MATH2111 Higher Several Variable Calculus
Differentiable Functions

Dr. Jonathan Kress

School of Mathematics and Statistics
University of New South Wales

Semester 1, 2014

Differentiability of $f : \mathbb{R}^n \to \mathbb{R}^m$

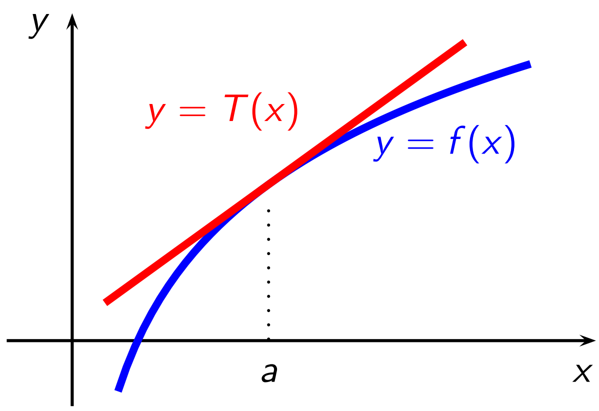
$f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ means there is a "good" straight line[1] approximation to f near a called the tangent line. This approximating function is given by

$$T(x) \;=\; f(a) + f'(a)(x-a) \;=\; f(a) - f'(a)a + f'(a)x \;=\; y_0 + L(x).$$

where, for each a, $y_0 = f(a) - f'(a)a$ is a fixed number and $L : \mathbb{R} \to \mathbb{R}$ is the linear map given by $L(x) = f'(a)x$.

$f'(a)$ is called the derivative of f at a and is the slope of the "good" straight line approximation. It can be found by calculating the a limit.

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



[1] not vertical

Affine maps

Definition

The function $T : \mathbb{R}^n \to \mathbb{R}^m$ is affine means there is $\mathbf{y}_0 \in \mathbb{R}^m$ and a linear map (ie matrix) $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$T(\mathbf{x}) = \mathbf{y}_0 + \mathbf{L}(\mathbf{x}).$$

An affine function $T : \mathbb{R} \to \mathbb{R}$ has the form

$$T(x) = b + mx, \qquad \text{for constants } m, b \in \mathbb{R}.$$

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at a if there is a "good" affine approximation to f at a given by

$$T(x) = \underbrace{f(a) - f'(a)a}_{\mathbf{y}_0} + \underbrace{f'(a)x}_{\mathbf{L}(x)}$$

and "good" means

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

Good affine approximation

Need to rewrite the definition of "good" in a way that can be used for $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$.

$$\begin{aligned}
& f'(a) &=& \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \\
\Leftrightarrow\quad & 0 &=& \lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x - a)}{x - a} \\
\Leftrightarrow\quad & 0 &=& \lim_{x \to a} \frac{f(x) - T(x)}{x - a} \\
\Leftrightarrow\quad & 0 &=& \lim_{x \to a} \left| \frac{f(x) - T(x)}{x - a} \right| \\
\Leftrightarrow\quad & 0 &=& \lim_{x \to a} \frac{|f(x) - T(x)|}{|x - a|}
\end{aligned}$$

$$T(x) = f(a) + f'(a)(x - a) = f(a) + L(x - a).$$

Differentiability of $f : \mathbb{R}^n \to \mathbb{R}^m$

Definition

A function $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable $\mathbf{a} \in \Omega$ if there is a linear map $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - \mathbf{L}(\mathbf{x} - \mathbf{a})||}{||\mathbf{x} - \mathbf{a}||} = 0.$$

The matrix of the linear map $\mathbf{L}$ is called the derivative of $\mathbf{f}$ at $\mathbf{a}$ and is denoted $D_{\mathbf{a}}\mathbf{f}$.

We could use the ϵ-δ definition of the limit to give an alternative form.

Definition (Alternative)

A function $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable $\mathbf{a} \in \Omega$ if there is a linear map $\mathbf{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that for all $\epsilon > 0$ there exists $\delta > 0$ such that for $\mathbf{x} \in \Omega$

$$||\mathbf{x} - \mathbf{a}|| < \delta \;\Rightarrow\; ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - \mathbf{L}(\mathbf{x} - \mathbf{a})|| < \epsilon||\mathbf{x} - \mathbf{a}||.$$

Differentiability examples

Suppose $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation given by $\mathbf{T}(\mathbf{x}) = A_{\mathbf{T}}\mathbf{x}$. Is it differentiable and if so, what is it's derivative?

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{||\mathbf{T}(\mathbf{x}) - \mathbf{T}(\mathbf{a}) - \mathbf{T}(\mathbf{x} - \mathbf{a})||}{||\mathbf{x} - \mathbf{a}||} = \lim_{\mathbf{x} \to \mathbf{a}} \frac{0}{||\mathbf{x} - \mathbf{a}||} = 0.$$

Hence $\mathbf{T}$ is differentiable and $D_{\mathbf{a}}\mathbf{T} = A_{\mathbf{T}}$.

Differentiability examples

For $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ with

$$\mathbf{f}(x,y) = \begin{pmatrix} x^2 + 2xy \\ x + y^2, \end{pmatrix} \qquad L = \begin{pmatrix} 4 & 2 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

show that $\mathbf{f}$ is differentiable at $\mathbf{a}$ and that the matrix of its derivative is $D_{\mathbf{a}}\mathbf{f} = L$.

$$\begin{aligned} \mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - L(\mathbf{x} - \mathbf{a}) &= \begin{pmatrix} x^2 + 2xy \\ x + y^2 \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \end{pmatrix} - \begin{pmatrix} 4 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x-1 \\ y-1 \end{pmatrix} \\ &= \begin{pmatrix} x^2 + 2xy - 4x - 2y + 3 \\ y^2 + 1 - 2y \end{pmatrix} \end{aligned}$$

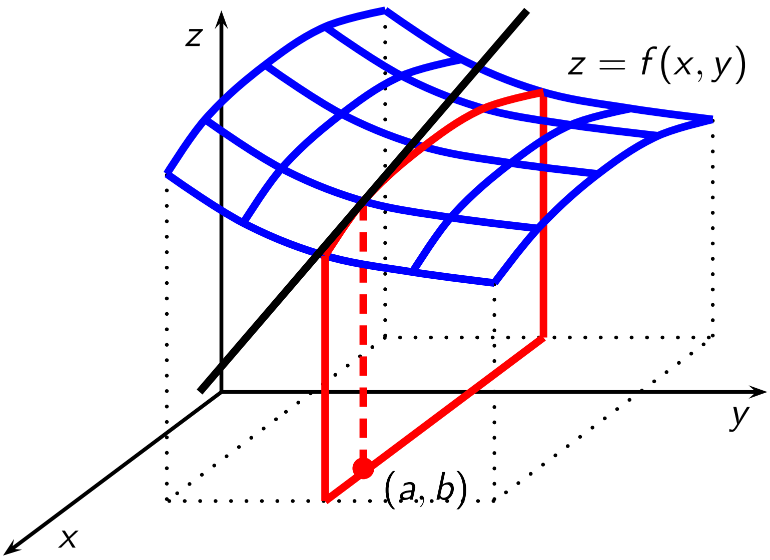
So for $\mathbf{f}$ to be differentiable at $\mathbf{a}$ with derivative L, we need

$$\lim_{(x,y)\to(1,1)} \frac{\sqrt{(x^2 + 2xy - 4x - 2y + 3)^2 + (y^2 + 1 - 2y)^2}}{\sqrt{(x-1)^2 + (y-1)^2}} = 0.$$

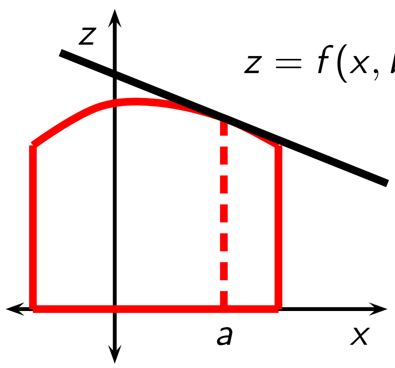
This is true, but takes a bit of work.

Partial derivatives

If we fix a value of y we can calculate the rate of change of $f(x,y)$ as only x changes. This is called the partial derivative of $f(x,y)$ with respect to x.



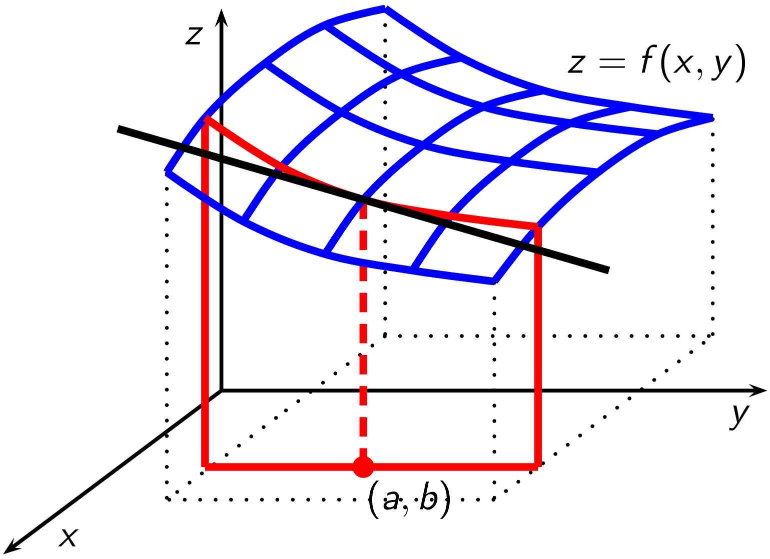
In the $y = b$ slice we can find the slope of the tangent line at the point $x = a$.



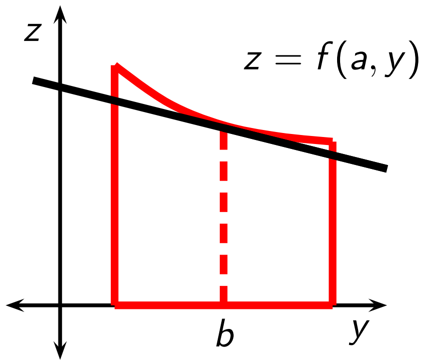
$$D_1 f(a,b) = f_1(a,b) = f_x(a,b) = \frac{\partial f}{\partial x}(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h}.$$

Partial derivatives

If we fix a value of x we can calculate the rate of change of $f(x, y)$ as only y changes. This is called the partial derivative of $f(x, y)$ with respect to y.



In the $x = a$ slice we can find the slope of the tangent line at the point $y = b$.



$$D_2 f(a, b) = f_2(a, b) = f_y(a, b) = \frac{\partial f}{\partial y}(a, b) = \lim_{h \to 0} \frac{f(a, b + h) - f(a, b)}{h}.$$

Partial derivatives

Just as in one variable calculus, we rarely use the definition to calculate a derivative, we use the 'rules' of differentiation remembering to treat some variables as constants.

If $z = f(x, y) = x^2 y + x^3 + e^{2y}$, find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$.

$$\frac{\partial z}{\partial x} = 2xy + 3x^2, \qquad \frac{\partial z}{\partial y} = x^2 + 2e^{2y}$$

Find $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ if $f(x, y) = (x^2 + y^3)^{\frac{1}{2}}$.

$$\frac{\partial f}{\partial x} = \frac{1}{2}(x^2 + y^3)^{-\frac{1}{2}} 2x = x(x^2 + y^3)^{-\frac{1}{2}}$$

$$\frac{\partial f}{\partial y} = \frac{1}{2}(x^2 + y^3)^{-\frac{1}{2}} 3y^2 = \frac{3}{2} y^2 (x^2 + y^3)^{-\frac{1}{2}}$$

Find $\dfrac{\partial G}{\partial b}$ if $G(a, b, c) = a^2 b^3 c^4 + bc$. $\qquad \dfrac{\partial G}{\partial b} = 3a^2 b^2 c^4 + c\,.$

Partial derivatives

We can also calculate higher partial derivative, but unlike one variable calculus, there are a number of possibilities.

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) \quad \text{is denoted} \quad \frac{\partial^2 f}{\partial x^2} \quad \text{or} \quad f_{xx} \quad \text{or} \quad f_{11}$$
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \quad \text{is denoted} \quad \frac{\partial^2 f}{\partial x \partial y} \quad \text{or} \quad f_{yx} \quad \text{or} \quad f_{12}$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) \quad \text{is denoted} \quad \frac{\partial^2 f}{\partial y \partial x} \quad \text{or} \quad f_{xy} \quad \text{or} \quad f_{21}$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) \quad \text{is denoted} \quad \frac{\partial^2 f}{\partial y^2} \quad \text{or} \quad f_{yy} \quad \text{or} \quad f_{22}$$

For $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ with coordinates x_i and standard basis vectors $\mathbf{e}_i$

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h}.$$

Partial derivatives

For $f(x, y) = x^2 y + 2$,

$$\frac{\partial f}{\partial x} = 2xy, \qquad \frac{\partial f}{\partial y} = x^2$$

$$\frac{\partial^2 f}{\partial x^2} = 2y, \qquad \frac{\partial^2 f}{\partial y^2} = 0,$$

$$\begin{aligned} \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}\left(2xy\right) = 2x, \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}\left(x^2\right) = 2x. \end{aligned}$$

Notice that, as expected, the two mixed partial derivatives are equal.

Partial derivatives

Theorem (Clariaut's theorem)

If $f, \dfrac{\partial f}{\partial x_i}, \dfrac{\partial f}{\partial x_j}, \dfrac{\partial^2 f}{\partial x_i x_j}, \dfrac{\partial^2 f}{\partial x_j x_i}$ *all exist and are continuous on an open set around* $\mathbf{a}$ *then*

$$\frac{\partial^2 f}{\partial x_i x_j}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j x_i}(\mathbf{a}).$$

That is, the partial derivatives commute.

Here's an example where they don't commute.

Calculate $f_{xy}(0,0)$ and $f_{yx}(0,0)$ for

$$f(x,y) = \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2} & \text{for } (x,y) \neq (0,0) \\ 0 & \text{for } (x,y) = (0,0). \end{cases}$$

Partial derivatives

$$f(x,y) = \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2} & \text{for } (x,y) \neq (0,0) \\ 0 & \text{for } (x,y) = (0,0). \end{cases}$$

Away from $(0,0)$, f is a well defined rational function of its arguments.

$$f_x(x,y) = \frac{y(x^4-y^4+4x^2y^2)}{(x^2+y^2)^2} \quad \text{and} \quad f_y(x,y) = \frac{x(x^4-y^4-4x^2y^2)}{(x^2+y^2)^2}.$$

At $(0,0)$ we need to use the definition to calculate the partial derivatives.

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

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

$$f_{xy}(0,0) = \lim_{h\to 0} \frac{f_x(0,h)-f_x(0,0)}{h} = \lim_{h\to 0} \frac{\dfrac{h(0^4-h^4+0)}{h^4} - 0}{h} = -1.$$

$$f_{yx}(0,0) = \lim_{h\to 0} \frac{f_y(h,0)-f_y(0,0)}{h} = \lim_{h\to 0} \frac{\dfrac{h(h^4-0^4-0)}{h^4} - 0}{h} = 1 \neq f_{xy}(0,0).$$

Jacobian matrix

Definition

If all partial derivatives of $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ exist at $\mathbf{a} \in \Omega$, then the Jacobian matrix of $\mathbf{f}$ at $\mathbf{a}$ is

$$J_{\mathbf{a}}\mathbf{f} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \frac{\partial f_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}.$$

Theorem

For $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ and an interior point $\mathbf{a} \in \Omega$. If $\mathbf{f}$ is differentiable at $\mathbf{a}$ then all partial derivatives $\dfrac{\partial f_j}{\partial x_i}$ of the components of $\mathbf{f}$ exist at $\mathbf{a}$ and $D_{\mathbf{a}}\mathbf{f} = J_{\mathbf{a}}\mathbf{f}$.

That is, where $\mathbf{f}$ is differentiable, its derivative is given by its Jacobian matrix.

Jacobian matrix

The Jacobian matrix may exist even when the function is not differentiable.

Example: $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x,y) = \begin{cases} 0 & \text{for } x = 0 \text{ or } y = 0, \\ -1 & \text{otherwise.} \end{cases}$

Clearly $\dfrac{\partial f}{\partial x}(0,0) = \dfrac{\partial f}{\partial y}(0,0) = 0$. However, the affine function

$$T(x,y) = f(0,0) + J_{(0,0)}f \begin{pmatrix} x \\ y \end{pmatrix} = 0 + (0\ 0)\begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

is not a "good" approximation to $f(x,y)$ near $(0,0)$.

Notice that in this example, f is not continuous. Should a differentiable function be continuous?

Differentiable $\Rightarrow$ continuous

Lemma

For $\mathbf{x} \in \mathbb{R}^n$ *and* L *an* $m \times n$ *matrix,* $\lim_{\mathbf{x}\to\mathbf{0}} ||L\mathbf{x}|| = 0$.

Proof.

Let $\mathbf{r}_i$ be the i^{th} row of L and so the i^{th} row of $L\mathbf{x}$ is $\mathbf{r}_i \cdot \mathbf{x}$. Then, using the Cauchy-Schwarz inequality ($|\mathbf{a}\cdot\mathbf{b}| \le ||\mathbf{a}||\ ||\mathbf{b}||$),

$$||L\mathbf{x}|| = \sqrt{\sum_{i=1}^{m} (\mathbf{r}_i \cdot \mathbf{x})^2} \le \sqrt{\sum_{i=1}^{m} ||\mathbf{r}_i||^2 ||\mathbf{x}||^2} = ||\mathbf{x}|| \sqrt{\sum_{i=1}^{m} ||\mathbf{r}_i||^2}.$$

So,

$$0 \le \lim_{\mathbf{x}\to\mathbf{0}} ||L\mathbf{x}|| \le \sqrt{\sum_{i=1}^{m} ||\mathbf{r}_i||^2}\ \lim_{\mathbf{x}\to\mathbf{0}} ||\mathbf{x}|| = 0.$$

Hence, $\lim_{\mathbf{x}\to\mathbf{0}} ||L\mathbf{x}|| = 0$. □

Differentiable $\Rightarrow$ continuous

Theorem

Suppose $\Omega \in \mathbb{R}^n$ *is open and* $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable on* Ω. *Then* $\mathbf{f}$ *is continuous on* Ω.

Proof.

