Mathematics for Analytics and Finance

Sami Najafi

MSIS2402/2405

Module 4

Partial Derivative

The partial derivative measures how a function changes as one of its variables changes, while holding other variables constant. This is useful in business, especially when analyzing how changes in one factor (e.g., price or advertising spend) affect outcomes (e.g., profit or sales), while keeping other factors fixed.

- The partial derivatives with respect to x is:

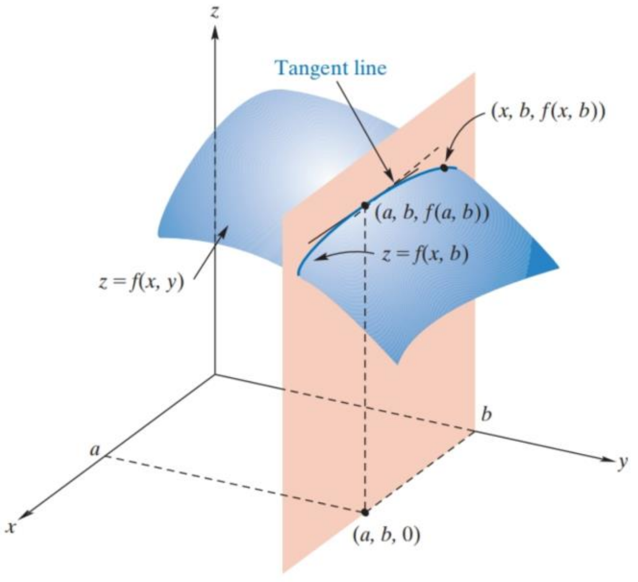
$$f_x(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$

This shows how f changes as x changes, with y fixed.

- The partial derivatives with respect to x is:

$$f_y(x, y) = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}$$

This shows how f changes as y changes, with x fixed.



Geometric interpretation of $f_x(a, b)$.

Example

Finding Partial Derivatives

If $f(x, y) = xy^2 + x^2y$, find $f_x(x, y)$ and $f_y(x, y)$. Also, find $f_x(3,4)$ and $f_y(3,4)$.

Solution

The partial derivatives are

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

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

Thus, the solutions are

$$f_x(3, 4) = 40$$

$$f_y(3, 4) = 33$$

Example

Partial Derivatives of a Function of Four Variables

If $p = g(r, s, t, u) = \frac{rsu}{rt^2+s^2t}$, find $\frac{\partial p}{\partial s}, \frac{\partial p}{\partial t}$ and $\left.\frac{\partial p}{\partial t}\right|_{(0,1,1,1)}$.

Solution

By partial differentiating, we get

$$\frac{\partial p}{\partial s} = \frac{(rt^2 + s^2t)(ru) - (rsu)(2st)}{(rt^2 + s^2t)^2} = \frac{ru(rt - s^2)}{t(rt + s^2)^2}$$

$$\frac{\partial p}{\partial t} = -rsu(rt^2 + s^2t)^{-2}(2rt + s^2) \Rightarrow \frac{\partial p}{\partial s} = -\frac{rsu(2rt + s^2)}{(rt^2 + s^2t)^2}$$

$$\left.\frac{\partial p}{\partial t}\right|_{(0,1,1,1)} = 0$$

Applications of Partial Derivatives

Interpretation of rate of change

$\frac{\partial f(\boldsymbol{x})}{\partial x_i}$: the rate of change of f with respect to x_i keeping all other variables constant.

For example, consider a manufacturer producing x units of product $\boldsymbol{X}$ and y units of product $\boldsymbol{Y}$. Then the total cost c of these units depends on both x and y forming a joint-cost function.

- $\frac{\partial c}{\partial x}$ is approximately additional the cost of producing one more unit of $\boldsymbol{X}$ assuming y is fixed.
- $\frac{\partial c}{\partial y}$ is approximately the additional cost of producing one more unit of $\boldsymbol{Y}$ assuming x is fixed.

Example

Marginal Costs

A company manufactures two types of skis, the Lightning and the Alpine models. Suppose the joint-cost function for producing x pairs of the Lightning model and y pairs of the Alpine model per week is

$$c(x, y) = 0.07x^2 + 75x + 85y + 6000$$

where c is expressed in dollars. Determine the marginal costs $\frac{\partial c}{\partial x}$ and $\frac{\partial c}{\partial y}$ when $x = 100$ and $y = 50$ and interpret the results.

