Introduction to Continuous Optimization

Lê Xuân Thanh

Contents

Analysis of multivariate functions

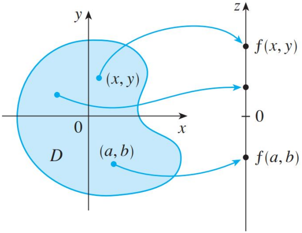
- Domain, range, graph
- Limit and continuity
- Partial derivatives
- Hessian matrix

Multivariate function

Definition

A function f of n variables is a rule that assigns each vector $\mathbf{x} = (x_1, \ldots, x_n)$ in a given set $D \subseteq \mathbb{R}_n$ to a unique real number denoted by $f(\mathbf{x})$ or $f(x_1, \ldots, x_n)$.

- D is the *domain* of f
- $\{f(\mathbf{x}) \mid \mathbf{x} \in D\}$ is the *range* of f
- $n = 1$: univariate function
- $n = 2$: bivariate function
- $n \geq 3$: multivariate function



Examples

Problem: Find the domain and range of $f(x, y) = \sqrt{9 - x^2 - y^2}$.

Solution.

- The domain of f is

$$D = \{(x, y) \in \mathbb{R}_2 \mid 9 - x^2 - y^2 \geq 0\} = \{(x, y) \in \mathbb{R}_2 \mid x^2 + y^2 \leq 9\}$$

  which is a disk with center $(0, 0)$ and radius 3.

- The range of f is

$$\{z \mid z = \sqrt{9 - x^2 - y^2}, (x, y) \in D\} = [0, 3].$$

Exercises

What are the domain and range of the following functions?
(i) $f(x) = \sqrt{x^2 - 4}$ (ii) $f(x, y) = \sqrt{x^2 - y^2}$
(iii) $f(x, y, z) = \cos(x + y + z)$

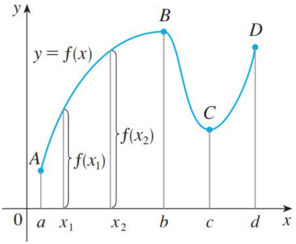
Graph of multivariate function

Definition

Let f be a function of n variables with domain $D \subseteq \mathbb{R}_n$.
The *graph* of f is the set

$$\{(x_1, \ldots, x_n, y) \mid (x_1, \ldots, x_n) \in D, y = f(x_1, \ldots, x_n)\} \subseteq \mathbb{R}_{n+1}.$$

Note:
The graph of a univariate function can be visualized as a curve in plane.



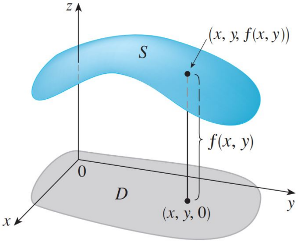
Graph of multivariate function

Definition

Let f be a function of n variables with domain $D \subseteq \mathbb{R}_n$.
The *graph* of f is the set

$$\{(x_1, \ldots, x_n, y) \mid (x_1, \ldots, x_n) \in D, y = f(x_1, \ldots, x_n)\} \subseteq \mathbb{R}_{n+1}.$$

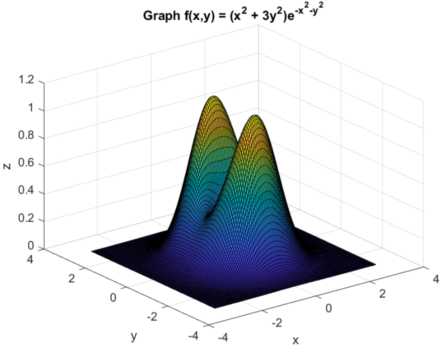
Note:
The graph of a bivariate function
can be visualized as a surface in space.



Plotting graphs of bivariate functions: Example 1

Plot the graph of $f(x,y) = (x^2 + 3y^2)e^{-x^2-y^2}$

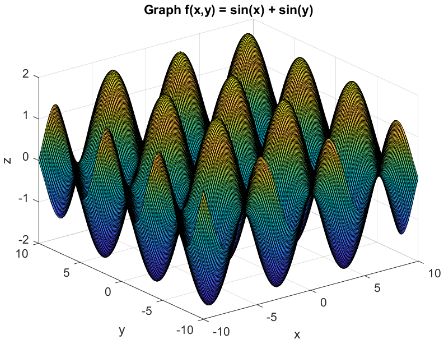
```
[x,y] = meshgrid(-3:0.05:3,
-3:0.05:3);

z = (x.^2 + 3*y.^2)
.*exp(-x.^2 - y.^2);

surf(x,y,z);

title('Graph f(x,y) =
(x^2 + 3y^2)e^{-x^2-y^2}');

xlabel('x');
ylabel('y');
zlabel('z');

grid on;
```



Plotting graphs of bivariate functions: Example 2

Plot the graph of $f(x, y) = \sin x + \sin y$

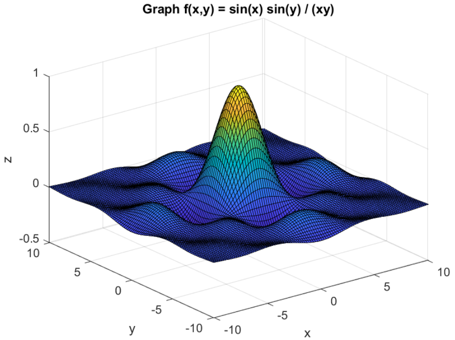
```
[x,y] = meshgrid(-10:0.1:10,
-10:0.1:10);

z = sin(x) + sin(y);

surf(x,y,z);

title('Graph f(x,y) =
sin(x) + sin(y)');

xlabel('x');
ylabel('y');
zlabel('z');

grid on;
```



Plotting graphs of bivariate functions: Example 3

Plot the graph of $f(x, y) = \frac{\sin x \sin y}{xy}$

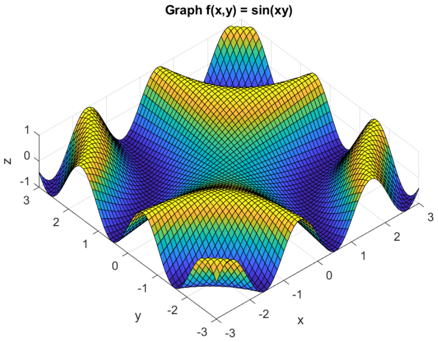
```
[x,y] = meshgrid(-10.1:0.2:10.1,
-10.1:0.2:10.1);

z = sin(x) .* sin(y)./(x .* y);

surf(x,y,z);

title('Graph f(x,y) =
sin(x) sin(y) / (xy)');

xlabel('x');
ylabel('y');
zlabel('z');

grid on;
```



Plotting graphs of bivariate functions: Example 4

Plot the graph of $f(x, y) = \sin(xy)$

```
[x,y] = meshgrid(-3:0.1:3,
-3:0.1:3);

z = sin(x.*y);

surf(x,y,z);

title('Graph f(x,y)
= sin(xy)');

xlabel('x');
ylabel('y');
zlabel('z');

grid on;
```



Analysis of multivariate functions

- Domain, range, graph
- Limit and continuity
- Partial derivatives
- Hessian matrix

Limit and continuity

- Case of univariate functions
- Case of bivariate functions
- Case of multivariate functions

Limit of univariate functions

Let $f : \mathbb{R} \to \mathbb{R}$ be a univariate function, and $a \in R$.

