Chapter 7

Extremal Problems

No matter in theoretical context or in applications many problems can be formulated as problems of finding the maximum or minimum of a function. Whenever this is the case, advanced calculus comes into play. In Section 7.1 relevant terminologies are reviewed. In Section 7.2 unconstrained extremal problems are treated. In Section 7.3 Taylor's Expansion Theorem for functions of several variables is established and subsequently applied to the study of local extrema. Next we turn to constrained extremal problems. Problems with a single constraint is studied in Section 7.4 where the basic Hölder Inequality is proved as an application. Problems with multiple constraints are studied in Section 7.5, where the theorem on reduction to principal axes is proved as an application.

7.1 Extrema and Local Extrema

Recall in calculus we learned how to use derivatives to find the maximum and minimum of a function over some interval. The terminologies introduced there still make sense for functions of several variables. Let us recall these definitions but now in the higher dimensional setting.

Let f be a real-valued function defined in a non-empty set S in $\mathbb{R}^n$. A point $p \in S$ is called a **minimum point** and **maximum point** of f if

$$f(p) \leq f(x) \ , \quad \text{and} \ \ f(p) \geq f(x), \quad \forall x \in S \ ,$$

respectively. The corresponding value is called the **minimum** or **minimum value** (resp. **maximum** or **maximum value**) of f. This value is called a **strict minimum**(resp. **strict maximum**) if the inequality sign becomes strict unless $x = p$ in the definition above. The point p is called a **local minimum point** and **local maximum point** of f

if there exists some ball B containing p such that $f(p) \leq f(x)$ and $f(p) \geq f(x)$) for all $x \in S \cap B$ respectively. A minimum or maximum point is called an **extremum** or an **extremum point**. A strict extremum point and a local extremum point are similarly defined. In some occasions it is necessary to stress the difference between an extremum and a local extremum, the terminologies a **global extremum** or an **absolute extremum** will be used for an extremum. Whenever we say f attains its maximum or minimum in S, it means that there is a maximum point or minimum point for f in S.

An interior point x of S is called a **critical point** if all (first) partial derivatives of f exist and vanish at x, that is, its gradient is a zero vector $\nabla f(x) = (0, \cdots, 0)$. The following theorem tells us how to find interior extremum points.

Theorem 7.1. *Let p be an interior point of f. If it is a local extremum where the partial derivatives of f exist, it must be a critical point of f.*

Proof. Assume that p is a local minimum, say. For a fixed $j \in \{1, \cdots, n\}$, the function $\varphi(t) = f(p + te_j)$ is well-defined for all small t and has a local minimum at 0. Hence $\varphi'(0) = 0$ according to calculus. By definition, we have

$$\begin{aligned}
\frac{\partial f}{\partial x_j}(p) &= \lim_{t \to 0} \frac{f(p + te_j) - f(p)}{t} \\
&= \lim_{t \to 0} \frac{\varphi(t) - \varphi(0)}{t} \\
&= \varphi'(0) \\
&= 0 \ ,
\end{aligned}$$

so p is a critical point of f. □

Some remarks are in order.

First, two conditions are required for a critical point, namely, the existence of the partial derivatives and the vanishing of the gradient at this point. For instance, the function $f(x) = |x|$ attains minimum at 0 but 0 is not a critical point of f. Why? f is not differentiable at 0.

Next, consider the function $g(x) = x$ in $[0, 1]$. It attains maximum and minimum at 1 and 0 respectively. However, $g'(x) \equiv 1$. It shows that the interior point condition in this theorem is necessary.

Third, consider the function $h(x) = x^3$. It has a critical point at 0, but 0 is neither a minimum nor a maximum point. A non-extremal critical point is called a **saddle** or a **saddle point**. Now let us look at some examples in two variables.

Example 7.1. Consider the function

$$f(x,y) = x^2 + y^2 + 5 \ , \quad (x,y) \in \overline{B} \equiv \overline{B_1((0,0))} \ .$$

We have $\nabla f(x,y) = (2x, 2y) = (0,0)$ if and only if $(x,y) = (0,0)$. So $(0,0)$ is the unique critical point of f in $\overline{B}$. It is clear that it is the minimum of f over $\overline{B}$ with minimal value 5. On the other hand, it is also clear that the maximum of f attains at every point on the boundary of $\overline{B}$ with maximal value 6.

Example 7.2. Consider the function

$$g(x,y) = \sqrt{x^2 + y^2} \ , \quad (x,y) \in \overline{B} \ ,$$

where $\overline{B}$ is given as in the previous example. Although it is clear that $(0,0)$ is the minimum point with minimal value 0, it is not a critical point, for $g(x,0) = |x|$ and the partial derivative g_x does not exist at $(0,0)$.

Example 7.3. Consider

$$h(x,y) = x^2 - y^2 \ , \quad (x,y) \in \mathbb{R}^2 \ .$$

Since $h(x,0) \to \infty$ as $x \to \infty$ and $h(0,y) \to -\infty$ as $y \to \infty$, this function is unbounded from above and from below. Consequently it has no absolute minimum nor absolute maximum. On the other hand, $\nabla h(x,y) = (2x, -2y)$ shows that $(0,0)$ is the unique critical point of h. As $h(x,0) > 0$ for all non-zero x and $h(0,y) < 0$ for all non-zero y, h assumes both positive and negative values around origin. Therefore, the critical point $(0,0)$ is neither a local minimum nor a local maximum. It is a saddle.

Example 7.4. Consider

$$j(x,y) = \frac{3}{4}y^2 + \frac{1}{24}y^3 - \frac{1}{32}y^4 - x^2 \ , \quad (x,y) \in \mathbb{R}^2 \ .$$

It admits three critical points, namely,

$$(0,-3), \quad (0,0), \quad (0,4) \ ,$$

where $(0,0)$ is a local minimum and $(0,-3), (0,4)$ are local maximum. The function tends to $-\infty$ as $|(x,y)| \to \infty$, so it does not have any absolute minimum. On the other hand, it does have an absolute maximum. Since every interior maximum must be a critical point. This absolute maximum point is either $(0,-3)$ or $(0,4)$. Observing $j(0,-3) = 99/32$ is less than $j(0,4) = 20/3$, we conclude that $(0,4)$ is the absolute maximum of j.

7.2 Unconstrained Extremal Problems

In an extremal problem one is asked to find the extremum of a given function over some set. Problems of this type come up both in theory and applications. For practical problems, people certainly would like to find the extremum points, determine the extremal values, and design algorithms to locate them in an efficient way. The existence of an extremum is usually not much a question as it is clear from the context. However, at the theoretical level the existence of an extremum is an important issue. When the function or the set under consideration is very complex, it is not clear whether the extremum is attained or not. In view of this, we separate the issue into two questions:

Question A. Is there any extremum ?

