Chapter 4

Maxima and minima

4.1 Linearization and differential

Review of one variable function

When we study complicated functions, we can find its derivative at a point and study the tangent line instead. Thus, a tangent approximation is meaningful and call it linearization. Note that any nice curve, if enlarged, will look like a line. The linear function

$$L(x) = f(a) + f'(a)(x - a)$$

is called the **linearization** of f at a.

Differential

The geometric meaning of differential is given in Figure **??**.

$$f(x) = f(a) + f'(a)(x - a) + R_1(x, a)$$

where

$$\lim_{x \to a} \frac{R_1(x, a)}{x - a} = 0.$$

Taylor Polynomial

Taylor series involves infinitely many terms and require the function to be infinitely differentiable. However, if the function is differentiable only a few times, how can we approximate it?

Consider

$$y = P_1(x) := f(a) + f'(a)(x-a)$$

This is linear approximation to $f(x)$. Similarly, we can consider

$$y = P_2(x) := f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$$

which has same derivative up to second order. By the same way one can find a polynomial $P_n(x)$ of degree n which has the same derivatives at a up to n-th order. This polynomial will be a good approximation, called the **a Taylor polynomial of degree** n.

Hence the Taylor polynomial of degree k is given by

$$P_k(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

4.1.1 Taylor theorem

When f is differentiable at $\mathbf{a}$, the linear approximation of $f(\mathbf{x})$ is $f(\mathbf{a}) + \mathbf{D}f(\mathbf{a})(\mathbf{x}-\mathbf{a})$. Here the error $R_1 = |f(\mathbf{a}) - f(\mathbf{a}) - \mathbf{D}f(\mathbf{a})(\mathbf{x}-\mathbf{a})|$ satisfies

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_1}{\|\mathbf{x}-\mathbf{a}\|} = 0.$$

What if we want higher order approximation?

Theorem 4.1.1 (Taylor theorem one variable-integral remainder)**.** *If* $f\colon \mathbb{R} \to \mathbb{R}$ *has continuous* k*- th partial derivatives*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + R_k(x,a) \quad (4.1)$$

where $R_k(x,a)/(x-a)^k \to 0$ *as* $x \to a$*.*

Proposition 4.1.2. *If* f *is differentiable up to order* $k+1$*, then there exists a number* z *between* a *and* x *such that*

$$R_k(x,a) = \frac{f^{(k+1)}(z)}{(k+1)!}(x-a)^{k+1}. \quad (4.2)$$

Example 4.1.3. Find the Taylor polynomial of order 5 for $f(x) = \cos x$ at

$x = \pi/2$ and estimate the remainder.

$$R_5(x, \frac{\pi}{2}) = \frac{f^{(6)}(z)}{6!}(x - \frac{\pi}{2})^6.$$

Taylor theorem in several variable -1st order

Suppose $f\colon X \subset \mathbb{R}^2 \to \mathbb{R}$ is class C^1. Then the tangent plane at (a, b) is a good approx. That is

$$f(x, y) \approx p_1(x, y,$$

where

$$p_1(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

We can generalize this to a function of several variables.

Theorem 4.1.4 (Taylor theorem in several variable -1st order)**.** *If* $f\colon \mathbb{R}^n \to \mathbb{R}$ *is differentiable at* $\mathbf{a}$*, then*

$$f(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_1(\mathbf{x}, \mathbf{a}),$$

where $R_1(\mathbf{x}, \mathbf{a})/\|\mathbf{x} - \mathbf{a}\| \to 0$ *as* $\mathbf{x} \to \mathbf{a}$.

In scalar form, we have

$$f(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{n} f_{x_i}(\mathbf{a})(x_i - a_i) + R_1(\mathbf{x}, \mathbf{a}).$$

Example 4.1.5. Find the equation of tangent plane to $f(\mathbf{x}) = x_1 + 2x_1x_2 + x_3 + x_1x_4 + x_1x_2x_3^2$ at $\mathbf{x} = (1, 1, 2, 2)$.

$$f_{x_1} = 1 + 2x_2 + x_4 + x_2x_3^2 = 9, \quad f_{x_2} = 2x_1 + x_1x_3^2 = 4$$
$$f_{x_3} = 1 + 2x_1x_2x_3 = 5, \quad f_{x_4} = x_1 = 1$$

Thus

$$p_1 = 11 + 9(x_1 - 1) + 4(x_2 - 1) + 5(x_3 - 2) + (x_4 - 2).$$

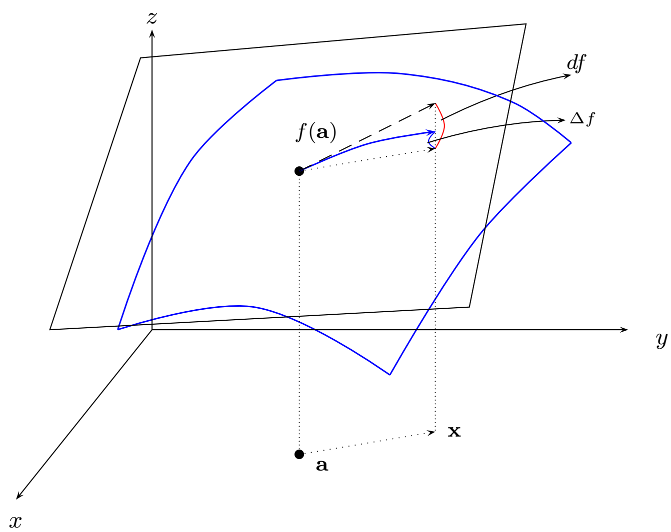


Figure 4.1: True increase Δf vs. differential df

Differential in several variable

Definition 4.1.6. Let $f\colon X \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable and $\mathbf{a} \in X$. The differential of f is

$$df(\mathbf{a},\mathbf{h}) = \frac{\partial f}{\partial x_1}(\mathbf{a})h_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{a})h_n.$$

The significance of differential is that for small $\mathbf{h}$

$$df \approx \Delta f := f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}).$$

Here $\mathbf{h} = (h_1, \cdots, h_n)$ denote small change in the variables and it is also written as $\mathbf{h} = \Delta\mathbf{x} = (\Delta x_1, \cdots, \Delta x_n)$. Hence the differential is also written as

$$df = \frac{\partial f}{\partial x_1}dx_1 + \cdots + \frac{\partial f}{\partial x_n}dx_n.$$

Example 4.1.7. Find the differential of $f(x,y,z) = e^{x+y}\sin(yz)$.

$$\begin{aligned} df &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \\ &= e^{x+y}\sin(yz)\,dx + e^{x+y}(\sin(yz) + z\cos(yz))\,dy + e^{x+y}y\cos(yz)\,dz. \end{aligned}$$

Example 4.1.8. Find the Δf at $(2,-1)$ when $f(x,y) = x - y + 2x^2 + xy^2$.

$$\begin{aligned}
\Delta f &= f(2+\Delta x, -1+\Delta y) - f(2,-1) \\
&= (2+\Delta x) - (-1+\Delta y) + 2(2+\Delta x)^2 + (2+\Delta x)(-1+\Delta y)^2 - (13) \\
&= 3 + \Delta x - \Delta y + 2(4 + 4\Delta x + \Delta^2 x) + (2+\Delta x)(1 - 2\Delta y + \Delta^2 y) - 13 \\
&= 10\Delta x - 5\Delta y + 2(\Delta x)^2 - 2\Delta x\Delta y + 2(\Delta y)^2 + \Delta x(\Delta y)^2.
\end{aligned}$$

On the other hand

$$\begin{aligned}
df &= f_x(2,-1)\Delta x + f_y(2,-1)\Delta y \\
&= (1 + 4x + y^2)_{(2,-1)}\Delta x + (-1 + 2xy)_{(2,-1)}\Delta y \\
&= (1 + 8 + 1)\Delta x + (-1 - 4)\Delta y.
\end{aligned}$$

Thus df coincides with Δf up to the linear factor of Δx and Δy.