Definition

We say $\lim_{x \to a} f(x) = L$ if

$\forall \varepsilon > 0\ \exists \delta > 0$ such that

$|x - a| \leq \delta \implies |f(x) - L| \leq \varepsilon$

Definition

We say $\lim_{x \to a^-} f(x) = L$ if

$\forall \varepsilon > 0\ \exists \delta > 0$ such that

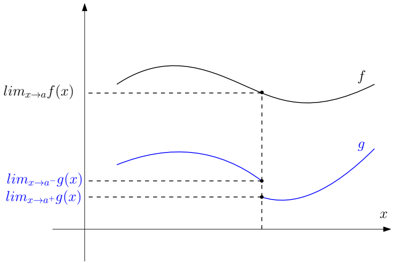
$0 \leq a - x \leq \delta \implies |f(x) - L| \leq \varepsilon$

Definition

We say $\lim_{x \to a^+} f(x) = L$ if

$\forall \varepsilon > 0\ \exists \delta > 0$ such that

$0 \leq x - a \leq \delta \implies |f(x) - L| \leq \varepsilon$



Limit of univariate functions

Exercises:

(i) Compute $\lim_{x\to 1^-} f(x)$ and $\lim_{x\to 1^+} f(x)$ where

$$f(x) = \begin{cases} 1-x & \text{if } x \leq 1, \\ x^2 & \text{if } x > 1. \end{cases}$$

(ii) Compute $\lim_{x\to 0^-} f(x)$ and $\lim_{x\to 0^+} f(x)$ where $f(x) = |x|$.

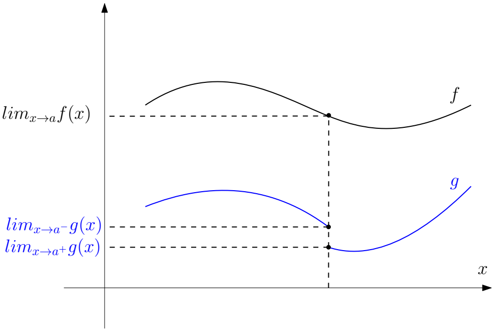
Continuity of univariate functions

Definition

- Function f is continuous at $a \in \mathbb{R}$ if $\lim_{x \longrightarrow a} f(x) = f(a)$.
- Function f is continuous on $D \subset \mathbb{R}$ if it is continuous at every $x \in D$.

Note:

$$\lim_{x \longrightarrow a} f(x) = L \quad \Leftrightarrow \quad \lim_{x \longrightarrow a^-} f(x) = \lim_{x \longrightarrow a^+} f(x) = L$$



Limit and continuity

- Case of univariate functions
- Case of bivariate functions
- Case of multivariate functions

Limit of bivariate functions

Let $f\colon \mathbb{R}_2 \to \mathbb{R}$ be a bivariate function, and $\mathbf{a} = (a_1, a_2) \in \mathbb{R}_2$.

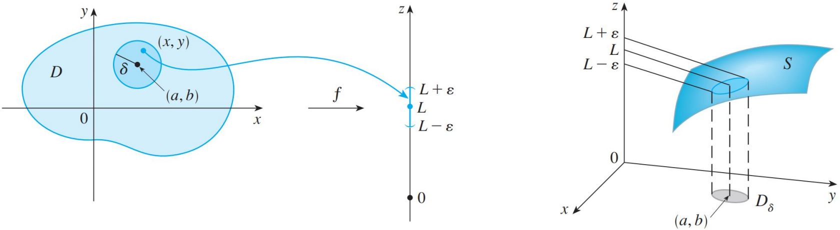
Notation: The distance between $\mathbf{x} = (x_1, x_2) \in \mathbb{R}_2$ and $\mathbf{a}$ is

$$|\mathbf{x} - \mathbf{a}| = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2}$$

Definition

We say $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L$ if $\forall \varepsilon > 0\ \exists \delta > 0$ such that

$$|\mathbf{x} - \mathbf{a}| < \delta \quad \Longrightarrow \quad |f(\mathbf{x}) - L| < \varepsilon.$$



Limit of bivariate functions

Example: Find $\lim_{(x,y)\to(0,0)} f(x,y)$ with $f(x,y) = \frac{3x^2y}{x^2+y^2}$.

Solution.

We will show that

$$\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2} = 0.$$

Indeed, let $\varepsilon > 0$, and $\delta = \frac{\varepsilon}{3}$. For any (x, y) satisfying

$$|(x,y) - (0,0)| = \sqrt{x^2+y^2} < \delta = \frac{\varepsilon}{3}$$

we have

$$\left|\frac{3x^2y}{x^2+y^2} - 0\right| = \left|\frac{3x^2y}{x^2+y^2}\right| \le 3|y| \le 3\sqrt{x^2+y^2} < \varepsilon.$$

Continuity of bivariate functions

Definition

- f is called *continuous at* $\mathbf{a}$ if

$$\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a}).$$

- f is called *continuous on* D if it is continuous at every point in D.

Properties:

- Sums, differences, products, quotients of continuous multivariate functions are continuous on their domains.
- If $f : \mathbb{R}_2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are continuous, then the composite function $g \circ f$ is continuous.

Example:

$$g(x, y) = \begin{cases} \frac{3x^2 y}{x^2+y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

is continuous on $\mathbb{R}_2$.

Limit and continuity

- Case of univariate functions
- Case of bivariate functions
- Case of multivariate functions

Limit and continuity of multivariate functions

Let $f\colon \mathbb{R}_n \to \mathbb{R}$ be a multivariate function, and $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}_n$.

Notation: The distance between $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_n$ and $\mathbf{a}$ is

$$|\mathbf{x} - \mathbf{a}| = \sqrt{(x_1 - a_1)^2 + \ldots + (x_n - a_n)^2}$$

Definition

We say $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L$ if $\forall \varepsilon > 0\ \exists \delta > 0$ such that

$$|\mathbf{x} - \mathbf{a}| < \delta \quad \implies |f(\mathbf{x}) - L| < \varepsilon.$$

Definition

- f is called *continuous at* $\mathbf{a}$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a}).$$

- f is called *continuous on D* if it is continuous at every point in D.

Analysis of multivariate functions

- Domain, range, graph
- Limit and continuity
- Partial derivatives
- Hessian matrix

Partial derivatives

- Case of univariate functions
- Case of bivariate functions
- Case of multivariate functions

Tangent line

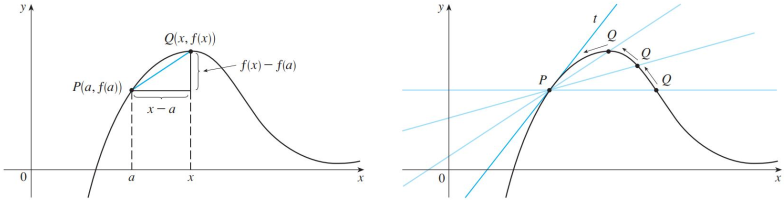
Let $f : \mathbb{R} \to \mathbb{R}$ be a univariate function, and $a \in \mathbb{R}$.

Definition

The tangent line to the graph of $y = f(x)$ at the point $P = (a, f(a))$ is the line through P with slope

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

provided that this limit exists.



Derivative

Definition

The derivative of f at a is

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{\Delta x \to 0} \frac{f(a + \Delta x) - f(a)}{\Delta x}$$

provided that this limit exists.

Note: $f'(a)$ is the slope of the tangent line to the graph of $y = f(x)$ at the point $P = (a, f(a))$

Exercise:

Compute derivative of $f(x) = x^2 - 8x + 9$ at $x = a$.

Partial derivatives

- Case of univariate functions
- Case of bivariate functions
- Case of multivariate functions

Partial derivatives at a point

Let $f\colon \mathbb{R}_2 \to \mathbb{R}$ be a bivariate function, and $(a,b) \in \mathbb{R}_2$.