Solution

The marginal costs are

$$\frac{\partial c}{\partial x} = 0.14x + 75 \quad \text{and} \qquad \frac{\partial c}{\partial y} = 85$$

Thus,

$$\left.\frac{\partial c}{\partial x}\right|_{(100,50)} = 0.14(100) + 75 = \$89$$

and

$$\left.\frac{\partial c}{\partial y}\right|_{(100,50)} = \$85$$

Example

Marginal Production

A toy manufacturer has a production function: $P = \sqrt{lk}$, where l is the number of labor-hours per week, k is the capital (in hundreds of dollars per week), and P is weekly production of the toy in gross (1 gross = 144 units). Find the marginal production function with respect to the number of labor-hour and the marginal production function with respect to the capital. Evaluate these two functions at $l = 400$ and $k = 16$. Interpret the results.

Solution

Since $P = (lk)^{1/2}$, thus:

$$\frac{\partial P}{\partial l} = \frac{1}{2}(lk)^{-1/2}k = \frac{k}{2\sqrt{lk}} \quad \text{and} \quad \frac{\partial P}{\partial k} = \frac{l}{2\sqrt{lk}}$$

$$\left.\frac{\partial P}{\partial l}\right|_{l=400,k=16} = \frac{1}{10} \quad \text{and} \quad \left.\frac{\partial P}{\partial k}\right|_{l=400,k=16} = \frac{5}{2}$$

Implicit Partial Differentiation

Sometimes, functions are not given explicitly. Instead, variables are interrelated in complex ways. Implicit differentiation helps us find the rates of change even in these cases.

Example: If $\frac{xz^2}{x+y} + y^2 = 0$, evaluate $\frac{\partial z}{\partial x}$ when $x = -1$, $y = 2$, and $z = 2$.

Solution

Using partial differentiation, we get

$$\frac{\partial}{\partial x}\left(\frac{xz^2}{x+y}\right)+\frac{\partial}{\partial x}\left(y^2\right)=\frac{\partial}{\partial x}(0)$$

$$2xz(x+y)\frac{\partial z}{\partial x}+z^2(x+y)-xz^2=0$$

$$\frac{\partial z}{\partial x}=-\frac{yz}{2x(x+y)} \quad z\neq 0$$

$$\left.\frac{\partial z}{\partial x}\right|_{(-1,2,2)}=2$$

Second-Order Partial Derivatives

When $\frac{\partial f(\boldsymbol{x})}{\partial x_i} \equiv f_i(\boldsymbol{x})$ is differentiated with respect to x_j, the result is a second-order partial derivative with respect to x_i and x_j:

$$\boxed{\frac{\partial}{\partial x_j}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_i}\right) \equiv \frac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i} \equiv f_{ji}(\boldsymbol{x})}$$

- First-order partial derivative: Measures how the function changes with one variable while keeping others fixed.
- Second-order partial derivative: Measures how the rate of change itself changes.

Young Theorem: For any twice differentiable function $f(\boldsymbol{x})$, the mixed second-order partial derivatives are equal:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_j \partial x_i} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$$

Example: Find the four second-order partial derivatives of $f(x, y) = x^2 y + x^2 y^2$.

Solution

$f_x(x, y) = 2xy + 2xy^2$

$f_{xx}(x, y) = 2y + 2y^2$ and $f_{xy}(x, y) = 2x + 4xy$

$f_y(x, y) = x^2 + 2x^2y$

$f_{yy}(x, y) = 2x^2$ and $f_{yx}(x, y) = 2x + 4xy$

Example

Second-Order Partial Derivative of an Implicit Function

Determine $\frac{\partial^2 z}{\partial x^2}$ if $z^2 = xy$.