Question B. How can we find these extrema ?

Here we kick off with the case in which a function defined in a closed, bounded set. It may be regarded as the simplest case since Question A has a positive answer. Recall that a closed, bounded set is called a **compact set**. The following result is a basic one.

Theorem 7.2 (Extremal Value Theorem). *Let S be a compact set in $\mathbb{R}^n$ and f a continuous function in S. Then it must attain its minimum point and maximum point in S.*

Backing up by this theorem, we consider Question B. As in many occasions, the set under consideration is of the form $S = \overline{G} \equiv G \cup \partial G$ where G is a bounded, open set. Then $\overline{G}$ is a compact set. When f is differentiable in G, by Theorem 7.1 any interior local extremum must be a critical point of f. Hence by solving the system of equations given by $\partial f/\partial x_j(x) = 0$, $j = 1, \cdots, n$, we can determine all critical points of f which are candidates for extrema. Conceivably there may be some extrema on the boundary of G. Together with these boundary extrema, we can determine which are maxima or minima by comparing their values. In many cases, finding boundary extrema reduces to an extremal problem of lower dimension. The following two examples illustrate how one can proceed.

Example 7.5. Find the maximum and minimum of the function

$$f(x, y) = x^2 - xy - y^2 + x - 2y + 3 \ , \ (x, y) \in \overline{R} \ ,$$

where R is the square $[-1, 1] \times [-1, 1]$. By Extremal Value Theorem its extrema must be attained. To find the critical points, we calculate its gradient and set it to zero:

$$\frac{\partial f}{\partial x} = 2x - y + 1 = 0, \quad \frac{\partial f}{\partial y} = -x - 2y - 2 = 0 \ .$$

This system admits a single solution $(-4/5, -3/5)$ which lies inside R, so it is the unique critical point of f. On the other hand, the boundary of this rectangle consists of four line segments $l_j, j = 1, \cdots, 4$. Along $l_1, (1, y), y \in [-1, 1]$, the function becomes $\varphi(t) \equiv f(1, t) = -t^2 - 3t + 5$. By looking at φ' we see that φ is strictly decreasing for $t \in [-1, 1]$, so only $(1, -1)$ and $(1, 1)$ are candidates for extremum points for f. Next along l_2, the function $\eta(t) \equiv f(t, 1) = t^2, t \in [-1, 1]$, has a minimum at $t = 0$, so $(0, 1)$ together with the end points $(1, 1), (1, -1)$ are candidates for extreme points. A similar analysis on l_3 and l_4 shows that other possible extremum points are $(-1, -1/2), (-1, 1), (-1, -1)$. Now a direct evaluation gives

$$f\left(-\frac{4}{5}, -\frac{3}{5}\right) = \frac{122}{25}, \quad f(1, 1) = 1, \quad f(0, 1) = 0,$$

$$f(-1, 1) = 1, \quad f\left(-1, \frac{1}{2}\right) = \frac{15}{4}, \quad f(-1, -1) = 3, \quad f(1, -1) = 6 \ .$$

Hence the minimum of f over $\overline{R}$ is attained at $(0, 1)$ with minimum value 0, and the maximum is attained at $(1, -1)$ with maximal value 6.

Example 7.6. A delivery company accepts only rectangular boxes the sum of whose length and girth (the perimeter of a cross-section) does not exceed 270 cm. Find the dimensions of an acceptable box of largest volume. Let x, y and z be the length, width and height of the box respectively. We want to maximize the volume which is equal to xyz under the condition $x + 2y + 2z = 270$. By writing the condition as $x = 270 - 2y - 2z$, we remove x in xyz and the problem reduces to maximize the function

$$g(y, z) = (270 - 2y - 2z)yz \ ,$$

over the triangle $\overline{D}$ where

$$D = \{(y, z) : \ y, z > 0, \quad 2y + 2z < 270 \ \} \ .$$

By solving $\partial g/\partial y = 0$ and $\partial g/\partial z = 0$ we get four solutions $(0, 0), (0, 135), (135, 0)$ and $(45, 45)$. Only $(45, 45)$ is the critical point in D. It is obvious that g vanishes along the boundary of D. Since g is a non-negative function, $(45, 45)$ is the maximum and the boundary consists of minimum points. The dimensions of the desired box are 90 (length), 45 (width) and 45 (height).

When the underlying set is not compact, Question A must be studied prior to Question B. Let us look at some one dimensional cases. Consider the following differentiable functions defined on $\mathbb{R}$:

$$f_1(x) = x^3, \quad f_2(x) = 3 - x^2, \quad f_3(x) = (1 + x^2)^{-1}, \quad f_4(x) = x^2 e^{-x^2} \ .$$

It is clear that f_1 goes to $\pm\infty$ as $x \to \pm\infty$, hence it does not have any extremum. The function f_2 is bounded from above but goes to $-\infty$ as $|x| \to \infty$. It has a unique maximum point at 0 but no minimum. The function f_3 is bounded between 0 and 1, tends to 0 as $|x| \to \infty$ but never attains 0. It has a unique maximum at 0 but no minimum. Finally, f_4 is a non-negative function which tends to 0 at infinity, but it has two maximum points ± 1 and a unique minimum at 0. The situation is much more complicated in higher dimensions.

From these examples we can see that the behavior of the functions at infinity plays an important role in the attempt to answer Question A. To obtain any meaningful result boundary behavior must be taken into account.

Theorem 7.3. *Let f be a continuous function in $\mathbb{R}^n$. Suppose that there exist some real numbers α, a ball B and a point $p \in B$ such that $f(x) \geq \alpha$ (resp. $f(x) \leq \alpha$) for all x outside B and $f(p) < \alpha$ (resp. $f(p) > \alpha$) Then the minimum (resp. maximum) of f over $\mathbb{R}^n$ is attained in B.*

Proof. If suffices to prove the case of minimum. Applying Theorem 7.1 to the function f restricted to the compact set $\overline{B}$, there is some $w \in \overline{B}$ such that $f(w) \leq f(x)$ for all $x \in \overline{B}$. In particular, taking $x = p$ we see that $f(w) \leq f(p) < \alpha$. Now, for x lying outside B, $f(x) \geq \alpha > f(w)$. We conclude that w is the minimum point of f over the entire space. $\square$

A function f in $\mathbb{R}^n$ is called **tends to ∞ uniformly as** $x \to \infty$ if for each $M > 0$ there corresponds some R such that $f(x) \geq M$ for all $x \in \mathbb{R}^n \setminus B_R(0)$. Similarly it **tends to $-\infty$ uniformly** if for each $M > 0$ there corresponds some R such that $f(x) \leq -M$ for all $x \in \mathbb{R}^n \setminus B_R(0)$. It is clear we have

Corollary 7.4. *Let f be a continuous function which tends to ∞ uniformly as $x \to \infty$. Then it attains its minimum.*

Proof. Let $\alpha = f(0) + 1$. As f tends to ∞ uniformly, there is some R such that $f(x) > \alpha$ for all $x, |x| \geq R$. The corollary follows from Theorem 7.2.