Defintion

- The partial derivative of f with respect to x at (a,b) is

$$\frac{\partial f}{\partial x}(a,b) = \lim_{\Delta x \to 0} \frac{f(a+\Delta x, b) - f(a,b)}{\Delta x}$$

  provided that the limit exists.

- The partial derivative of f with respect to y at (a,b) is

$$\frac{\partial f}{\partial y}(a,b) = \lim_{\Delta y \to 0} \frac{f(a, b+\Delta y) - f(a,b)}{\Delta y}$$

  provided that the limit exists.

- The gradient vector of f at (a,b) is

$$\nabla f(a,b) = \left(\frac{\partial f}{\partial x}(a,b), \frac{\partial f}{\partial y}(a,b)\right).$$

Connection to classical derivative

- $\frac{\partial f}{\partial x}(a,b) = \lim_{\Delta x \to 0} \dfrac{f(a+\Delta x, b) - f(a,b)}{\Delta x} = g'(a)$

  with $g(x) = f(x, b)$

- $\frac{\partial f}{\partial y}(a,b) = \lim_{\Delta y \to 0} \dfrac{f(a, b+\Delta y) - f(a,b)}{\Delta y} = h'(b)$

  with $h(y) = f(a, y)$

Geometrical interpretation

Let S be graph of f, and $P = (a, b, f(a, b))$.

$C_1 = S \cap \{y = b\}, \quad C_2 = S \cap \{x = a\}$.

T_1 = tangent line (on $\{y = b\}$) of C_1 at P.

T_2 = tangent line (on $\{x = a\}$) of C_2 at P.

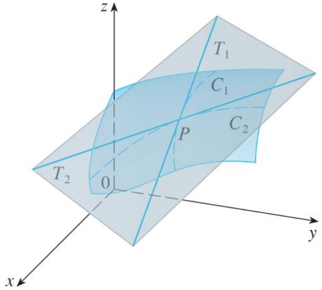
- $\frac{\partial f}{\partial x}(a, b)$ = slope of T_1.
- $\frac{\partial f}{\partial y}(a, b)$ = slope of T_2.

Tangent plane to S at P is
the plane containing T_1 and T_2:

$$z - f(a, b) = \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b).$$

Normal vector to tangent plane:

$$\left(\frac{\partial f}{\partial x}(a, b), \frac{\partial f}{\partial y}(a, b), -1\right) = (\nabla f(a, b), -1)$$



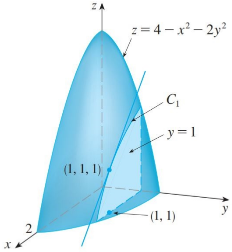
Example 1

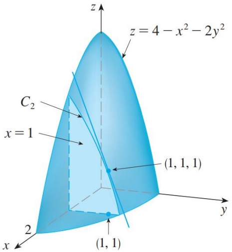
$$f(x,y) = 4 - x^2 - 2y^2, \quad (a,b) = (1,1).$$

Let $g(x) = f(x,1)$ and $h(y) = f(1,y)$.

Then

- $g(x) = 2 - x^2$, and $\frac{\partial f}{\partial x}(1,1) = g'(1) = -2$
- $h(y) = 3 - 2y^2$, and $\frac{\partial f}{\partial y}(1,1) = h'(1) = -4$





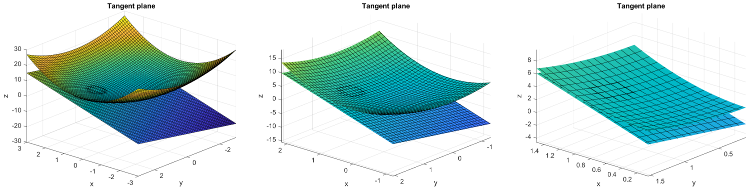
Example 2

Consider elliptic paraboloid $z = f(x, y) = 2x^2 + y^2$.

$$\frac{\partial f}{\partial x}(x, y) = 4x \quad \Longrightarrow \quad \frac{\partial f}{\partial x}(1, 1) = 4$$
$$\frac{\partial f}{\partial y}(x, y) = 2y \quad \Longrightarrow \quad \frac{\partial f}{\partial y}(1, 1) = 2$$

Tangent plane to graph of f at $(1, 1, 3)$:

$$z - 3 = 4(x - 1) + 2(y - 1) \quad \Leftrightarrow \quad 4x + 2y - z - 3 = 0$$



Partial derivatives as functions

Let $f: \mathbb{R}_2 \to \mathbb{R}$ be a bivariate function.

Defintion

- The partial derivative of f with respect to x is

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

  provided that the limit exists.

- The partial derivative of f with respect to y is

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

  provided that the limit exists.

- The gradient of f is

$$\nabla f(x,y) = \left(\frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)\right).$$

Example 1

For $f(x, y) = \sin\left(\frac{x}{1+y}\right)$:

$$\frac{\partial f}{\partial x} = \cos\left(\frac{x}{1+y}\right)\frac{\partial}{\partial x}\left(\frac{x}{1+y}\right) = \cos\left(\frac{x}{1+y}\right)\frac{1}{1+y}$$

$$\frac{\partial f}{\partial y} = \cos\left(\frac{x}{1+y}\right)\frac{\partial}{\partial y}\left(\frac{x}{1+y}\right) = -\cos\left(\frac{x}{1+y}\right)\frac{x}{(1+y)^2}$$

Example 2

If $x^3 + y^3 + z^3 + 6xyz = 1$,
then taking derivative w.r.t. x of both sides gives

$$3x^2 + 3z^2\frac{\partial z}{\partial x} + 6yz + 6xy\frac{\partial z}{\partial x} = 0,$$

and consequently

$$\frac{\partial z}{\partial x} = -\frac{x^2 + 2yz}{z^2 + 2xy}.$$

Similarly, we have

$$\frac{\partial z}{\partial y} = -\frac{y^2 + 2xz}{z^2 + 2xy}.$$

Partial derivatives

- Case of univariate functions
- Case of bivariate functions
- Case of multivariate functions

Partial derivatives at a point

Let $f : \mathbb{R}_n \to \mathbb{R}$ be a multivariate function, and $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}_n$.

Defintion

The partial derivative of f with respect to x_i at $(a_1, \ldots, a_n)$ is

$$\begin{aligned}&\frac{\partial}{\partial x_i} f(a_1, \ldots, a_n)\\ &= \lim_{\Delta x_i \to 0} \frac{f(a_1, \ldots, a_{i-1}, a_i + \Delta x_i, a_{i+1}, \ldots, a_n) - f(a_1, \ldots, a_n)}{\Delta x_i}\end{aligned}$$

provided that the limit exists.

Defintion

The gradient of f at $(a_1, \ldots, a_n)$ is

$$\nabla f(a_1, \ldots, a_n) = \left(\frac{\partial}{\partial x_1} f(a_1, \ldots, a_n), \ldots, \frac{\partial}{\partial x_n} f(a_1, \ldots, a_n) \right).$$

Partial derivatives as functions

Let $f: \mathbb{R}_n \to \mathbb{R}$ be a multivariate function.

Defintion

The partial derivative of f with respect to x_i is

$$\begin{aligned}&\frac{\partial}{\partial x_i} f(x_1, \ldots, x_n)\\ =& \lim_{\Delta x_i \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + \Delta x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{\Delta x_i}\end{aligned}$$

provided that the limit exists.