Solution

By implicit differentiation,

$$\frac{\partial}{\partial x}\left(z^2\right) = \frac{\partial}{\partial x}(xy) \Rightarrow \frac{\partial z}{\partial x} = \frac{y}{2z} \quad z \neq 0$$

Differentiating both sides with respect to *x*, we obtain

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{1}{2}yz^{-1}\right) \Rightarrow \frac{\partial^2 z}{\partial x^2} = -\frac{1}{2}yz^{-2}\frac{\partial z}{\partial x}$$

Substituting $\frac{\partial z}{\partial x} = \frac{y}{2z}$,

$$\frac{\partial^2 z}{\partial x^2} = -\frac{1}{2}yz^{-2}\left(\frac{y}{2z}\right) = -\frac{y^2}{4z^3} \qquad z \neq 0$$

The Chain Rule

Chain Rule

If $f(\boldsymbol{x})$ and $\boldsymbol{x}$ are functions with respect to a variable r then:

$$\frac{\partial f(\boldsymbol{x})}{\partial r} = \boldsymbol{\nabla} f(\boldsymbol{x}) \cdot \frac{\partial \boldsymbol{x}}{\partial r} = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x})}{\partial x_i} \frac{\partial x_i}{\partial r}$$

Where,

- $\boldsymbol{\nabla} f(\boldsymbol{x}) = \left[\frac{\partial f}{\partial x_i}\right]_{n\times 1}$ is called the gradient, which tells us how the function changes with each variable x_i.
- $\frac{\partial \boldsymbol{x}}{\partial r} = \left[\frac{\partial x_i}{\partial r}\right]_{n\times 1}$ shows how the input variables x_i change with respect to r.

For $\boldsymbol{x} = (x, y)$:

$$\frac{\partial f(x,y)}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}$$

Chain Rule

Proof (for $\boldsymbol{x} = (x, y)$):

$$f_r'(x,y) = \lim_{h\to 0} \frac{f(x(r+h), y(r+h)) - f(x(r), y(r))}{h}$$

$$= \lim_{h\to 0} \frac{f(x(r+h), y(r+h)) - f(x(r+h), y(r)) + f(x(r+h), y(r)) - f(x(r), y(r))}{h}$$

$$= \lim_{h\to 0} \left\{ \left(\frac{f(x(r+h), y(r)) - f(x(r), y(r))}{x(r+h) - x(r)} \right) \left(\frac{x(r+h) - x(r)}{h} \right) + \left(\frac{f(x(r+h), y(r+h)) - f(x(r+h), y(r))}{y(r+h) - y(r)} \right) \left(\frac{y(r+h) - y(r)}{h} \right) \right\}$$

$$= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}$$

Example

Chain Rule in Exponential and Logarithmic Functions

Given that $z = e^{xy}$, $x = r - 4s$, and $y = r - s$, find $\frac{\partial z}{\partial r}$ in terms of r and s.

Solution

By the chain rule,

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = (x+y)e^{xy}$$

Since $x = r - 4s$ and $y = r - s$,

$$\left.\frac{\partial z}{\partial r}\right|_{\substack{x=r-4s \\ y=r-s}} = (2r-5s)e^{r^2-5rs+4s^2}$$

Directional (Fréchet) Derivative

Definition: The directional derivative of a function f at a point $\boldsymbol{x}$, in the direction of a unit vector $\boldsymbol{u}$, measures the rate of change of f as we move in that direction.

Formula:

$$Df_{\boldsymbol{u}}(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{u}) - f(\boldsymbol{x})}{h}$$

Result: For a differentiable function f, the directional derivative can be calculated as:

$$Df_{\boldsymbol{u}}(\boldsymbol{x}) = \boldsymbol{\nabla} f(\boldsymbol{x}) \cdot \boldsymbol{u}$$

where $\boldsymbol{\nabla} f(\boldsymbol{x})$ is the gradient of f at $\boldsymbol{x}$, and $\cdot$ denotes the dot product.

Directional (Fréchet) Derivative

Proof (for interested students):

From the definition,

$$D f_{\boldsymbol{u}}(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{u}) - f(\boldsymbol{x})}{h}.$$

To compute this derivative, assume: $w(h) = f(\boldsymbol{x} + h\boldsymbol{u})$. So,

$$D f_{\boldsymbol{u}}(\boldsymbol{x}) = \lim_{h \to 0} \frac{w(h) - w(0)}{h} = w'(0)$$

By using the Chain Rule, assuming $\boldsymbol{z}(h) = \boldsymbol{x} + h\boldsymbol{u}$ we have:

$$w'(h) = \boldsymbol{\nabla} f(\boldsymbol{z}) \cdot \frac{\partial \boldsymbol{z}}{\partial h} = \boldsymbol{\nabla} f(\boldsymbol{x} + h\boldsymbol{u}) \cdot \boldsymbol{u}$$

So, replacing $h = 0$ in the above gives:

$$D f_{\boldsymbol{u}}(\boldsymbol{x}) = w'(0) = \boldsymbol{\nabla} f(\boldsymbol{x}) \cdot \boldsymbol{u}$$

Remark: Setting $w(h) = f(\boldsymbol{x} + h\boldsymbol{u})$ is a very important technique. Make sure not to forget it! It is used a lot.