If $\mathbf{f}$ is differentiable at $\mathbf{a}$ then there is a matrix L such that

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - L(\mathbf{x}-\mathbf{a})||}{||\mathbf{x}-\mathbf{a}||} = 0 \quad \Rightarrow \quad \lim_{\mathbf{x}\to\mathbf{a}} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - L(\mathbf{x}-\mathbf{a})|| = 0.$$

Now,

$$\begin{aligned} \lim_{\mathbf{x}\to\mathbf{a}} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})|| &= \lim_{\mathbf{x}\to\mathbf{a}} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - L(\mathbf{x}-\mathbf{a}) + L(\mathbf{x}-\mathbf{a})|| \\ &\le \lim_{\mathbf{x}\to\mathbf{a}} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - L(\mathbf{x}-\mathbf{a})|| + ||L(\mathbf{x}-\mathbf{a})|| \\ &= 0 + 0 \ = \ 0. \end{aligned}$$

So $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$ and hence $\mathbf{f}$ is continuous at $\mathbf{a}$. □

Differentiability

Theorem

Suppose $\Omega \subset \mathbb{R}^n$ is open and $\mathbf{f} : \Omega \to \mathbb{R}^m$. If $\dfrac{\partial f_j}{\partial x_i}$ exists and is continuous on Ω for all $i = 1, \dots, n$ and $j = 1, \dots, m$, then $\mathbf{f}$ is differentiable on Ω.

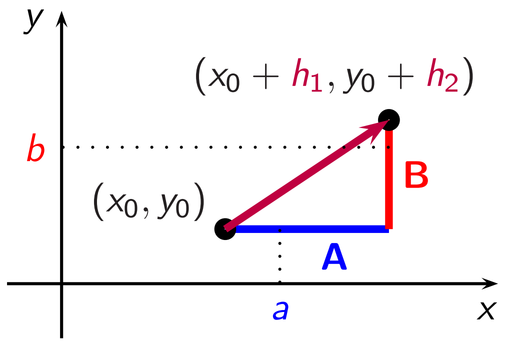
Example: Consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ with $f(x, y) = (x^2 + 2xy, x + y^2)$.

The Jacobian exists and is given by $J_{(x,y)} f = \begin{pmatrix} 2x + 2y & 2x \\ 1 & 2y \end{pmatrix}$.

Each entry is continuous on $\mathbb{R}^2$ and hence f is differentiable on $\mathbb{R}^2$ with derivative $D_{(x,y)} f = J_{(x,y)} f$.

Notation: We often write $Jf(x, y)$ instead of $J_{(x,y)} f$ or even just Jf. Eg, $Jf(1, 1) = \begin{pmatrix} 4 & 2 \\ 1 & 2 \end{pmatrix}$. Similarly form Df, $D_{(x,y)} f$, $Df(x, y)$.

Sketch of proof for $f : \mathbb{R}^2 \to \mathbb{R}$



The MVT says that along side **A** there is $a \in (x_0, x_0 + h_1)$ such that

$$f(x_0 + h_1, y_0) - f(x_0, y_0) = \frac{\partial f}{\partial x}(a, y_0) h_1.$$

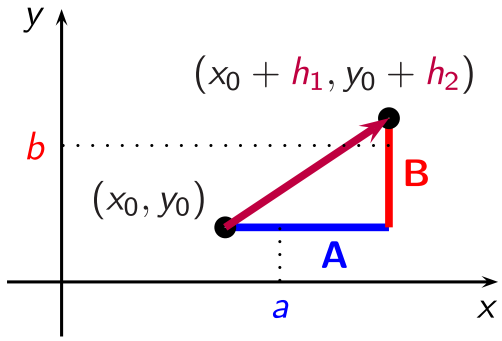
Continuity of $\dfrac{\partial f}{\partial x}$ says $\forall \epsilon_1 > 0$ we can choose h_1 small enough so that

$$\left| \frac{\partial f}{\partial x}(a, y_0) - \frac{\partial f}{\partial x}(x_0, y_0) \right| < \epsilon_1 \quad \Rightarrow \quad \frac{\partial f}{\partial x}(a, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) + \epsilon_1'$$

where $-\epsilon_1 < \epsilon_1' < \epsilon_1$. So,

$$f(x_0 + h_1, y_0) - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) h_1 + \epsilon_1' h_1.$$

Sketch of proof for $f : \mathbb{R}^2 \to \mathbb{R}$



Similarly, the MVT says that along side **B** there is $b \in (y_0, y_0 + h_1)$ such that

$$f(x_0+h_1, y_0+h_2) - f(x_0+h_1, y_0) = \frac{\partial f}{\partial y}(x_0+h_1, b)h_2.$$

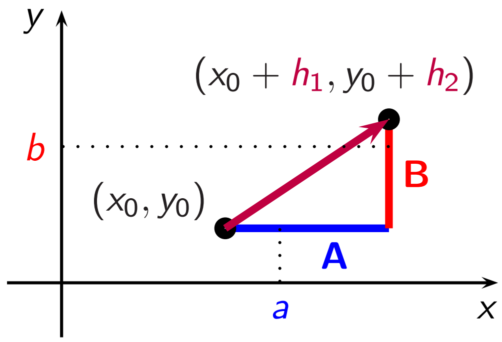
Continuity of $\dfrac{\partial f}{\partial y}$ says $\forall \epsilon_2 > 0$ we can choose $||(h_1, h_2)||$ small enough so that

$$\left| \frac{\partial f}{\partial y}(x_0 + h_1, b) - \frac{\partial f}{\partial y}(x_0, y_0) \right| < \epsilon_2 \quad \Rightarrow \quad \frac{\partial f}{\partial y}(x_0 + h_1, b) = \frac{\partial f}{\partial y}(x_0, y_0) + \epsilon_2'$$

where $-\epsilon_2 < \epsilon_2' < \epsilon_2$. So,

$$f(x_0 + h_1, y_0 + h_2) - f(x_0 + h_1, y_0) = \frac{\partial f}{\partial y}(x_0, y_0)h_2 + \epsilon_2' h_2.$$

Sketch of proof for $f : \mathbb{R}^2 \to \mathbb{R}$



$$f(x_0 + h_1, y_0) - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)h_1 + \epsilon_1' h_1.$$

$$f(x_0 + h_1, y_0 + h_2) - f(x_0 + h_1, y_0) = \frac{\partial f}{\partial y}(x_0, y_0)h_2 + \epsilon_2' h_2.$$

So,

$$\begin{aligned} & f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0) \\ &= f(x_0 + h_1, y_0 + h_2) - f(x_0 + h_1, y_0) + f(x_0 + h_1, y_0) - f(x_0, y_0) \\ &= \frac{\partial f}{\partial x}(x_0, y_0)h_1 + \epsilon_1' h_1 + \frac{\partial f}{\partial y}(x_0, y_0)h_2 + \epsilon_2' h_2 \\ &= Jf(x_0, y_0) \cdot (h_1, h_2) + (\epsilon_1', \epsilon_2') \cdot (h_1, h_2) \end{aligned}$$

Sketch of proof for $f : \mathbb{R}^2 \to \mathbb{R}$

For any $\epsilon_1 > 0$ and $\epsilon_2 > 0$ we can choose $h_1 > 0$ and $h_2 > 0$ such that

$$\begin{aligned}
0 &\leq \frac{|f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0) - Jf(x_0, y_0) \cdot (h_1, h_2)|}{||(h_1, h_2)||} \\
&= \frac{|(\epsilon_1', \epsilon_2') \cdot (h_1, h_2)|}{||(h_1, h_2)||} \\
&\leq \frac{||(\epsilon_1', \epsilon_2')||\ ||(h_1, h_2)||}{||(h_1, h_2)||} \\
&= ||(\epsilon_1', \epsilon_2')|| \\
&\leq ||(\epsilon_1, \epsilon_2)||
\end{aligned}$$

So,

$$\lim_{(h_1, h_2) \to (0,0)} \frac{|f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0) - Jf(x_0, y_0) \cdot (h_1, h_2)|}{||(h_1, h_2)||} = 0.$$

Hence f is differentiable at (x_0, y_0) with derivative $Jf(x_0, y_0)$.

Gradient of f

For $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, the Jacobian, if it exists, is a $1 \times n$ matrix

$$Jf = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

Often we think of this as a vector called the gradient of f. That is,

$$\operatorname{grad}(f) = \nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right).$$

$\left[\text{Think of } \nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n} \right).\right]$

Example: $f : \mathbb{R}^4 \to \mathbb{R}$ $\quad f(x, y, z, t) = xyz + \cos(x + 3t)$.

$$\nabla f = \Big(yz - \sin(x + 3t), xz, xy, -3\sin(x + 3t) \Big),$$

$$\nabla f(1, 2, 3, 0) = (6 - \sin 1, 3, 2, -3\sin 1).$$

Affine approximation

Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \Omega$. The best affine approximation to f at $\mathbf{a}$ can be written in terms of the gradient vector.

$$T(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

For $n = 1$:

$$T(x) = f(a) + f'(a)(x - a).$$

For $n = 2$: $\quad (\mathbf{a} = (a, b))$

$$\begin{aligned} T(x, y) &= f(a, b) + \nabla f(a, b) \cdot ((x, y) - (a, b)) \\ &= f(a, b) + \left(\frac{\partial f}{\partial x}(a, b), \frac{\partial f}{\partial y}(a, b) \right) \cdot (x - a, y - b) \\ &= f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) \end{aligned}$$

$z = T(x, y)$ is the tangent plane to $z = f(x, y)$ at $(x, y) = (a, b)$.

Tangent planes

Find the equation of the tangent plane to the graph of $f(x, y) = x^2 + y^4 + e^x$ at the point $(1, 0)$.

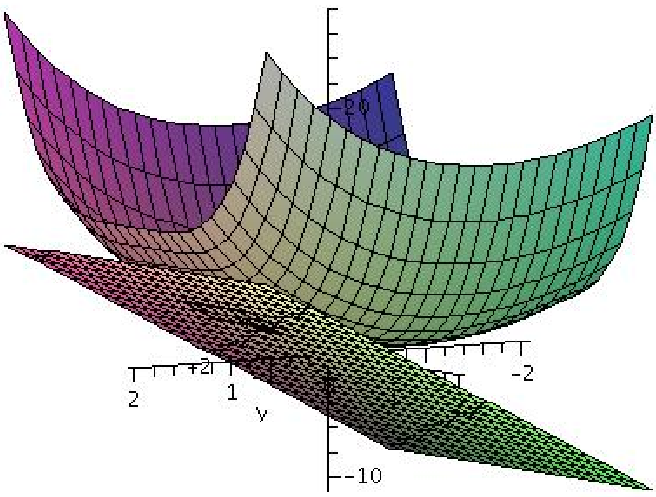
$$\nabla f(x, y) = \Big(f_x(x, y), f_y(x, y) \Big) = (2x + e^x, 4y^3).$$

So

$$f(1, 0) = 1 + e, \qquad \nabla f(1, 0) = (2 + e, 0)$$

and the tangent plane is

$$\begin{aligned} z &= f(1, 0) + \nabla f(1, 0) \cdot (x - 1, y - 0) \\ &= 1 + e + (2 + e)(x - 1) + 0y \\ &= -1 + (2 + e)x. \end{aligned}$$



Chain rule

First look at the composition of two affine maps $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^m \to \mathbb{R}^p$.

$$T_1(\mathbf{x}) = \mathbf{y}_1 + L_1\mathbf{x}, \quad \text{and} \quad T_2(\mathbf{x}) = \mathbf{y}_2 + L_2\mathbf{x}.$$

The derivatives of these affine maps are $DT_1 = L_1$ and $DT_2 = L_2$. What is the derivative of $T_3 = T_2 \circ T_1$?

$$\begin{aligned} T_3(\mathbf{x}) = (T_2 \circ T_1)(\mathbf{x}) &= T_2(T_1(\mathbf{x})) \\ &= T_2(\mathbf{y}_1 + L_1\mathbf{x}) \\ &= \mathbf{y}_2 + L_2(\mathbf{y}_1 + L_1\mathbf{x}) \\ &= \mathbf{y}_2 + L_2\mathbf{y}_1 + L_2L_1\mathbf{x} \\ &= \mathbf{y}_3 + L_3\mathbf{x} \end{aligned}$$

where $\mathbf{y}_3 = \mathbf{y}_2 + L_2\mathbf{y}_1$ and $L_3 = L_2L_1$ and so $D(T_2 \circ T_1) = L_2L_1$.

So the composition of two affine maps is an affine map and the derivative of the composition is the the product of the derivatives.

Chain rule

Consider some differentiable functions $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ with best affine approximations T_1 and T_2.

It seems plausible that $g \circ f$ is differentiable with best affine approximation $T_2 \circ T_1$. In that case we would have, $D(g \circ f) = Dg\ Df$.

Theorem (Chain rule)

Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ and $g : \Omega' \subset \mathbb{R}^m \to \mathbb{R}^p$, with $f(\Omega) \subset \Omega'$. If f and g are differentiable, then so is $g \circ f : \Omega \to \mathbb{R}^p$ and

$$D_{\mathbf{a}}(g \circ f) = D_{f(\mathbf{a})}g\ D_{\mathbf{a}}f,$$

or alternatively,

$$D(g \circ f)(\mathbf{a}) = Dg(f(\mathbf{a}))\ Df(\mathbf{a}).$$

See Marsden and Tromba for a proof in the case when Df and Dg are continuous and the Marsden and Tromba internet supplement for a more general proof.

Chain rule

Example

Let
$$x = r\cos\theta, \qquad y = r\sin\theta \qquad\qquad (*)$$
and $g(x,y) = xy^2$. What is $\dfrac{\partial g}{\partial r}$?

Since we have explicit expressions, we could calculate directly as
$$\frac{\partial}{\partial r} g(x(r,\theta), y(r,\theta)) = \frac{\partial}{\partial r}\left(r\cos\theta r^2\sin^2\theta\right) = 3r^2\cos\theta\sin^2\theta,$$
or we could use the chain rule:
$$\frac{\partial g}{\partial r} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial r} = y^2\cos\theta + 2xy\sin\theta = 3r^2\cos\theta\sin^2\theta.$$

How does this come from the chain rule stated on the previous slide?

Note that $(*)$ is really a map $f : \mathbb{R}^2 \to \mathbb{R}^2$ and g as a function of r and θ is really $g \circ f$.

Chain rule

We have $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}$ given by
$$f(r,\theta) = (f_1(r,\theta), f_2(r,\theta)) = (r\cos\theta, r\sin\theta) \;\Rightarrow\; Df = \begin{pmatrix} \dfrac{\partial f_1}{\partial r} & \dfrac{\partial f_1}{\partial \theta} \\ \dfrac{\partial f_2}{\partial r} & \dfrac{\partial f_2}{\partial \theta} \end{pmatrix}$$
and
$$g(x,y) = xy^2 \;\Rightarrow\; Dg = \begin{pmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{pmatrix}.$$
So, the derivative of $g \circ f : \mathbb{R}^2 \to \mathbb{R}$ is
$$D(g\circ f) = Dg\ Df = \begin{pmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial f_1}{\partial r} & \dfrac{\partial f_1}{\partial \theta} \\ \dfrac{\partial f_2}{\partial r} & \dfrac{\partial f_2}{\partial \theta} \end{pmatrix}$$
$$= \begin{pmatrix} \dfrac{\partial g}{\partial x}\dfrac{\partial f_1}{\partial r} + \dfrac{\partial g}{\partial y}\dfrac{\partial f_2}{\partial r} & \dfrac{\partial g}{\partial x}\dfrac{\partial f_1}{\partial \theta} + \dfrac{\partial g}{\partial y}\dfrac{\partial f_2}{\partial \theta} \end{pmatrix}$$

Chain rule

Suppose

$$z = e^{x^2+y} \quad \text{and} \quad x = \cos t,\ y = \sin t.$$

Find $\dfrac{dz}{dt}$ at $t = 0$.

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = e^{x^2+y}\ 2x(-\sin t) + e^{x^2+y}\cos t.$$

At $t = 0$,

$$x = 1, \qquad y = 0,$$

so

$$\left.\frac{dz}{dt}\right|_{t=0} = e^{1+0}.2.1.0 + e^{1+0}.1 = e.$$

Chain rule

Define

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x,y) = e^{x^2+y},$$
$$g : \mathbb{R} \to \mathbb{R}^2, \qquad g(t) = (\cos t, \sin t).$$

Both f and g are differentiable because the partial derivatives of their components exist and are continuous everywhere.

$$Df = Jf = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix}, \qquad Dg = Jg = \begin{pmatrix} \dfrac{\partial g_1}{\partial t} \\ \dfrac{\partial g_2}{\partial t} \end{pmatrix} = \begin{pmatrix} \dfrac{dg_1}{dt} \\ \dfrac{dg_2}{dt} \end{pmatrix}.$$

$$D(f\circ g) = J(f\circ g) = Jf\ Jg = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix}\begin{pmatrix} \dfrac{dg_1}{dt} \\ \dfrac{dg_2}{dt} \end{pmatrix} = \frac{\partial f}{\partial x}\frac{dg_1}{dt} + \frac{\partial f}{\partial y}\frac{dg_2}{dt}.$$

Chain rule

Suppose f depends on x, y, z and w and x, y, z and w depend on r, s and t. Write out the chain rule for $\dfrac{\partial f}{\partial s}$.

$$\begin{aligned}
\frac{\partial f}{\partial s} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial s} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial s} \\
\frac{\partial f}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial r} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial r} \\
\frac{\partial f}{\partial t} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial t}
\end{aligned}$$

[Of course we are assuming differentiability of the underlying maps.]