Example 4.1.9. Let $f(x,y) = x + y + e^{x+y^2}$. Find Δf at $\mathbf{a} = (1,1)$. Also find df and compare.

$$\begin{aligned}
\Delta f &= f(1+\Delta x, 1+\Delta y) - f(1,1) \\
&= 1 + \Delta x + 1 + \Delta y + e^{1+\Delta x+(1+\Delta y)^2} - (1 + 1 + e^{1+(1)^2}) \\
&= \Delta x + \Delta y + e^{2+\Delta x+2\Delta y+\Delta y^2} - e^2 \\
&= \Delta x + \Delta y + e^2(e^{1+\Delta x+2\Delta y+\Delta y^2} - 1) \\
&= \Delta x + \Delta y + e^2(\Delta x + 2\Delta y + \Delta y^2 + \Delta x^2 + 4\Delta y^2 + \cdots)
\end{aligned}$$

On the other hand, from $f_x = 1 + e^{x+y^2}$, and $f_y = 1 + 2ye^{x+y^2}$, we get

$$\begin{aligned}
df &= f_x(1,1)\Delta x + f_y(1,1)\Delta y \\
&= (1 + e^2)\Delta x + (1 + 2e^2)\Delta y \\
&= \Delta x + \Delta y + e^2(\Delta x + 2\Delta y).
\end{aligned}$$

Again df and Δf coincide up to linear factor of Δx and Δy.

From these examples you see computing the approximate value df is easier than computing the exact value Δf.

Example 4.1.10. A box is to be made with dimension $3 \times 4 \times 6$ inches with a possible error in measuring is the same in all direction. We would like to know

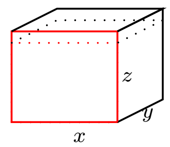


Figure 4.2: Volume of a box

how accurate we must measure the length so that the error in the volume is less than 0.1 in^3.

Sol.

Since the exact volume is $V = xyz, \quad x = 3, y = 4, z = 6$, we have

$$\begin{aligned} dV &= V_x dx + V_y dy + V_z dz \\ &= yz\,dx + xz\,dy + xy\,dz \\ &\quad \text{since the possible error in measuring the length is } dx = dy = dz \\ &= (24 + 18 + 12)dx \le 0.1. \end{aligned}$$

Hence the measurement in each dimension must be accurate within $dx \le 0.1/54 = 0.0019$

Example 4.1.11. The volume of cylindrical can is $V(r, h) = \pi r^2 h$. Find the change of volume when the radius or the height changes.

$$dV = V_r\,dr + V_h\,dh = 2\pi rh\Delta r + \pi r^2 \Delta h.$$

Taylor theorem -second order formula

Theorem 4.1.12 (Taylor theorem 2nd order formula)**.** *Suppose $f\colon X \subset \mathbb{R}^n \to \mathbb{R}$ is class $\mathcal{C}^2$, then*

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^{n} h_i f_{x_i}(\mathbf{a}) + \frac{1}{2}\sum_{i,j=1}^{n} h_i h_j f_{x_i x_j}(\mathbf{a}) + R_2(\mathbf{x}, \mathbf{a}),$$

where $\mathbf{h} = \mathbf{x} - \mathbf{a}$, $h_i = x_i - a_i$ and $R_2(\mathbf{x}, \mathbf{a})/\|\mathbf{h}\|^2 \to 0$ as $\mathbf{h} \to 0$.

Let $n = 2$ and try to find a quadratic polynomial $p(x, y)$ which has same derivatives up to second order as f at a given point $\mathbf{a} = (a, b)$. Let

$$p(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

and we require p have the same derivatives as f up to second order:

$$\begin{aligned} p(a,b) &= f(a,b) \\ p_x(a,b) &= f_x(a,b) \quad p_y(a,b) = f_y(a,b) \\ p_{xx}(a,b) &= f_{xx}(a,b), \quad p_{xy}(a,b) = f_{xy}(a,b), \quad p_{yy}(a,b) = f_{yy}(a,b). \end{aligned}$$

Hence

$$\begin{aligned} p(x,y) &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &\quad + \frac{1}{2} f_{xx}(a,b)(x-a)^2 + f_{xy}(a,b)(x-a)(y-b) + \frac{1}{2} f_{yy}(a,b)(y-b)^2. \end{aligned}$$

In matrix form,

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{a}) + (\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \\ &\quad + \frac{1}{2}(h_1, \cdots, h_n) \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ & \cdots & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}. \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \quad + R_2(\mathbf{x}, \mathbf{a}) \\ &= f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{a})\mathbf{h} + R_2. \end{aligned}$$

(In the third term, treat Df as a column vector for consistency of derivative.) Here $Hf(\mathbf{a})$ is the Hessian defined by

$$Hf(\mathbf{a}) = \begin{pmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{pmatrix}$$

and $P_2(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{a})\mathbf{h}$ is the second degree Taylor polynomial at $\mathbf{a}$.

Proof. Let $g(t) = f(\mathbf{a} + t\mathbf{h})$ and use Taylor theorem in one variable and chain rule. We have

$$g(1) = g(0) + g'(0) + \frac{g''(0)}{2!} + R_2$$

and

$$R_2 = \int_0^1 \frac{(t-1)^2}{2!} g'''(t)dt.$$

By Chain rule, we see

$$g'(t) = Df \cdot \frac{d\mathbf{x}}{dt} = \sum_{i=1}^n f_{x_i}(\mathbf{a}+t\mathbf{h})h_i, \quad g''(t) = \frac{d\mathbf{x}}{dt}^t D^2 f \cdot \frac{d\mathbf{x}}{dt} = \sum_{i,j=1}^n f_{x_i x_j}(\mathbf{a}+t\mathbf{h})h_i h_j,$$

and

$$g'''(t) = \sum_{i,j,k=1}^n f_{x_i x_j x_k}(\mathbf{x}_0 + t\mathbf{h})h_i h_j h_k.$$

Hence we have

$$R_2(\mathbf{x}, \mathbf{a}) = \sum_{i,j,k=1}^n \int_0^1 \frac{(1-t)^2}{2} f_{x_i x_j x_k}(\mathbf{a} + t\mathbf{h})h_i h_j h_k \, dt.$$

and

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^n h_i f_{x_i}(\mathbf{a}) + \frac{1}{2} \sum_{i,j=1}^n h_i h_j f_{x_i x_j}(\mathbf{a}) + R_2(\mathbf{x}, \mathbf{a}).$$

Here the integrand of $R_2(\mathbf{x}, \mathbf{a})$ is continuous. So if $\|\mathbf{h}\|$ is small, it is less than M. So $|R_2(\mathbf{h}, \mathbf{x})| \le M\|h\|^3$. In other words $|R_2(\mathbf{x}, \mathbf{a})|/\|h\|^2 \to 0$ as $\mathbf{h} \to 0$. □

Example 4.1.13. Find 2nd order Taylor approximation of $f(x, y) = e^{x+y}$ near $\mathbf{a} = (0, 0)$.

sol. Partials of f are

$$f_x(0,0) = f_y(0,0) = e^0 = 1$$
$$f_{xx}(0,0) = f_{xy}(0,0) = f_{yy}(0,0) = e^0 = 1.$$

$$f(x, y) = 1 + x + y + \frac{1}{2}(x^2 + 2xy + y^2) + R_2.$$

As $(x, y) \to (0, 0)$, $R_2/\|(x, y)\|^2 \to 0$.

Repeat this with $\mathbf{a} = (1, 1)$.

■

Example 4.1.14. Find 2nd order Taylor approximation of $f(x, y) = \cos x e^{x+y}$ near $\mathbf{a} = (0, 0)$.

sol. Partials of f are

$$f_x = -\sin x e^{x+y} + \cos x e^{x+y},\ f_y = \cos x e^{x+y}$$
$$f_{xx} = -2\sin x e^{x+y},\ f_{xy} = (-\sin x + \cos x)e^{x+y},\ f_{yy} = \cos x e^{x+y}.$$

Since $f(0,0) = 1$, $f_x(0,0) = 1$, $f_y(0,0) = 1$, $f_{xx}(0,0) = 0$, $f_{xy}(0,0) = 1$, $f_{yy}(0,0) = 1$, we see

$$f(x,y) = 1 + x + y + \frac{1}{2}(2xy + y^2) + R_2.$$

Using matrix form,

$$\begin{aligned} p_2(x,y) &= f(0,0) + Df(0,0)\mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf(0,0)\mathbf{h} \\ &= 1 + \begin{bmatrix} 1, 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} h_1, h_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}. \end{aligned}$$

Here $h_1 = x - 0$, $h_2 = y - 0$.