Defintion

The gradient of f is

$$\nabla f(x_1, \ldots, x_n) = \left(\frac{\partial}{\partial x_1} f(x_1, \ldots, x_n), \ldots, \frac{\partial}{\partial x_n} f(x_1, \ldots, x_n) \right).$$

Analysis of multivariate functions

- Domain, range, graph
- Limit and continuity
- Partial derivatives
- Hessian matrix

Second partial derivatives

Let $f\colon \mathbb{R}_n \to \mathbb{R}$ be a multivariate function, and $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}_n$.

Definition

The second partial derivatives of f at $\mathbf{a}$ are

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(a_1, \ldots, a_n)\right) \quad \text{for } i, j = 1, \ldots, n.$$

Definition

As functions, the second partial derivatives of f are

$$\frac{\partial^2 f}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) \quad \text{for } i, j = 1, \ldots, n.$$

Note: If $i = j$, then we denote $\frac{\partial^2 f}{\partial x_i^2}$ instead of $\frac{\partial^2 f}{\partial x_i \partial x_i}$.

Second partial derivatives: Examples

For $f(x, y) = x^3 + x^2y^3 - 2y^2$ we have

$$\frac{\partial f}{\partial x} = 3x^2 + 2xy^3$$
$$\frac{\partial f}{\partial y} = 3x^2y^2 - 4y$$

and

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(3x^2 + 2xy^3) = 6x + 2y^3$$
$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(3x^2 + 2xy^3) = 6xy^2$$
$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(3x^2y^2 - 4y) = 6xy^2$$
$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(3x^2y^2 - 4y) = 6x^2y - 4$$

Exercises: Compute second partial derivatives of $f(x, y) = \sin(3x + y^2)$ and $f(x, y, z) = e^{xy} \log z$.

Hessian matrix

Definition

The Hessian matrix of f at $\mathbf{a}$ is

$$H(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{a}) \end{bmatrix}$$

Exercises: Compute the Hessian matrix of

- $f(x_1, x_2) = 3x_1^2 + 4x_1x_2 + x_2^2$
- $f(x, y) = x^4 + y^4 - 4xy + 1$

Contents

Positive and negative definite matrices

Let $A \in \mathbb{R}^{n\times n}$ be a square matrix.

Definition

- A is called *positive definite* if $\mathbf{x}^t A\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$.
- A is called *negative definite* if $\mathbf{x}^t A\mathbf{x} < 0$ for all $\mathbf{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$.

Examples:

- $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ is positive definite since

  $$\mathbf{x}^t A\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2+2x_2^2+2x_1x_2 = x_1^2+x_2^2+(x_1+x_2)^2$$

  which is positive for all $\mathbf{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$.
- $A = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}$ is negative definite.

Leading principal minors

Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ be a square matrix.

Definition

The i^{th} *leading principal minor* of A is the upper left i-by-i corner of A.

Examples:

- Leading principal minor of order 1: a_{11}
- Leading principal minor of order 2: $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$
- Leading principal minor of order 3: $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$
- ...

Sylvester's criterion

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix.

Theorem

- A is called *positive definite* if and only if all of its leading principal minors have positive determinant.
- A is called *negative definite* if and only if
  - all of its leading principal minors of odd orders have negative determinant, and
  - all of its leading principal minors of even orders have positive determinant.

Examples:

- $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ is positive definite since $\det(a_{11}) = 2 > 0$ and $\det(A) = 3 > 0$.
- $A = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}$ is negative definite since $\det(a_{11}) = -2 < 0$ and $\det(A) = 3 > 0$.

Exercises

Determine the definiteness of the following matrices:

$$\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}, \quad \begin{bmatrix} -2 & 2 \\ 2 & -4 \end{bmatrix}, \quad \begin{bmatrix} 2 & 4 \\ 4 & 3 \end{bmatrix},$$

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix},$$

$$\begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} -5 & 1 & 1 \\ 1 & -7 & 1 \\ 1 & 1 & -5 \end{bmatrix}.$$

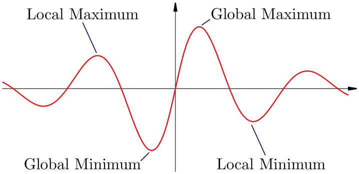
Contents

Optima

Given a multivariate function $f(\mathbf{x})$ with domain $D \subset \mathbb{R}^n$.

Definition

- f has a *global maximum* (so-called *absolute maximum*) at $\mathbf{a} \in D$ if $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in D$.
- f has a *global minimum* (so-called *absolute minimum*) at $\mathbf{a} \in D$ if $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in D$.
- f has a *local maximum* at $\mathbf{a} \in D$ if there exists $r > 0$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in D \cap \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y} - \mathbf{a}| < r\}$.
- f has a *local minimum* at $\mathbf{a} \in D$ if there exists $r > 0$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in D \cap \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y} - \mathbf{a}| < r\}$.

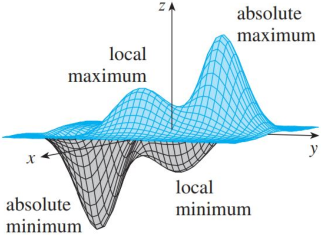


Optima

Given a multivariate function $f(\mathbf{x})$ with domain $D \subset \mathbb{R}^n$.

Definition

- f has a *global maximum* (so-called *absolute maximum*) at $\mathbf{a} \in D$ if $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in D$.
- f has a *global minimum* (so-called *absolute minimum*) at $\mathbf{a} \in D$ if $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in D$.
- f has a *local maximum* at $\mathbf{a} \in D$ if there exists $r > 0$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in D \cap \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y} - \mathbf{a}| < r\}$.
- f has a *local minimum* at $\mathbf{a} \in D$ if there exists $r > 0$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in D \cap \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y} - \mathbf{a}| < r\}$.



Contents

First order criterion

Suppose that $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ has partial derivatives.

Theorem

If $\mathbf{a} = (a_1, \ldots, a_n)$ is a local maximizer of local minimizer of f, then

$$\frac{\partial f}{\partial x_1}(\mathbf{a}) = \ldots = \frac{\partial f}{\partial x_n}(\mathbf{a}) = 0.$$

Proof.

Let $g(x_1) = f(x_1, a_2, \ldots, a_n)$.

Since f has local maximum (or local minimum) at $\mathbf{a} = (a_1, \ldots, a_n)$,

respectively g has local maximum (or local minimum) at a_1.

By Fermat's theorem, we have $g'(a_1) = 0$. Note that

$$g'(a_1) = \frac{\partial f}{\partial x_1}(\mathbf{a}).$$

Similar to other partial derivatives.

Stationary points and saddle points

Suppose that $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ has partial derivatives.

Definition

$\mathbf{a} = (a_1, \ldots, a_n)$ is a *stationary point* of f if $\nabla f(\mathbf{a}) = \mathbf{0}$, i.e.,

$$\frac{\partial f}{\partial x_1}(\mathbf{a}) = \ldots = \frac{\partial f}{\partial x_n}(\mathbf{a}) = 0.$$

Remark:
Extreme points (local maximizers and local minimizers) are stationary. The inverse claim does not hold.

Definition

$\mathbf{a} = (a_1, \ldots, a_n)$ is a *saddle point* of f
if it is a stationary point but not an extreme point of f.

First order criterion: Example 1

For $f(x, y) = x^2 + y^2 - 2x - 6y + 14$ we have

$$\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0$$

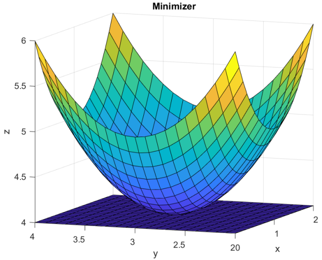
$$\Leftrightarrow \quad \begin{cases} 2x - 2 = 0 \\ 2y - 6 = 0 \end{cases}$$

$$\Leftrightarrow \quad (x, y) = (1, 3).$$

Since $f(1, 3) = 4$ and

$$f(x, y) = (x - 1)^2 + (y - 3)^2 + 4 \geq 4,$$

$(1, 3)$ is (global) minimizer of f.