Chain rule

Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable and define $F : \mathbb{R}^2 \to \mathbb{R}$ by

$$F(x, y) = g(3x - 4y^2).$$

Show that any such function F must be a solution of the PDE

$$8y\frac{\partial F}{\partial x} + 3\frac{\partial F}{\partial y} = 0. \qquad (*)$$

Let $h : \mathbb{R}^2 \to \mathbb{R}$ be defined by $h(x, y) = 3x - 4y^2$. So $F = g \circ h$ and

$$\begin{pmatrix} \dfrac{\partial F}{\partial x} & \dfrac{\partial F}{\partial y} \end{pmatrix} = D_{(x,y)}F = D_{h(x,y)}g \ D_{(x,y)}h$$

$$= \left(g'(3x - 4y^2)\right)\begin{pmatrix} \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y} \end{pmatrix} = g'(3x - 4y^2)(3 \quad -8y)$$

So

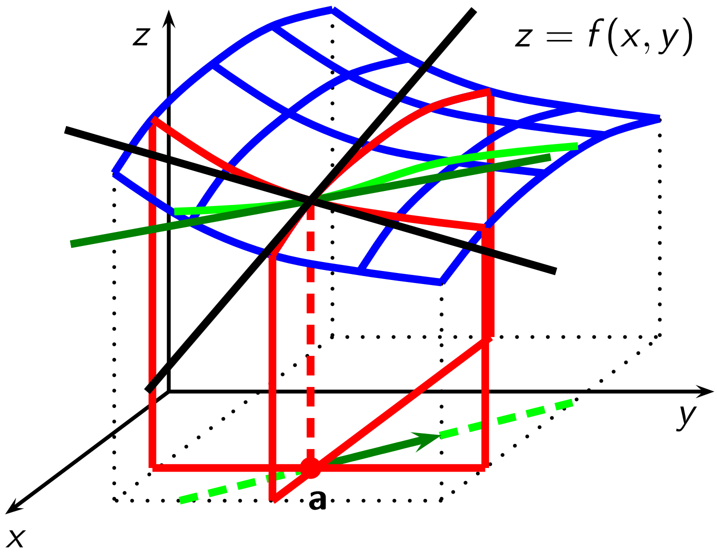
$$\frac{\partial F}{\partial x} = 3g'(3x - 4y^2) \quad \text{and} \quad \frac{\partial F}{\partial y} = -8yg'(3x - 4y^2)$$

and it is now easy to check that F satisfies $(*)$.

Directional derivative

For $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, the partial derivative $\dfrac{\partial f}{\partial x_i}$ measures the rate of change of f in the x_i-direction.

We can also ask for the rate of change in a non-coordinate direction.



Definition

The directional derivative of $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ in the direction of the unit vector $\hat{\mathbf{u}}$ at $\mathbf{a} \in \Omega$ is

$$D_{\hat{\mathbf{u}}} f(\mathbf{a}) = f'_{\hat{\mathbf{u}}}(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\hat{\mathbf{u}}) - f(\mathbf{a})}{t}.$$

Directional derivatives

Let $\mathbf{r} : I \subset \mathbb{R} \to \mathbb{R}^n$ (with 0 an interior point of I) be given by $\mathbf{r}(t) = \mathbf{a} + t\hat{\mathbf{u}}$. Then the directional derivative of f at $\mathbf{a}$ in the direction $\hat{\mathbf{u}}$ is

$$D_{\hat{\mathbf{u}}} f(\mathbf{a}) = f'_{\hat{\mathbf{u}}}(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{r}(t)) - f(\mathbf{r}(0))}{t}.$$

If we write $F = f \circ \mathbf{r}$ then

$$f'_{\hat{\mathbf{u}}}(\mathbf{a}) = \lim_{t \to 0} \frac{F(t) - F(0)}{t} = F'(0).$$

For differentiable f, the chain rule says

$$F'(0) = Df(\mathbf{a})\ D\mathbf{r}(0) = \nabla f(\mathbf{a}) \cdot \hat{\mathbf{u}}.$$

Theorem

Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$ and that $\hat{\mathbf{u}}$ is a unit vector. Then $f'_{\hat{\mathbf{u}}}(\mathbf{a})$ exists and

$$f'_{\hat{\mathbf{u}}}(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \hat{\mathbf{u}}.$$

Directional derivatives

For a differentiable function f, the Cauchy-Schwarz inequality gives

$$||f'_{\hat{\mathbf{u}}}(\mathbf{a})|| = ||\hat{\mathbf{u}} \cdot \nabla f(\mathbf{a})|| \leq ||\hat{\mathbf{u}}|| \, ||\nabla f(\mathbf{a})|| = ||\nabla f(\mathbf{a})|| \, .$$

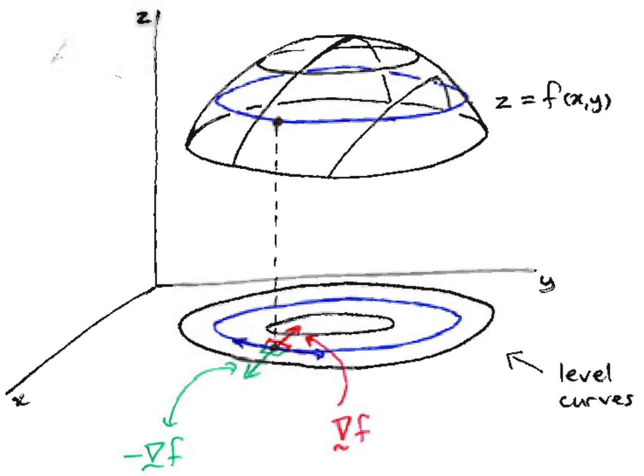
Equality occurs when $\hat{\mathbf{u}}$ is proportional to $\nabla f(\mathbf{a})$.

- The maximum rate of change of f at $\mathbf{a}$ occurs in the direction of $\nabla f(\mathbf{a})$.
- The minimum rate of change of f at $\mathbf{a}$ occurs in the direction of $-\nabla f(\mathbf{a})$.

Also,

$$f'_{\hat{\mathbf{u}}}(\mathbf{a}) = 0 \;\Leftrightarrow\; \hat{\mathbf{u}} \perp \nabla f(\mathbf{a}).$$

Directions normal to $\nabla f(\mathbf{a})$ are directions in which f is not changing, that is, tangent to a level set of f.



Directional derivatives

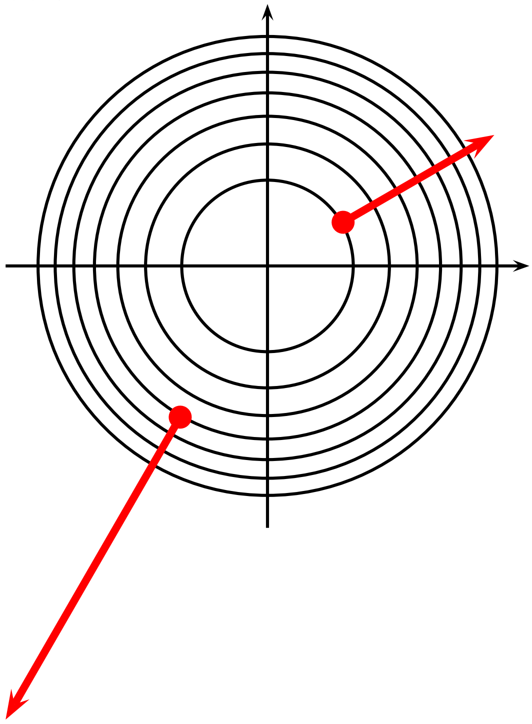
Consider $f : \mathbb{R}^2 \to \mathbb{R}$,

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

(a) Find ∇f.
(b) Sketch some level curves of f
(c) Indicate ∇f at some points on these curves.

(a) $\nabla f = (2x, 2y)$.

(b) The level curves $f(x, y) = 1, 2, 3, 4, 5, 6, 7$ are plotted below.



Directional derivatives

Find the directional derivative of f in the direction $(5,1)$ at the point $(2,1)$ where

$$f(x,y) = x^3 + 2y^2.$$

The function f is differentiable because its partial derivatives exist and are continuous and so we can calculate the directional derivative using the gradient vector.

$$\nabla f = (3x^2, 4y) \quad \Rightarrow \quad \nabla f(2,1) = (12,4).$$

A unit vector in the direction $(5,1)$ is

$$\hat{\mathbf{u}} = \frac{1}{\sqrt{26}}(5,1)$$

so

$$f'_{\hat{\mathbf{u}}}(2,1) = \hat{\mathbf{u}} \cdot \nabla f(2,1) = \frac{1}{\sqrt{26}}(5,1)\cdot(12,4) = \frac{64}{\sqrt{26}}.$$

Tangent planes

Consider the surface in $\mathbb{R}^3$ defined by the equation

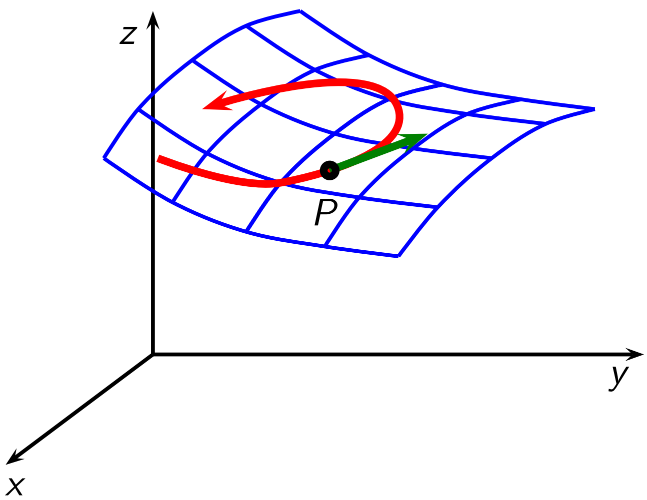
$$\phi(x,y,z) = c$$

for some constant c and differentiable function ϕ and let

$$\mathbf{r}(t) = \Big(f(t), g(t), h(t)\Big)$$

be a differentiable curve lying in the surface with tangent vector given by

$$\mathbf{r}'(t) = \Big(f'(t), g'(t), h'(t)\Big)$$



Since all points along $\mathbf{r}(t)$ lie in the surface,

$$\phi\Big(f(t),g(t),h(t)\Big) = c \;\Rightarrow\; \Big(\phi\circ\mathbf{r}\Big)(t) = c \;\Rightarrow\; D_{\mathbf{r}(t)}\phi\; D_t\mathbf{r} = 0 \;\Rightarrow\; \nabla\phi\cdot\mathbf{r}'(t) = 0.$$

Hence all curves passing through a point P on the surface have tangent vector normal to $\nabla\phi$ and so they all lie in a common plane called the tangent plane at P.

Tangent planes

Find the tangent plane to the surface

$$x^2 + y^2 + z^2 = 6$$

at the point $(1, 2, -1)$.

The surface is

$$\phi(x, y, z) = 6$$

where $\phi : \mathbb{R}^3 \to \mathbb{R}$ is the differentiable function given by

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

So a normal to the tangent plane at (x, y, z) on the surface is

$$\nabla\phi = (2x, 2y, 2z).$$

At $(1, 2, -1)$ the normal is

$$\nabla\phi(1, 2, -1) = (2, 4, -2)$$

and hence an equation for the tangent plane at $(1, 2, -1)$ is

$$2x + 4y - 2z = 12.$$

Tangent lines

Find the tangent line to the curve

$$3x^2 + 2y^2 = 14$$

at the point $(2, 1)$.

The curve is $\phi(x, y) = 14$ with $\phi : \mathbb{R}^2 \to \mathbb{R}$ a differentiable function given by

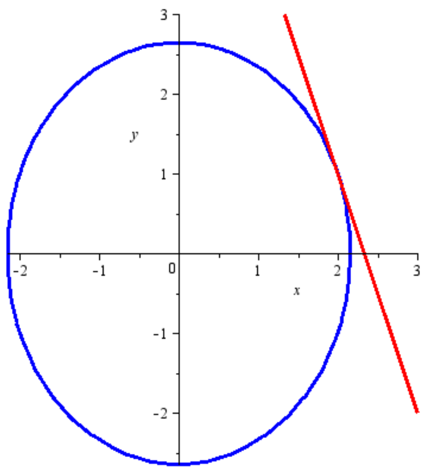
$$\phi(x, y) = 3x^2 + 2y^2.$$

A normal at (x, y) on the curve is $\nabla\phi = (6x, 4y)$ and at $(2, 1)$,

$$\nabla\phi(2, 1) = (12, 4).$$

Hence a Cartesian equation for the tangent line is

$$12x + 4y = 28.$$



Note that we don't need to solve for y to find the tangent line.

[Exercise: check using a 'first year' method with $y = \sqrt{7 - \frac{3}{2}x^2}$.]

Tangent planes

Consider the surface S_1 in $\mathbb{R}^3$ defined by

$$S_1 = \{(x, y, z) : x^3 + 2y^2 - z = 0\}.$$

At the point $(2, 1, 10)$ find

(i) a parametric equation of the normal line and

(ii) a Cartesian equation of the tangent plane.

The surface is the 0 level set of the differentiable function $\phi : \mathbb{R}^3 \to \mathbb{R}$ given by $\phi(x, y, z) = x^3 + 2y^2 - z$.

So a normal to the surface at (x, y, z) is given by $\nabla\phi = (3x^2, 4y^2, -1)$ and at $(2, 1, 10)$ by $\nabla\phi(2, 1, 10) = (12, 4, -1)$.

(i) $\mathbf{r}(t) = (2, 1, 10) + t(12, 4, -1), \quad t \in \mathbb{R}$.

(ii) $12x + 4y - z = 18$.

Tangent planes

Find the best affine approximation to $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x, y) = x^3 + 2y^2$ at the point $(2, 1)$ and compare this with the equation of the tangent plane to S_1.

The partial derivatives of f exist and are continuous everywhere. So f is differentiable and

$$Df = Jf = \begin{pmatrix} 3x^2 & 4y \end{pmatrix} \quad \text{or} \quad \nabla f = (3x^2, 4y).$$

The best affine approximation at $(2, 1)$ is

$$\begin{aligned} T(x, y) &= f(2, 1) + \nabla f(2, 1) \cdot (x - 2, y - 1) \\ &= 10 + (12, 4) \cdot (x - 2, y - 1) \\ &= 10 + 12(x - 2) + 4(y - 1) \\ &= -18 + 12x + 4y. \end{aligned}$$

Note that the graph of T give by $z = T(x, y)$ is

$$z = -18 + 12x + 4y \quad \Rightarrow \quad 12x + 4y - z = 18$$

Tangent planes

Find the curves obtained by the intersection of $S_1 = \{(x, y, z) : x^3 + 2y^2 - z = 0\}$ with the planes (i) $x = 2$, and (ii) $y = 1$.

Find the tangent vectors to these curves at the point $(2, 1, 10)$ and hence give a parametric equation for the tangent plane to S_1 at $(2, 1, 10)$.

(i) $\mathbf{r}_1(t) = (2, t, 8 + 2t^2)$, $\quad \mathbf{r}_1 : \mathbb{R} \to \mathbb{R}^3$.
(ii) $\mathbf{r}_2(t) = (t, 1, t^3 + 2)$, $\quad \mathbf{r}_2 : \mathbb{R} \to \mathbb{R}^3$.

Tangent vectors to the curves are

$$\mathbf{r}_1'(t) = (0, 1, 4t), \qquad \text{and} \qquad \mathbf{r}_2'(t) = (1, 0, 3t^2)$$

and at $(2, 1, 10)$ these are

$$\mathbf{r}_1'(1) = (0, 1, 4), \qquad \text{and} \qquad \mathbf{r}_2'(2) = (1, 0, 12).$$

So the tangent plane is given by

$$\mathbf{r}(s, t) = (2, 1, 10) + t(0, 1, 4) + s(1, 0, 12).$$

Tangent planes

Consider $g : \mathbb{R}^3 \to \mathbb{R}$ with

$$g(x, y, z) = 3x^2 - 24x + 3y^2 - 10y + 3z^2 - 59z + 333$$

and the surface S_2 defined as the 0 level set of g, that is,

$$S_2 = \{(x, y, z) : g(x, y, z) = 0\}.$$

(i) Describe S_2.
(ii) Show that S_2 touches S_1 tangentially at $(2, 1, 10)$.
(iii) Solve $g(x, y, z) = 0$ for z in terms of x and y for (x, y) "near" $(2, 1)$. [That is find $f : \mathbb{R}^2 \to \mathbb{R}$ with $z = f(x, y)$ near $(2, 1)$.]
(iv) Find the best affine approximation to f near $(2, 1)$.
(v) What fact involving ∇g makes it possible to find f?

Tangent planes

(i)

$$g(x,y,z) = 3x^2 - 24x + 3y^2 - 10y + 3z^2 - 59z + 333$$

and

$$S_2 = \{(x,y,z) : 3x^2 - 24x + 3y^2 - 10y + 3z^2 - 59z + 333 = 0\}.$$

Completing the squares x, y and z,

$$g(x,y,z) = 3(x-4)^2 + 3(y-\tfrac{5}{2})^2 + 3(z-\tfrac{59}{6})^2 - \tfrac{143}{6}.$$

So S_2 is implicitly defined by the equation

$$3(x-4)^2 + 3(y-\tfrac{5}{2})^2 + 3(z-\tfrac{59}{6})^2 = \tfrac{143}{6}$$

which is a sphere of radius $\sqrt{\frac{143}{18}}$ centred at $(4, \frac{5}{2}, \frac{59}{6})$.

Tangent planes

(ii)

$$g(x,y,z) = 3x^2 - 24x + 3y^2 - 10y + 3z^2 - 59z + 333$$

and

$$S_2 = \{(x,y,z) : 3x^2 - 24x + 3y^2 - 10y + 3z^2 - 59z + 333 = 0\}.$$

First check that $g(2,1,10) = 0$ so that $(2,1,10)$ lies on S_2.

[We previously found that a normal to the tangent plane of S_1 at $(2,1,10)$ was $\nabla\phi(2,1,10) = (12,4,-1)$.]

Now, a normal to the tangent plane of S_2 is given by

$$\nabla g = \Big(6(x-4), 6(y-\tfrac{5}{3}), 6(z-\tfrac{59}{6})\Big) \quad \Rightarrow \quad \nabla g(2,1,10) = (-12,-4,1).$$

Since one normal is a multiple of the other, the two tangent planes are parallel.

