \qquad \text{subject to } \boldsymbol{g}(\boldsymbol{x}) \geqslant \boldsymbol{0}, \qquad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \tag{30}$$

where $\boldsymbol{g}$: $\mathbb{R}^N \mapsto \mathbb{R}^M$ and h: $\mathbb{R}^N \mapsto \mathbb{R}^K$. All functions are assumed to be having continuous second order derivatives.

Remark 7. Note that one can replace the K equality constraints $\boldsymbol{h}(\boldsymbol{x}) = 0$ by $2\,K$ inequality constraints $\boldsymbol{h}(\boldsymbol{x}) \geqslant \mathbf{0}$ and $\boldsymbol{h}(\boldsymbol{x}) \leqslant \mathbf{0}$.

The following set of conditions are called KKT (Karush-Kuhn-Tucker) conditions.

- Sufficient conditions. $\boldsymbol{x}_0$ is a local minimizer if there are $\boldsymbol{\lambda}_0 \in \mathbb{R}^M$ and $\boldsymbol{\mu}_{\mathbf{0}} \in \mathbb{R}^K$ such that

    1. **(Feasibility)** $\boldsymbol{g}(\boldsymbol{x}_0) \geqslant \mathbf{0}, \boldsymbol{h}(\boldsymbol{x}_0) = \mathbf{0}$;

    2. **(Criticality)** $\mathrm{grad}_{\boldsymbol{x}} L(\boldsymbol{x}_0, \boldsymbol{\lambda}_0, \boldsymbol{\mu}_0) = \mathbf{0}$ where

    $$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\boldsymbol{x}) - \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{\mu}^T \boldsymbol{h}(\boldsymbol{x}) \tag{31}$$

    and $\mathrm{grad}_x L := \begin{pmatrix} \frac{\partial L}{\partial x_1} \\ \vdots \\ \frac{\partial L}{\partial x_N} \end{pmatrix}$;

    3. $\boldsymbol{\lambda}_{\mathbf{0}} \geqslant \mathbf{0}$;

    4. **(Strict complementarity)** $\lambda_i\, g_i(\boldsymbol{x}_0) = 0$ for every $i = 1, 2, ..., M$; Furthermore for each i, exactly one of λ_i, g_i is 0.

    5. **(Second order condition)** Let $A \subseteq \{1, 2, ..., M\}$ be the set of "active" inequality constraints, that is $i \in A \Longleftrightarrow g_i(\boldsymbol{x}_0) = 0$. Then for every $\boldsymbol{v}$ such that $\forall i \in A, \quad \boldsymbol{v}^T (\mathrm{grad}\, g_i)(\boldsymbol{x}_0) = 0$,

    $$\boldsymbol{v}^T \left(\frac{\partial^2 L}{\partial x_i \partial x_j} \right) \boldsymbol{v} > 0. \tag{32}$$

- Necessary conditions. Change strictly complementarity to "complementarity": $\lambda_i\, g_i(\boldsymbol{x}_0) = 0$ for every $i = 1, 2, ..., M$; And change the >0 in (32) to $\geqslant 0$.

Remark 8. The (first order) KKT conditions take the form of solving a system of nonlinear equations. As a consequence one can invoke popular methods such as Newton's method to find the critical points. This is the idea behind the so-called "Interior point revolution" in Optimization Theory which lies behind much progress in the past half century in linear and convex programming.

Problem 2. (S. S. Rao, *Engineering Optimization: Theory and Practice*, 2009) Consider

$$\max f(x, y) = (x - 1)^2 + y^2 \tag{33}$$

subject to

$$g_1(x, y) = x^3 - 2\,y \leqslant 0, \qquad g_2(x, y) = x^3 + 2\,y \leqslant 0. \tag{34}$$

Determine whether the KKT conditions are satisfied at the maximizer.