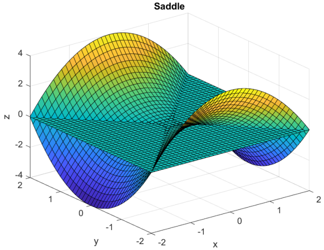
First order criterion: Example 2

For $f(x,y) = y^2 - x^2$ we have

$$\frac{\partial f}{\partial x} f(x,y) = \frac{\partial f}{\partial y}(x,y) = 0$$

$$\Leftrightarrow \quad \begin{cases} -2x = 0 \\ 2y = 0 \end{cases}$$

$$\Leftrightarrow \quad (x,y) = (0,0).$$



Since $f(x,0) < 0 = f(0,0)$ for $x \neq 0$ and $f(0,y) > 0 = f(0,0)$ for $y \neq 0$, $(0,0)$ is NEITHER minimizer NOR maximizer of f. It is a saddle point of f.

Contents

Second order criterion: multivariate cases

Given $f(\mathbf{x}) = f(x_1, \dots, x_n)$ having continuous second partial derivatives. Let $\mathbf{a} \in \mathbb{R}^n$ and H the Hessian matrix of f at $\mathbf{a}$:

$$H_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{a}) \qquad \forall i, j = 1, \dots, n.$$

Theorem

Suppose that $\frac{\partial}{\partial x_i} f(\mathbf{a}) = 0$ for all $i = 1, \dots, n$.

- If H is positive definite, then $\mathbf{a}$ is a local minimizer of f.
- If H is negative definite, then $\mathbf{a}$ is a local maximizer of f.

Second order criterion: bivariate cases

Given $f(x, y)$ having continuous second partial derivatives.
Let $(a, b) \in \mathbb{R}^2$ and $H = H(a, b)$ the Hessian matrix of f at (a, b):

$$H = H(a, b) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a, b) & \frac{\partial^2 f}{\partial x \partial y}(a, b) \\ \frac{\partial^2 f}{\partial y \partial x}(a, b) & \frac{\partial^2 f}{\partial y^2}(a, b) \end{pmatrix}.$$

Theorem

Suppose that $\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(a, b) = 0$.

- If $\det(H) > 0$ and $\frac{\partial^2 f}{\partial x^2}(a, b) > 0$, then (a, b) is a local minimizer of f.
- If $\det(H) > 0$ and $\frac{\partial^2 f}{\partial x^2}(a, b) < 0$, then (a, b) is a local maximizer of f.
- If $\det(H) < 0$, then (a, b) is a saddle point of f.

Note: $\det(H) = 0$ gives NO information about optimality of (a, b).

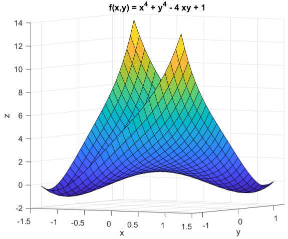
Second order criterion: Example

For $f(x,y) = x^4 + y^4 - 4xy + 1$ we have

$$\begin{aligned} & \frac{\partial f}{\partial x}(x,y) = \frac{\partial f}{\partial y}(x,y) = 0 \\ \Leftrightarrow \quad & 4x^3 - 4y = 4y^3 - 4x = 0 \\ \Leftrightarrow \quad & (x,y) \in \{(0,0),(1,1),(-1,-1)\}. \end{aligned}$$

Hessian matrix of f:

$$H(x,y) = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}.$$



- $\det(H(0,0)) = -16 < 0$,
  hence $(0,0)$ is a saddle point of f
- $\det(H(1,1)) = 128 > 0$ and
  $\frac{\partial f}{\partial x}(1,1) = 12 > 0$,
  hence $(1,1)$ is a local minimizer of f
- Similarly, $(-1,-1)$ is a local minimizer

Contents

Formulation

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathrm{dom}(f)$ be the domain of f.

Unconstrained Optimization Problem:

$$\min \mid \max \quad f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathrm{dom}(f)$$

Constrained Optimization Problem:

$$\min \mid \max \quad f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in C \subsetneq \mathrm{dom}(f)$$

Explicit form of feasible set:

$$C = \{\mathbf{x} \in \mathrm{dom}(f) \mid g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m; h_j(\mathbf{x}) = 0, j = 1, \ldots, \ell\}$$

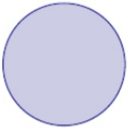
Solution existence

$$\min \mid \max \quad f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in C \subsetneq \mathrm{dom}(f) \subset \mathbb{R}^n$$

Extreme value theorem

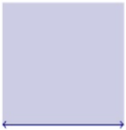
If f is continuous and C is compact, then f has a global minimizer and a global maximizer in C.

Recall: compact set in $\mathbb{R}^n$ = closed + bounded



$x^2 + y^2 \leq 1$
closed

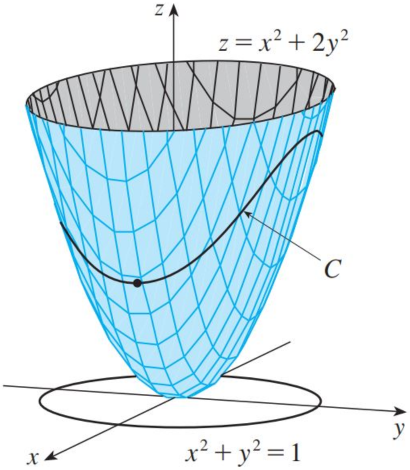
$x^2 + y^2 < 1$
not closed



$y \geq 0$
closed

An illustration of extreme value theorem

$$\min \mid \max \quad x^2 + 2y^2 \quad \text{s.t.} \quad x^2 + y^2 \leq 1$$



Contents

Graphical method

- Concept of level sets
- Level-set method

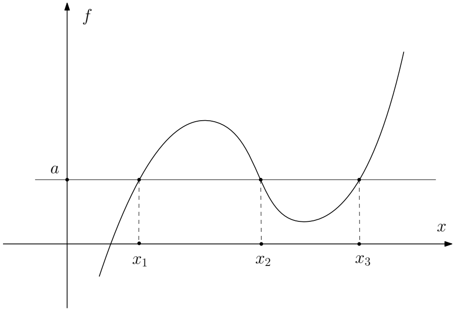
Level sets

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ with domain $\mathrm{dom}(f) \subset \mathbb{R}^n$, and $a \in \mathbb{R}$.

Definition

The level set $[f = a]$ is defined as

$$[f = a] := \{\mathbf{x} \in \mathrm{dom}(f) \mid f(\mathbf{x}) = a\}.$$



$$[f = a] = \{x_1, x_2, x_3\}$$

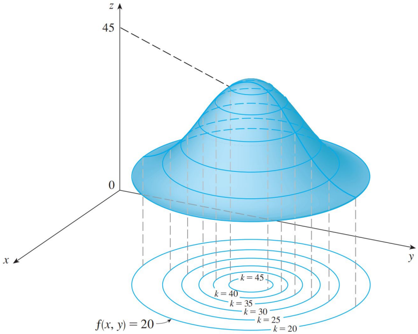
Level sets

Let $f : \mathbb{R}^n \to \mathbb{R}$ with domain $\mathrm{dom}(f) \subset \mathbb{R}^n$, and $a \in \mathbb{R}$.