Tangent planes

(iii)

$$g(x,y,z) = 3x^2 - 24x + 3y^2 - 10y + 3z^2 - 59z + 333$$

and

$$S_2 = \{(x,y,z) : 3x^2 - 24x + 3y^2 - 10y + 3z^2 - 59z + 333 = 0\}.$$

$$\begin{aligned} & 3(x-4)^2 + 3(y-\tfrac{5}{2})^2 + 3(z-\tfrac{59}{6})^2 - \tfrac{143}{6} = 0 \\ \Rightarrow \quad & 3(z-\tfrac{59}{6})^2 = \tfrac{143}{6} - 3(x-4)^2 - 3(y-\tfrac{5}{2})^2 \\ \Rightarrow \quad & z = \tfrac{59}{6} + \sqrt{\tfrac{143}{18} - (x-4)^2 - (y-\tfrac{5}{2})^2} \end{aligned}$$

(iv)

The best affine approximation is given by the tangent plane that has already been found.

$$T(x,y) = 10 + 12(x-2) + 4(y-1).$$

Taylor series

Taylor's theorem says for a suitably continuous and differentiable function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = P_{n,a}(x) + R_{n,a}(x)$$

where $P_{n,a}(x)$ is the polynomial

$$P_{n,a}(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \cdots + \frac{1}{n!}f^{(n)}(a)(x-a)^n$$

and the remainder $R_{n,a}(x)$ is

$$R_{n,a}(x) = \frac{1}{(n+1)!}f^{(n+1)}(z)(x-a)^{n+1}$$

for some z between x and a. When $R_{n,a}(x)$ is "small enough",

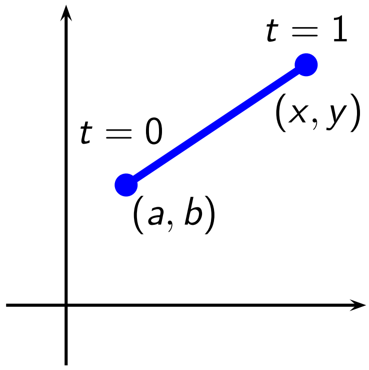
$$f(x) \simeq P_{n,a}(x)$$

and $P_{0,a}(x)$, $P_{1,a}(x)$, $P_{2,a}(x)$, $P_{3,a}(x)$, ... are the the best constant, affine, quadratic, cubic, ... approximations to $f(x)$.

Taylor series

Taylor's theorem can be generalised to $f : \mathbb{R}^n \to \mathbb{R}$.

Consider $f : \mathbb{R}^2 \to \mathbb{R}$ and try to write $f(x, y)$ in terms of f and it's derivatives at (a, b). Let



$$g(t) = f(u, v), \quad u = a + t(x - a), \; v = b + t(y - b).$$

For g continuous on $[0, t]$, Taylor's theorem says

$$g(t) = g(0) + R_0(t) \quad \text{where} \quad R_0(t) = g'(z_0)t$$

for some z_0 between 0 and t provided g is differentiable on $[0, t]$, and

$$g(t) = g(0) + g'(0)t + R_1(t) \quad \text{where} \quad R_1(t) = \frac{1}{2!}g''(z_1)t^2$$

for some z_1 between 0 and t provided g' is differentiable on $[0, t]$, and

$$g(t) = g(0) + g'(0)t + \frac{1}{2!}g''(0)t^2 + R_2(t) \quad \text{where} \quad R_2(t) = \frac{1}{3!}g'''(z_2)t^3$$

for some z_2 between 0 and t provided g'' is differentiable on $[0, t]$, and so on.

Taylor series

$$u = a + t(x - a), \quad v = b + t(y - b) \quad \Rightarrow \quad \frac{du}{dt} = x - a, \; \frac{dv}{dt} = y - b.$$

$$g(t) = f(u, v) \qquad \begin{aligned} g'(t) &= f_1(u, v)\frac{du}{dt} + f_2(u, v)\frac{dv}{dt} \\ &= f_1(u, v)(x - a) + f_2(u, v)(y - b) \end{aligned}$$

$$\begin{aligned} g''(t) &= \frac{d}{dt}\Big(f_1(u, v)(x - a) + f_2(u, v)(y - b)\Big) \\ &= \left(f_{11}(u, v)\frac{du}{dt} + f_{12}(u, v)\frac{dv}{dt}\right)(x - a) \\ &\qquad + \left(f_{21}(u, v)\frac{du}{dt} + f_{22}(u, v)\frac{dv}{dt}\right)(y - b) \\ &= f_{11}(u, v)(x - a)^2 + 2f_{12}(u, v)(x - a)(y - b) + f_{22}(u, v)(y - b)^2 \end{aligned}$$

$$\begin{aligned} g'''(t) = f_{111}(u, v)(x - a)^3 + 3f_{112}(u, v)(x - a)^2(y - b) + 3f_{122}(u, v)(x - a)(y - b)^2 \\ + f_{222}(u, v)(y - b)^3. \end{aligned}$$

Taylor series

$$u = a + t(x-a), \quad v = b + t(y-b) \quad \text{and} \quad g(t) = f(u,v).$$

Recall that the 0^{th} order form of Taylor's theorem (MVT) says, for g continuous on $[0,t]$ and differentiable on $(0,t)$,

$$g(t) = g(0) + R_0(t) \quad \text{where} \quad R_0(t) = g'(z_0)t.$$

Now, using

$$g(t) = f(u,v), \qquad g'(t) = f_1(u,v)(x-a) + f_2(u,v)(y-b)$$

gives the multivariable version

$$\begin{aligned} f(x,y) = g(1) &= P_0(1) + R_0(1) \\ &= f(a,b) + f_1(c_0,d_0)(x-a) + f_2(c_0,d_0)(y-b) \end{aligned}$$

for some (c_0,d_0) on the line segment between (a,b) and (x,y).

$$\Big[(c_0,d_0) = (a + z_0(x-a), b + z_0(y-b))\Big]$$

Taylor series

The 1^{st} order form of Taylor's theorem says, for g' continuous on $[0,t]$ and g' differentiable on $(0,t)$,

$$g(t) = g(0) + g'(0)t + R_1(t) \quad \text{where} \quad R_1(t) = \frac{1}{2!}g''(z_1)t^2.$$

Now, using

$$g(t) = f(u,v), \qquad g'(t) = f_1(u,v)(x-a) + f_2(u,v)(y-b)$$

$$g''(t) = f_{11}(u,v)(x-a)^2 + 2f_{12}(u,v)(x-a)(y-b) + f_{22}(u,v)(y-b)^2$$

gives the multivariable version

$$\begin{aligned} f(x,y) = g(1) &= P_1(1) + R_1(1) \\ &= f(a,b) + f_1(a,b)(x-a) + f_2(a,b)(y-b) \\ &\quad + \frac{1}{2}\Big(f_{11}(c_1,d_1)(x-a)^2 + 2f_{12}(c_1,d_1)(x-a)(y-b) \\ &\qquad + f_{22}(c_1,d_1)(y-b)^2\Big) \end{aligned}$$

for some (c_1,d_1) on the line segment between (a,b) and (x,y).

$$\Big[(c_1,d_1) = (a + z_1(x-a), b + z_1(y-b))\Big]$$

Taylor series

Taylor's theorem says, for g'' continuous on $[0,t]$ and g'' differentiable on $(0,t)$,

$$g(t) = g(0) + g'(0)t + \tfrac{1}{2!}g''(0)t^2 + R_2(t) \quad \text{where} \quad R_2(t) = \tfrac{1}{3!}g'''(z_2)t^3.$$

$$\begin{aligned}
g(t) &= f(u,v), \qquad\qquad g'(t) = f_1(u,v)(x-a) + f_2(u,v)(y-b) \\
g''(t) &= f_{11}(u,v)(x-a)^2 + 2f_{12}(u,v)(x-a)(y-b) + f_{22}(u,v)(y-b)^2 \\
g'''(t) &= f_{111}(u,v)(x-a)^3 + 3f_{112}(u,v)(x-a)^2(y-b) \\
&\qquad + 3f_{122}(u,v)(x-a)(y-b)^2 + f_{222}(u,v)(y-b)^3.
\end{aligned}$$

gives the multivariable version (for some (c_2,d_2) between (a,b) and (x,y)),

$$\begin{aligned}
f(x,y) = g(1) &= P_2(1) + R_2(1) \\
&= f(a,b) + f_1(a,b)(x-a) + f_2(a,b)(y-b) + \frac{1}{2}\Big(f_{11}(a,b)(x-a)^2 \\
&\quad +2f_{12}(a,b)(x-a)(y-b) + f_{22}(a,b)(y-b)^2\Big) \\
&\quad + \frac{1}{3!}\Big(f_{111}(c_2,d_2)(x-a)^3 + 3f_{112}(c_2,d_2)(x-a)^2(y-b) \\
&\qquad + 3f_{122}(c_2,d_2)(x-a)(y-b)^2 + f_{222}(c_2,d_2)(y-b)^3\Big).
\end{aligned}$$

Taylor series

Definition

$f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is C^r on an open set $\Omega \subset \mathbb{R}^n$ if all partial derivatives of f of order $\leq r$ exist and are continuous.

Theorem (Taylor's Theorem)

Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be C^r on the open set Ω. Let $\mathbf{a} \in \Omega$ be such that the line segment joining $\mathbf{a}$ and $\mathbf{x}$ lies in Ω. Then

$$f(\mathbf{x}) = P_{r,\mathbf{a}}(\mathbf{x}) + R_{r,\mathbf{a}}(\mathbf{x})$$

where, for some point $\mathbf{z}$ on the line segment joining $\mathbf{x}$ and $\mathbf{a}$,

$$P_{r,\mathbf{a}}(\mathbf{x}) = f(\mathbf{a}) + \sum_{k=1}^{r-1} \frac{1}{k!} D^k f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a})^k, \qquad R_{r,\mathbf{a}}(\mathbf{x}) = \frac{1}{r!} D^r f(\mathbf{z}) \cdot (\mathbf{x}-\mathbf{a})^r.$$

Note that the $D^r f(\mathbf{z}) \cdot (\mathbf{x}-\mathbf{a})^r$ is not a dot product. It represents the terms that we have found in the last few slides and their generalisations.

Taylor series

Find the Taylor polynomial of order 2 about $\left(1, -\frac{\pi}{2}\right)$ for $f(x,y) = \sin(x^2y)$.

$$\begin{array}{rcl|c} & & & \text{at } \left(1, -\frac{\pi}{2}\right) \\ \hline f(x,y) & = & \sin(x^2y) & -1 \\ f_x(x,y) & = & 2xy\cos(x^2y) & 0 \\ f_y(x,y) & = & x^2\cos(x^2y) & 0 \\ f_{xx}(x,y) & = & 2y\cos(x^2y) - 4x^2y^2\sin(x^2y) & \pi^2 \\ f_{xy}(x,y) & = & 2x\cos(x^2y) - 2x^3y\sin(x^2y) & -\pi \\ f_{yy}(x,y) & = & -x^4\sin(x^2y) & 1 \end{array}$$

$$\begin{aligned} P_{2,(1,-\frac{\pi}{2})}(x,y) &= -1 + 0(x-1) + 0\left(y - \left(-\frac{\pi}{2}\right)\right) + \frac{1}{2}\Bigg(\pi^2(x-1)^2 \\ &\quad +2(-\pi)(x-1)\left(y - \left(-\frac{\pi}{2}\right)\right) + \left(y - \left(-\frac{\pi}{2}\right)\right)^2\Bigg) \\ &= -1 + \frac{\pi^2}{2}(x-1)^2 - \pi(x-1)\left(y - \left(-\frac{\pi}{2}\right)\right) + \frac{1}{2}\left(y - \left(-\frac{\pi}{2}\right)\right)^2. \end{aligned}$$

Taylor series

Find the Taylor polynomial of order 2 about $(4,8)$ for $f(x,y) = \sqrt{x}\sqrt[3]{y}$.

$$\begin{array}{rcl|c} & & & \text{at } (4,8) \\ \hline f(x,y) & = & x^{\frac{1}{2}}y^{\frac{1}{3}} & 4 \\ f_x(x,y) & = & \frac{1}{2}x^{-\frac{1}{2}}y^{\frac{1}{3}} & \frac{1}{2} \\ f_y(x,y) & = & \frac{1}{3}x^{\frac{1}{2}}y^{-\frac{2}{3}} & \frac{1}{6} \\ f_{xx}(x,y) & = & -\frac{1}{4}x^{-\frac{3}{2}}y^{\frac{1}{3}} & -\frac{1}{16} \\ f_{xy}(x,y) & = & \frac{1}{6}x^{-\frac{1}{2}}y^{-\frac{2}{3}} & \frac{1}{48} \\ f_{yy}(x,y) & = & -\frac{2}{9}x^{\frac{1}{2}}y^{-\frac{5}{3}} & -\frac{1}{72} \end{array}$$

$$\begin{aligned} P_{2,(4,8)}(x,y) &= 4 + \frac{1}{2}(x-4) + \frac{1}{6}(y-8) \\ &\quad + \frac{1}{2}\left(-\frac{1}{16}(x-4)^2 + 2.\frac{1}{48}(x-4)(y-8) + \left(-\frac{1}{72}\right)(y-8)^2\right). \end{aligned}$$

Taylor series

Use Taylor polynomials for $\sqrt{x}\sqrt[3]{y}$ about the point $(4,8)$ to approximate to $\sqrt{3.98}\sqrt[3]{8.03}$ using

(i) the constant and linear terms, and

(ii) terms up to second order.

(i)
$$\begin{aligned} f(3.98, 8.03) &\simeq P_{1,(4,8)}(3.98, 8.03) \\ &= 4 + \frac{1}{2}(3.98-4) + \frac{1}{6}(8.03-8) \\ &= 3.995 \end{aligned}$$

(ii)
$$\begin{aligned} f(3.98, 8.03) &\simeq P_{2,(4,8)}(3.98, 8.03) \\ &= 4 + \frac{1}{2}(3.98-4) + \frac{1}{6}(8.03-8) + \frac{1}{2}\left(-\frac{1}{16}(3.98-4)^2\right. \\ &\qquad \left. + 2 \times \frac{1}{48}(3.98-4)(8.03-8) + \left(-\frac{1}{72}\right)(8.03-8)^2\right) \\ &= 3.99496875 \end{aligned}$$

[Maple gives $3.99496873\ldots$]

Taylor series

Find the Taylor polynomial of

$$f(x,y) = \sin xe^{y/2}$$

including terms up to order 3 about $(0,0)$.

$$\begin{aligned} \sin xe^{y/2} &= \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots\right)\left(1 + \frac{y}{2} + \frac{\left(\frac{y}{2}\right)^2}{2!} + \frac{\left(\frac{y}{2}\right)^3}{3!} + \frac{\left(\frac{y}{2}\right)^4}{4!} + \cdots\right) \\ &= x + \frac{xy}{2} - \frac{x^3}{6} + \frac{xy^2}{8} + \cdots \end{aligned}$$

So,

$$P_{3,(0,0)}(x,y) = x + \frac{xy}{2} - \frac{x^3}{6} + \frac{xy^2}{8}.$$

Maxima, minima and saddle points

Definition

Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. Then

- $\mathbf{a} \in \Omega$ is an absolute or global maximum of f if $f(\mathbf{a}) \geq f(\mathbf{x})$ for all $\mathbf{x} \in \Omega$.
- $\mathbf{a} \in \Omega$ is an absolute or global minimum of f if $f(\mathbf{a}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \Omega$.
- $\mathbf{a} \in \Omega$ is a local maximum of f if there is an open set A containing $\mathbf{a}$ such that $f(\mathbf{a}) \geq f(\mathbf{x})$ for all $\mathbf{x} \in \Omega \cap A$.
- $\mathbf{a} \in \Omega$ is a local minimum of f if there is an open set A containing $\mathbf{a}$ such that $f(\mathbf{a}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \Omega \cap A$.
- $\mathbf{a} \in \Omega$ is a stationary point of f if f is differentiable at $\mathbf{a}$ and $\nabla f(\mathbf{a}) = \mathbf{0}$.
- $\mathbf{a} \in \Omega$ is a saddle point of f if it is a stationary point of f but is neither a local maximum nor minimum point of f.

Maxima, minima and saddle points

Theorem

Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. Then local and maxima and minima can only occur at $\mathbf{a} \in \Omega$ where $\mathbf{a}$ satisfies one of the following:

(1) $\mathbf{a}$ *is a stationary point,*

(2) $\mathbf{a}$ *lies on the boundary of Ω or*

(3) f *is not differentiable at $\mathbf{a}$.*

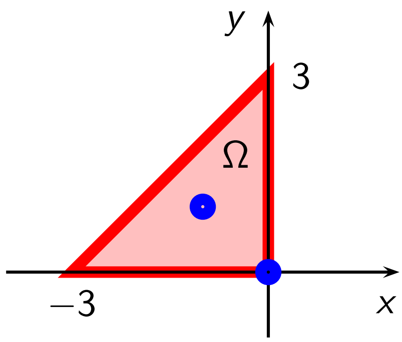
Definition

Points satisfying at least one of (1), (2) or (3) in the theorem above are called critical points.

Maxima, minima and saddle points

Consider Ω, the region of $\mathbb{R}^2$ bounded by $x = 0$, $y = 0$ and $y = x + 3$. Find the maximum and minimum values of $f : \Omega \to \mathbb{R}$, given by,

$$f(x, y) = x^3 - y^3 - 3xy.$$



f is continuous and differentiable on Ω which is compact. Hence $f(\Omega)$ is compact and so maximum and minimum values exist and are attained on Ω.

Since f is differentiable everywhere, the extrema must occur at (1) stationary points f or (2) boundary points of Ω.

Stationary points of f occur when

$$\nabla f = \mathbf{0} \;\Leftrightarrow\; (3x^2 - 3y, -3y^2 - 3x) = (0, 0) \;\Leftrightarrow\; y = x^2 \text{ and } x = -y^2$$

$$\Rightarrow\; y = x^2 \;\Rightarrow\; x^4 + x = 0 \;\Rightarrow\; (x^3 + 1)x = 0.$$

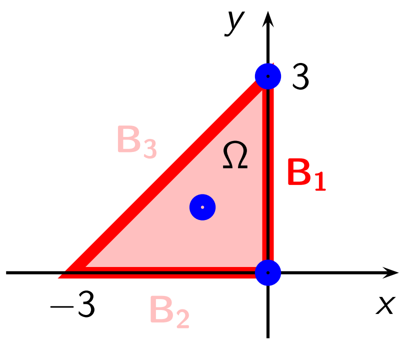
Hence the only stationary points of f are $(0, 0)$ and $(-1, 1)$. Also note that

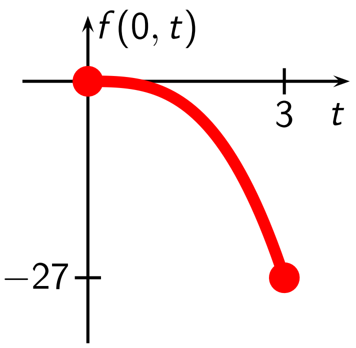
$$f(0, 0) = 0 \quad \text{and} \quad f(-1, 1) = 1.$$

Maxima, minima and saddle points

Divide the boundary into 3 pieces. First consider $\mathbf{B_1}$.

$$\begin{aligned} \mathbf{B_1} &= \{(0, t) : 0 \le t \le 3\}, \\ \mathbf{B_2} &= \{(t, 0) : -3 \le t \le 0\}, \\ \mathbf{B_3} &= \{(t, t+3) : -3 \le t \le 0\}. \end{aligned}$$





On $\mathbf{B_1}$

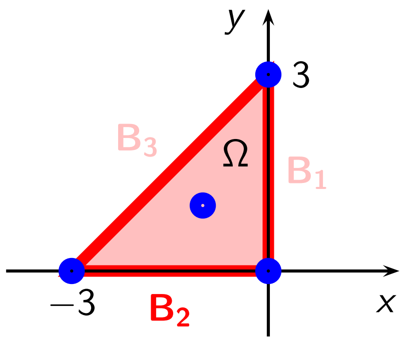
$$f(0, t) = 0^3 - t^3 - 0 = -t^3$$

for $t \in [0, 3]$.