□

Theorem 4.1.15. *(1) Remainder of Taylor's theorem 4.1.4 is given by*

$$R_1(\mathbf{x},\mathbf{a}) = \sum_{i,j=1}^{n} \int_0^1 (1-t) f_{x_i x_j}(\mathbf{a} + t\mathbf{h}) h_i h_j \, dt = \sum_{i,j=1}^{n} \frac{1}{2} f_{x_i,x_j}(\mathbf{c}) h_i h_j.$$

Here $\mathbf{c}$ *is a point between* $\mathbf{a}$ *and* $\mathbf{x}$*.*

(2) Remainder of Taylor's theorem 4.1.12 is

$$R_2(\mathbf{x},\mathbf{a}) = \sum_{i,j,k=1}^{n} \int_0^1 \frac{(1-t)^2}{2} f_{x_i x_j x_k}(\mathbf{a}+t\mathbf{h}) h_i h_j h_k \, dt = \sum_{i,j,k=1}^{n} \frac{1}{3!} f_{x_i,x_j,x_k}(\mathbf{c}) h_i h_j h_k$$

Here $\mathbf{c}$ *is a point between* $\mathbf{a}$ *and* $\mathbf{x}$*.*

Proof. **[MVT for integral]** *If h and g are continuous on $[a,b]$ and $g > 0$ then for some $c \in [a,b]$ the following holds*

$$\int_a^b h(t)g(t)\,dt = h(c)\int_a^b g(t)\,dt$$

.

(1) From thm 4.1.4 we see

$$R_1(\mathbf{h}, \mathbf{x}_0) = \sum_{i,j=1}^{n} \int_0^1 (1-t) f_{x_i x_j}(\mathbf{x}_0 + t\mathbf{h}) h_i h_j \, dt.$$

If we set $h(t) = f_{x_i x_j}(\mathbf{x}_0 + t\mathbf{h})$ and $g(t) = (1-t)h_i h_j$ and use MVT. Then

$$\begin{aligned}
&\int_0^1 (1-t) f_{x_i x_j}(\mathbf{x}_0 + t\mathbf{h}) h_i h_j \, dt \\
&\qquad = f_{x_i x_j}(\mathbf{c}) \int_0^1 (1-t) h_i h_j \, dt \\
&\qquad = \frac{1}{2} f_{x_i x_j}(\mathbf{c}) h_i h_j.
\end{aligned}$$

(2) similar □

Second order Taylor formula gives second order approximation.

Example 4.1.16. Find the second order approx. of $f(x, y) = \sin x \sin y$. What is error if $|x| \le 0.1$, $|y| \le 0.1$.

sol. partials of $f(x, y)$ are

$$\begin{aligned}
f(0,0) &= 0, & f_x(0,0) &= 0, & f_y(0,0) &= 0 \\
f_{xx}(0,0) &= 0, & f_{xy}(0,0) &= 1, & f_{yy}(0,0) &= 0.
\end{aligned}$$

Hence

$$f(x, y) = 0 + 0 + 0 + \frac{1}{2}\Big(x^2(0) + 2xy + y^2(0)\Big) + R_2.$$

So $\sin x \sin y \approx xy$ and the error is

$$\begin{aligned}
|R_2| &= |\frac{1}{6}(x^3 f_{xxx} + 3x^2 y f_{xxy} + 3xy^2 f_{xyy} + y^3 f_{yyy})|_{(c_1, c_2)} \\
&\le \frac{1}{6}\Big((0.1)^3 + 3(0.1)^1 + 3(0.1)^1 + (0.1)^1\Big) \\
&\le \frac{8}{6}(0.1)^3 \le 0.00134.
\end{aligned}$$

■

Example 4.1.17. Find second order approx. of $f(x, y) = e^x \cos y$ at $(0, 0)$.

sol. partials of $f(x,y)$ are

$$f(0,0)=0, \qquad f_x(0,0)=1, \qquad f_y(0,0)=0$$
$$f_{xx}(0,0)=1, \qquad f_{xy}(0,0)=0, \qquad f_{yy}(0,0)=-1.$$

Hence

$$f(\mathbf{h}) = 1 + h_1 + \frac{1}{2}\Big(h_1^2 - h_2^2\Big) + R_2,$$

where $R_2/\|\mathbf{h}\|^2 \to 0$ as $\|\mathbf{h}\| \to 0$.

■

Example 4.1.18. Find approx. value of $(3.98-1)^2/(5.97-3)^2$ compare with exact value.

sol. Let $f = (x-1)^2/(y-3)^2$. Desired value is close to $f(4,6) = 1$. partials of $f(x,y)$ are

$$f(4,6)=1, \qquad f_x(4,6)=\frac{2}{3}, \qquad f_y(4,6)=-\frac{2}{3},$$
$$f_{xx}(4,6)=\frac{2}{9}, \qquad f_{xy}(4,6)=-\frac{4}{9}, \qquad f_{yy}(4,6)=\frac{2}{3}.$$

Hence linear approx. is

$$1+\frac{2}{3}(-0.02)-\frac{2}{3}(-0.03) = 1.00666$$

while quadratic approx is

$$1+\frac{2}{3}(-0.02)-\frac{2}{3}(-0.03)+\frac{2}{9}\frac{(-0.02)^2}{2}-\frac{4}{9}(-0.02)(-0.03)+\frac{2}{3}\frac{(-0.03)^2}{2} = 1.00674.$$

A more exact value is 1.00675.(calculator value)

■

4.2 Extrema of real valued functions

Local Max, Min

Definition 4.2.1. We say $f\colon X \subset \mathbb{R}^n \to \mathbb{R}$ has **local minimum** at $\mathbf{a} \in U$ if there is a neighborhood U of $\mathbf{a}$ such that $f(\mathbf{x}) \le f(\mathbf{a})$ for all $\mathbf{x} \in U$. Similarly, we say f has a **local maximum** at $\mathbf{a} \in U$ if there is a neighborhood U of $\mathbf{a}$ such that $f(\mathbf{a}) \le f(\mathbf{x})$ for all $\mathbf{x} \in U$.

global minimum and **global maximum** are clearly defined. A critical point which is either local max or min is called a **saddle**.

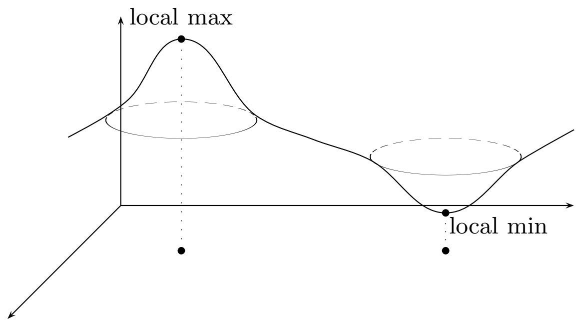


Figure 4.3: Near extreme

Theorem 4.2.2 (First derivative test for local extrema)**.** *If $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a} \in \mathbb{R}^n$ and assumes an extreme value, then $Df(\mathbf{a}) = 0$.*

Proof. Suppose f has local maximum at $\mathbf{a}$. Then for any $\mathbf{h} \in \mathbb{R}^n$, the function $g(t) = f(\mathbf{a} + t\mathbf{h})$ has a local minimum. Hence

$$g'(0) = D_{\mathbf{h}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{h} = 0$$

Sine this holds for every $\mathbf{h}$, $\nabla f(\mathbf{a}) = \mathbf{0}$, i.e, $\mathbf{a}$ is critical point of f. □

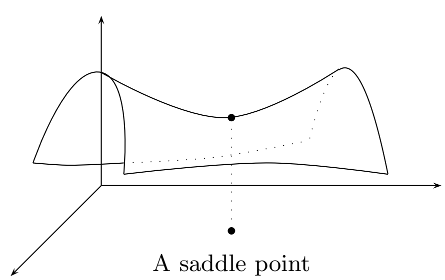


Figure 4.4: Near Saddle point

Definition 4.2.3. A point $\mathbf{a} \in \mathbb{R}^n$ is called a **critical point** if f is not differentiable or $\nabla f(\mathbf{a}) = \mathbf{0} = (0, \ldots, 0)$.

Example 4.2.4. Find the (local) maximum and minimum of $f = x^2 + y^2$.

Example 4.2.5. Find the extrema of $f = x^2 - y^2$ (if any).

Sol. $Df = (2x, -2y)$. Hence the only critical point is $(0, 0)$. But we see it is not an extreme point.

Example 4.2.6. Find critical points of $z = x^2y + y^2x$ and investigate their behavior.

sol. From

$$z_x = 2xy + y^2 = 0, \quad z_y = 2xy + x^2 = 0$$

We obtain $x^2 = y^2$. For $x = y$, we get $2y^2 + y^2 = 0$ and $(x, y) = (0, 0)$. For $x = -y$, we again get $x = y = 0$. Now for $x = y$, $z = 2x^3$. Not a extreme. So saddle.