$\square$

Corollary 7.5. *Let $f = g + h$ where g and h are continuous in $\mathbb{R}^n$. Suppose that*

(a) $g(x) \geq \rho |x|^\alpha \ , \quad |x| \geq R,$ *for some* $\rho, \alpha > 0$.

(b) $\lim_{|x|\to\infty} \dfrac{|h(x)|}{|x|^\alpha} = 0 \ .$

Then f attains its minimum in $\mathbb{R}^n$.

Proof. First of all, fix an R such that $|h(x)| \le \frac{\rho}{2}|x|^\alpha$ for all $x, |x| \ge R$. We have

$$\begin{aligned} f(x) = g(x) + h(x) &\ge \rho|x|^\alpha - \frac{\rho}{2}|x|^\alpha \\ &\ge \frac{\rho}{2}|x|^\alpha, \quad \forall x, |x| \ge R \ , \end{aligned}$$

which shows that f tends to ∞ uniformly. □

One obtains corresponding statements for the maximum in these corollaries by looking at $-f$ instead of f.

Example 7.7. Find the distance from the origin to the plane $x + y + z = 1$. Well, the square of the distance, given by $x^2 + y^2 + z^2$, shares the same minimum point with the distance but is easier to calculate. Using $z = 1 - x - y$ to remove the variable z, the problem is reduced to finding the minimum of the function

$$f(x, y) = x^2 + y^2 + (1 - x - y)^2 = 2x^2 + 2y^2 + 2xy - 2x - 2y + 1$$

over the plane. By the splitting

$$g = 2x^2 + 2y^2 + 2xy \ge x^2 + y^2 \ , \quad h = -2x - 2y + 1 \ ,$$

Corollary 7.4 ($\alpha = 2$) asserts that the minimum is attained. To find it we set $\nabla f = (0, 0)$ to get

$$\begin{cases} 4x + 2y = 2, \\ 4y + 2x = 2 \ , \end{cases}$$

This is linear system whose solution is given by $x = y = 1/3$. Therefore, the minimum of the square of the distance function attains at the point $(1, 1, 1)/3$ and the distance is $1/\sqrt{3}$. Of course, it is the same as calculated via the formula in Chapter 2.

Lastly, we consider the problem where f is defined in some open set, continuous in its interior but not necessarily up to the boundary. Things could be very complicated and one needs to deal with the problem case by case.

A continuous function f defined in some open G is said to **tend to** ∞ (resp. $-\infty$) **uniformly at the boundary** if for each $M > 0$ there exists some open subset G' with compact closure in G such that $f(x) \ge M$ (resp. $f(x) \le -M$) for all $x \in G \setminus G'$. In practise G' is usually taken to be

$$G' = \{x \in G : \ \mathrm{dist}(x, \partial G) > \rho \ , |x| < R\},$$

for some small $\rho > 0$ and large R.

Theorem 7.6. *A function defined in the open G which tends to ∞ (resp. $-\infty$) uniformly at the boundary attains its minimum (resp. maximum).*

The proof is similar to that of Theorem 7.2 which we omit.

Example 7.8. * Find the minimum of the function $h(x,y) = xy + 2x - \log x^2 y$ over the first quadrant. Here the set is given by the unbounded, open set $D \equiv \{(x,y) : \ x, y > 0\}$ and h blows up at its boundary. We claim that h tends to ∞ uniformly as $(x,y) \to \partial D$. The proof is a bit delicate. Given $M > 0$, we first fix a small number $\delta_1 \in (0,1)$ such that $h(x,y) \geq M$ whenever $(x,y) \in D$ and $y \leq \delta_1$. For, we use the fact that $t - \log t \geq 1$ for all $t > 0$, we have

$$h(x,y) = xy + 2x - 2\log x - \log y \geq -\log y \ ,$$

so such δ_1 always exists. Next, we look at the subset $\{(x,y) : \ x > 0, y \geq \delta_1\}$. We have

$$h(x,y) \geq -2\log x - \log \delta_1$$

so there is some $\delta_2 \in (0,1)$ such that $h(x,y) \geq M$ for all $x \in (0,\delta_2), y \geq \delta_1$. Summing up, the function is greater than or equal to M in the subset $S \equiv \{(x,y) : \ y \in (0,\delta_1] \text{ or } x \in (0,\delta_2]\}$. Now, consider the set $D_1 = D \setminus S$. Each $(x,y) \in D_1$ satisfies $x > \delta_2$ and $y > \delta_2$. Therefore,

$$\begin{aligned}
h(x,y) &= \frac{xy}{2} + \frac{xy}{2} + 2x - 2\log x - \log y \\
&\geq \frac{\delta_2}{2} y + \frac{\delta_1}{2} x + 2x - 2\log x - \log y \\
&\geq \left(2 + \frac{\delta_1}{2}\right) x + \frac{\delta_2}{2} y - 2\log x - \log y \\
&\geq \rho(x+y) - 2\log x - \log y \\
&\geq \rho(x+y) - 3\log(x+y) \ , \quad \rho = \min\left\{2 + \frac{\delta_1}{2}, \frac{\delta_2}{2}\right\} \ .
\end{aligned}$$

Using the fact $\log r/r \to 0$ as $r \to \infty$, we can fix a large R such that $h(x,y) \geq M$ for all $(x,y) \in D' \equiv D_1 \setminus B_R(0)$. More precisely,

$$D' = \left\{(x,y) : \ \sqrt{x^2+y^2} < R, \ x > \delta_2, \ y > \delta_1\right\} \ .$$

We conclude that h tends to ∞ uniformly as x approaches ∂D.

To determine the minimum we take partial derivatives. It is readily found $(2, 1/2)$ is the unique minimum. Since this is the only critical point, it must be the minimum point. The minimum value of h is given by $h(2,1/2) = 5 + \log 2$. From this example you see that verifying the boundary behavior could be very nasty.

7.3 Taylor's Expansion

In some applications people are interested in determining whether a critical point is a local minimum point or a local maximum point. In the single variable case, the second derivative test is an effective method. Of course, in order to apply this test one has to assume the function under consideration is twice differentiable. It is natural to wonder if the second derivative test could be extended to the multi-dimensional situation. In fact, this is true. Recalling the proof of the second derivative test is based on Taylor's expansion, our first task is to obtain the multi-dimensional version of the Taylor's Expansion Theorem.