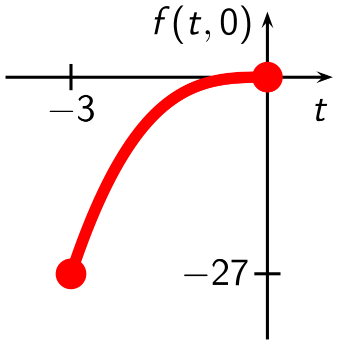
So the max on $\mathbf{B_1}$ is at $(0, 0)$ where $f(0, 0) = 0$ and the min is at $(0, 3)$ where $f(0, 3) = -27$.

Maxima, minima and saddle points

Next consider $\mathbf{B_2}$.



$$
\begin{aligned}
\mathbf{B_1} &= \{(0,t) : 0 \le t \le 3\},\\
\mathbf{B_2} &= \{(t,0) : -3 \le t \le 0\},\\
\mathbf{B_3} &= \{(t,t+3) : -3 \le t \le 0\}.
\end{aligned}
$$



On $\mathbf{B_2}$

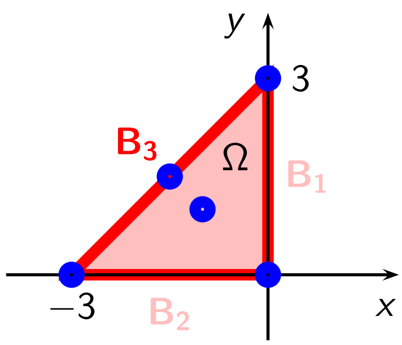
$$f(t,0) = t^3 - 0^3 - 0 = t^3$$

for $t \in [-3,0]$.

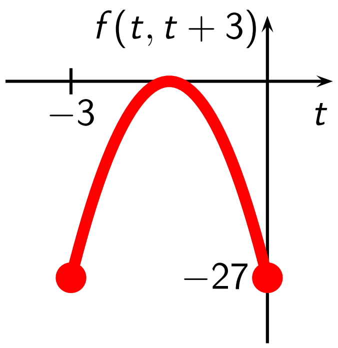
So the max on $\mathbf{B_2}$ is at $(0,0)$ where $f(0,0) = 0$ and the min is at $(-3,0)$ where $f(-3,0) = -27$.

Maxima, minima and saddle points

Lastly consider $\mathbf{B_3}$.



$$
\begin{aligned}
\mathbf{B_1} &= \{(0,t) : 0 \le t \le 3\},\\
\mathbf{B_2} &= \{(t,0) : -3 \le t \le 0\},\\
\mathbf{B_3} &= \{(t,t+3) : -3 \le t \le 0\}.
\end{aligned}
$$



On $\mathbf{B_3}$

$$f(t,t+3) = t^3 - (t+3)^3 - 3t(t+3) = -3(4t^2 + 12t + 9)$$

for $t \in [-3,0]$. Now, $g(t) = f(t,t+3)$ has a stationary point when

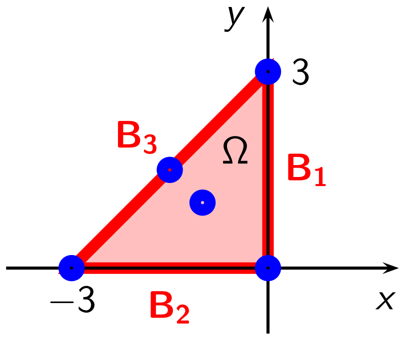
$$8t + 12 = 0 \;\Rightarrow t = -\tfrac{3}{2}.$$

Extreme values can occur on $\mathbf{B_3}$ at the end points (already considered) or the stationary point where

$$f\left(-\tfrac{3}{2}, \tfrac{3}{2}\right) = 0.$$

Maxima, minima and saddle points

So we have a number of candidate points for the extreme values of f.



$$\begin{aligned} f(-1,1) &= 1 \\ f(0,0) &= 0 \\ f(0,3) &= -27 \\ f(-3,0) &= -27 \\ f(-1.5,1.5) &= 0 \end{aligned}$$

Hence the maximum of f on Ω is 1 and the minimum value of f on Ω is -27.

Classification of stationary points

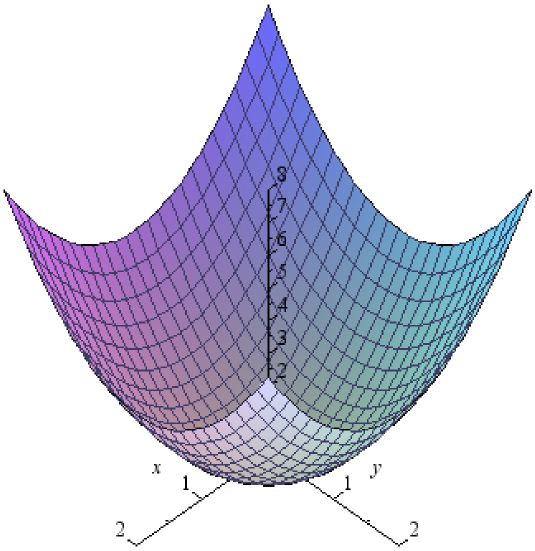
The following functions $f : \mathbb{R}^2 \to \mathbb{R}$ have a stationary point at $(0,0)$.

Is it a local maximum, minimum or saddle point?

(i) $f(x,y) = x^2 + y^2$
(ii) $f(x,y) = -x^2 - y^2$
(iii) $f(x,y) = x^2 - y^2$
(iv) $f(x,y) = xy$
(v) $f(x,y) = x^2 + y^4$
(vi) $f(x,y) = x^2 - y^4$
(vii) $f(x,y) = x^2 - 6xy + y^2$
(viii) $f(x,y) = 3x^2 - 2xy + 3y^2$

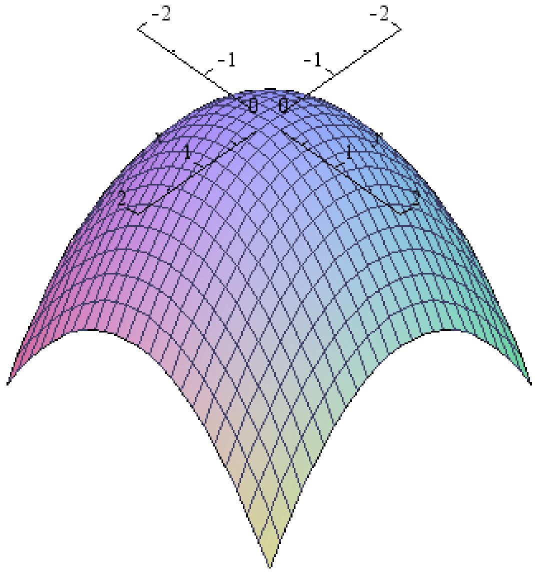
Classification of stationary points

(i) $f(x, y) = x^2 + y^2$



Local minimum at $(0, 0)$.

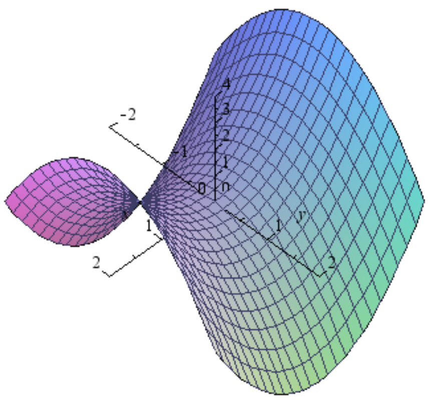
(ii) $f(x, y) = -x^2 - y^2$



Local maximum at $(0, 0)$.

Classification of stationary points

(iii) $f(x, y) = x^2 - y^2$



Along $y = 0$, $f(x, 0) = x^2$ and $(0, 0)$ is a local minimum.

Along $x = 0$, $f(0, y) = -y^2$ and $(0, 0)$ is a local maximum.

For all $\epsilon > 0$,

$$\left(\frac{\epsilon}{2}, 0\right) \in B((0,0), \epsilon) \text{ with } f\left(\frac{\epsilon}{2}, 0\right) = \frac{\epsilon^2}{4}$$

and

$$\left(0, \frac{\epsilon}{2}\right) \in B((0,0), \epsilon) \text{ with } f\left(0, \frac{\epsilon}{2}\right) = -\frac{\epsilon^2}{4}.$$

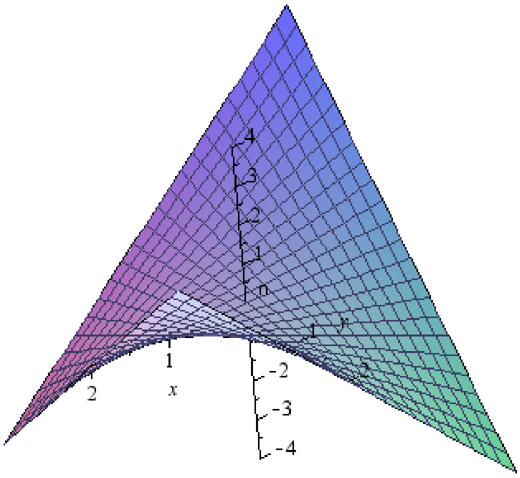
So,

$$f\left(0, \frac{\epsilon}{2}\right) < f(0,0) < f\left(\frac{\epsilon}{2}, 0\right).$$

That is, $(0, 0)$ is a stationary point that is neither a local max nor min and hence is a saddle point.

Classification of stationary points

(iv) $f(x,y) = xy$



Along $y = x$,

$$f(x,x) = x^2$$

which has a local minimum at $(0,0)$.

Along $y = -x$,

$$f(x,-x) = -x^2$$

which has a local maximum at $(0,0)$.

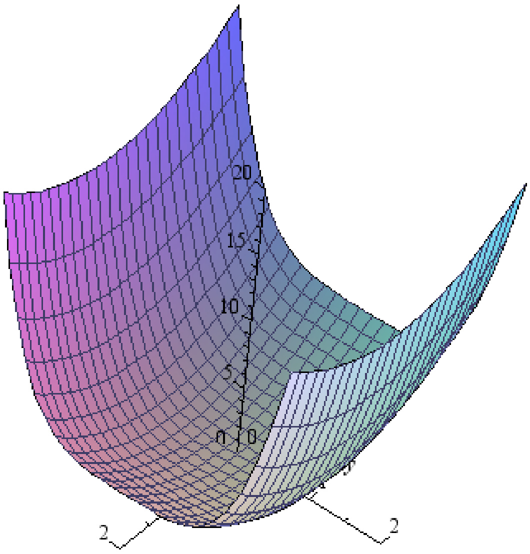
So $(0,0)$ is neither a local maximum nor local minimum. Hence f has a saddle point at $(0,0)$.

Note that

$$f(x,y) = \frac{1}{4}\Big((x+y)^2 - (x-y)^2\Big).$$

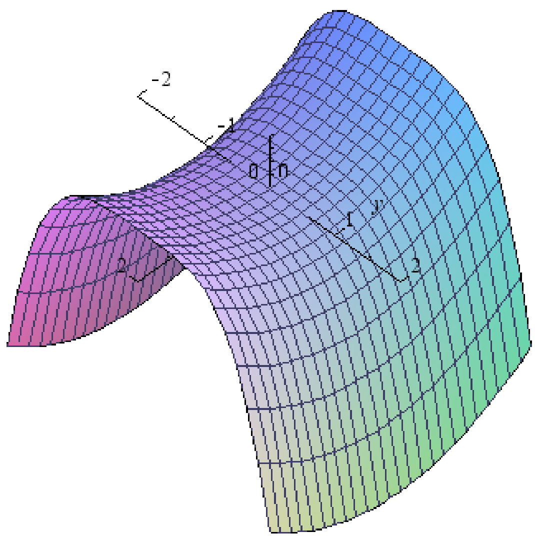
Classification of stationary points

(v) $f(x,y) = x^2 + y^4$



Local minimum at $(0,0)$.

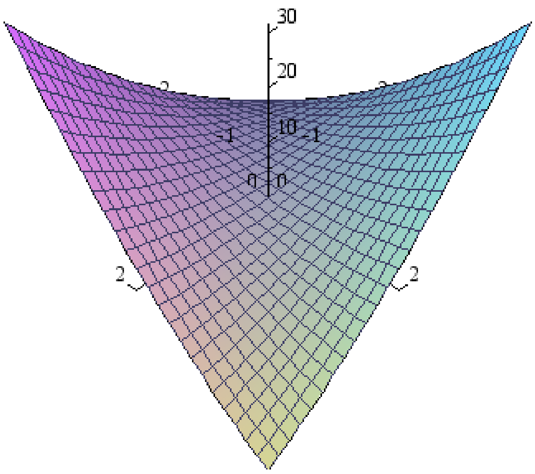
(iv) $f(x,y) = x^2 - y^4$



Saddle point at $(0,0)$.

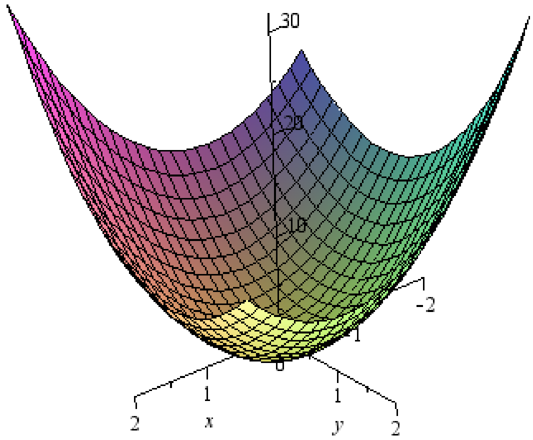
Classification of stationary points

(vii) $f(x,y) = x^2 - 6xy + y^2$



Saddle point at $(0,0)$.

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



Local minimum at $(0,0)$.

Classification of stationary points

(vii)

$$f(x,y) = x^2 - 6xy + y^2$$
$$= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -3 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Let

$$H = \begin{pmatrix} 1 & -3 \\ -3 & 1 \end{pmatrix}.$$

H has eigenvalues and eigenvectors

$$\lambda_1 = -2, \qquad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$
$$\lambda_2 = 4, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

So we can orthoganally diagonalise H.

Let

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and then

$$P^{-1} = P^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

So

$$P^T H P = D = \begin{pmatrix} -2 & 0 \\ 0 & 4 \end{pmatrix}.$$

Now make a change of variables

$$\begin{pmatrix} x \\ y \end{pmatrix} = P \begin{pmatrix} X \\ Y \end{pmatrix}.$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = P \begin{pmatrix} X \\ Y \end{pmatrix}. \quad \Rightarrow \quad (x \quad y) = (X \quad Y)\, P^T.$$

So,

$$\begin{aligned} f(x,y) = (x \quad y)\, H \begin{pmatrix} x \\ y \end{pmatrix} &= (X \quad Y)\, P^T H P \begin{pmatrix} X \\ Y \end{pmatrix} \\ &= (X \quad Y) \begin{pmatrix} -2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \\ &= -2X^2 + 4Y^2 \end{aligned}$$

Note that

$$\begin{pmatrix} X \\ Y \end{pmatrix} = P^T \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(x+y) \\ \frac{1}{\sqrt{2}}(y-x) \end{pmatrix}.$$

So,

$$f(x,y) = -2\left(\frac{1}{\sqrt{2}}(x+y)\right)^2 + 4\left(\frac{1}{\sqrt{2}}(y-x)\right)^2 = -(x+y)^2 + 2(x-y)^2.$$

(viii)

$$f(x,y) = (x \quad y) \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

The eigenvalues and eigenvectors of

$$H = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$$

are

$$\lambda_1 = 2, \qquad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$\lambda_2 = 4, \qquad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Diagonalising and rotating the coordinates leads to

$$f(x,y) = 2X^2 + 4Y^2 = (x+y)^2 + 2(x-y)^2.$$

Classification of stationary points

The 'Taylor series' of f at a stationary point (a, b) is

$$\begin{aligned} f(x,y) &= f(a,b) + \cancel{\nabla f(a,b)\cdot\big((x,y)-(a,b)\big)} \\ &\quad + \frac{1}{2!}\begin{pmatrix} x-a & y-b \end{pmatrix}\begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a,b) & \frac{\partial^2 f}{\partial y \partial x}(a,b) \\ \frac{\partial^2 f}{\partial x \partial y}(a,b) & \frac{\partial^2 f}{\partial y^2}(a,b) \end{pmatrix}\begin{pmatrix} x-a \\ y-b \end{pmatrix} \\ &\quad + \cdots \big(\text{terms involving higher powers of } (x-a) \text{ and } (y-b)\big) \end{aligned}$$

since $\nabla f(a, b) = (0, 0)$.

For (x, y) close to (a, b) the nature of the stationary point will be determined by the eigenvalues of the matrix

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

Classification of stationary points

Suppose $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ is C^2 and has a stationary point at (a, b), that is, $\nabla f(a, b) = 0$. So Taylor's theorem says that

$$f(x,y) = f(a,b) + R_{1,(a,b)}(x,y)$$

where the remainder term is given by

$$R_{1,(a,b)}(x,y) = \frac{1}{2!}\begin{pmatrix} x-a & y-b \end{pmatrix} H \begin{pmatrix} x-a \\ y-b \end{pmatrix}$$

$$\text{where} \qquad H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(c,d) & \frac{\partial^2 f}{\partial y \partial x}(c,d) \\ \frac{\partial^2 f}{\partial x \partial y}(c,d) & \frac{\partial^2 f}{\partial y^2}(c,d) \end{pmatrix}$$

for some point (c, d) between (a, b) and (x, y).