Definition

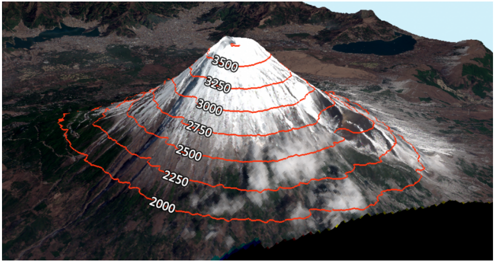
The level set $[f = a]$ is defined as

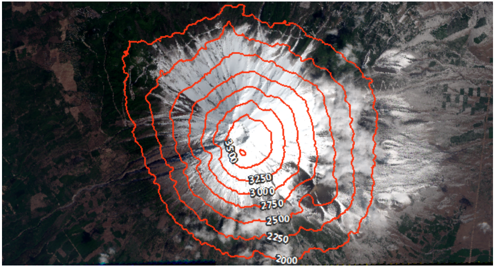
$$[f = a] := \{\mathbf{x} \in \mathrm{dom}(f) \mid f(\mathbf{x}) = a\}.$$



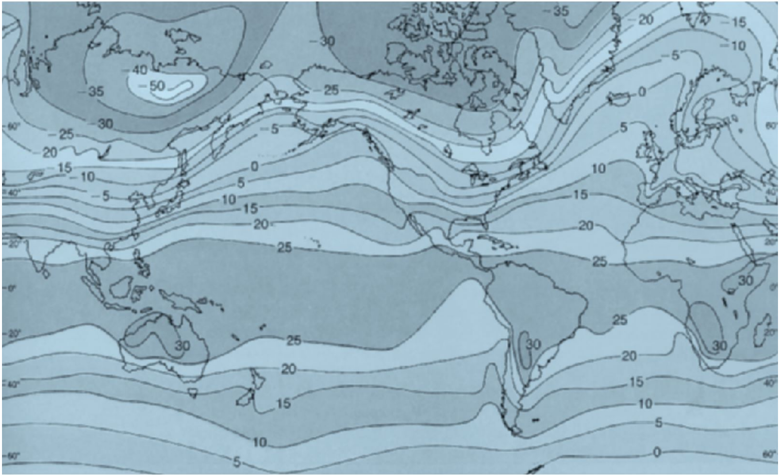
Level sets of a bivariate function are called contours

Contours in topographic maps of mountainous regions



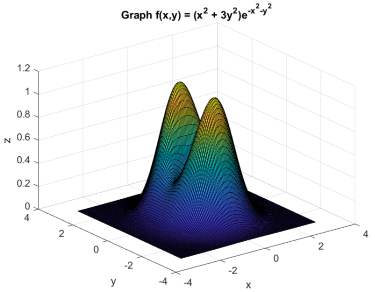


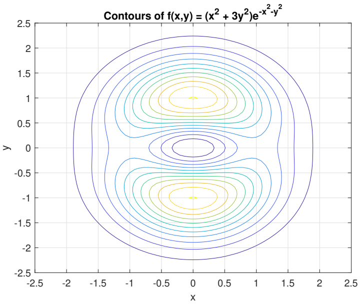
Contours as isothermals in temperature maps



Contours of surfaces: Example 1

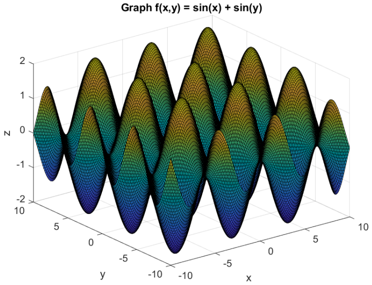
$$f(x, y) = (x^2 + 3y^2)e^{-x^2-y^2}$$

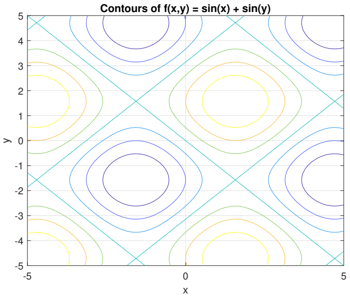




Contours of surfaces: Example 2

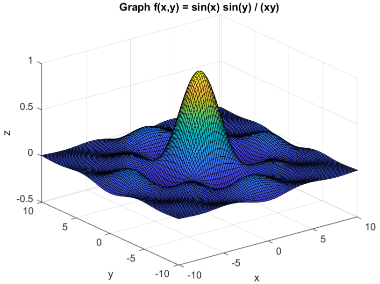
$$f(x, y) = \sin x + \sin y$$

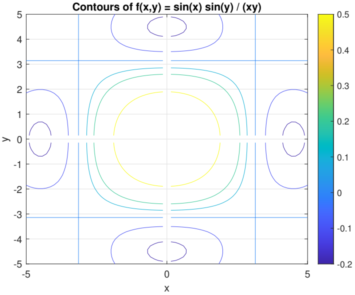




Contours of surfaces: Example 3

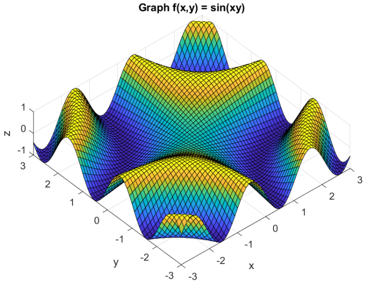
$$f(x, y) = \frac{\sin x \sin y}{xy}$$

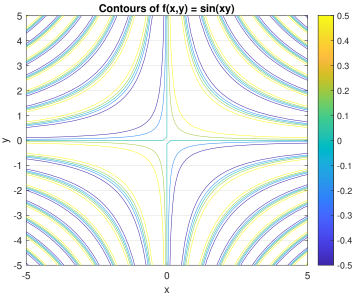




Contours of surfaces: Example 4

$$f(x, y) = \sin(xy)$$





Graphical method

- Concept of level sets
- Level-set method

Level-set method

$$(P) \qquad \min \quad f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in C \subsetneq \mathrm{dom}(f)$$

- Step 1: Plot the feasible set C.
  - If $C = \emptyset$, then (P) is infeasible.
- Step 2: Plot level sets $[f = \alpha]$ for $\alpha \in \mathbb{R}$.
- Step 3: Decrease α whenever $[f = \alpha] \cap C \neq \emptyset$.
  - If $\alpha \to -\infty$, then $f_{min} = -\infty$.
  - Otherwise, $f_{min} =$ the smallest α such that $[f = \alpha] \cap C \neq \emptyset$.

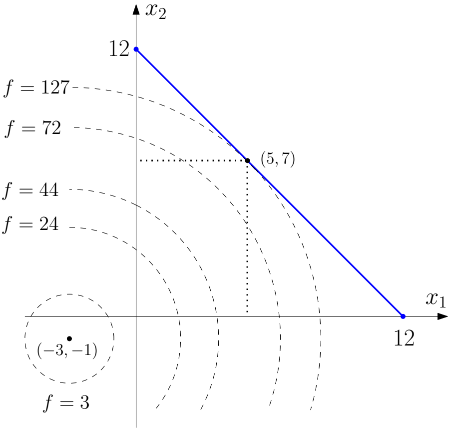
Level-set method

$$(P) \qquad \max \quad f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in C \subsetneq \mathrm{dom}(f)$$

- Step 1: Plot the feasible set C.
  - If $C = \emptyset$, then (P) is infeasible.
- Step 2: Plot level sets $[f = \alpha]$ for $\alpha \in \mathbb{R}$.
- Step 3: Increase α whenever $[f = \alpha] \cap C \neq \emptyset$.
  - If $\alpha \to +\infty$, then $f_{max} = +\infty$.
  - Otherwise, f_{max} = the largest α such that $[f = \alpha] \cap C \neq \emptyset$.