To simplify notation we introduce a differential operator: for a fixed $a \in \mathbb{R}^n$,

$$a \cdot \nabla f = \sum_{j=1}^{n} a_j \frac{\partial f}{\partial x_j} \ .$$

Powers of the differential operator is understood as repeated applications. For instance,

$$\begin{aligned}(a \cdot \nabla)^2 f &= \sum_{k=1}^{n} a_k \frac{\partial}{\partial x_k} \left(\sum_{j=1}^{n} a_j \frac{\partial}{\partial x_j} \right) f \\ &= \sum_{j,k=1}^{n} a_j a_k \frac{\partial^2 f}{\partial x_j x_k} \ ,\end{aligned}$$

$$\begin{aligned}(a \cdot \nabla)^3 f &= \sum_{k=1}^{n} a_k \frac{\partial}{\partial x_k} \left(\sum_{j=1}^{n} a_j \frac{\partial}{\partial x_j} \right)^2 f \\ &= \sum_{i,j,k=1}^{n} a_i a_j a_k \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} \ .\end{aligned}$$

When $n = 2$, we have

$$(a \cdot \nabla)^2 f = a_1^2 \frac{\partial^2 f}{\partial x^2} + 2 a_1 a_2 \frac{\partial^2 f}{\partial x \partial y} + a_2^2 \frac{\partial^2 f}{\partial y^2} \ ,$$

and

$$(a \cdot \nabla)^3 f = a_1^3 \frac{\partial^3 f}{\partial x^3} + 3 a_1^2 a_2 \frac{\partial^3 f}{\partial x^2 \partial y} + 3 a_1 a_2^2 \frac{\partial^3 f}{\partial x \partial y^2} + a_2^3 \frac{\partial^3 f}{\partial y^3} \ ,$$

and etc.

Theorem 7.7. *Let f be a function defined in some ball $B \subset \mathbb{R}^n$. Suppose that all partial derivatives of f up to order $k+1, k \geq 0$, are continuous in B. Then for $x, p \in B$,*

$$\begin{aligned} f(x) &= f(p) + (x-p)\cdot\nabla f(p) + \frac{((x-p)\cdot\nabla)^2 f(p)}{2!} + \cdots + \frac{((x-p)\cdot\nabla)^k f(p)}{k!} \\ &\quad + \frac{((x-p)\cdot\nabla)^{(k+1)} f(c)}{(k+1)!} \ , \end{aligned}$$

where c is a point on the line segment connecting x and p.

Proof. * Recall the one dimensional Taylor's Expansion Theorem. For φ on some interval (a,b) and $0 \in (a,b)$. Suppose that φ has continuous derivatives up to order $k+1$ on (a,b). Then

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \cdots + \frac{\varphi^{(k)}(0)}{k!}t^k + \frac{\varphi^{(k+1)}(c)}{(k+1)!}t^{k+1} \ , \tag{7.1}$$

where c lies between t and 0. For the given f, $\varphi(t) \equiv f(p + t(x-p))$ is a function of t. Note the line segment connecting x to p lies inside B. The regularity assumption on f can be translated to the regularity of φ. By the chain rule we see that the one dimensional Taylor's expansion above is valid. It suffices to express the derivatives of φ in terms of the partial derivatives of f. Indeed, by the Chain Rule

$$\varphi'(t) = (x-p)\cdot\nabla f(p + t(x-p))$$

and in general

$$\varphi^{(j)}(t) = ((x-p)\cdot\nabla)^j f(p + t(x-p)) \ , \quad j \geq 1 \ .$$

The theorem follows from the expansion formula for φ above. $\square$

Formula (7.1) will be proved in MATH2060 under more precise assumptions.

Taking $k = 1$ in the Taylor's Expansion Theorem,

$$f(x) = f(p) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(p)(x_j - p_j) + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(c)(x_i - p_i)(x_j - p_j) \ . \tag{7.2}$$

When $n = 2$, writing $x = (x, y)$ and $p = (x_0, y_0)$, we have

$$\begin{aligned} f(x,y) &= f(p_0) + \frac{\partial f}{\partial x}(p)(x - x_0) + \frac{\partial f}{\partial y}(p)(y - y_0) \\ &\quad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(c)(x - x_0)^2 + \frac{\partial^2 f}{\partial x \partial y}(c)(x - x_0)(y - y_0) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(c)(y - y_0)^2. \end{aligned}$$

The symmetric matrix

$$\nabla^2 f \equiv \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right) \ , \quad i, j = 1, \cdots, n \ ,$$

is called the **Hessian matrix** of f. By the Principal Axis Theorem, see Chapter 2, every Hessian matrix can be diagonalized by an orthogonal matrix. The diagonal entries of the resulting diagonal matrix are precisely the eigenvalues of the Hessian matrix. Now we can state the second derivative test in higher dimensions.

Theorem 7.8 (**The Second Derivative Test**). *Let p be a critical point of f. Suppose that all partial derivatives up to order two are continuous near p. Let $\lambda_j, j = 1, \cdots, n$ be the eigenvalues of the Hessian matrix of f at p.*

(a) p is a strict local minimum point if all λ_j's are positive,

(b) p is a strict local maximum point if all λ_j's are negative,

(c) p is a saddle if there are two λ_j's with different sign.

A local maximum (resp. local minimum) p is called a **strict local maximum** (resp. **strict local minimum**) if there exists an open set $U \subset G$ containing p such that $f(p) > f(x)$ (resp. $f(p) < f(x)$) for all $x \neq p, x \in U$. A critical point p is called a **saddle** if for each $r > 0$, there are points $x, y \in B_r(p)$ such that $f(x) < f(p) < f(y)$.