Can the eigenvalues of H be used to determine whether f has a local max, min or saddle point at (a, b)? H is made of partial derivatives evaluated at an unknown point (c, d). Can we determine the nature of the stationary point using partial derivatives calculated at (a, b)? Yes, on a sufficiently small ball. Why?

Maxima, minima and saddle points

Definition

For $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ the Hessian of f at $\mathbf{a}$ is the $n \times n$ matrix

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

Classification of stationary points

The signs of the eigenvalues of

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

can be determined from the signs of the trace[2] and determinant of $H(f, (a, b))$.

$$\mathrm{Tr}\Big(H(f, (a, b))\Big) = \text{ sum of eigenvalues}$$

and

$$\det\Big(H(f, (a, b))\Big) = \text{ product of eigenvalues.}$$

These are continuous functions of the entries in the matrix which are continuous by the assumption that f is C^2. Hence there must be a open ball around (a, b) on which the trace and determinant (and hence the eigenvalues) of the Hessian have the same signs as those of the Hessian at (a, b).

[2] The trace of a matrix is the sum of its diagonal entries.

Maxima, minima and saddle points

Find the eigenvalues of the Hessian of f at $(0,0)$ for each of the functions we considered last lecture.

$$H(f,(0,0)) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(0,0) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(0,0) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(0,0) & \frac{\partial^2 f}{\partial x_2^2}(0,0) \end{pmatrix}.$$

(i) $f(x,y) = x^2 + y^2$

$$H(f,(0,0)) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Eigenvalues are $2, 2$.

(ii) $f(x,y) = -x^2 - y^2$

$$H(f,(0,0)) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Eigenvalues are $-2, -2$.

Maxima, minima and saddle points

(iii) $f(x,y) = x^2 - y^2$

$$H(f,(0,0)) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Eigenvalues are $2, -2$.

(iv) $f(x,y) = xy$

$$H(f,(0,0)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Eigenvalues are $1, -1$.

(v) $f(x,y) = x^2 + y^4$

$$H(f,(0,0)) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Eigenvalues are $2, 0$.

(vi) $f(x,y) = x^2 - y^4$

$$H(f,(0,0)) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Eigenvalues are $2, 0$.

(vii) $f(x,y) = x^2 - 6xy + y^2$

$$H(f,(0,0)) = \begin{pmatrix} 2 & -6 \\ -6 & 2 \end{pmatrix}.$$

Eigenvalues are $-4, 8$.

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

$$H(f,(0,0)) = \begin{pmatrix} 6 & -2 \\ -2 & 6 \end{pmatrix}.$$

Eigenvalues are $4, 8$.

Classification of stationary points

Definition

An $n \times n$ matrix H is

$$\begin{array}{lcl} \text{positive definite} & \Leftrightarrow & \text{all eigenvalues are} > 0 \\ \text{positive semidefinite} & \Leftrightarrow & \text{all eigenvalues are} \geq 0 \\ \text{negative definite} & \Leftrightarrow & \text{all eigenvalues are} < 0 \\ \text{negative semidefinite} & \Leftrightarrow & \text{all eigenvalues are} \leq 0 \end{array}$$

Theorem (Alternative test — Sylvester's criterion)

If H_k is the upper left $k \times k$ submatrix of H and $\triangle_k = \det H_k$ then H is

$$\begin{array}{lcl} \textit{positive definite} & \Leftrightarrow & \triangle_k > 0 \textit{ for all } k \\ \textit{positive semidefinite} & \Rightarrow & \triangle_k \geq 0 \textit{ for all } k \\ \textit{negative definite} & \Leftrightarrow & \triangle_k < 0 \textit{ for all odd } k \textit{ and} \\ & & \triangle_k > 0 \textit{ for all even } k \\ \textit{negative semidefinite} & \Rightarrow & \triangle_k \leq 0 \textit{ for all odd } k \textit{ and} \\ & & \triangle_k \geq 0 \textit{ for all even } k \end{array}$$

Classification of stationary points

Theorem

Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is C^2 and $\nabla f(\mathbf{a}) = \mathbf{0}$ at an interior point $\mathbf{a}$ of Ω. Then

- $H(f, \mathbf{a})$ *is positive definite* $\Rightarrow$ f *has a local minimum at* $\mathbf{a}$.
- $H(f, \mathbf{a})$ *is negative definite* $\Rightarrow$ f *has a local maximum at* $\mathbf{a}$.
- f *has a local minimum at* $\mathbf{a}$ $\Rightarrow$ $H(f, \mathbf{a})$ *is positive semidefinite.*
- f *has a local maximum at* $\mathbf{a}$ $\Rightarrow$ $H(f, \mathbf{a})$ *is negative semidefinite.*

Classification of stationary points

For $f : \mathbb{R}^2 \to \mathbb{R}$ with a stationary point at (a, b),

$$\triangle_1 = \frac{\partial^2 f}{\partial x^2}(a,b) \quad \text{and} \quad \triangle_2 = \frac{\partial^2 f}{\partial x^2}(a,b)\frac{\partial^2 f}{\partial y^2}(a,b) - \left(\frac{\partial^2 f}{\partial x \partial y}(a,b)\right)^2.$$

Then

- $\triangle_1 > 0$ and $\triangle_2 > 0$ (two positive eigenvalues) $\Rightarrow (a, b)$ is a local minimum.
- $\triangle_1 < 0$ and $\triangle_2 > 0$ (two negative eigenvalues) $\Rightarrow (a, b)$ is a local maximum.
- local minimum at $(a, b) \Rightarrow \triangle_1 \geq 0$ and $\triangle_2 \geq 0$ (no negative eigenvalues).
- local maximum at $(a, b) \Rightarrow \triangle_1 \leq 0$ and $\triangle_2 \geq 0$ (no positive eigenvalues).

Notes:

- $\triangle_2 < 0 \Rightarrow (a, b)$ is a saddle point (one positive and one negative eigenvalue).
- The semidefinite case can also be a saddle point.

Classification of stationary points

Find and classify the stationary points of

$$f(x, y) = x^3 + 6x^2 + 3y^2 - 12xy + 9x.$$

Stationary points occur when $\nabla f = \mathbf{0}$, that is,

$$(3x^2 + 12x - 12y + 9, 6y - 12x) = (0, 0)$$

$$\Rightarrow \begin{cases} 3x^2 + 12x - 12y + 9 = 0 & (1) \\ 6y - 12x = 0 & (2) \end{cases}$$

(2) $\Rightarrow$ $y = 2x$ which when substituted into (1) becomes

$$3(x - 3)(x - 1) = 0.$$

So $x = 1 \Rightarrow y = 2$ or $x = 3 \Rightarrow y = 6$.

So f has stationary points at $(1, 2)$ and $(3, 6)$.

Classification of stationary points

$$f(x,y) = x^3 + 6x^2 + 3y^2 - 12xy + 9x. \quad \Rightarrow \quad H(f,(x,y)) = \begin{pmatrix} 6x+12 & -12 \\ -12 & 6 \end{pmatrix}.$$

At $(1,2)$:

$$H(f,(1,2)) = \begin{pmatrix} 18 & -12 \\ -12 & 6 \end{pmatrix}$$

$$\begin{aligned} \triangle_2 &= 18 \times 6 - (-12) \times (-12) \\ &= -36 < 0. \end{aligned}$$

So $(1,2)$ is a saddle point of f.

At $(3,6)$:

$$H(f,(3,6)) = \begin{pmatrix} 30 & -12 \\ -12 & 6 \end{pmatrix}$$

$$\begin{aligned} \triangle_1 &= 30 > 0, \\ \triangle_2 &= 30 \times 6 - (-12) \times (-12) \\ &= 36 > 0. \end{aligned}$$

So $(3,6)$ is a local minimum point of f.

Classification of stationary points

Find and classify the stationary points of

$$f(x,y,z) = yx^2 + zy^2 + z^2 - 2yx - 2zy + y - z.$$

$$\nabla f = \mathbf{0} \ \Rightarrow \ \begin{cases} 2xy - 2y = 0 & (1) \\ x^2 + 2zy - 2x - 2z + 1 = 0 & (2) \\ y^2 + 2z - 2y - 1 = 0 & (3) \end{cases}$$

(1) is $2y(x-1) = 0$ so there are two cases

$y = 0$:
(3) $\Rightarrow z = \frac{1}{2}$.
(2) $\Rightarrow x = 0$ or $x = 2$.

So $(0,0,\frac{1}{2})$ and $(2,0,\frac{1}{2})$ are stationary points.

$x = 1$:
(2) $\Rightarrow z = 0$ or $y = 1$.
For $z = 0$, (3) $\Rightarrow y = 1 \pm \sqrt{2}$.
For $y = 1$, (3) $\Rightarrow z = 1$.

So, $(1, 1 \pm \sqrt{2}, 0)$ and $(1,1,1)$ are stationary points.

f has 5 stationary points: $(0,0,\frac{1}{2})$, $(2,0,\frac{1}{2})$, $(1,1+\sqrt{2},0)$, $(1,1-\sqrt{2},0)$, $(1,1,1)$.

Classification of stationary points

To classify we need $H(f,(x,y)) = \begin{pmatrix} 2y & 2x-2 & 0 \\ 2x-2 & 2z & 2y-2 \\ 0 & 2y-2 & 2 \end{pmatrix}$.

$$H(f,(0,0,\tfrac{1}{2})) = \begin{pmatrix} 0 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & -2 & 2 \end{pmatrix}$$

$$\triangle_1 = 0, \quad \triangle_2 = \begin{vmatrix} 0 & -2 \\ -2 & 1 \end{vmatrix} = -4$$

$$\triangle_3 = \begin{vmatrix} 0 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & -2 & 2 \end{vmatrix} = -8$$

$(0,0,\frac{1}{2})$ is a saddle point as the Hessian is neither positive semidefinite nor negative semidefinite.
[Eigenvalues are 1, −2, 4.]

$$H(f,(1,1,1)) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\triangle_1 = 2$$
$$\triangle_2 = 4$$
$$\triangle_3 = 8$$

$(1,1,1)$ is a local minimum point as the Hessian is positive definite.
[Eigenvalues are 2, 2, 2.]

Classification of stationary points

$$H(f,(2,0,\tfrac{1}{2})) = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 1 & -2 \\ 0 & -2 & 2 \end{pmatrix}$$

$$\triangle_1 = 0$$
$$\triangle_2 = -4$$
$$\triangle_3 = -8$$

$(2,0,\frac{1}{2})$ is a saddle point as the Hessian is neither positive semidefinite nor negative semidefinite.

[E'values are 1, −2, 4.]

$$H(f,(1,1+\sqrt{2},0)) = \begin{pmatrix} 2+2\sqrt{2} & 0 & 0 \\ 0 & 0 & 2\sqrt{2} \\ 0 & 2\sqrt{2} & 2 \end{pmatrix}$$

$$\triangle_1 = 2+2\sqrt{2}$$
$$\triangle_2 = 0$$
$$\triangle_3 = -16-16\sqrt{2}$$

$(1,1+\sqrt{2},0)$ is a saddle point as the Hessian is neither positive semidefinite nor negative semidefinite.

[E'values are −2, 4, $2+2\sqrt{2}$.]

$$H(f,(1,1-\sqrt{2},0)) = \begin{pmatrix} 2-2\sqrt{2} & 0 & 0 \\ 0 & 0 & -2\sqrt{2} \\ 0 & -2\sqrt{2} & 2 \end{pmatrix}$$

$$\triangle_1 = 2-2\sqrt{2}$$
$$\triangle_2 = 0$$
$$\triangle_3 = -16+16\sqrt{2}$$

$(1,1-\sqrt{2},0)$ is a saddle point as the Hessian is neither positive semidefinite nor negative semidefinite.

[E'values are −2, 4, $2-2\sqrt{2}$.]

Lagrange multipliers

We wish to find the extreme values of a function subject to a constraint (or constraints).

We want to solve problems like:

(a) Find the extreme values of $2x + 3y$ subject to the constraint $x^2 + y^2 = 4$.
(b) Find the minimum value of $x^2 + y^2$ subject to the constraint $2x + 3y = 20$.
(c) Find the minimum value of $x^2 + y^2$ subject to the constraint $xy = 16$.

In the first case, the set of points satisfying the constraint

$$\Omega = \{(x, y) : x^2 + y^2 = 4\}$$

is compact and the function we are applying to those points

$$f(x, y) = 2x + 3y$$

is continuous. So we are guaranteed that $f(\Omega)$ has extreme values.

For the other two cases the existence of a minimum value may need to be considered on a case by case basis.

We will attempt to find candidate points for the extreme values.

Lagrange multipliers

Consider two differentiable functions

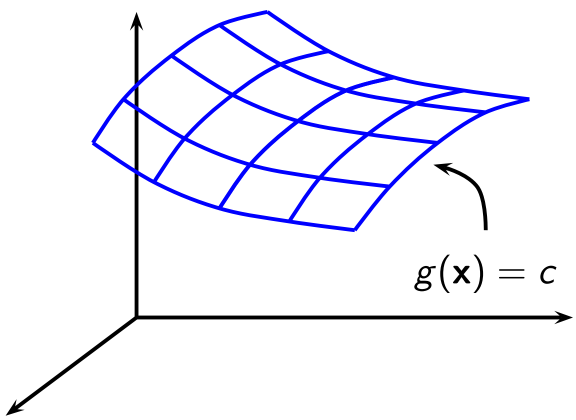
$$f : \mathbb{R}^n \to \mathbb{R} \quad \text{and} \quad g : \mathbb{R}^n \to \mathbb{R}$$

and try to find extreme values of f subject to the constraint

$$g(\mathbf{x}) = c$$

for some constant c.

What is the maximum or minimum value of f on this surface?



Lagrange multipliers

For differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ look for points where f has a maximum or minimum value on the hypersurface

$$S = \{\mathbf{x} \in \mathbb{R}^n \ : \ g(\mathbf{x}) = c\}.$$

Let $\mathbf{r} : I \subset \mathbb{R} \to \mathbb{R}^n$ be a curve in the hypersurface S, that is

$$g(r_1(t), r_2(t), \ldots, r_n(t)) = c, \quad \text{that is} \quad (g \circ \mathbf{r})(t) = c.$$

Points that maximise or minimise f on S should also maximise or minimise f on any curve passing through those points. So we look for stationary points of $h = f \circ \mathbf{r}$.

$$h'(t) = 0 \ \Rightarrow \ D(f \circ \mathbf{r})(t) = 0 \ \Rightarrow \ \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) = 0.$$

We want this condition to hold for all curves through the candidate point and hence ∇f must be normal to the tangent plane to S. That is, provided $\nabla g \neq \mathbf{0}$, there must be a scalar function λ (Lagrange multiplier) such that

$$\nabla f = \lambda \nabla g.$$

Lagrange multipliers

Theorem

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are differentiable and

$$S = \{\mathbf{x} \in \mathbb{R}^n \ : \ g(\mathbf{x}) = c\}$$

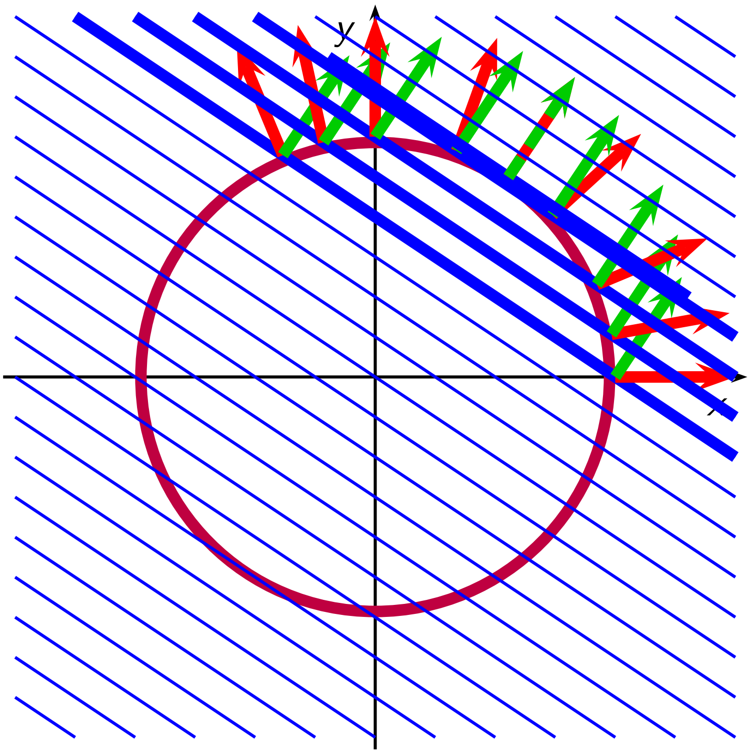
defines a smooth surface in $\mathbb{R}^n$. If a local maximum or minimum value of f on S occurs at $\mathbf{a}$ then $\nabla f(\mathbf{a})$ and $\nabla g(\mathbf{a})$ are parallel. Thus if $\nabla g(\mathbf{a}) \neq \mathbf{0}$, then there exists $\lambda \in \mathbb{R}$ such that

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

Note that this theorem only gives us candidate points for where to look for maxima and minima. There is no guarantee that a maximum or minimum of f on S exists.

Lagrange multipliers

Find the maximum and minimum values of $2x + 3y$ subject to $x^2 + y^2 = 4$.



The constraint $x^2 + y^2 = 4$ is purple. Some contours of $2x + 3y$ are blue.

$\mathbf{2x + 3y = 7.2111\ldots}$
$\mathbf{2x + 3y = 7}$
$\mathbf{2x + 3y = 6}$
$\mathbf{2x + 3y = 5}$
$\mathbf{2x + 3y = 4}$

The constraint set is compact and f is continuous. Hence f attains a maximum and minimum value on the constraint set.