□

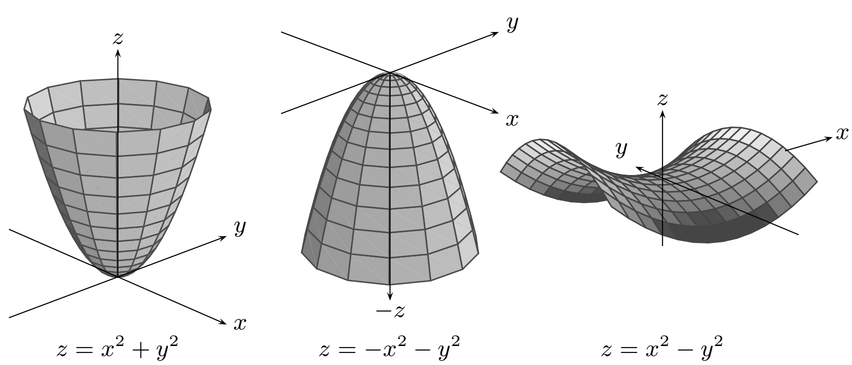


Figure 4.5: Graphs of critical points

Example 4.2.7. Find the extrema of $z = 2(x^2 + y^2)e^{-x^2-y^2}$.

sol.

$$\begin{aligned} z_x &= [4x + 2(-2x)(x^2 + y^2)]e^{-(x^2+y^2)} \\ &= 4x(1 - x^2 - y^2)e^{-(x^2+y^2)} \\ z_y &= 4y(1 - x^2 - y^2)e^{-(x^2+y^2)}. \end{aligned}$$

Solving these, we obtain $x = y = 0$ or $x^2 + y^2 = 1$. We can check the points on the crater's rim are points of local maximum.

■

Investigate the property of a critical point by Hessian

Example 4.2.8. Find the extrema of $f = x^2 + xy + y^2 + 2x - 2y + 5$.

Sol. First we find the critical point by setting $Df(x, y) = 0$.

$$\begin{aligned} f_x &= 2x + y + 2 = 0 \\ f_y &= x + 2y - 2 = 0. \end{aligned}$$

Thus $(-2, 2)$ is the only critical point. To determine whether this point is a max or min(or neither), we do as follows: With $(\Delta x, \Delta y) = (h, k)$,

$$\Delta f = h^2 + hk + k^2.$$

We have three possibilities:

- If the quantity Δf is nonnegative for all small values of h and k, then $(-2, 2)$ yields a local min.
- Similarly, if Δf is nonpositive for all small values of h and k, then $(-2, 2)$ yields a local max.
- If neither holds, then it yields a saddle.

In this particular example, the point clearly yields a local min.

In general, let us look at the Δf more carefully. From Taylor formula

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + Df(\mathbf{x}_0) + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h} + R_2(\mathbf{x}_0, \mathbf{h}), \quad \frac{R_2(\mathbf{x}_0, \mathbf{h})}{\|\mathbf{h}\|^2} \to 0$$

Hence

$$\begin{aligned} \Delta f &= f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) \\ &= Df(\mathbf{x}_0) + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h} + R_2(\mathbf{x}_0, \mathbf{h}) \\ &= \frac{1}{2}\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h} + o(\|\mathbf{h}\|^2) \\ &= (\frac{1}{2} - \epsilon)\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h}. \end{aligned}$$

Thus the point $\mathbf{x}_0$ is

- a point of local min. if $\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h} \geq 0$ for all small values of $\mathbf{h}$
- a point of local max. if $\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h} \leq 0$ for all small values of $\mathbf{h}$
- a saddle if $\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h}$ assumes both positive value and negative value.

More generally we have second derivative test:

Definition 4.2.9. A function $Q : \mathbb{R}^n \to \mathbb{R}$ of the form

$$Q(h_1, \cdots, h_n) = \sum_{i,j=1}^{n} b_{ij} h_i h_j = \mathbf{h}^T B\mathbf{h}$$

is called a **quadratic form**. It is symmetric, if $b_{ij} = b_{ji}$. A quadratic form(function) $Q : \mathbb{R}^n \to \mathbb{R}$ is said to be

(1) **positive definite** if $Q(\mathbf{h}) \geq 0$, $\forall \mathbf{h} \in \mathbb{R}^n$, and $Q(\mathbf{h}) = 0$ implies $\mathbf{h} = 0$.

(2) **negative definite** if $Q(\mathbf{h}) \leq 0$, $\forall \mathbf{h} \in \mathbb{R}^n$, and $Q(\mathbf{h}) = 0$ implies $\mathbf{h} = 0$.

We note that the quadratic form(function) $Q(\mathbf{h}) = \mathbf{h}^T B\mathbf{h}$ is positive definite if and only if the matrix B is positive definite.

Recall the **Hesssian** at (x_0, y_0). The Hessian of f naturally defines a quadratic form

$$Q(\mathbf{h}) = \frac{1}{2}(h_1, \cdots, h_n) \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ & \cdots & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.$$

If $\mathbf{x}_0$ is a critical point, then $DF(\mathbf{x}_0) = 0$. Hence

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h} + R_2(\mathbf{x}_0, \mathbf{h}).$$

Thus to study the behavior of a critical point, it suffices to study the quadratic form $Q(\mathbf{h}) = \frac{1}{2}\mathbf{h}^T Hf(\mathbf{x}_0)\mathbf{h}$.

Theorem 4.2.10. *[Second derivative test] Suppose f is $\mathcal{C}^2$ and $\mathbf{a} = (x_0, y_0)$ is a critical point of f.*

(1) If the Hessian $Hf(\mathbf{a})$ is positive definite, then f has a relative minimum at $\mathbf{a}$.

(2) If $Hf(\mathbf{a})$ is negative definite, then f has a relative maximum at $\mathbf{a}$.

(3) If $Hf(\mathbf{a})$ is neither positive nor negative definite, then f has a saddle point at $\mathbf{a}$.

Example 4.2.11. Consider $f(x, y) = x^2 + y^2$. $(0, 0)$ is the critical point. We see

$$f(x, y) = f(0, 0) + (h_1^2 + h_2^2) + 0.$$

Since $Q(h_1, h_2) = h_1^2 + h_2^2$ is positive definite, $(0, 0)$ is a local min.

Lemma 4.2.12. *If $B = [b_{ij}]$ is $n \times n$ real matrix and if*

$$Q : \mathbb{R}^n \to \mathbb{R}, (h_1, \cdots, h_n) \mapsto \frac{1}{2} \sum b_{ij} h_i h_j$$

is positive-definite, then there is $M > 0$ such that for all $\mathbf{h}$,

$$Q(\mathbf{h}) \geq M \|\mathbf{h}\|^2.$$

Proof. For $\|\mathbf{h}\| = 1$, set $g(\mathbf{h}) = Q(\mathbf{h})$. Then g is continuous function on a closed set, hence have a positive minimum, say M. Because Q is quadratic,

$$Q(\mathbf{h}) = Q(\frac{\mathbf{h}}{\|\mathbf{h}\|} |\mathbf{h}\|) = Q(\frac{\mathbf{h}}{\|\mathbf{h}\|}) |\mathbf{h}\|^2 = g(\frac{\mathbf{h}}{\|\mathbf{h}\|}) \|\mathbf{h}\|^2 \geq M |\mathbf{h}\|^2$$

for any $\mathbf{h} \neq 0$. ☐

Proof of Theorem 4.2.10.

$$f(\mathbf{x}_0) - f(\mathbf{x}_0) = Q(\mathbf{x}_0)(\mathbf{h}) + R_2,$$

where $R_2/\|\mathbf{h}\|^2 \to 0$. Hence we can say $|R_2| < \epsilon \|\mathbf{h}\|^2$ for some small ϵ when $0 < \|\mathbf{h}\| < \delta$. Since $Q(\mathbf{x}_0)$ is positive definite,

$$Q(\mathbf{x}_0)(\mathbf{h}) \geq M \|\mathbf{h}\|^2, \forall \mathbf{h}.$$

Hence

$$f(\mathbf{x}_0) - f(\mathbf{x}_0) = Q(\mathbf{x}_0)(\mathbf{h}) + R_2 \geq (M - \epsilon) \|\mathbf{h}\|^2,$$

for $0 < \|\mathbf{h}\| < \delta$ and so we have a strict min at $\mathbf{x}_0$.

Determinant test for Positive definiteness

Then how do we know Positive definiteness ?

Lemma 4.2.13. *Let* $B = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ *and* $Q(\mathbf{h}) = \frac{1}{2}\mathbf{h}^T B\mathbf{h}$. *Then* $Q(\mathbf{h})$ *is positive-definite if and only if* $a > 0$ *and* $ac - b^2 > 0$.

sol. We have

$$Q(\mathbf{h}) = \frac{1}{2}[h_1, h_2] \begin{bmatrix} a, b \\ b, c \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \frac{1}{2}(ah_1^2 + 2bh_1h_2 + ch_2^2)$$

$$Q(\mathbf{h}) = \frac{1}{2}a\left(h_1 + \frac{b}{a}h_2\right)^2 + \frac{1}{2}\left(c - \frac{b^2}{a}\right)h_2^2.$$

Suppose Q is positive definite. Then setting $h_2 = 0$, we see $a > 0$. Next setting $h_1 = 0$, we get $ac - b^2 > 0$. The converse also hold.