Directional (Fréchet) Derivative

Example: Compute the derivative of $f(\boldsymbol{x}) = x_1 x_2$ at $\boldsymbol{x} = (1,0)$ in the direction of $\boldsymbol{u} = (-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$.

Solution:

$$\boldsymbol{\nabla} f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right) = (x_2, x_1)$$

$$Df_{\boldsymbol{u}}(\boldsymbol{x}) = \boldsymbol{\nabla} f(\boldsymbol{x}) \cdot \boldsymbol{u} = \frac{\partial f}{\partial x_1} u_1 + \frac{\partial f}{\partial x_2} u_2 = -\frac{x_2}{\sqrt{2}} - \frac{x_1}{\sqrt{2}} = -\frac{1}{\sqrt{2}}$$

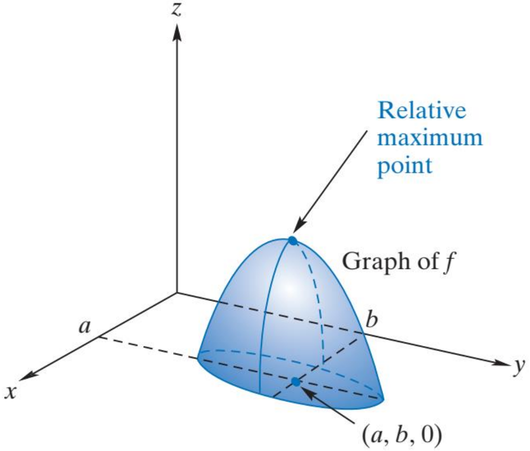
Unconstrained Optimization

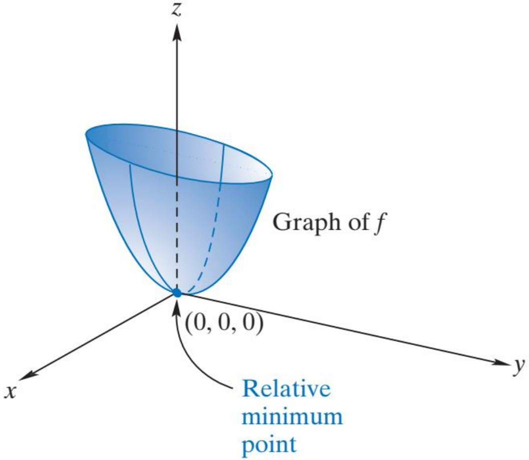
Relative Maximum and Minimum

Definition: A function $z = f(\boldsymbol{x})$ has a relative maximum at the point $\boldsymbol{x}^*$ if, for all points $\boldsymbol{x}$ in the neighborhood of $\boldsymbol{x}^*$:

$$f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$$

For a relative minimum, we replace $\geq$ by $\leq$ in the above inequality.





First Order Necessary Condition

If $\boldsymbol{x}^*$ is a local (relative) minimum or maximum of $f(\boldsymbol{x})$, then

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} = 0, \qquad i = 1, \ldots, n$$

First Order Necessary Condition

Proof (for interested students):

Suppose $\boldsymbol{x}^*$ to be a minimum point (the proof for maximum is similar) and

$$w(h) = f(\boldsymbol{x}^* + h\boldsymbol{u})$$

Since $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^* + h\boldsymbol{u})$ for h small enough, we have $w(0) \leq w(h)$. So, by First Order Necessary Condition, $w'(0) = 0$. From Directional Derivative's slide, we found that:

$$w'(0) = \boldsymbol{\nabla} f(\boldsymbol{x}) \cdot \boldsymbol{u} = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x})}{\partial x_i} u_i = 0.$$

But, since this must hold for every $\boldsymbol{u}$, in particular $\boldsymbol{u} = \boldsymbol{e}_i$, we must have:

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} = 0, \qquad i = 1, \dots, n,$$

Remark: $Df_{\boldsymbol{u}}(\boldsymbol{x}^*) = 0$ means that f does not change by moving a bit from $\boldsymbol{x}^*$ in any direction ($\boldsymbol{x}^*$ is a stationary point).