Lagrange multipliers

Extreme values of $f(x, y) = 2x + 3y$ subject to $g(x, y) = x^2 + y^2 = 4$ occur when

$$\nabla f = \lambda \nabla g \Rightarrow (2,\ 3) = \lambda(2x,\ 2y).$$

So,

$$\left.\begin{aligned} 2 &= 2x\lambda && (1)\\ 3 &= 2y\lambda && (2)\\ x^2 + y^2 &= 4 && (3) \end{aligned}\right\} \Rightarrow \quad \lambda = \frac{1}{x} = \frac{3}{2y} \quad \Rightarrow \quad y = \frac{3x}{2}.$$

Substituting into the constraint equation (3) gives

$$x^2 + \left(\frac{3x}{2}\right)^2 = 4 \quad \Rightarrow \quad \frac{13x^2}{4} = 4 \quad \Rightarrow \quad (x, y) = \left(\pm\frac{4}{\sqrt{13}}, \pm\frac{6}{\sqrt{13}}\right).$$

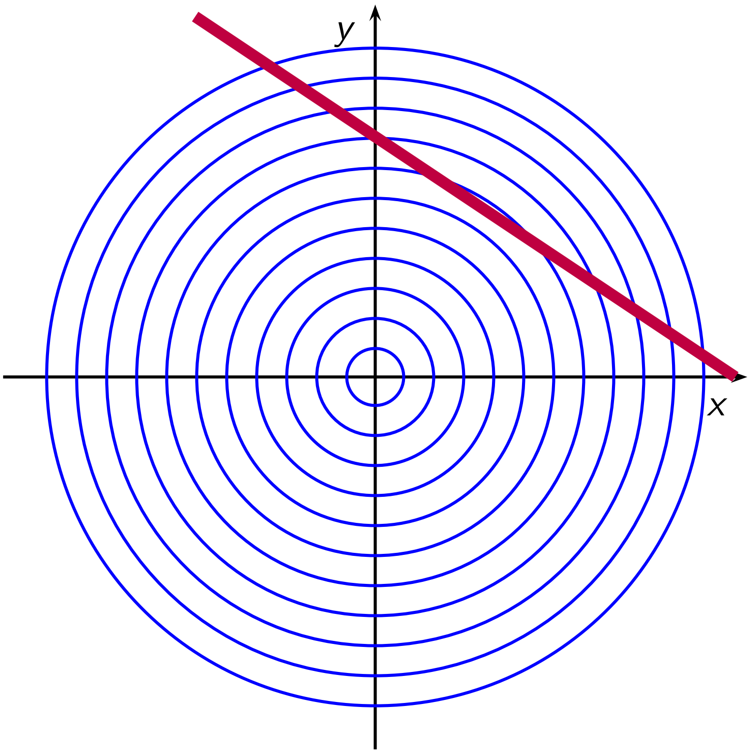
Evaluating f at the two candidate points,

$$f\left(\frac{4}{\sqrt{13}}, \frac{6}{\sqrt{13}}\right) = 4\sqrt{13} \quad \text{and} \quad f\left(-\frac{4}{\sqrt{13}}, -\frac{6}{\sqrt{13}}\right) = -4\sqrt{13}.$$

These are the maximum and minimum values of $f(x, y)$ subject to $g(x, y) = 4$.

Lagrange multipliers

Find the maximum and minimum values of $x^2 + y^2$ subject to $2x + 3y = 20$.



The constraint

$$2x + 3y = 20$$

is purple. Some contours of $x^2 + y^2$ are blue.

Lagrange multipliers

For extreme values of $f(x, y) = x^2 + y^2$ subject to $g(x, y) = 2x + 3y = 20$,

$$\nabla f = \lambda \nabla g \ \Rightarrow \ (2x,\ 2y) = \lambda(2,\ 3).$$

So,

$$\left.\begin{aligned} 2x &= 2\lambda \quad (1) \\ 2y &= 3\lambda \quad (2) \\ 2x + 3y &= 20 \quad (3) \end{aligned}\right\} \quad \Rightarrow \quad \lambda = x = \frac{2y}{3} \quad \Rightarrow \quad y = \frac{3x}{2}.$$

Substituting into the constraint equation (3) gives

$$2x + 3\left(\frac{3x}{2}\right) = 20 \quad \Rightarrow \quad x = \frac{40}{13} \quad \Rightarrow \quad (x, y) = \left(\frac{40}{13}, \frac{60}{13}\right).$$

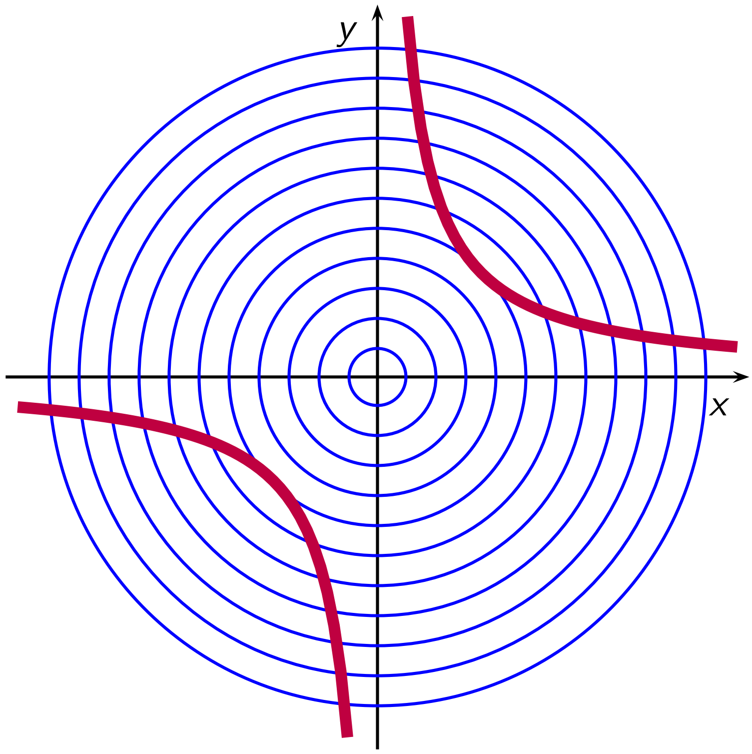
Evaluating f at this candidate point,

$$f\left(\frac{40}{13}, \frac{60}{13}\right) = \frac{400}{13}.$$

It is clear that there is no maximum and this is the minimum.

Lagrange multipliers

Find the maximum and minimum values of $x^2 + y^2$ subject to $xy = 16$.



The constraint

$$xy = 16$$

is purple. Some contours of $x^2 + y^2$ are blue.

Lagrange multipliers

For extreme values of $f(x, y) = x^2 + y^2$ subject to $g(x, y) = xy = 16$,

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

So,

$$\left.\begin{aligned} 2x &= y\lambda \quad (1)\\ 2y &= x\lambda \quad (2)\\ xy &= 16 \quad (3) \end{aligned}\right\} \quad\Rightarrow\quad \lambda = \frac{2x}{y} = \frac{2y}{x} \quad\Rightarrow\quad y^2 = x^2 \quad\Rightarrow\quad y = \pm x.$$

Substituting into the constraint equation (3) gives

$$\pm x^2 = 16 \quad\Rightarrow\quad x = \pm 4 \quad\Rightarrow\quad (x, y) = (\pm 4, \pm 4).$$

Evaluating f at these this candidate points,

$$f(4, 4) = 32 \quad \text{and} \quad f(-4, -4) = 32.$$

It is clear that there is no maximum and this is the minimum.

Lagrange multipliers

Find the maximum and minimum values (if they exist) of $f(x,y,z) = \dfrac{1}{xyz}$ on the ellipsoid $g(x,y,z) = 9x^2 + y^2 + z^2 = 1$ in the region where $x > 0,\ y > 0,\ z > 0$.

$$\nabla f = \lambda \nabla g \ \Rightarrow \ \left(-\frac{1}{x^2yz},\ -\frac{1}{xy^2z},\ -\frac{1}{xyz^2}\right) = \lambda(18x,\ 2y,\ 2z).$$

$$\left.\begin{array}{lr} -\dfrac{1}{x^2yz} = 18\lambda x & (1) \\ -\dfrac{1}{xy^2z} = 2\lambda y & (2) \\ -\dfrac{1}{xyz^2} = 2\lambda z & (3) \\ 9x^2 + y^2 + z^2 = 1 & (4) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} -\dfrac{1}{xyz} = 18\lambda x^2 = 2\lambda y^2 = 2\lambda z^2 \\ \quad \Rightarrow\ 9x^2 = y^2 = z^2 \\ (4) \Rightarrow\ 3z^2 = 1 \ \Rightarrow\ z = \dfrac{1}{\sqrt{3}} \\ \quad \Rightarrow (x,y,z) = \left(\dfrac{1}{3\sqrt{3}}, \dfrac{1}{\sqrt{3}}, \dfrac{1}{\sqrt{3}}\right). \end{array}\right.$$

Evaluating f at this candidate point,

$$f\left(\frac{1}{3\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = 9\sqrt{3}.$$

It is clear that there is no maximum and this is the minimum.

Lagrange multipliers

If $\mathbf{a}$ is a maximum or minimum point of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ subject to r independent[3] constraints

$$g_1(\mathbf{x}) = 0,\ g_2(\mathbf{x}) = 0, \ldots, g_r(\mathbf{x}) = 0$$

that define a smooth surface

$$S = \{\mathbf{x} \in \mathbb{R}^n \ : \ g_1(\mathbf{x}) = 0,\ g_2(\mathbf{x}) = 0, \ldots, g_r(\mathbf{x}) = 0\},$$

then there must exist constants $\lambda_1, \lambda_2, \ldots, \lambda_r$ such that

$$\nabla f(\mathbf{a}) = \lambda_1 \nabla g_1(\mathbf{a}) + \lambda_2 \nabla g_2(\mathbf{a}) + \cdots + \lambda_r \nabla g_r(\mathbf{a}).$$

As for the single constraint case, if S is compact the existence of a maximum and minimum is guaranteed. In other cases, there may be no maximum or minimum points.

[3] The gradient vectors, $\nabla g_1, \nabla g_2, \ldots, \nabla g_r$ must be linearly independent.

Lagrange multipliers

Example: Find the extreme values of $f(x,y,z) = x + y + z$ subject to the two constraints

$$g_1(x,y,z) = x^2 + y^2 = 2 \quad \text{and} \quad g_2(x,y,z) = x + z = 1.$$

To find candidate points for extrema, we solve

$$\nabla f(x,y,z) = \lambda_1 \nabla g_1(x,y,z) + \lambda_2 \nabla g_2(x,y,z)$$

in conjuction with the constraints. That is,

$$\left.\begin{aligned} 1 &= 2\lambda_1 x + \lambda_2 && (1)\\ 1 &= 2\lambda_1 y && (2)\\ 1 &= \lambda_2 && (3)\\ x^2 + y^2 &= 2 && (4)\\ x + z &= 1 && (5) \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} &\lambda_2 = 1,\ \lambda_1 \neq 0\\ &\quad \Rightarrow\ (x,y,z) = (0, \pm\sqrt{2}, 1)\\ &\quad \Rightarrow\ f(0,\pm\sqrt{2},1) = 1 \pm \sqrt{2}. \end{aligned}\right.$$

Since the constraint surface is compact and f is continuous, minimum and maximum values exist and hence are $1 - \sqrt{2}$ and $1 + \sqrt{2}$.

Lagrange multipliers

Example: Find the points on the surface

$$S = \{(x,y,z)\ : z^2 = x^2 y - y^2 + 4\}$$

that are closest to the origin. That is, we want to minimize $\sqrt{x^2 + y^2 + z^2}$ subject to

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

It is simpler to minimise the square of the distance to the origin, so we look for extreme values of

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

Solving

$$\nabla f(x,y,z) = \lambda \nabla g(x,y,z), \qquad g(x,y,z) = 4$$

gives the following set of candidate points:

$$\{(x,y,z)\ :\ x = 0 \text{ and } y^2 + z^2 = 4\} \cup \{(\pm 1.1433\ldots, 1.4505\ldots, 0)\}.$$

Now, $f(\pm 1.1433\ldots, 1.4505\ldots, 0) = 1.8469\ldots < 2$.
Hence the points on S closest to the origin are $(\pm 1.1433\ldots, 1.4505\ldots, 0)$.

Lagrange multipliers

$\nabla g = (-2xy, -x^2+2y, 2z)$, $\nabla f = (2x, 2y, 2z)$

$\nabla f = \lambda \nabla g \Rightarrow$ $-2xy = 2\lambda x$ ...① and
$-x^2+2y = 2\lambda y$ ...②
$2z = 2\lambda z$ ...③
$z^2 - x^2y + y^2 = 4$ ④

---

① $\Rightarrow$ either $x=0$ or $y=-\lambda$

② $\Rightarrow$ $2y = 2\lambda y$

So $y=0$ or $\lambda=1$

④ $\Rightarrow$ $z=\pm 2$ subcase of $\lambda=1$

④ $\Rightarrow$ $z^2+y^2=4$ circle of radius 2 about 0 in the plane $x=0$

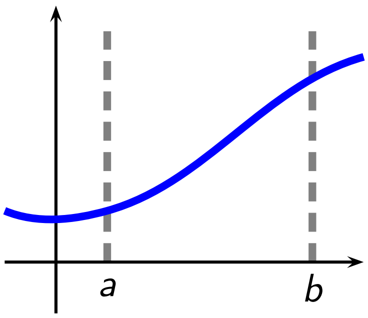
③ $\Rightarrow$ either $z=0$ or $\lambda=1$

② is $-x^2+2y=-2y^2$
④ is $-x^2y+y^2=4$

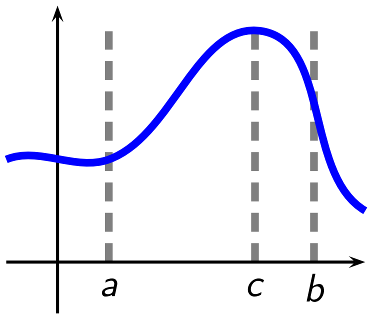
② $\Rightarrow$ $x^2=0$ already considered

$\Rightarrow 2y^3+y^2+4=0$
only one real root $y \simeq -1.4505$
$\Rightarrow x \simeq \pm 1.1433$

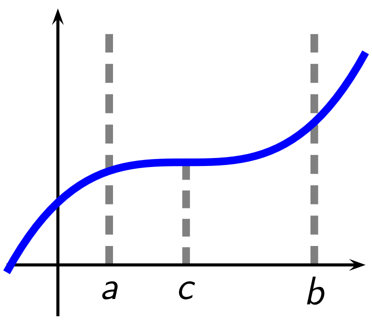
Inverse function theorem for $f : \mathbb{R} \to \mathbb{R}$



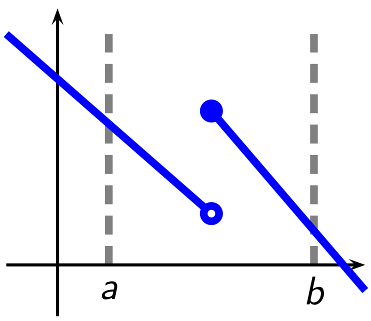
f invertible on (a, b)



$f'(c) = 0$, f not invertible on (a, b)



$f'(c) = 0$, f invertible on (a, b)



f not invertible on (a, b)

Inverse function theorem for $f : \mathbb{R} \to \mathbb{R}$

From first year...

Theorem (Inverse function theorem)

If $f : \mathbb{R} \to \mathbb{R}$ is differentiable on an interval $I \subset \mathbb{R}$ and $f'(x) \neq 0$ for all $x \in I$, then f is invertible on I and the inverse f^{-1} is differentiable with

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}.$$

That is, if $y = f(x)$ then f^{-1} exists and is differentiable with $x = f^{-1}(y)$ and

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}.$$

Inverse function theorem

Consider an affine function $T : \mathbb{R} \to \mathbb{R}$ given by

$$T(x) = mx + b.$$

T is differentiable on $\mathbb{R}$ with $T'(x) = m$. When $m \neq 0$, T is invertible and

$$T(x) = mx + b \;\Rightarrow\; x = m^{-1}T(x) - m^{-1}b$$

so

$$T^{-1}(x) = m^{-1}x - m^{-1}b.$$

T^{-1} is differentiable and

$$(T^{-1})'(x) = m^{-1}.$$

If T is a good affine approximation to f near c then it seems plausible that on a small enough interval around c, the existence of T^{-1} guarantees the existence of f^{-1} with good affine approximation T^{-1}.

We would expect $(f^{-1})'(f(x)) = (f'(x))^{-1}$.

Inverse function theorem

Consider an affine function $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\mathbf{T}(\mathbf{x}) = L\mathbf{x} + \mathbf{b}$$

$\mathbf{T}$ is differentiable on $\mathbb{R}^n$ with $D\mathbf{T} = L$. When $\det L \neq 0$, $\mathbf{T}$ is invertible and

$$\mathbf{T}(\mathbf{x}) = L\mathbf{x} + \mathbf{b} \;\Rightarrow\; \mathbf{x} = L^{-1}\mathbf{T}(\mathbf{x}) - L^{-1}\mathbf{b}$$

so

$$\mathbf{T}^{-1}(\mathbf{x}) = L^{-1}\mathbf{x} - L^{-1}\mathbf{b}.$$

If $\mathbf{T}$ is a good affine approximation to $\mathbf{f}$ near $\mathbf{c}$ then it seems plausible that on a small enough ball around $\mathbf{c}$, the existence of $\mathbf{T}^{-1}$ guarantees the existence of $\mathbf{f}^{-1}$ with good affine approximation $\mathbf{T}^{-1}$.

We would expect $D_{\mathbf{c}}\mathbf{f} = L$ then $D_{\mathbf{f}(\mathbf{c})}\left(\mathbf{f}^{-1}\right) = L^{-1}$.

Inverse function theorem

Theorem

Let $\Omega \subset \mathbb{R}^n$ be open, $\mathbf{f} : \Omega \to \mathbb{R}^n$ be C^1 and suppose $\mathbf{a} \in \Omega$.

If $D\mathbf{f}(\mathbf{a})$ is invertible (as a matrix) then $\mathbf{f}$ is invertible on an open set U containing $\mathbf{a}$. That is,

$$\mathbf{f}^{-1} : \mathbf{f}(U) \to U$$

exists.

Furthermore, $\mathbf{f}^{-1}$ is C^1 and for $\mathbf{x} \in U$,

$$D_{\mathbf{f}(\mathbf{x})}\mathbf{f}^{-1} = \left(D_{\mathbf{x}}\mathbf{f}\right)^{-1}.$$

Note that this says $\mathbf{f}^{-1}$ has a good affine approximation at $\mathbf{f}(\mathbf{a})$ given by

$$\mathbf{f}^{-1}(\mathbf{x}) \simeq \mathbf{a} + \left(D_{\mathbf{a}}\mathbf{f}\right)^{-1}\left(\mathbf{x} - \mathbf{f}(\mathbf{a})\right).$$

Inverse function theorem

Example: Can the map $x = r\cos\theta,\ y = r\sin\theta$ be inverted?