Proof. Let $R = (r_{ij})$ be the orthogonal matrix that diagonalizes the Hessian matrix of f. Introducing the orthogonal change of coordinates $x_i - p_i = \sum_j r_{ij} y_j$ and observing that $|x - p| = |y|$, (7.2) becomes

$$\begin{aligned} f(x) &= f(p) + \frac{1}{2} \sum_{i,j,k,m} \frac{\partial^2 f}{\partial x_i \partial x_j}(c) r_{ik} y_k r_{jm} y_m + T \\ &= f(p) + \frac{1}{2} \sum_{i,j,k,m} \frac{\partial^2 f}{\partial x_i \partial x_j}(p) r_{ik} y_k r_{jm} y_m + T_1 + T \\ &= f(p) + \sum_{k=1}^{n} \lambda_k y_k^2 + T_1 + T \ , \end{aligned}$$

where

$$T_1 = \frac{1}{2} \sum_{i,j,k,m} \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(c) - \frac{\partial^2 f}{\partial x_i \partial x_j}(p) \right) r_{ik} y_k r_{jm} y_m \ ,$$

and $|T|/|y|^2 \to 0$ as $y \to 0$. From $|T_1|/|y|^2 \to 0$ as $y \to 0$ we see that (a) and (b) hold. To prove (c), let the eigenvalues be placed in ascending order where λ_1 is negative and λ_n is positive. At points of the form $x = (x_1, p_2, \cdots, p_n)$, $y = (x_1 - p_1, 0, \cdots, 0)$, so

$$f(x) = f(p) + \lambda_1 y_1^2 + T_1 + T \ ,$$

where $(|T_1|+|T|)/|y_1|^2 \to 0$ as x tends to p. Clearly $f(x) < 0 = f(p)$ is negative for $x \neq p$ but close to it. On the other hand, by considering the point $x = (p_1, \cdots, p_{n-1}, x_n)$ and $y = (0, \cdots, 0, x_n - p_n)$, we have

$$f(x) = f(p) + \lambda_n y_n^2 + T_1 + T \ .$$

As $\lambda_n > 0$, it follows that $f(x) > f(p)$ for points x close to p which is of the form $x = (p_1, \cdots, p_{n-1}, x_n), x_n \neq p_n$. Hence p is neither a local minimum point nor a local maximum point.

□

In the two dimensional case, the eigenvalues of the Hessian matrix are found by solving a quadratic equation. Knowledge from linear algebra would be helpful in higher dimension. However, we will not elaborate on this point.

Example 7.9. Study the local extrema of the function

$$f(x,y) = 3x^2 - 6xy - 3y^2 + 2y^3 \ , \quad (x,y) \in \mathbb{R}^2 \ .$$

We have

$$f_x = 6x - 6y = 0 \ , \quad f_y = -6x - 6y + 6y^2 = 0 \ ,$$

which shows that there are two critical points $(0,0)$ and $(2,2)$. Next, the Hessian matrix of f is given by

$$\begin{bmatrix} 6 & -6 \\ -6 & -6+12y \end{bmatrix} \ .$$

Its eigenvalues at $(0,0)$ are $\pm\sqrt{72}$. According to the Second Derivative Test, $(0,0)$ is a saddle. On the other hand, the eigenvalues at $(2,2)$ are given by $12 \pm 2\sqrt{18}$ which are positive, hence $(2,2)$ is a local minimum point of f.

7.4 Constrained Extremal Problems I

In the previous section we have discussed how to find extrema of a given function in a set which is the closure of an open set in $\mathbb{R}^n$. Here we are concerned with extremal problems satisfying some constraints in these sets.

To fix the ideas, let us examine the following special case. Let f be a function defined in an open set $G \subset \mathbb{R}^3$ and g is another function in G such that the set

$$\Sigma \equiv \{(x,y,z) \in G : \ g(x,y,z) = 0\}$$

is non-empty. We would like to find the extremal points of f in Σ. The method of Lagrange multipliers provides an effective way to find local extremal points lying in the interior of Σ.

Theorem 7.9. *Settings as above, let $p \in \Sigma$ be a local extremum of f on Σ. Suppose that $\nabla g(p) \neq (0,0,0)$. There exists some $\lambda \in \mathbb{R}$ such that*

$$\nabla f(p) - \lambda \nabla g(p) = (0,0,0)\ .$$

The number λ is called the **Lagrange multiplier** of the problem. Here p is a **local minimum** (resp. **local maximum**) of f on Σ means $f(p) \leq f(x)$ or $f(p) \geq f(x)$ for $x \in \Sigma \cap B$ for some ball B containing p.

Proof. Let $p = (x_0, y_0, z_0)$. As $\nabla g(p) \neq (0,0,0)$ by assumption, without loss of generality let $\partial g/\partial z(p) \neq 0$, by Theorem 6.2 there is a differentiable function φ from a disk D containing (x_0, y_0) to some ball B containing p such that $\varphi(x_0, y_0) = z_0$ and

$$g(x, y, \varphi(x, y)) = 0, \quad (x, y) \in D\ .$$

Now, since p is a local extremum of f over Σ, the function

$$F(x, y) \equiv f(x, y, \varphi(x, y))\ , \quad (x, y) \in D\ ,$$

has a local extremum at (x_0, y_0). By the Chain Rule,

$$\frac{\partial F}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial \varphi}{\partial x} = 0\ ,$$

and

$$\frac{\partial F}{\partial y} = \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial \varphi}{\partial y} = 0\ ,$$

at (x_0, y_0). In other words,

$$\nabla f(p) \cdot (1, 0, \varphi_x(x_0, y_0)) = 0, \qquad \nabla f(p) \cdot (0, 1, \varphi_y(x_0, y_0)) = 0\ .$$

We see that ∇f is perpendicular to the tangent space of the surface Σ at p, so it must point to the normal direction or vanishes. As the normal direction of the surface is given by ∇g, we conclude that $\nabla f(p)$ and $\nabla g(p)$ must be linearly dependent, so that there exists some λ such that $\nabla f(p) = \lambda \nabla g(p)$.

□

The situation becomes clear if we take the constraint to be $g(x, y, z) = z$. Then $g = 0$ is simply the x-y-plane. A critical point p_0 of $f(x, y, z)$ on this plane satisfies $f_x(p_0) = f_y(p_0) = 0$ but there is not restriction on $f_z(p_0)$. The gradient of f at the critical point is of the form $(0, 0, f_z(p_0))$ which points in the z-direction, so $(0, 0, f_z(p_0)) = f_z(p_0)(0, 0, 1)$ where λ is equal to $f_z(p_0)$ in this case.

In view of Theorem 7.9 we call a point $p_0 \in G$ a **critical point** of f under the constraint $g = 0$ if $\nabla g(p_0) \neq (0,0,0)$ and $\nabla f(p_0) = \lambda \nabla g(p_0)$ for some real number λ. Theorem 7.9 asserts that local extrema are critical points of f under $g = 0$. Like the unconstrained case, there are non-extremal critical points such as saddles.