Example

Finding Critical Points

Find the critical points of the function $f(x, y, z) = 2x^2 + xy + y^2 + 100 - z(x + y - 100)$

Solution

Since

$$f_x(x, y, z) = 4x + y - z = 0$$

$$f_y(x, y, z) = x + 2y - z = 0$$

$$f_z(x, y, z) = -x - y + 100 = 0$$

we solve the system and get

$$\begin{cases} x = 25 \\ y = 75 \\ z = 175 \end{cases}$$

The Matrix of Second Derivatives

Definition: The second-order partial derivatives matrix (also known as Hessian) is a square matrix:

$$\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{\nabla}^2 f(\boldsymbol{x}) = \left[f_{x_i x_j}(\boldsymbol{x})\right]_{n\times n}.$$

$\boldsymbol{\nabla}^2 f(\boldsymbol{x})$ gives information about the concavity of f. To interpret $\boldsymbol{\nabla}^2 f(\boldsymbol{x})$, for all nonzero directions $\boldsymbol{u}$, we use the quadratic form:

$$\boldsymbol{u}^T \boldsymbol{\nabla}^2 f(\boldsymbol{x}) \boldsymbol{u} = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{x_i x_j}(\boldsymbol{x}) u_i u_j$$

- If $\boldsymbol{u}^T \boldsymbol{\nabla}^2 f(\boldsymbol{x}) \boldsymbol{u} \geq 0$, then f is convex or flat at $\boldsymbol{x}$
  - $\boldsymbol{\nabla}^2 f(\boldsymbol{x})$ is called positive semidefinite (PSD)
- If $\boldsymbol{u}^T \boldsymbol{\nabla}^2 f(\boldsymbol{x}) \boldsymbol{u} > 0$, then f is strictly convex at $\boldsymbol{x}$
  - $\boldsymbol{\nabla}^2 f(\boldsymbol{x})$ is called positive definite (PD)
- If $\boldsymbol{u}^T \boldsymbol{\nabla}^2 f(\boldsymbol{x}) \boldsymbol{u} \leq 0$, then f is concave or flat at $\boldsymbol{x}$
  - $\boldsymbol{\nabla}^2 f(\boldsymbol{x})$ is called negative semidefinite (PSD)
- If $\boldsymbol{u}^T \boldsymbol{\nabla}^2 f(\boldsymbol{x}) \boldsymbol{u} < 0$, then f is strictly concave at $\boldsymbol{x}$
  - $\boldsymbol{\nabla}^2 f(\boldsymbol{x})$ is called negative definite (ND)

Second Order Necessary Conditions

Let $f(\boldsymbol{x})$ be twice differentiable.

1. If $\boldsymbol{x}^*$ is a local interior minimum of $f(\boldsymbol{x})$, then $\boldsymbol{\nabla}^2 f(\boldsymbol{x}^*)$ is positive semidefinite.
2. If $\boldsymbol{x}^*$ is a relative maximum, then $\boldsymbol{\nabla}^2 f(\boldsymbol{x}^*)$ is negative semidefinite.

Second Order Necessary Conditions

Proof (for interested students):

Suppose $\boldsymbol{x}^*$ to be a min point (the proof for max is similar) and

$$w(h) = f(\boldsymbol{x}^* + h\boldsymbol{u})$$

Since $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^* + h\boldsymbol{u})$ for h near zero, $w(0) \leq w(h)$. So, by Second Order Condition, $w''(0) \geq 0$:

$$w'(h) = \sum_{i=1}^{n} f_{x_i}(\boldsymbol{x} + h\boldsymbol{u})u_i \Rightarrow w''(h) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} f_{x_i x_j}(\boldsymbol{x} + h\boldsymbol{u})u_j\right)u_i \Rightarrow w''(0) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{x_i x_j}(\boldsymbol{x})u_i u_j \geq 0$$

or:

$$\boldsymbol{s}^T \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s} \geq 0.$$

(Note that $\frac{\partial f_{x_i}(\boldsymbol{x}+h\boldsymbol{u})}{\partial h}$ is computed in the same way as $\frac{\partial f(\boldsymbol{x}+h\boldsymbol{u})}{\partial h}$, which results in: $\frac{\partial f_{x_i}(\boldsymbol{x}+h\boldsymbol{u})}{\partial h} = \sum_{j=1}^{n} f_{x_i x_j}(\boldsymbol{x} + h\boldsymbol{u})u_j$.)