■

Similarly, we have negative definite if $a < 0$ and $ac - b^2 > 0$.

Theorem 4.2.14 (Second derivative test)**.** *Suppose* f *is* $\mathcal{C}^2$ *on an open subset* U *of* $\mathbb{R}^2$ *and* $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ *holds, i.e.,* (x_0, y_0) *is a critical point.) Let* $D = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - (f_{xy}(x_0, y_0))^2$. *Then the following holds:*

(1) f *has a local min. if* $f_{xx}(x_0, y_0) > 0$ *and* $D > 0$
(2) f *has a local max. if* $f_{xx}(x_0, y_0) < 0$ *and* $D > 0$
(3) f *has a saddle point if* $D < 0$
(4) If $D = 0$ *then we say* f *has a* **degenerate** *critical point.*

Example 4.2.15. Classify the critical points of the following functions.

(1) $g_1(x, y) = 3x^2 + 6xy + 9y^2$

(2) $g_2(x, y) = -2x^2 + xy - y^2$

(3) $g_3(x, y) = x^2 - xy + 2y^2$

sol. All the critical points are $(0, 0)$. For g_1, we see $D = 3 \cdot 9 - 3^2 = 18 > 0$. Hence $(0, 0)$ is a local min of g_1.

For g_2, we have $D = (-2)(-1) - 1/4 = 7/4 > 0$ and $a = -2 < 0$, we see g_2 has local maximum at $(0, 0)$.

For g_3, $D = 2 \cdot 1 - 1/4 = 3/4 > 0$ and $a = 1 > 0$, hence g_3 has local minimum at $(0,0)$.

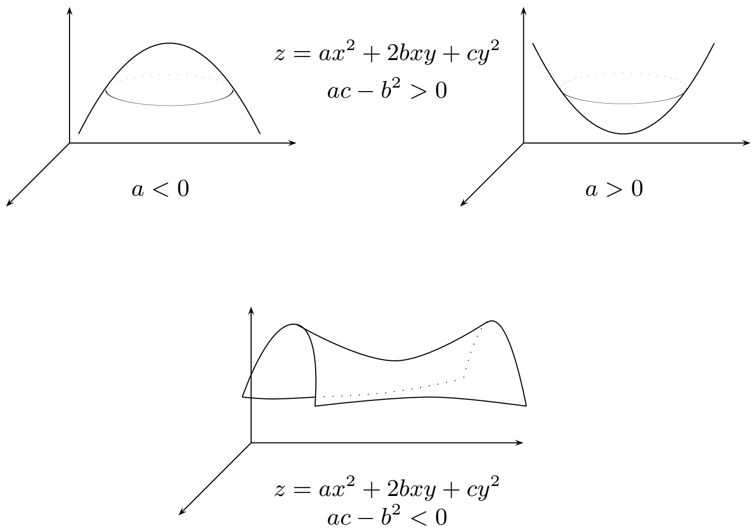


Figure 4.6: Graphs of quadratic functions

Example 4.2.16. 'Investigate' the behavior of "critical points" of the following functions:

(1) $f(x,y) = x^4 + y^4$

(2) $g(x,y) = x^3 + y^3$

sol.

(1) $(0,0)$ is the only critical point of f and $D = 0$ at $(0,0)$. This is a degenerate case. So the test fails. It is easy to check $f \geq 0$ for all (x,y) and $f(0,0) = 0$. So $(0,0)$ is local minimum

(2) Again $D = 0$. Hence the test fails. Instead we check the behavior of g directly. We see

$$\begin{aligned} g > 0 \quad &\text{if } \ xy > 0 \\ g < 0 \quad &\text{if } \ xy < 0. \end{aligned}$$

Hence $(0,0)$ is a saddle of g.

Proof of pos. definiteness when $D > 0$.

$$Hf = \frac{1}{2}[h_1, h_2] \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Thus we need to check definiteness of the matrix of the form $B = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$.

Example 4.2.17. Let $f(x, y) = x^3 + xy^2 + x^2 + y^2 + 3z^2$. Identify critical points and determine local max or min.

Example 4.2.18. Locate relative maxima minima saddle of

$$f(x, y) = \log(x^2 + y^2 + 1).$$

sol. $\nabla f = 0$ gives $(0, 0)$ as a critical point. Second derivatives are

$$f_{xx}(0, 0) = 2 = f_{yy}(0, 0), \quad f_{xy}(0, 0) = 0.$$

Hence $D = 2 \cdot 2 = 4 > 0$. Hence min.

■

Example 4.2.19. The graph of $g = 1/xy$ is a surface S. Find the point on S closest to $(0, 0)$.

sol. Each point on the surface is $(x, y, 1/xy)$. Hence

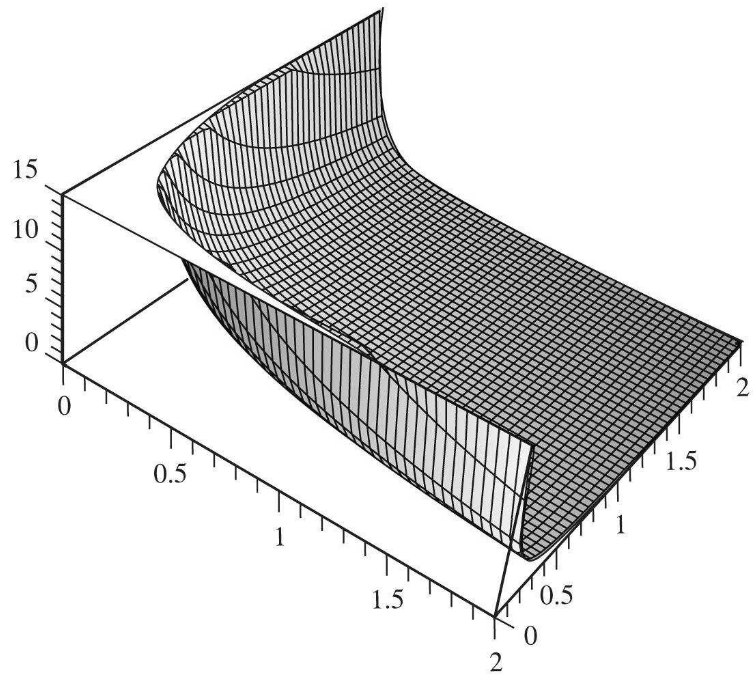
$$d^2 = x^2 + y^2 + \frac{1}{x^2y^2}.$$

We find the point which minimize $f(x, y) = d^2(x, y)$ rather than d itself. Solving

$$f_x = 2x - \frac{2}{x^3y^2} = 0, \quad f_y = 2y - \frac{2}{x^2y^3} = 0,$$

we obtain $x^4y^2 = 1$ and $x^2y^4 = 1$. From the first eq. we get $y^2 = 1/x^4$. Substitute into second equation, we get $x^6 = 1$. So $x = \pm 1$ and $y = \pm 1$. Considering the geometry, one can easily see that all these four points give minimum ($d = \sqrt{3}$).(As x or y approaches ∞, $f \to \infty$). So f has no max.

■



Example 4.2.20. Find the critical points of $f(x, y) = (x^2 - y^2)e^{(-x^2-y^2)/2}$and determine if they are local max. or min. or neither.

sol. To find the critical points, we need to solve the following system

$$f_x = [2x - x(x^2 - y^2)]e^{(-x^2-y^2)/2} = 0$$
$$f_y = [-2y - y(x^2 - y^2)]e^{(-x^2-y^2)/2} = 0.$$

From these, we see

$$x[2 - (x^2 - y^2)] = 0, \qquad y[-2 - (x^2 - y^2)] = 0.$$

Hence

$$(x, y) = (0, 0), \quad (\pm\sqrt{2}, 0), \quad (0, \pm\sqrt{2}).$$

On the other hand, the second derivatives are

$$f_{xx} = [2 - 5x^2 + x^2(x^2 - y^2) + y^2]e^{(-x^2-y^2)/2},$$
$$f_{yy} = [5y^2 - 2 + y^2(x^2 - y^2) - x^2]e^{(-x^2-y^2)/2},$$
$$f_{xy} = xy(x^2 - y^2)e^{(-x^2-y^2)/2}.$$

Since $D(0,0) = -4$, the point $(0,0)$ is a saddle. While $D(\pm\sqrt{2},0) = 16/e^2 > 0$ and $f_{xx}(\pm\sqrt{2},0) = -4/e$. So $(\pm\sqrt{2},0)$ is local min. Since $D(0,\pm\sqrt{2}) = 16/e^2 > 0$ and $f_{xx}(0,\pm\sqrt{2}) = 4/e$, $(0,\pm\sqrt{2})$ is local max. Graph is as Fig 4.7.