Example 7.10. Let us go back to Example 7.7 in which we found the distance from the origin to the plane $x+y+z=1$ by reducing it to a minimization problem for some function over the xy-plane. Now, regarding $x+y+z=1$ as a constraint, we use the Lagrange Multiplier to tackle the problem. We minimize the distance square

$$f(x,y,z)=x^2+y^2+z^2 \ ,$$

subject to the constraint

$$g(x,y,z)=x+y+z-1=0 \ .$$

By Theorem 7.8, there is some scalar λ such that $\nabla f=\lambda\nabla g$, that is,

$$2x-\lambda=0, \quad 2y-\lambda=0, \quad 2z-\lambda=0 \ ,$$

at any local extremum (x,y,z). Summing these equations up, we get $2(x+y+z)=3\lambda$. Using the constraint it yields $\lambda=2/3$. Therefore, $x=y=z=1/3$, that is, $(1,1,1)/3$ is the only critical point. Since the function $x^2+y^2+z^2$ tends to ∞ uniformly as $(x,y,z)\to\infty$, this critical point must be the minimum point. The distance is therefore given by

$$\sqrt{\frac{1}{3^2}+\frac{1}{3^2}+\frac{1}{3^2}}=\frac{1}{\sqrt{3}} \ .$$

This approach enables us once again to find the general formula of the distance from a point to a hyperplane, see exercise.

Example 7.11. The total cost of a product depends on labor and capital. Assume that the cost of producing one unit of a certain product is given by the productivity function

$$P(x,y)=20x^{0.5}y^{0.5} \ ,$$

where x stands for the number of units of labor and y for the number of units of capital. Each unit of labor costs \$40 and each unit of capital costs \$80. Now the total investment is \$1,000,000. How can we allocate the labor and capital to obtain the maximal production? Here we want to maximize the function P subject to the constraint g:

$$40x+80y=1,000,000.$$

We have $\nabla P=\lambda\nabla g$, that is,

$$10x^{-0.5}y^{0.5}=40\lambda \ , \quad 10x^{0.5}y^{-0.5}=80\lambda \ .$$

Eliminating λ from these two equations, we get $x=3y$. Plugging in the constraint, $x=10,000$ and $y=5,000$. We conclude that in order to produce the maximum units of product, one should allocate $10,000$ units of labor and $5,000$ units of capital in the production. (Let me justify the solution we got is really the maximum for those with mathematical mind. Indeed, we were considering the maximization of a non-negative, continuous function P over the line segment $40x + 80y = 1,000,000$ whose endpoints are $(25,000, 0)$ and $(0, 12, 500)$. This is a compact subset in the plane with P vanishing at its endpoints. By Maximum-Minimum Theorem its maximum must attain inside the segment.)

Observe that there are three equations in $\nabla f = \lambda \nabla g$ and yet in order to solve for four unknowns x, y, z and λ we need one more equation. In fact, the last equation comes from the constraint $g = 0$. Therefore, there are always four equations for four unknowns. Of course, now the difficulty is how to solve this system. It usually involves a system of *nonlinear* equations whose solvability resists skills from linear algebra which work only for linear systems. One needs to solve the system case by case. Exercises have been carefully chosen so that solvability comes rather handy. In real life one has to relies on computer software to find the critical points numerically.

Example 7.12. Find all extrema of

$$f(x, y, z) = xyz\ , \qquad \text{subject to} \quad g(x, y, z) \equiv \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0\ , a, b, c > 0.$$

At a local extremum (x, y, z) we have

$$\begin{cases} yz = \lambda \dfrac{2x}{a^2}, \\[2ex] xz = \lambda \dfrac{2y}{b^2}, \\[2ex] xy = \lambda \dfrac{2z}{c^2}\ . \end{cases}$$

Multiply the first equation by x, the second by y and the third by z and then sum up. Using the constraint one arrives at $xyz = 2\lambda/3$. In particular, it shows that $\lambda, x, y, z \neq 0$. Putting $xyz = 2\lambda/3$ back to these equations, we conclude that all local extrema are given by $(\pm a, \pm b, \pm c)/\sqrt{3}$. Since f is a continuous function over the ellipsoid which is compact, its global extrema must exist. By comparing the values of f at these local extrema, we see that the minima are given by

$$\frac{1}{\sqrt{3}}(-a, b, c)\ , \quad \frac{1}{\sqrt{3}}(a, -b, c)\ , \quad \frac{1}{\sqrt{3}}(a, b, -c)\ , \quad \frac{1}{\sqrt{3}}(-a, -b, -c)\ ,$$

and the maxima by

$$\frac{1}{\sqrt{3}}(a, b, c)\ , \quad \frac{1}{\sqrt{3}}(-a, -b, c)\ , \quad \frac{1}{\sqrt{3}}(-a, b, -c)\ , \quad \frac{1}{\sqrt{3}}(a, -b, -c)\ .$$

Therefore, the minimal and maximal values of f are given by

$$f_{min} = -\frac{abc}{3\sqrt{3}}\ , \quad f_{max} = \frac{abc}{3\sqrt{3}},$$

respectively.

This example has a geometric interpretation, namely, to inscribe a rectangular box with faces parallel to the x, y, z-axes inside an ellipsoid so that its volume is the largest. The volume is equal to $8xyz = \frac{8}{3\sqrt{3}}abc$.

Although it is possible to remove one variable among x, y and z and reduce the problem to an unconstrained one, the reduction is very tedious. The method of Lagrange multiplier is much more effective.

Many useful inequalities in analysis can be proved by the method of Lagrange multipliers. To illustrate this approach we establish the following fundamental Inequality, which could be regarded as a far reaching generalization of the Cauchy-Schwarz Inequality ($p = 2$).

For $p > 1$, its conjugate number q is given by $1/p + 1/q = 1$.

Theorem 7.10 (Hölder's Inequality). ** For $a, b \in \mathbb{R}^n$, both with non-negative components,*

$$\sum_{j=1}^n a_j b_j \le \left(\sum_{j=1}^n a_j^p \right)^{1/p} \left(\sum_{j=1}^n b_j^q \right)^{1/q} ,$$

and equality sign in this inequality holds if and only if either (a) a or b is $(0, 0, \cdots, 0)$ or (b)

$$(b_1^q, \cdots, b_n^q) = \lambda(a_1^p, \cdots, a_n^p)$$

for some $\lambda \neq 0$.