Level-set method: Example

$$\min \quad f(x_1, x_2) = (x_1+3)^2+(x_2+1)^2-1 \quad \text{s.t.} \quad x_1+x_2 = 12, x_1 \geq 0, x_2 \geq 0$$



$$f_{min} = 127 \text{ at } (x_1, x_2) = (5, 7)$$

Contents

Lagrange's method[1]

- Case of one equality constraint
- Case of multiple equality constraints
- Case of mixed constraints

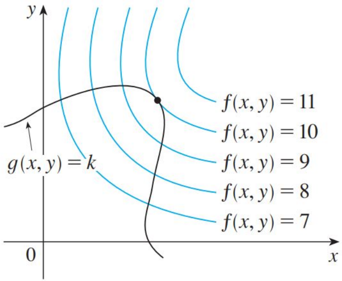
[1] Joseph-Louis Lagrange (25.01.1736–10.04.1813): an Italian mathematician and astronomer, later naturalized French

Intuition

$$\text{maximize} \quad f(x,y) \quad \text{subject to} \quad g(x,y) = k$$

Geometry:

- Find largest value of c such that $[f = c]$ is tangent to $[g = k]$ at some point (x_0, y_0)
- In that situation: $\nabla f(x_0, y_0) \;//\; \nabla g(x_0, y_0)$, i.e., $\exists \lambda \in \mathbb{R}$ s.t. $\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$



Method

Problem:

$$\text{minimize} \mid \text{maximize} \quad f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) = k, \quad \mathbf{x} \in \mathbb{R}^n$$

Method:

- Solve the so-called Karush-Kuhn-Tucker system w.r.t. $\mathbf{x}$ and λ:

$$\begin{aligned} \nabla f(\mathbf{x}) - \lambda \nabla g(\mathbf{x}) &= \mathbf{0} \\ g(\mathbf{x}) &= k \end{aligned}$$

- Let S be KKT points (i.e., solutions $\mathbf{x}$ of KKT system)
- Evaluate f at all KKT points
  - $\text{argmin}_S f$ solves the minimum problem
  - $\text{argmax}_S f$ solves the maximum problem

Proof

Problem:

$$(P) \qquad \text{minimize} \mid \text{maximize} \quad f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) = k, \quad \mathbf{x} \in \mathbb{R}^n$$

Lagrange function: $\mathcal{L}(\mathbf{x}, \lambda) := f(\mathbf{x}) - \lambda\,[g(\mathbf{x}) - k]$

Note: λ is called Lagrange multiplier

Restated:

$$(P_\lambda) \qquad \text{minimize} \mid \text{maximize} \quad \mathcal{L}(\mathbf{x}, \lambda)$$

Arguments:

- Observe that

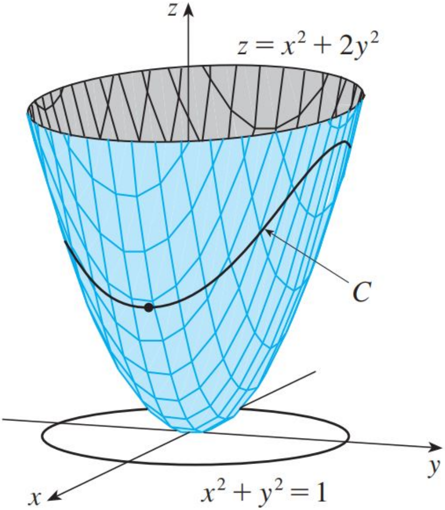
$$\text{opt}_{\lambda \in \mathbb{R}}\{f(\mathbf{x}) - \lambda\,[g(\mathbf{x}) - k]\} = \begin{cases} \text{opt } f(\mathbf{x}) & \text{if } g(\mathbf{x}) = k \\ \infty & \text{otherwise} \end{cases}$$

- Therefore: $\bar{\mathbf{x}}$ solves (P) $\implies$ $\exists\ \bar{\lambda}$ such that $\nabla \mathcal{L}(\bar{\mathbf{x}}, \bar{\lambda}) = \mathbf{0}$
- Note that

$$\nabla \mathcal{L}(\mathbf{x}, \lambda) = \mathbf{0} \quad \Leftrightarrow \quad \begin{cases} \nabla f(\mathbf{x}) - \lambda \nabla g(\mathbf{x}) & = \mathbf{0} \\ g(\mathbf{x}) & = k \end{cases}$$

Example 1

Problem: min $\quad f(x, y) = x^2 + 2y^2 \quad$ s.t. $\quad x^2 + y^2 = 1$



Example 1

Problem: min $\quad f(x,y) = x^2 + 2y^2 \quad$ s.t. $\quad x^2 + y^2 = 1$

Solution.

- Lagrange function

$$\mathcal{L}(x,y,\lambda) = (x^2 + 2y^2) - \lambda(x^2 + y^2 - 1)$$

- $\nabla \mathcal{L}(x,y,\lambda) = \mathbf{0}$ gives KKT system

$$2x - 2\lambda x = 0 \tag{1}$$
$$4y - 2\lambda y = 0 \tag{2}$$
$$x^2 + y^2 - 1 = 0 \tag{3}$$

- (1) gives $x = 0$ or $\lambda = 1$
  - If $x = 0$, then (3) gives $y = 1$ or $y = -1$
  - If $\lambda = 1$, then (2) gives $y = 0$, and (3) gives $x = 1$ or $x = -1$
- KKT points (extreme points): $(0,1), (0,-1), (1,0), (-1,0)$
- $f(0,1) = 2, \quad f(0,-1) = 2, \quad f(1,0) = 1, \quad f(-1,0) = 1$
  $\Rightarrow \quad f_{min} = 1$ at $(1,0)$ and $(-1,0)$

Example 2

Problem: max $\quad f(x,y,z) = xyz \quad \text{s.t.} \quad 2xy + 2yz + xz = 12, \quad x,y,z \geq 0$

Solution.

- Lagrange function

$$\mathcal{L}(x,y,z,\lambda) = xyz - \lambda(2xy + 2yz + xz - 12)$$

- $\nabla\mathcal{L}(x,y,z,\lambda) = \mathbf{0}$ gives KKT system

$$\begin{aligned} yz - \lambda(2y + z) &= 0 && (4) \\ xz - \lambda(2x + 2z) &= 0 && (5) \\ xy - \lambda(2y + x) &= 0 && (6) \\ 2xy + 2yz + xz - 12 &= 0 && (7) \end{aligned}$$

- Observe that $\lambda \neq 0$ (Why?)
- Multiply (4) by x, (5) by y, using $\lambda \neq 0$, we have

$$2xy + xz = 2xy + 2yz \quad \Leftrightarrow \quad xz = 2yz \quad \Rightarrow \quad x = 2y$$

- Multiply (4) by x, (6) by z, using $\lambda \neq 0$, we have

$$2xy + xz = 2yz + xz \quad \Leftrightarrow \quad xy = yz \quad \Rightarrow \quad x = z$$

- Put $x = 2y = z$ into (7), we obtain $y = 1$, $x = z = 2$

Lagrange's method

- Case of one equality constraint
- Case of multiple equality constraints
- Case of mixed constraints

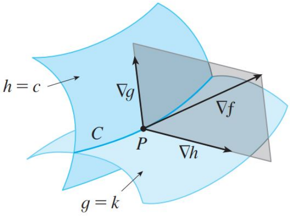
Intuition

Problem: optimize $f(x,y,z)$ subject to $g(x,y,z) = k$, $h(x,y,z) = c$
Geometry:

- Curve $C := [g = k] \cap [h = c]$
- If (x_0, y_0, z_0) is an optimizer of f over C, then at (x_0, y_0, z_0) we have
  - $\nabla f \perp C$
  - $\nabla g \perp [g = k] \quad \Rightarrow \quad \nabla g \perp C$
  - $\nabla h \perp [h = c] \quad \Rightarrow \quad \nabla h \perp C$

So ∇f, ∇g, ∇h are coplanar at (x_0, y_0, z_0), and therefore

$$\exists \lambda, \mu \in \mathbb{R} : \nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0) + \mu \nabla h(x_0, y_0, z_0)$$



Method

Problem:

optimize $f(\mathbf{x})$ subject to $g_1(\mathbf{x}) = c_1, \dots, g_k(\mathbf{x}) = c_k, \mathbf{x} \in \mathbb{R}^n$

Method:

- Lagrange function

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i=1}^{k} \lambda_i \left[g_i(\mathbf{x}) - c_i \right]$$

Note: $\lambda_1, \dots, \lambda_k$ are called Lagrange multipliers

- Solve the KKT system w.r.t. $\mathbf{x}$ and $\lambda = (\lambda_1, \dots, \lambda_k)$:

$$\nabla \mathcal{L}(\mathbf{x}, \lambda) = \mathbf{0} \quad \Leftrightarrow \quad \begin{cases} \nabla f(\mathbf{x}) - \sum_{i=1}^{k} \lambda_i \nabla g_i(\mathbf{x}) & = \mathbf{0} \\ g_i(\mathbf{x}) & = c_i \quad (i = 1, \dots, k) \end{cases}$$

- Let S be the set of all KKT points (solutions $\mathbf{x}$ to the KKT system)
- Evaluate f at all KKT points
  - $\operatorname{argmin}_S f$ solves the minimum problem
  - $\operatorname{argmax}_S f$ solves the maximum problem

Example

Problem: max $\quad x + 2y + 3z \quad$ s.t. $\quad x^2 + y^2 = 1, x - y + z = 1$

Solution.

- Lagrange function

$$\mathcal{L}(x, y, z, \lambda_1, \lambda_2) = (x + 2y + 3z) - \lambda_1(x^2 + y^2 - 1) - \lambda_2(x - y + z - 1)$$

- $\nabla\mathcal{L}(x, y, z, \lambda_1, \lambda_2) = \mathbf{0}$ gives KKT system

$$1 - 2\lambda_1 x - \lambda_2 = 0 \tag{8}$$
$$2 - 2\lambda_1 y + \lambda_2 = 0 \tag{9}$$
$$3 - \lambda_2 = 0 \tag{10}$$
$$x^2 + y^2 - 1 = 0 \tag{11}$$
$$x - y + z - 1 = 0 \tag{12}$$

- (10) & (8) give $x = -\frac{1}{\lambda_1}$; (10) & (9) give $y = \frac{5}{2\lambda_1}$
- Substitute to (11), then use (12), we obtain extreme points

$$(x, y, z) = \left(-\frac{2}{\sqrt{29}}, \frac{5}{\sqrt{29}}, 1 + \frac{7}{\sqrt{29}}\right), \left(\frac{2}{\sqrt{29}}, -\frac{5}{\sqrt{29}}, 1 - \frac{7}{\sqrt{29}}\right)$$

- Evaluate objective function at extreme points, we conclude that maximum $= 3 + \sqrt{29}$ at $\left(-\frac{2}{\sqrt{29}}, \frac{5}{\sqrt{29}}, 1 + \frac{7}{\sqrt{29}}\right)$

Lagrange's method

- Case of one equality constraint
- Case of multiple equality constraints
- Case of mixed constraints

Intuition

(Minimization) problem:

$$\min \quad f(\mathbf{x}) \quad \text{s.t.} \quad g_1(\mathbf{x}) \leq 0, \ldots, g_m(\mathbf{x}) \leq 0, \quad h_1(\mathbf{x}) = 0, \ldots, h_p(\mathbf{x}) = 0$$

Note: Finding $\max f(\mathbf{x})$ is equivalent to finding $\min(-f(\mathbf{x}))$

Lagrange function: $\mathcal{L}(\mathbf{x}, \lambda, \mu) := f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x})$

Note: $\lambda_i \geq 0$ and $\mu_j \in \mathbb{R}$ are called Lagrange multipliers

Observation:

$$\max_{\lambda \in \mathbb{R}_+^m, \mu \in \mathbb{R}^p} \mathcal{L}(\mathbf{x}, \lambda, \mu) = \begin{cases} f(\mathbf{x}) & \text{if } g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0 \\ +\infty & \text{otherwise} \end{cases}$$

KKT system:

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \mu) &= \mathbf{0} \\ g_i(\mathbf{x}) \leq 0, \quad \lambda_i \geq 0, \quad \lambda_i g_i(\mathbf{x}) &= 0 \quad \forall i = 1, \ldots, m \\ h_j(\mathbf{x}) &= 0 \quad \forall j = 1, \ldots, p \end{aligned}$$

Method

Problem:

$$\min \quad f(\mathbf{x}) \quad \text{s.t.} \quad g_1(\mathbf{x}) \leq 0, \ldots, g_m(\mathbf{x}) \leq 0, \quad h_1(\mathbf{x}) = 0, \ldots, h_p(\mathbf{x}) = 0$$

Lagrange function: $\mathcal{L}(\mathbf{x}, \lambda, \mu) := f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x})$

(on $\lambda \in \mathbb{R}_+^m$ and $\mu \in \mathbb{R}^p$)

Method: Evaluate f at KKT points of the system

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \mu) &= \mathbf{0} \\ g_i(\mathbf{x}) \leq 0, \quad \lambda_i \geq 0, \quad \lambda_i g_i(\mathbf{x}) &= 0 \quad \forall i = 1, \ldots, m \\ h_j(\mathbf{x}) &= 0 \quad \forall j = 1, \ldots, p \end{aligned}$$

In shorten form:

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda, \mu) &= \mathbf{0} \\ \nabla_{\lambda} \mathcal{L}(\mathbf{x}, \lambda, \mu) &\leq \mathbf{0} \\ \lambda^T \nabla_{\lambda} \mathcal{L}(\mathbf{x}, \lambda, \mu) &= 0 \\ \nabla_{\mu} \mathcal{L}(\mathbf{x}, \lambda, \mu) &= \mathbf{0} \end{aligned}$$

Example

Problem: $\min \quad f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 + x_4^2 \quad \text{s.t.} \quad x_4 \le \frac{1}{4}, \quad x_1 + x_2 + x_3 + x_4 = 1$

Solution.

- Lagrange function (with $\lambda \ge 0$ and $\mu \in \mathbb{R}$)

$$\mathcal{L}(\mathbf{x}, \lambda, \mu) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \lambda(x_4 - \frac{1}{4}) + \mu(x_1 + x_2 + x_3 + x_4 - 1)$$

- KKT system

$$\begin{aligned}
2x_1 + \mu &= 0 && (13)\\
2x_2 + \mu &= 0 && (14)\\
2x_3 + \mu &= 0 && (15)\\
2x_4 + \lambda + \mu &= 0 && (16)\\
x_4 &\le \frac{1}{4} && (17)\\
\lambda &\ge 0 && (18)\\
\lambda(x_4 - \frac{1}{4}) &= 0 && (19)\\
x_1 + x_2 + x_3 + x_4 &= 1 && (20)
\end{aligned}$$

has unique solution $(\mathbf{x}^*, \lambda^*, \mu^*) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, -\frac{1}{2})$. Hence $f_{min} = f(\mathbf{x}^*) = \frac{1}{4}$.

Thank you for your attention!