Second Order Sufficient Conditions

Let $f(\boldsymbol{x})$ be twice differentiable.

1. If $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ and $\nabla^2 f(\boldsymbol{x}^*)$ is positive semidefinite, then $\boldsymbol{x}^*$ is a local interior minimum of $f(\boldsymbol{x})$.
2. If $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ and $\nabla^2 f(\boldsymbol{x}^*)$ is negative semidefinite, then $\boldsymbol{x}^*$ is a local interior maximum of $f(\boldsymbol{x})$.

Second Order Sufficient Conditions

Proof (for interested students):

Suppose $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ and $\nabla^2 f(\boldsymbol{x}^*)$ is positive semidefinite. Then assuming $w(h) = f(\boldsymbol{x}^* + h\boldsymbol{s})$, we have:

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0} \Rightarrow w'(0) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{s} = 0,$$

$$\nabla^2 f(\boldsymbol{x}^*) \geq 0 \Rightarrow w''(0) = \boldsymbol{s}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{s} \geq 0.$$

Therefore, by Second Order Sufficient Condition (SOSC) for single variable functions, $w(h)$ has a local minimum at $h = 0$, i.e., $w(0) \leq w(h)$. Considering that $w(h) = f(\boldsymbol{x}^* + h\boldsymbol{u})$, this means that:

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^* + h\boldsymbol{u})$$

For any h (near zero) and any direction $\boldsymbol{s}$. This means that $f(\boldsymbol{x})$ has a (local) minimum at $\boldsymbol{x}^*$.

Second Order Condition for Two Variables

For two variables, the Second Order Conditions simplify to the below conditions:

Let $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \mathbf{0}$ and:

$$D(\boldsymbol{x}^*) = f_{xx}(\boldsymbol{x}^*) f_{yy}(\boldsymbol{x}^*) - f_{xy}^2(\boldsymbol{x}^*)$$

Then:

1. If $D(\boldsymbol{x}^*) > 0$ and $f_{xx}(\boldsymbol{x}^*) < 0$, then f has a relative maximum at $\boldsymbol{x} = \boldsymbol{x}^*$.
2. If $D(\boldsymbol{x}^*) > 0$ and $f_{xx}(\boldsymbol{x}^*) > 0$, then f has a relative minimum at $\boldsymbol{x} = \boldsymbol{x}^*$.
3. If $D(\boldsymbol{x}^*) < 0$, then f has a saddle point at $\boldsymbol{x} = \boldsymbol{x}^*$.
4. If $D(\boldsymbol{x}^*) = 0$, the test is inconclusive.

For more than 2 variables, use R.

Second Derivative Test for Two Variables

Proof (for interested students):

If $\boldsymbol{x} = (x, y)$, then

$$\boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{H}(\boldsymbol{x})\boldsymbol{u} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = u_1^2 f_{xx} + 2u_1 u_2 f_{xy} + u_2^2 f_{yy}$$

This expression can be seen as a quadratic form based on u_1(or alternatively based on u_2). The sign of this expression depends on the discriminant:

If $\boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{H}(\boldsymbol{x})\boldsymbol{u} > \boldsymbol{0}$, then $\Delta = 4u_2^2\left(f_{xy}^2 - f_{xx}f_{yy}\right) < 0$ and $f_{xx} > 0$.

If $\boldsymbol{u}^{\boldsymbol{T}}\boldsymbol{H}(\boldsymbol{x})\boldsymbol{u} < \boldsymbol{0}$, then $\Delta = 4u_2^2\left(f_{xy}^2 - f_{xx}f_{yy}\right) < 0$ and $f_{xx} < 0$.

The term $f_{xx}f_{yy} - f_{xy}^2$ is known as the Determinant of $\boldsymbol{H}(\boldsymbol{x})$ and is denoted by $D(\boldsymbol{x})$ or $|H(\boldsymbol{x})|$. (For those interested, additional details on determinants can be found in the supplementary slides.)

Example

Applying the Second-Derivative Test

Examine $f(x, y) = x^3 + y^3 - xy$ for relative maxima or minima by using the second derivative test.