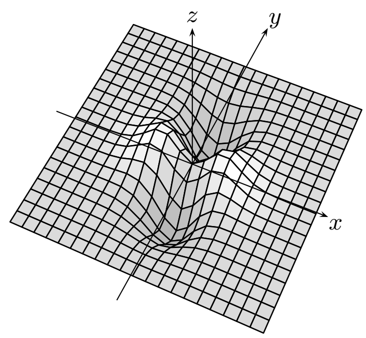


Figure 4.7: $f(x,y) = (x^2 - y^2)e^{(-x^2-y^2)/2}$

Global maxima and Minima

Definition 4.2.21. Suppose $f : D \subset \mathbb{R}^n \to \mathbb{R}$ is real valued function. A point $\mathbf{x}_0 \in D$ is a point of **absolute maximum** if $f(\mathbf{x}_0) \geq f(\mathbf{x})$ for all $\mathbf{x} \in D$. Similarly, it is a point of **absolute minimum** if $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$.

Strategy of finding Global maxima and Minima

(1) Find all critical points

(2) Compute values at critical points

(3) Find max or min on the boundary ∂U(by parametrization)

(4) Compare all values obtained in (2) and (3).

Example 4.2.22. Find the maximum and the minimum of $f(x,y) = x^2 + y^2 - x - y + 1$ in $D = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

sol. First we compute the critical points of f. Since

$$f_x = 2x - 1 = 0, \quad f_y = 2y - 1 = 0,$$

the point $(1/2, 1/2)$ is the only critical point. Since $f_{xx} = 2$, $f_{xy} = 0$, $f_{yy} = 2$, $f_{xx}f_{yy} - f_{xy}^2 = 4 > 0$, $f_{xx} = 2 > 0$, the point $(1/2, 1/2)$ is gives minimum by second derivative test. Now check the boundary D: $x^2 + y^2 = 1$. Use parametrization $x = \cos t$, $y = \sin t$, $0 \le t \le 2\pi$.

$$g(t) = \sin^2 t + \cos^2 t - \sin t - \cos t + 1 = 2 - \sin t - \cos t.$$

See $g'(t) = -\cos t + \sin t = 0$ hence $t = \pi/4, 5\pi/4$ are critical points. We have to check the end points $t = 0, 2\pi$ also. Hence the values are

$$g(0) = 1, \quad g(\pi/4) = 2 - \sqrt{2}.$$

$$g(5\pi/4) = 2 + \sqrt{2}, \quad g(2\pi) = 1.$$

Comparing, we see maximum is at $t = 5\pi/4$, $(x, y) = (-\sqrt{2}/2, -\sqrt{2}/2)$ and min at $\pi/4$.

□

Example 4.2.23. (See the book p.254) Find the maximum and the minimum of $f(x, y) = x^2 - xy + y^2 + 1$ in T where T is a square bounded by four lines $x = -1, x = 2, y = -1, y = 2$.

Existence of max and min

Definition 4.2.24. A set $D \subset R^n$ is **bounded** if $\|\mathbf{x}\| \le M$ for all $\mathbf{x} \in D$. It is **closed** if it contains all boundary points. (fig 4.8)

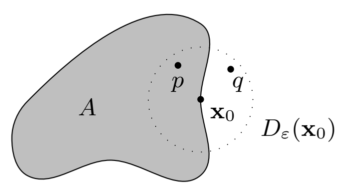


Figure 4.8: A neighborhood $D_\epsilon(\mathbf{x}_0)$ of a boundary point $\mathbf{x}_0$ contains contains both points of A and points not in A

Example 4.2.25. (1) $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \le 1\}$

(2) $D_0 = D - \{(0, 0)\}$

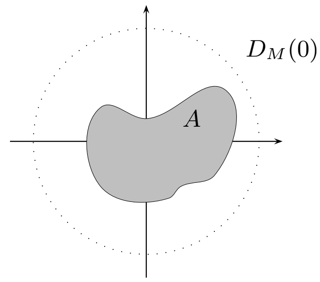


Figure 4.9: A is contained in a neighborhood $D_M(\mathbf{0})$.

(3) $S = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$

Theorem 4.2.26 (Existence of maximum and minimum). *If $f: D \to \mathbb{R}$ is continuous function defined on a* **closed bounded** *(it is also called* **compact***) set $D \subset \mathbb{R}^n$. Then there exist two points $\mathbf{x}_0$ and $\mathbf{x}_1 \in D$ such that for all $\mathbf{x} \in D$, the following holds:*

$$f(\mathbf{x}_0) \leq f(\mathbf{x}) \leq f(\mathbf{x}_1).$$

4.3 Constrained Extrema and Lagrange multiplier

Constrained Extrema

Suppose we are going to find a maximum of some function $f(x, y, z)$ constrained to a closed bounded set defined by another function $g(x, y, z) = c$. For example, we want to design an open box of volume $4\, ft^3$ which require the minimum amount of material. Let $A(x, y, z)$ be the surface area of the box. Then we want to find

$$\min A(x, y, z) = \min\{2(xy + yz) + zx\}$$

subject to to the condition

$$V = xyz = 4.$$

This condition is called the **constraint equation.**

A naive way of solving this problem is to eliminate one variable from the constraint, say, we let $z = 4/xy$ and substitute into the object function

$A(x, y, z)$ to get.

$$A(x, y, \frac{4}{xy}) = 2(xy + y\frac{4}{xy}) + x\frac{4}{xy} = 2xy + \frac{8}{x} + \frac{4}{y}$$

....

But what if the constraint equation is not easy? We study a more systematic way.

Lagrange multiplier method

Theorem 4.3.1 (Lagrange multiplier method)**.** *Assume* $f\colon \mathbb{R}^n \to \mathbb{R}$ *and* $g\colon \mathbb{R}^n \to \mathbb{R}$ *are of* $\mathcal{C}^1$ *class. And the restriction of* f *to the level set* $S = \{\mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = c\}$ *(written as* $f|_S$*) has a (local) maximum or minimum at* $\mathbf{x}_0 \in S$ *with* $\nabla g(\mathbf{x}_0) \neq 0$*. Then there is a scalar* λ *(***Lagrange multiplier***) such that*

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0).$$

Before proving the theorem, we see how to solve a minimization(max) problem with a constraint.

Example 4.3.2. Let us solve the above problem by Lagrange multiplier method. Let

$$\begin{aligned} A(x, y, z) &= 2(xy + yz) + zx, \\ g(x, y, z) &= xyz - 4. \end{aligned}$$

By the Lagrange multiplier method, we have

$$\nabla A = \lambda \nabla g \Rightarrow (2y + z, 2x + z, 2y + x) = \lambda(yz, zx, xy).$$

This gives three equations in four unknowns, x, y, z and λ. Appending the constraint equation, we have four by four system:

$$\begin{aligned} 2y + z &= \lambda yz \\ 2x + 2z &= \lambda zx \\ 2y + x &= \lambda xy \\ xyz &= 4. \end{aligned}$$

Since λ is not essential, we usually eliminate λ using any of the three equations.

Thus we get

$$\lambda = \frac{2y+z}{yz} = \frac{2x+2z}{zx} = \frac{2y+x}{xy}.$$

From these we get

$$\frac{2}{z} + \frac{1}{y} = \frac{2}{z} + \frac{2}{x} = \frac{2}{x} + \frac{1}{y}.$$

Hence

$$x = 2y, \quad z = 2y$$

Substituting into last eq. $(2y)y(2y) = 4$ hence $y = y = 1, x = z = 2$.