Proof. * The idea of proof is to minimize the function

$$f(x) = \sum_{j=1}^n |x_j|^q$$

over the hyperplane H:

$$\sum_{j=1}^n a_j x_j = 1,$$

where $a \neq (0, \cdots, 0)$. It is clear that f tends to infinity uniformly as $|x| \to \infty$, so the minimum must attain. At this minimum point z we have

$$\nabla f(x) - \lambda \nabla \left(\sum_{j=1}^{n} a_j x_j - 1 \right) = (0, \cdots, 0) \ .$$

The equations are given by

$$q|z_j|^{q-1} \mathrm{sgn} z_j = \lambda a_j \ , \quad j = 1, \cdots, n \ .$$

(Here we have used the formula $d|x|/dx = \mathrm{sgn}\ x$ where the sign function sgn $x = 1, -1$ or 0 according to $x > 0, x < 0$ or $x = 0$.) We need to determine λ. Multiply the j-th equation by z_j and them sum up in j to get

$$\lambda = q \sum_{k=1}^{n} |z_k|^q > 0 \ .$$

It follows that all z_j are non-negative and

$$q z_j^{q-1} = \lambda a_j \ ,$$

hence

$$z_j^q = \left(\frac{\lambda}{q} a_j \right)^p \ .$$

Summing it up,

$$\frac{\lambda}{q} = \sum_{j=1}^{n} \left(\frac{\lambda}{q} a_j \right)^p \ .$$

After some manipulations we get

$$\lambda = q \left(\sum_{k=1}^{n} a_k^p \right)^{-q/p} \ ,$$

and

$$z_j = \frac{a_j^{p/q}}{\sum_{k=1}^{n} a_k^p} \ , \quad j = 1, \cdots, n \ .$$

Therefore, the minimum point is unique and the minimum value of this minimization problem is given by

$$\sum_{k=1}^{n} z_j^q = \left(\sum_{k=1}^{n} a_k^p \right)^{-q/p} \ .$$

In other words,

$$\left(\sum_{k=1}^{n} a_k^p \right)^{-q/p} \leq \sum_{j=1}^{n} |x_j|^q \ , \quad \text{whenever } \sum_{j=1}^{n} a_j x_j = 1 \ ,$$

and equality holds if and only if

$$x_j = \frac{a_j^{p/q}}{\sum_{k=1}^n a_k^p}\ , \quad j = 1, \cdots, n\ . \tag{7.3}$$

To deduce the inequality from this result, given a non-zero vector $a = (a_1, \cdots, a_n), a_j \geq 0$, satisfying $a \cdot b > 0$, the point

$$x = (x_1, \cdots, x_n)\ , \quad x_j = \frac{b_j}{a \cdot b}\ , \quad j = 1, \cdots, n,$$

lies on H, so $f(z) \leq f(x)$, that is,

$$\begin{aligned} \left(\sum_{k=1}^n a_k^p\right)^{-q/p} &\leq \sum_{j=1}^n |x_j|^q \\ &= \sum_{j=1}^n \frac{|b_j|^q}{(a \cdot b)^q}\ , \end{aligned}$$

from which the (strict) inequality follows. Furthermore, the equality case follows from (7.3). When $a \cdot b = 0$, the inequality clearly holds. $\square$

7.5 Constrained Extremal Problems II

Next, we consider the problem of multiple constraints in $\mathbb{R}^3$. For instance, let g and h be two differentiable functions defined in the open G in $\mathbb{R}^3$ and let

$$\Gamma \equiv \{(x, y, z) :\ g(x, y, z) = 0\ , \quad h(x, y, z) = 0\ ,\ (x, y, z) \in G\} \neq \phi\ .$$

Theorem 7.11. *Let f be a differential function in $G \subset \mathbb{R}^3$. If p is a local extremum of f over Γ defined above. Suppose that ∇g and ∇h are linearly independent at p. There exist λ and μ such that*

$$\nabla f(p) - \lambda \nabla g(p) - \mu \nabla h(p) = (0, 0, 0)\ .$$

Proof. Let $p = (x_0, y_0, z_0)$. According to Theorem 6.5, we may assume that there exists $\gamma = (\gamma_1, \gamma_2)$ from an open interval containing x_0 to Γ satisfying $\gamma(x_0) = (y_0, z_0)$ and

$$g(x, \gamma_1(x), \gamma_2(x)) = 0, \quad h(x, \gamma_1(x), \gamma_2(x)) = 0\ .$$

Then x_0 becomes an extremum point for the function $\varphi(x) \equiv f(x, \gamma_1(x), \gamma_2(x))$. By the Chain Rule

$$\varphi'(x_0) = \nabla f(p) \cdot (1, \gamma_1'(x_0), \gamma_2'(x_0)) = 0\ ,$$

which shows that $\nabla f(p)$ is perpendicular to the tangent vector of Γ at p. On the other hand, we also have

$$\nabla g(p)\cdot(1,\gamma_1'(x_0),\gamma_2'(x_0))=0\ ,\quad \nabla h(p)\cdot(1,\gamma_1'(x_0),\gamma_2'(x_0))=0\ .$$

By assumption, $\nabla g(p)$ and $\nabla h(p)$ are linearly independent, so they span the subspace perpendicular to $(1,\gamma_1'(x_0),\gamma_2'(x_0))$. As $\nabla f(p)$ belongs to this subspace, it can be expressed as a linear combination of $\nabla g(p)$ and $\nabla h(p)$, the desired result follows.

□

Example 7.13. The plane $H: x+y+z=12$ and the paraboloid $P: z=x^2+y^2$ intersects to form an ellipse. Find the highest and lowest points of the ellipse from the ground. The function to be optimized is the height function, which is given by $f(x,y,z)=z$. Taking $g=x+y+z$ and $h=x^2+y^2-z$, we consider the matrix formed by the gradients of the plane and the paraboloid:

$$\begin{bmatrix}1 & 1 & 1\\ 2x & 2y & -1\end{bmatrix}\ .$$

It is of rank 2 except when $x=y=-1/2$. However, when $(x,y)=(-1,-1)/2$, $z=12-x-y=13$ but $z=(1/2)^2+(1/2)^2=1/2$, so this point does not belong to the intersecting ellipse. In other words, ∇g and ∇h are always linearly independent at the ellipse. By Theorem 7.11,

$$\nabla f=\lambda\nabla g+\mu\nabla h\ ,$$

at any local extremum. Looking at each of the components,

$$0=\lambda+2\mu x,\quad 0=\lambda+2\mu y,\quad 1=\lambda-\mu\ .$$

It is clear that $\mu\neq 0$, so we have $x=y$. The constraints become $2x+z=12$ and $z=2x^2$, which are solved to get $x=2$ or -3. Therefore, there are two critical points $(2,2,8)$ and $(-3,-3,18)$. By observation the former is the lowest and the latter is the highest point.

We conclude this section by formulating the general case. Let $g_1,\cdots,g_m$ be m many differentiable functions defined in some open set $G\subset\mathbb{R}^n$ where $1\le m\le n-1$. Assume that

$$X=\{x\in G:\ g_j(x)=0, j=1,\ \cdots,m\}\neq\phi\ .$$

In the following we set

$$M=\begin{bmatrix}\dfrac{\partial g_1}{\partial x_1} & \cdots\cdots & \dfrac{\partial g_1}{\partial x_n}\\ \cdots & \cdots & \cdots\\ \dfrac{\partial g_m}{\partial x_1} & \cdots\cdots & \dfrac{\partial g_m}{\partial x_n}\end{bmatrix}\ .$$

Theorem 7.12. *Let f be differentiable in G. Let p be a local extremum of f over X. Suppose that M has rank m. There exist some $\lambda_j, j = 1, \cdots, m$, such that*

$$\nabla f(p) - \sum_{j=1}^{m} \lambda_j \nabla g_j(p) = (0, \cdots, 0) \ . \tag{7.4}$$

The proof of this theorem is essentially the same as the proof of the previous theorem and is left to you. Any point $p \in G$ at which M has rank m and which satisfies (7.4) is called a **critical point** of f under $g_j = 0, j = 1, \cdots, m$. Theorem 7.11 asserts that every local extremum is a critical point of f under $g_j = 0$.