Solution

We find critical points,

$$f_x(x,y) = 3x^2 - y = 0 \quad \text{and} \quad f_y(x,y) = 3y^2 - x = 0$$

which gives $(0,0)$ and $\left(\frac{1}{3},\frac{1}{3}\right)$. Now,

Thus,

$$f_{xx}(x,y) = 6x \qquad f_{yy}(x,y) = 6y \qquad f_{xy}(x,y) = -1$$

$$D(x,y) = (6x)(6y) - (-1)^2 = 36xy - 1$$

$D(x,y) < 0 \Rightarrow$ no relative extremum at $(0,0)$.

$D\left(\frac{1}{3},\frac{1}{3}\right) < 0$ and $f_{xx}\left(\frac{1}{3},\frac{1}{3}\right) > 0 \Rightarrow$ relative minimum at $\left(\frac{1}{3},\frac{1}{3}\right)$

Value of the function is

$$f\left(\frac{1}{3},\frac{1}{3}\right) = \left(\frac{1}{3}\right)^3 + \left(\frac{1}{3}\right)^3 - \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = -\frac{1}{27}$$

Example

Applying the Second-Derivative Test

Examine $f(x, y) = x^4 + (x - y)^4$ for relative extrema.

Solution

We find critical points at $(0,0)$ through

$$f_x(x,y) = 4x^3 + 4(x-y)^3 = 0 \qquad f_y(x,y) = -4(x-y)^3 = 0$$

$$f_{xx}(x,y) = 12x^2 + 12(x-y)^2 = 0 \qquad f_{yy}(x,y) = 12(x-y)^2 \qquad f_{xy}(x,y) = 0$$

$D(0,0) = 0 \Rightarrow$ no information.

f has a relative (and absolute) minimum at $(0,0)$.

Checking Matrix Definiteness in R

```
is.positive.definite(H)
is.negative.definite(H)

These are logical codes that return TRUE if the Hessian matrix is positive (/negative) definite.
Remark: These codes need installing the "matrixcal" package first
```

Example:

```
Q=matrix(c(4,1,-1,1,2,-1,-1,-1,1),nrow=3, byrow=TRUE)
is.positive.definite(Q)
[1] TRUE
```

Unconstrained Optimization in R

```
optim(initial, func, method, lower, upper)


initial: Initial values for the parameters to be optimized over.
func: A function to be minimized.
method (optional): the method (e.g., "Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN", "Brent")
lower, upper: the lower and upper bounds for the variable
```

Example: Minimize $f(x,y) = x^3 + y^3 - xy$

```
func =function(x){x[1]^3+x[2]^3-x[1]*x[2]}

optim(c(0,0), func)
$par
[1] 0.3333333 0.3333333

$value
[1] -0.03703704

grad(func,x)                    # calculates the gradient of the function at x

[1] -4.790924e-09 -5.519045e-09

hessian(func,x)                 # calculates the hessian of the function at x

     [,1] [,2]
[1,]    2   -1
[2,]   -1    2
```

Constrained Optimization

Equality Constraint

Lagrange Multiplier Theorem: Consider an optimization problem with multiple equality constraints:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\text{subject to:} \quad g_i(\boldsymbol{x}) = 0 \qquad i = 1, \dots, m.$$

where $f: \mathbb{R}^n \to \mathbb{R}$ and $g_i: \mathbb{R}^n \to \mathbb{R}, i = 1, \dots, m$, are continuously differentiable functions. Let $\boldsymbol{x}^*$ be a local minimum of the problem. Then there exists a unique vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_m)$, called the Lagrange multiplier vector, such that

$$\boldsymbol{\nabla} f(\boldsymbol{x}^*) - \sum_{i=1}^{m} \lambda_i \boldsymbol{\nabla} g_i(\boldsymbol{x}^*) = \boldsymbol{0}.$$

Remark: To apply the above theorem, we define the Lagrange function: $L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x}^*)$ where $\lambda_1, \lambda_2, \dots, \lambda_m$ as additional variables. This transforms the constrained optimization problem into an unconstrained optimization problem by considering $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ together. To find the optimal point, we solve:

$$\frac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial x_i} = 0, \qquad i = 1, \dots, n,$$

$$\frac{\partial L(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial \lambda_k} = 0, \qquad k = 1, \dots, m.$$

Example

Lagrange Multipliers

Find the critical points for $f(x, y) = 3x - y + 6$, subject to the constraint $x^2 + y^2 = 4$.