Proof. (of theorem 4.3.1.) First recall that $\nabla g(\mathbf{x}_0)$ is perpendicular to the level surface S at $\mathbf{x}_0$. Hence, if $\mathbf{c}(t)$ is any curve in S, then

$$\nabla g(\mathbf{x}_0) \cdot \mathbf{c}'(0) = 0.$$

Meanwhile, f has a local extreme at $\mathbf{x}_0$. Hence

$$0 = \left.\frac{df(\mathbf{c}(t))}{dt}\right|_{t=0} = \nabla f(\mathbf{x}_0) \cdot \mathbf{c}'(0).$$

Since the curve $\mathbf{c} \subset S$ is arbitrary, we conclude that $\nabla f(\mathbf{x}_0)$ is perpendicular to the surface S. Similarly, $\nabla g(\mathbf{x}_0)$ is also perpendicular to the surface S. Hence $\nabla f(\mathbf{x}_0)$ and $\nabla g(\mathbf{x}_0)$ are parallel(perpendicular to the same surface). Hence for some λ, $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$ holds. □

Theorem 4.3.3. *If f has maximum or minimum at a point $\mathbf{x}_0$ of S, then*

$$\nabla f \perp S.$$

Now a general minimization problem with a constraint is :

$$\begin{array}{ll} \text{Find the minimum of} & f(x,y,z) \\ \text{subject to} & g(x,y,z) = c. \end{array}$$

To solve it we solve system of equations with $n+1$ variables

$$\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x}) \tag{4.3}$$

$$g(\mathbf{x}) = c. \tag{4.4}$$

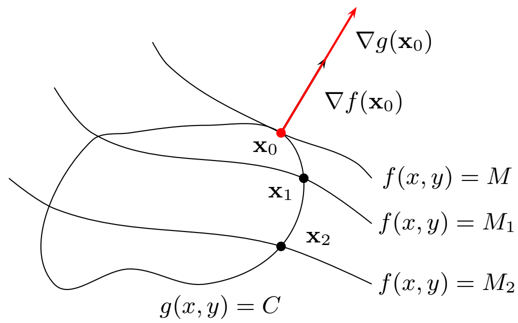


Figure 4.10: Lagrange multiplier method

Or

$$\begin{aligned} f_{x_1}(x_1,\cdots,x_n) &= \lambda g_{x_1}(x_1,\cdots,x_n) \\ &\vdots \\ f_{x_n}(x_1,\cdots,x_n) &= \lambda g_{x_n}(x_1,\cdots,x_n) \\ g(x_1,\cdots,x_n) &= c. \end{aligned} \tag{4.5}$$

Another interpretation of Lagrange multiplier method

Let

$$h(x_1,\cdots,x_n,\lambda) = f(x_1,\cdots,x_n) - \lambda[g(x_1,\cdots,x_n) - c]. \tag{4.6}$$

Lagrange multiplier method says : To find extreme points of $f|_S$, we should examine the extreme points of h. To see this, we set

$$\nabla_{\mathbf{x},\lambda} h = 0. \tag{4.7}$$

Then we see this is equivalent to solving equations (4.3). Thus Lagrange multiplier method is equivalent to solving **unconstrained extremal problem** (4.6) with extra variable λ.

Example 4.3.4. Find the extrema of $f(x,y) = x^2/4+y^2$ on the set $x^2+y^2 = 1$.

sol. Let $g(x,y) = x^2 + y^2$. Then constraint equation is $g(x,y) = 1$. Setting $\nabla f = \lambda \nabla g$, we have

$$\begin{cases} \frac{x}{2} &= 2\lambda x \\ 2y &= 2\lambda y \\ x^2 + y^2 &= 1. \end{cases}$$

From the first equation we see either $x = 0$ or $\lambda = 1/4$. So the critical points are $(0, \pm 1)$ corresponding to $\lambda = 1$ and $(\pm 1, 0)$ corresponding to $\lambda = 1/4$.

□

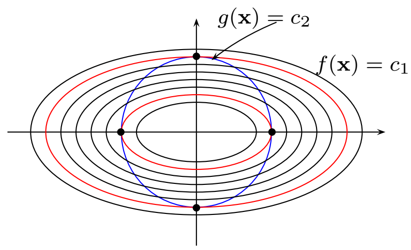


Figure 4.11: Level sets of g meets with the level set of f. Extreme occurs when two set meet tangentially

Example 4.3.5. Find max of $f(x, y) = x^2 - y^2$ on $S : x^2 + y^2 = 1$. (See figure) where the two level curves touch.

sol. Since $g(x, y) = x^2 + y^2 = 1$ and $\nabla f = (2x - 2y)$, $\nabla g = (2x, 2y)$ the equation is

$$\begin{aligned} f_x(x, y) = \lambda g_x(x, y) &\Longleftrightarrow 2x = \lambda 2x \\ f_y(x, y) = \lambda g_y(x, y) &\Longleftrightarrow -2y = \lambda 2y \\ g(x, y) = 1 &\Longleftrightarrow x^2 + y^2 = 1 \end{aligned}$$

From the first equation we get $x = 0$ or $\lambda = 1$. If $x = 0$, we see from third equation $y = \pm 1$. If $\lambda = 1$ then $y = 0$ and $x = \pm 1$. Now

$$\begin{aligned} f(0, 1) &= f(0, -1) = -1, \\ f(1, 0) &= f(-1, 0) = 1. \end{aligned}$$

Hence max is 1 min is -1.

□

Example 4.3.6. Find max of $f(x, y, z) = x + z$ subject to $x^2 + y^2 + z^2 = 1$.

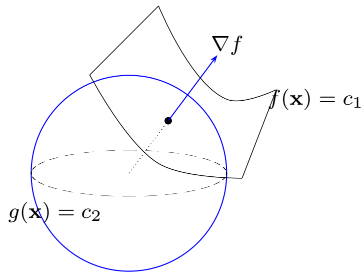


Figure 4.12: Level sets of g meets with the level set of f. Extreme occurs when two set meet tangentially

sol. Let $g(x, y, z) = x^2 + y^2 + z^2$. By the Lagrange multiplier method, we have $\nabla f = \lambda \nabla g$. Thus,

$$\begin{aligned} 1 &= 2x\lambda \\ 0 &= 2y\lambda \\ 1 &= 2z\lambda \\ 1 &= x^2 + y^2 + z^2. \end{aligned}$$

From first and third equation we see $\lambda \neq 0$ and $x = z$. Hence from second equation $y = 0$.

From fourth equation we obtain $x = z = \pm 1/\sqrt{2}$. Hence $(1/\sqrt{2}, 0, 1/\sqrt{2})$ and $(-1/\sqrt{2}, 0, -1/\sqrt{2})$.

$2/\sqrt{2}$ is max and $-2/\sqrt{2}$ is min.

■

Example 4.3.7. Find the maximum volume of rectangular box with fixed surface area $10m^2$.

sol. Let x, y, z be the dimension. Then volume is $f(x, y, z) = xyz$. But surface are is 10. Hence the condition $g(x, y, z) = 2(xy + yz + zx) = 10$ is the constraint.

$$\begin{aligned} yz &= \lambda(y + z), \\ xz &= \lambda(x + z), \\ yx &= \lambda(y + x), \\ 5 &= xy + yz + zx. \end{aligned}$$

Since $x > 0$, $y > 0$, $z > 0$, $y + z \neq 0$, $x + z \neq 0$. So eliminating λ we get $yz/(y + z) = xz/(x + z)$. Hence $x = y$. Similarly, $y = z$ and we see $x = y = z = \sqrt{5/3}$. i,e, $f(\sqrt{5/3}, \sqrt{5/3}, \sqrt{5/3}) = (5/3)^{3/2}$ are candidates for maximum or minimum.

Surface $S : xy + yz + zx = 5$ is not bounded. If function value $f(x, y, z)$ approaches 0 as any of x or y z approaches 0 or ∞ then $(5/3)^{3/2}$ is max.

□

Example 4.3.8. Find max(min) of $f(x, y) = x^2 - y^2$ on $x^2 + y^2 \leq 1$.

sol. Critical points of f. Since $f_x = 2x = 0$, $f_y = -2y = 0$, $(0, 0)$ is the only critical point. $f(0, 0) = 0$. But $D = f_{xx}f_{yy} - f_{xy}^2 = -4 < 0$ hence it is a saddle. We have seen in Example 4.3.5 that f has max and min 1 and -1

□

Several constraint

Theorem 4.3.9. *Let S be the surface determined by the following equations:*

$$
\begin{aligned}
g_1(\mathbf{x}) &= c_1, \\
&\cdots \\
g_k(\mathbf{x}) &= c_k.
\end{aligned}
$$

If $f\colon \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^1$ class and has a (local) maximum or minimum on S, where $\nabla g_1(\mathbf{x}_0), \cdots, \nabla g_k(\mathbf{x}_0)$ are linearly independent, then there exist scalars $\lambda_1, \ldots, \lambda_k$ such that

$$\nabla f(\mathbf{x}_0) = \lambda_1 \nabla g_1(\mathbf{x}_0) + \cdots + \lambda_k \nabla g_k(\mathbf{x}_0).$$

(Here k-vectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$ in $\mathbb{R}^n$ are called **linearly independent** *if the only way to satisfy the relation $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = 0$ for some scalars $a_1, \cdots, a_k$ is $a_1 = a_2 = \cdots = a_k = 0$.)*