As a last application let us prove the theorem on the diagonalization of symmetric matrices which have been used several times before.

Theorem 7.13 (**Principal Axis Theorem**). ** For every symmetric matrix A, there exists an orthogonal matrix R such that*

$$R'AR = D \ ,$$

where D is a diagonal matrix.

An orthogonal matrix is defined by the condition $R'R = I$ where R' is the transpose of R. In other words, the inverse of an orthogonal matrix is equal to its transpose matrix.

Proof. * We will give the proof for $n = 4$. The general case does not contain any additional difficulty. Let $A = (a_{ij}),\ i, j = 1, 2, 3, 4$, be a given symmetric matrix. Consider the quadratic function

$$f(x) = \sum_{i,j=1}^{4} a_{ij} x_i x_j \ .$$

First we minimize it over the sphere $g_0(x) = \sum_{j=1}^4 x_j^2 - 1 = 0$. This is a compact set so the minimum is attained at some $u^1 = (u_1^1, u_2^1, u_3^1, u_4^1)$. As $\nabla g(x) = 2x \neq (0, 0, 0, 0)$, Theorem 7.11 asserts that there exists some λ_1 such that

$$\nabla f(x) = \lambda_1 \nabla \left(\sum_j x_j^2 - 1 \right)$$

at $x = u^1$. In other words,

$$\sum_{j=1}^{4} a_{ij} u_j^1 = \lambda_1 u_i^1$$

for each $i = 1, \cdots, 4$. In matrix form

$$A \begin{bmatrix} u_1^1 \\ u_2^1 \\ u_3^1 \\ u_4^1 \end{bmatrix} = \lambda_1 \begin{bmatrix} u_1^1 \\ u_2^1 \\ u_3^1 \\ u_4^1 \end{bmatrix} ,$$

which shows that u^1 is an eigenvector of A with eigenvalue λ_1. Note that it is on the sphere so $|u^1| = 1$. Next we minimize f over two constraints, namely, $g_0(x) = 0$, and the hyperplane $g_1(x) \equiv u^1 \cdot x = 0$. Again this is a compact set and ∇g_0 and ∇g_1 are linearly independent on this set. Actually,

$$\nabla g_0 \cdot \nabla g_1 = 2x \cdot u^1 = 0,$$

that is, they are perpendicular to each other. By Theorem 7.11, letting $u^2 = (u_1^2, u_2^2, u_3^2, u_4^2)$ be its minimum point, there are two numbers λ_2, μ such that $\nabla f = \lambda \nabla g_0 + \mu \nabla g_1$, that is, for each $i = 1, 2, 3, 4$,

$$\sum_{j=1}^{4} a_{ij} u_j^2 \;\; = \;\; \lambda_2 u_i^2 + \mu u_i^1$$

Multiply this equation with u_i^1 and sum up over i,

$$\sum_{i,j=1}^{4} a_{ij} u_j^2 u_i^1 = \lambda_2 \sum_{i=1}^{4} u_i^2 u_i^1 + \mu \sum_{i=1}^{4} u_i^1 u_i^1 \; .$$

The left hand side vanishes as

$$\sum_{i,j=1}^{4} a_{ij} u_i^1 u_j^2 = \sum_{j=1}^{4} \lambda_1 u_j^1 u_j^2 = 0 \; .$$

On the other hand,

$$\sum_{i=1}^{4} u_i^2 u_i^1 = u^2 \cdot u^1 = 0 \; .$$

It forces $\mu = 0$, so u^2 is a unit eigenvector of A with eigenvalue λ_2. By construction it is perpendicular to u^1. At this point, more or less we know how to proceed further. Let $g_2(x) = u^2 \cdot x$ and minimize f over $g_i(x) = 0$ for $i = 0, 1, 2$. Since

$$\nabla g_0(x) = 2x, \quad \nabla g_1(x) = u^1, \quad \nabla g_2(x) = u^2 \; ,$$

form an orthogonal set, they must be linearly independent at every point on their common intersection. By Theorem 7.11 again the minimum point of f, u^3, satisfies

$$\sum_{j=1}^{4} a_{ij} u_j^3 = \lambda_3 u_i^3 + \mu_1 u_i^1 + \mu_2 u_i^2 \; ,$$

for some λ_3, μ_1, μ_2. Multiply this expression with u_i^1 and sum up in i yield $\mu_1 = 0$. Similarly, multiply it with u_i^2 and then sum up in i yield $\mu_2 = 0$. We conclude that u^3 is a unit eigenvector of A with eigenvalue λ_3.

To obtain the last eigenvector we fix a unit vector w that is perpendicular to u^1, u^2 and u^3. We claim that Au^4 is also perpendicular to $u^j, j = 1, 2, 3$. For,

$$\begin{aligned} \langle Aw, u^j \rangle &= \langle w, A'u^j \rangle \\ &= \langle w, Au^j \rangle \\ &= \lambda_j \langle w, u^j \rangle \\ &= 0 \ . \end{aligned}$$

It follows that Aw must be a scalar multiple of w, that is , $Aw = \lambda_4 w$ for some λ_4. Now we can take $u^4 = w$ and put u^1, u^2, u^3, u^4 together to form a 4×4-matrix R. Then

$$AR = DR \ ,$$

where

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} \ .$$

Since each column of R is a unit vector and the columns are perpendicular, R is an orthogonal matrix. We have completed the proof of the theorem.

□

Comments on Chapter 6

Optimization problem is one of the main applications of advanced calculus. You are supposed to master all topics in this chapter in particular the followings

- Finding all maximun/minimum points of a function over a compact set. You need to find all interior critical points as well as those critical points on the boundary of the set and then compare their corresponding values. When the set is unbounded such as the entire space or a half space, the behavior of the function at infinity must be taken care of.
- Second Derivative Test. Use the Hessian matrix to investigate the local behavior of a critical point.
- Lagrange multipliers in constrained problems: single and multiple constraints.

Supplementary Readings

5.1–5.4, 6.1, and 6.2 in [Au]. 14.7, 14.8, and 14.9 in [Thomas].