Solution

Constraint

$$g(x, y) = x^2 + y^2 - 4 = 0$$

Construct the function

$$L(x, y, \lambda) = f(x, y) - \lambda g(x, y) = 3x - y + 6 - \lambda\left(x^2 + y^2 - 4\right)$$

Setting $L_x = L_y = L_\lambda = 0$, we solve the equations to be

$$\begin{cases} 3 - 2x\lambda = 0 \\ -1 - 2y\lambda = 0 \\ -x^2 - y^2 + 4 = 0 \end{cases} \Rightarrow x = \frac{3}{2\lambda}, y = -\frac{1}{2\lambda}, \lambda = \pm\frac{\sqrt{10}}{4}$$

Example

Minimizing Costs

Suppose a firm has an order for 200 units of its product and wishes to distribute its manufacture between two of its plants, plant 1 and plant 2. Let q_1 and q_2 denote the outputs of plants 1 and 2, respectively, and suppose the total-cost function is given by

$$c = f(q_1, q_2) = 2q_1^2 + q_1 q_2 + q_2^2 + 200.$$

How should the output be distributed in order to minimize costs?

Solution

We minimize $f(q_1, q_2)$, given the constraint $q_1 + q_2 = 200$.

$$L(q_1, q_2, \lambda) = 2q_1^2 + q_1 q_2 + q_2^2 + 200 - \lambda(q_1 + q_2 - 200)$$

$$\begin{cases} \dfrac{\partial L}{\partial q_1} = 4q_1 + q_2 - \lambda = 0 \\ \dfrac{\partial L}{\partial q_2} = q_1 + 2q_2 - \lambda = 0 \\ \dfrac{\partial L}{\partial \lambda} = -q_1 - q_2 + 200 = 0 \end{cases} \Rightarrow q_1 = 50, q_2 = 150$$

Example

Lagrange Multipliers with Two Constraints

Find critical points for $f(x, y, z) = xy + yz$, subject to the constraints $x^2 + y^2 = 8$ and $yz = 8$.

Solution

Set

$$L(x, y, z, \lambda_1, \lambda_2) = xy + yz - \lambda_1\left(x^2 + y^2 - 8\right) - \lambda_2(yz - 8)$$

$$L_x = y - 2x\lambda_1 = 0 \Rightarrow \frac{y}{2x} = \lambda_1$$

$$L_y = x + z - 2y\lambda_1 - z\lambda_2 = 0$$

$$L_z = y - y\lambda_2 = 0 \Rightarrow \lambda_2 = 1$$

$$L_{\lambda_1} = -x^2 - y^2 + 8 = 0$$

$$L_{\lambda_2} = -yz + 8 = 0 \Rightarrow z = 8/y$$

We obtain 4 critical points:

$$(2,2,4), (2,-2,-4), (-2,2,4), (-2,-2,-4)$$

Constrained Optimization in R

```
constrOptim.nl(initial, func, heq=f1, hin=f2)
```

Remark: constrOptim.nl needs "Alabama" and "numDeriv" packages to be installed first

initial: Initial values for the parameters to be optimized over.

func: A function to be **minimized**.

heq=f1: the function for the equality constraints such that `hin[j]=0` for all j

hin=f2: the function for the inequality constraints such that `hin[j]>0` for all j

Remark: The initial point must satisfy the inequality constraints strictly. That is, at the initial point, we must have: `hin[j]>0`

Constrained Optimization in R

Example: Mixed Equality and Inequality Constraints

Maximize $f(x_1, x_2) = 20x_1 + 16x_2 - 2x_1^2 - x_2^2 - x_3^2$

subject to:

$$x_1 + x_2 \leq 5$$

$$x_1 + x_2 - x_3 = 0$$

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_3 \geq 0$$

Constrained Optimization

Solution

Write the function as minimizing $-f$.

```
func=function(x) -(20*x[1]+16*x[2]-2*x[1]^2-x[2]^2-x[3]^2)
p0=c(0.64,0.64,1.28)
g1=function(x){
  h=0
  h[1]=x[3]-x[1]-x[2]
  return (h)}
g2=function(x){
  h=0
  h[1]=5-x[1]-x[2]
  h[2]=x[1]
  h[3]=x[2]
  h[4]=x[3]
  return (h)}
y=constrOptim.nl(p0,func,heq=g1,hin=g2)
print(y$par)
[1] 2.333258 2.666742 5.000000
print(y$value)
[1] -46.33333 #Note that the optimal value is f*=+46.33 because we minimized -f.
```