Proof. Let $S_i = \{\mathbf{x} \in \mathbb{R}^n | g_i(\mathbf{x}) = c_i\}$ for $i = 1, \cdots, k$. Then $S = S_1 \cap \cdots \cap S_k$. Hence any vector tangent to S must be tangent to all S_i's. Let $\mathbf{x}_0 \in S$ be an extreme point of f. Then for any curve $\mathbf{x}$ with $\mathbf{x}(t_0) = \mathbf{x}_0$ contained in S, the function

$$\mathbf{F}(t) = f(\mathbf{x}(t))$$

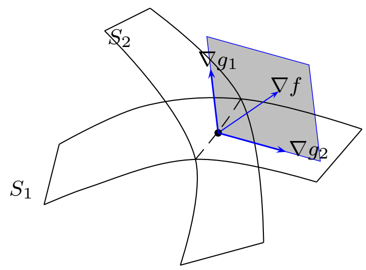


Figure 4.13: Several constraints; ∇f lies in the span of ∇g_1 and ∇g_2

assumes an extreme at t_0. Hence

$$0 = F'(t_0) = \nabla f(\mathbf{x}_0) \cdot \mathbf{x}'(t_0)$$

Hence $\nabla f(\mathbf{x}_0)$ is perpendicular to the tangent vector $\mathbf{x}'(t_0)$. But we also have

$$\begin{aligned} \nabla g_1(\mathbf{x}_0) \cdot \mathbf{x}'(t_0) &= 0, \\ \cdots &= \\ \nabla g_k(\mathbf{x}_0) \cdot \mathbf{x}'(t_0) &= 0, \end{aligned}$$

we must have $\nabla f(\mathbf{x}_0)$ lie in the hyperplane spanned by $\nabla g_1(\mathbf{x}_0), \cdots, \nabla g_k(\mathbf{x}_0)$.

□

Example 4.3.10. Find extreme points of $f = x + y + z$ subject to $x^2 + y^2 = 2$ and $x + z = 1$.

sol. Constraints are $g_1 = x^2 + y^2 - 2 = 0$ and $g_2 = x + z - 1 = 0$. Thus

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2.$$

Since

$$\begin{aligned} g_1 &= x^2 + y^2 - 2 \\ g_2 &= x + z - 1 \end{aligned}$$

we obtain

$$\begin{aligned} 1 &= \lambda_1 \cdot 2x + \lambda_2 \cdot 1 \\ 1 &= \lambda_1 \cdot 2y + \lambda_2 \cdot 0 \\ 1 &= \lambda_1 \cdot 0 + \lambda_2 \cdot 1 \\ 0 &= x^2 + y^2 - 2 \\ 0 &= x + z - 1 \end{aligned}$$

From third equation we obtain $\lambda_2 = 1$ and so $\lambda_1 \cdot 2x = 0$ and $\lambda_1 \cdot 2y = 1$. From second, we see $\lambda_1 \neq 0$, hence $x = 0$. Thus $y = \pm\sqrt{2}$ and $z = 1$. Hence possible extrema are $(0, \pm\sqrt{2}, 1)$. $(0, \sqrt{2}, 1)$ give max$(0, -\sqrt{2}, 1)$ give min.

■

Example 4.3.11. Suppose the cone $z^2 = x^2 + y^2$ is sliced by the plane $z = x + y + 2$ to create a conic section C. Find the points nearest to the curve C from the origin.

sol. Let $f = x^2 + y^2 + z^2$ be the square of the distance. The constraints are

$$\begin{cases} g_1 &= x^2 + y^2 - z^2 = 0 \\ g_2 &= x + y - z = -2. \end{cases}$$

Set

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2.$$

together with the constraint equations:

$$\begin{cases} 2x & = 2\lambda_1 x + \lambda_2 \\ 2y & = 2\lambda_1 y + \lambda_2 \\ 2z & = -2\lambda_1 z - \lambda_2 \\ x^2 + y^2 - z^2 & = 0 \\ x + y - z & = -2. \end{cases}$$

Eliminate λ_2 from the first two equations, we get

$$\lambda_2 = 2x - 2\lambda_1 x = 2y - 2\lambda_1 y$$

Thus

$$2(x-y)(1-\lambda_1)=0.$$

□

Example 4.3.12. Find the absolute maximum and minimum of $f = xy$ on $x^2+y^2 \le 1$.

sol. First find critical points.

$$\frac{\partial f}{\partial x}=y, \quad \frac{\partial f}{\partial y}=x$$

$(0,0)$ is the only critical point. Now consider on the unit circle $g = x^2+y^2 = 1$.

$$\nabla f = \lambda \nabla g \Rightarrow (y,x)=\lambda(2x,2y)$$

Thus we get $y = 4\lambda^2 y$, or $\lambda = \pm 1/2$ and $y = \pm x$. So $y = \pm 1/\sqrt{2} = x$. Checking f values at these points we see f has max $1/2$ and $-1/2$ is min. By checking second derivative, $(0,0)$ is saddle

□

Example 4.3.13. Find absolute maximum and minimum of $f = \frac{1}{2}x^2 + \frac{1}{2}y^2$ on $\frac{1}{2}x^2 + y^2 \le 1$.

sol. First find critical points.

$$\frac{\partial f}{\partial x}=x, \quad \frac{\partial f}{\partial y}=y$$

$(0,0)$ is the only critical point. Now consider on the unit circle $g = x^2+y^2 = 1$. Use Lagrange method.(Recall one could use parametrization as before).

$$\nabla f = \lambda \nabla g \Rightarrow (x,y)=\lambda(x,2y)$$

Thus we get

$$\begin{aligned} x &= \lambda x \\ y &= 2\lambda y \\ \frac{1}{2}x^2+y^2 &= 1 \end{aligned}$$

So $(0, \pm 1)$ and $(\pm\sqrt{2}, 0)$. Checking f values at these points we see f has max 1 at $(0, 0)$.

■

4.4 Some applications

Least Square approximation

We assume we have many data points like $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. We

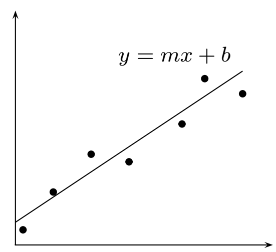


Figure 4.14: Least square fit by a line; linear regression

would like to find a simplest relation between the data; linear relation. Thus assume $f(x) = mx + b$ and want to find m, b so that $D(m, b) := \sum_{i=1}^n (y_i - f(x_i))^2$ is minimized!

$$D(m,b) = \sum_{i=1}^n (y_i - (mx_i + b))^2 = \sum_{i=1}^n y_i^2 - 2\sum_{i=1}^n y_i(mx_i + b) + \sum_{i=1}^n (mx_i + b)^2.$$

Hence

$$\begin{aligned} \frac{\partial D}{\partial m} &= -2\sum_{i=1}^n x_i y_i + 2\sum_{i=1}^n (mx_i + b)x_i \\ &= -2\sum_{i=1}^n x_i y_i + 2m\sum_{i=1}^n x_i^2 + 2b\sum_{i=1}^n x_i = 0 \end{aligned}$$

and

$$\begin{aligned} \frac{\partial D}{\partial b} &= -2\sum_{i=1}^n y_i + 2\sum_{i=1}^n (mx_i + b) \\ &= -2\sum_{i=1}^n y_i + 2m\sum_{i=1}^n x_i + 2nb = 0. \end{aligned}$$

Solving for m and b, we obtain

Proposition 4.4.1. *The least square approximation by linear function is obtained by*

$$\begin{aligned} m &= \frac{n\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\ b &= \frac{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n x_i y_i)}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}. \end{aligned}$$

Example 4.4.2. Find a linear regression of the data $(1,2),(2,1),(3,5),(4,3),(5,4)$.

$$\begin{aligned} D(m,b) &= (2-(m+b))^2+(1-(2m+b))^2+(5-(3m+b))^2+(3-(4m+b))^2+(4-(5m+b))^2 \\ &= 55 - 4(m+b) + (m+b)^2 - 2(2m+b) + (2m+b)^2 - 10(3m+b) \\ &\quad + (3m+b)^2 - 6(4m+b) + (4m+b)^2 - 8(5m+b) + (5m+b)^2 \\ &= 55 - 102m - 30b + 55m^2 + 30mb + 5b^2. \quad (4.8) \end{aligned}$$

$$\begin{aligned} \frac{\partial D}{\partial m} &= -102 + 110m + 30b = 0 \\ \frac{\partial D}{\partial b} &= -30 + 30m + 10b = 0. \end{aligned}$$

Hence $m = 3/5,\ b = 6/5$ and $y = 3/5x + 6/5$.