Define $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{f}\begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$$

Away from $(x, y) = (0, 0)$ (ie $r = 0$) $\mathbf{f}$ is differentiable with $D\mathbf{f} = J\mathbf{f}$ so

$$D\mathbf{f} = \begin{pmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and $\det(D\mathbf{f}) = r\cos^2\theta + r\sin^2\theta = r \neq 0$.

So $\mathbf{f}$ is locally invertible away from $r = 0$.

Eg, at $\mathbf{a} = \left(\sqrt{2}, \frac{\pi}{4}\right)$, $\mathbf{f}(\mathbf{a}) = (1, 1)$. So

$$D\mathbf{f}\left(\sqrt{2}, \frac{\pi}{4}\right) = \begin{pmatrix} \dfrac{1}{\sqrt{2}} & -1 \\ \dfrac{1}{\sqrt{2}} & 1 \end{pmatrix}$$

$$D\left(\mathbf{f}^{-1}\right)(1, 1) = \left(D\mathbf{f}\left(\sqrt{2}, \frac{\pi}{4}\right)\right)^{-1} = \begin{pmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ -\dfrac{1}{2} & \dfrac{1}{2} \end{pmatrix}.$$

Inverse function theorem

We can check that this matches what we get from directly inverting $\mathbf{f}$. In the first quadrant away from $\mathbf{0}$,

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}(y/x) \quad \Rightarrow \quad \mathbf{f}^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sqrt{x^2 + y^2} \\ \tan^{-1}(y/x) \end{pmatrix}$$

$$\Rightarrow \quad D\mathbf{f}^{-1} = \begin{pmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{pmatrix} \quad \Rightarrow \quad D\mathbf{f}^{-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ -\dfrac{1}{2} & \dfrac{1}{2} \end{pmatrix}.$$

An affine approximation to $\mathbf{f}^{-1}$ near $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is

$$\mathbf{f}^{-1}\begin{pmatrix} x \\ y \end{pmatrix} \simeq \mathbf{f}^{-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + D(\mathbf{f}^{-1})\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} x - 1 \\ y - 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sqrt{2} \\ \dfrac{\pi}{4} \end{pmatrix} + \begin{pmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ -\dfrac{1}{2} & \dfrac{1}{2} \end{pmatrix}\begin{pmatrix} x - 1 \\ y - 1 \end{pmatrix}.$$

Inverse function theorem

Suppose $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$\mathbf{f}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^3e^y + y - 2x \\ 2xy + 2x \end{pmatrix}.$$

Note $\mathbf{f}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

Show that $\mathbf{f}$ has a differentiable inverse near $(1, 0)$ and hence find an approximate solution to $\mathbf{f}^{-1}\begin{pmatrix} -1.2 \\ 2.1 \end{pmatrix}$, that is, an approximate solution to

$$\begin{aligned} x^3e^y + y - 2x &= -1.2, \\ 2xy + 2x &= 2.1. \end{aligned}$$

The partial derivatives of the components of $\mathbf{f}$ exist and are continuous everywhere. Hence $\mathbf{f}$ is differentiable on $\mathbb{R}^2$ and

$$D\mathbf{f} = J\mathbf{f} = \begin{pmatrix} 3x^2e^y - 2 & x^3e^y + 1 \\ 2y + 2 & 2x \end{pmatrix} \quad \Rightarrow \quad D\mathbf{f}(1, 0) = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}.$$

Inverse function theorem

Since $\det\left(D\mathbf{f}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = -2 \neq 0$, the Inverse Function Theorem says that $\mathbf{f}$ has a C^1 local inverse near $(1, 0)$ with derivative

$$D\mathbf{f}^{-1}\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \left(D\mathbf{f}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)^{-1} = -\frac{1}{2}\begin{pmatrix} 2 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix}.$$

Now, the best affine approximation to $\mathbf{f}^{-1}$ is

$$\mathbf{f}^{-1}\begin{pmatrix} u \\ v \end{pmatrix} \simeq \mathbf{f}^{-1}\begin{pmatrix} -1 \\ 2 \end{pmatrix} + D\mathbf{f}^{-1}\begin{pmatrix} -1 \\ 2 \end{pmatrix}\begin{pmatrix} u - (-1) \\ v - 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} u + 1 \\ v - 2 \end{pmatrix}$$

So now the approximate solution is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{f}^{-1}\begin{pmatrix} -1.2 \\ 2.1 \end{pmatrix} \simeq \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 1 & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} -0.2 \\ 0.1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -0.6 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 1.3 \\ -0.25 \end{pmatrix}$$

$$\left[\mathbf{f}\begin{pmatrix} 1.3 \\ -0.25 \end{pmatrix} \simeq \begin{pmatrix} -1.14 \\ 1.95 \end{pmatrix}, \ \mathbf{f}^{-1}\begin{pmatrix} -1.02 \\ 2.01 \end{pmatrix} \simeq \begin{pmatrix} 1.03 \\ -0.025 \end{pmatrix}, \ \mathbf{f}\begin{pmatrix} 1.03 \\ -0.025 \end{pmatrix} \simeq \begin{pmatrix} -1.019 \\ 2.009 \end{pmatrix}.\right]$$

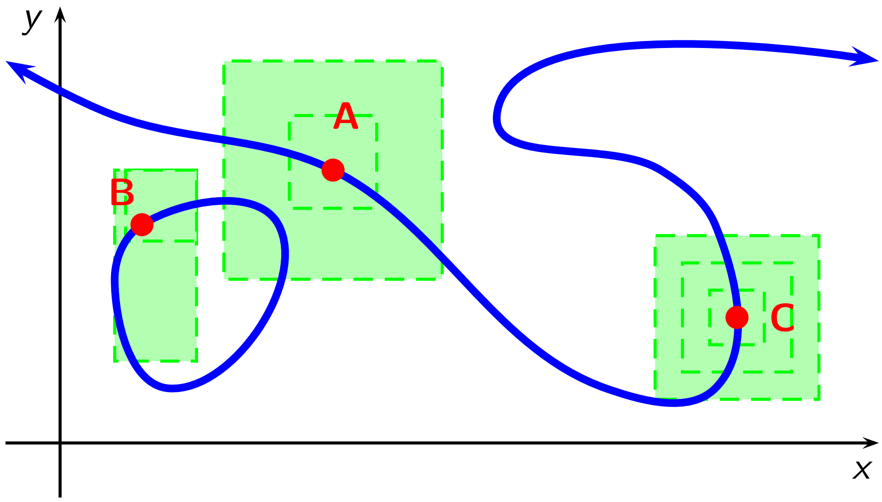
Implicit function theorem

Consider a C^1 function $g : \mathbb{R}^2 \to \mathbb{R}$, its 0 contour,

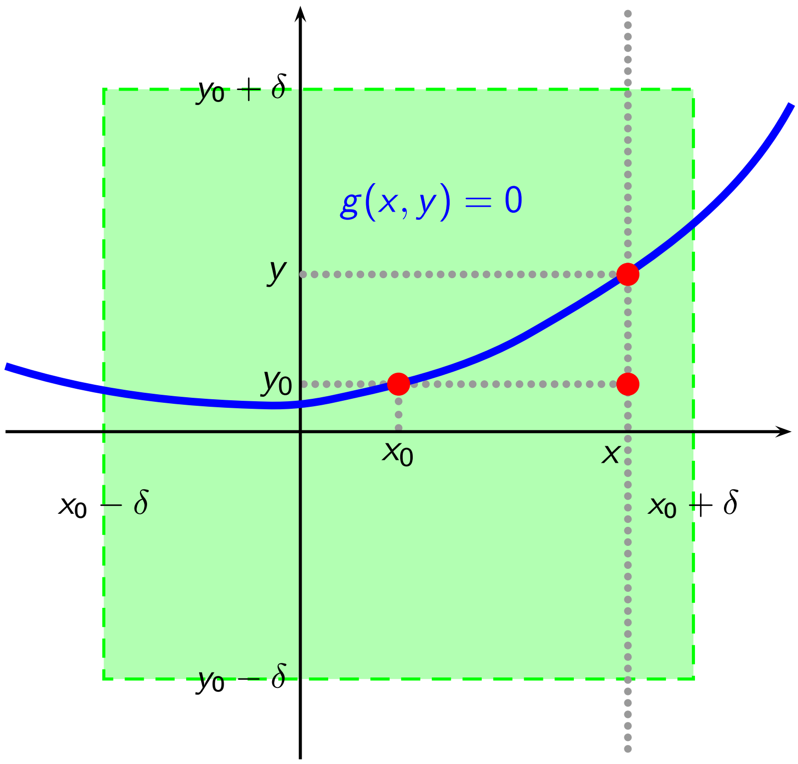
$$S = \{(x, y) \in \mathbb{R}^2 \ : \ g(x, y) = 0\}$$

and a point $(x_0, y_0) \in S$. When does S define y as a function of x near the point (x_0, y_0)?

At **A** and **B** but not **C**.



Implicit function theorem



Given

- $g : \mathbb{R}^2 \to \mathbb{R}$ is C^1,
- $g(x_0, y_0) = 0$ and
- $\dfrac{\partial g}{\partial y}(x_0, y_0) \neq 0$,

we want to show that there is a δ such that for

$$x \in (x_0 - \delta, x_0 + \delta)$$

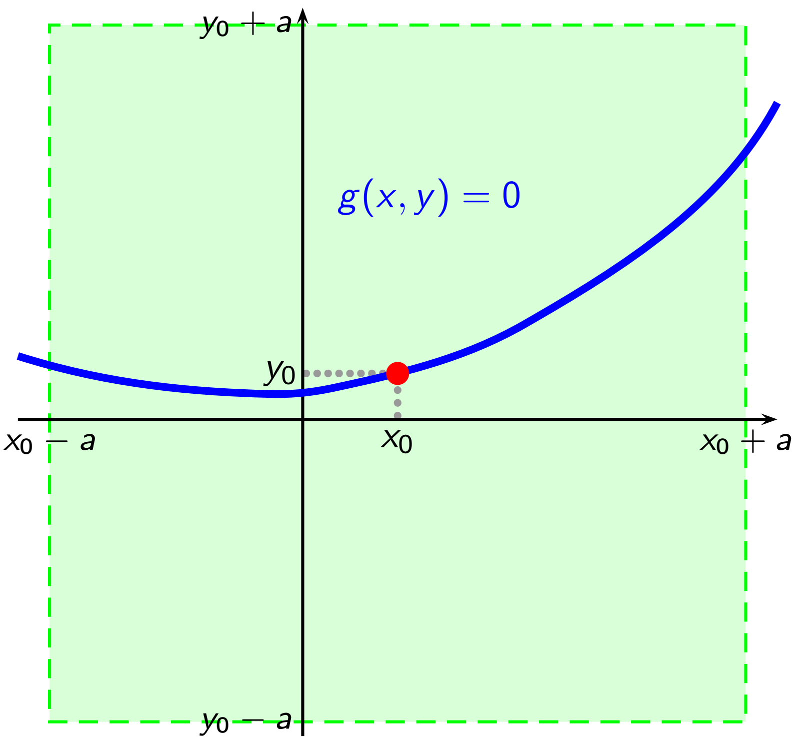
there is a unique

$$y \in (y_0 - \delta, y_0 + \delta)$$

satisfying

$$g(x, y) = 0.$$

Implicit function theorem



Assume that $\dfrac{\partial g}{\partial y}(x_0, y_0) > 0$.

As g is C^1, there are $a > 0$ and $b > 0$ such that for

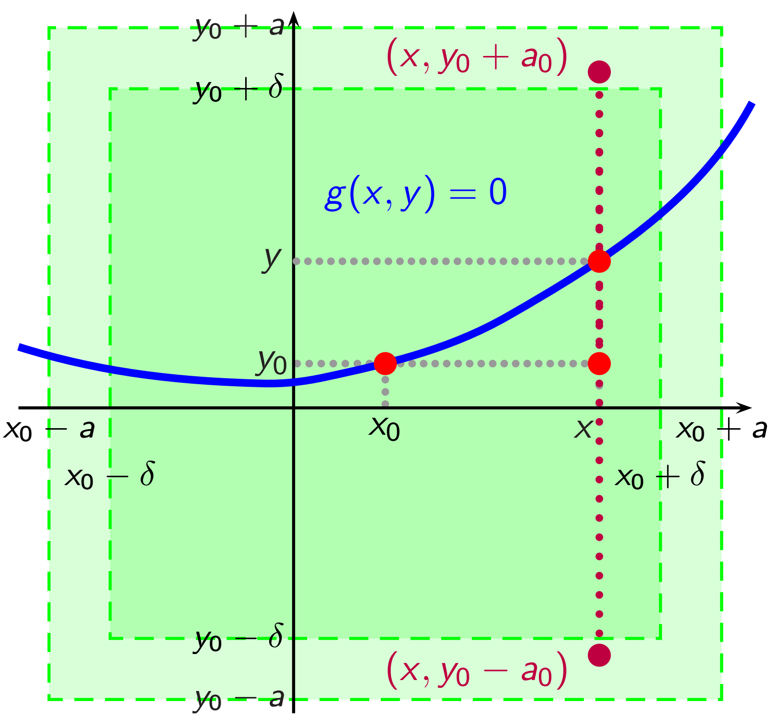
$$x \in (x_0 - a, x_0 + a)$$

$$y \in (y_0 - a, y_0 + a),$$

- $\dfrac{\partial g}{\partial y}(x, y) > b$
- $\left|\dfrac{\partial g}{\partial x}(x, y)\right| < M$

for some $M > 0$.

Implicit function theorem



Choose positive a_0 and δ so that

$$a_0 < a, \quad \delta < \min\left(a_0, \frac{ba_0}{M}\right).$$

$$\begin{aligned} g(x, y_0 + a_0) &> 0 - M\delta + ba_0 \\ &> -ba_0 + ba_0 = 0. \\ g(x, y_0 - a_0) &< 0 + M\delta - ba_0 \\ &< ba_0 - ba_0 = 0. \end{aligned}$$

IVT $\Rightarrow$ $\exists y \in (y_0 - a_0, y_0 + a_0)$ such that $g(x, y) = 0$.

$$\frac{\partial g}{\partial y} > 0 \;\Rightarrow\; y \text{ is unique.}$$

$$\text{MVT} \;\Rightarrow\; g(x, y_0 \pm a_0) = g(x_0, y_0) + \frac{\partial g}{\partial x}(c^{\pm}, y_0)(x - x_0) + \frac{\partial g}{\partial y}(x, d^{\pm})(a_0 - y_0).$$

Implicit function theorem

We have shown that given

- $g : \mathbb{R}^2 \to \mathbb{R}$ is C^1,
- $g(x_0, y_0) = 0$ and
- $\dfrac{\partial g}{\partial y}(x_0, y_0) \neq 0$,

there is a δ such that for

$$x \in (x_0 - \delta, x_0 + \delta)$$

there is a unique

$$y \in (y_0 - \delta, y_0 + \delta)$$

satisfying

$$g(x, y) = 0.$$

So, there is $f : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ such that

$$g(x, f(x)) = 0.$$

It can also be shown that f is C^1. Assuming f is differentiable, we can find f' by implicit differentiation and find

$$\begin{aligned} \frac{d}{dx}\Big(g(x, f(x))\Big) &= 0 \\ \Rightarrow \frac{\partial g}{\partial x}\frac{dx}{dx} + \frac{\partial g}{\partial y}\frac{dy}{dx} &= 0 \\ \Rightarrow \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} f'(x) &= 0 \end{aligned}$$

$$\Rightarrow \quad f'(x_0) = -\left(\frac{\partial g}{\partial y}(x_0, y_0)\right)^{-1} \frac{\partial g}{\partial x}(x_0, y_0).$$

Implicit Function Theorem

For the Implicit Function Theorem in higher dimensions, consider the following.

- Near which points does
$$x^2 + y^2 + z^2 = 1$$
define z as a function of x and y? That is, when does there exist f such that $z = f(x, y)$?
- Given
$$\begin{aligned} x + y + z &= 6 \\ 2x - y + 2z &= 8, \end{aligned}$$
you can find y and z given just the value of x. So there is a function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^2$ such that $\begin{pmatrix} y \\ z \end{pmatrix} = \mathbf{f}(x)$.

Typically, if there are n equations and r variables, we expect to be able to solve for n of variables in terms of the remaining $n - r$ variables near most points.

Implicit Function Theorem

Let $\mathbf{x} \in \mathbb{R}^m$ denote our known variables and let $\mathbf{u} \in \mathbb{R}^n$ denote our unknown variables. To solve for $\mathbf{u}$ in terms of $\mathbf{x}$ we expect to need n equations:

$$\begin{aligned} g_1(x_1, \ldots, x_m, u_1, \ldots, u_n) &= 0 \\ g_2(x_1, \ldots, x_m, u_1, \ldots, u_n) &= 0 \\ \vdots \quad &\quad \vdots \\ g_n(x_1, \ldots, x_m, u_1, \ldots, u_n) &= 0. \end{aligned}$$

We can write this more succinctly as

$$\mathbf{g}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$$

where $\mathbf{g} : \mathbb{R}^{m+n} \to \mathbb{R}^n$ is

$$\mathbf{g}(\mathbf{x}, \mathbf{u}) = (g_1(\mathbf{x}, \mathbf{u}), \ldots, g_n(\mathbf{x}, \mathbf{u})).$$

Solving this system of equations means finding a way of specifying what $\mathbf{u}$ is if we know $\mathbf{x}$. That is, we need to find a continuous function $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ satisfying

$$\mathbf{g}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}.$$

Implicit Function Theorem

Define the $n \times m$ matrix A and $n \times n$ matrix B in terms of $D\mathbf{g}$.

$$D\mathbf{g} = \left(\begin{array}{cccc|ccc} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_m} & \frac{\partial g_1}{\partial u_1} & \cdots & \frac{\partial g_1}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_m} & \frac{\partial g_n}{\partial u_1} & \cdots & \frac{\partial g_n}{\partial u_n} \end{array}\right) = [A|B]$$

Theorem (Implicit Function Theorem)

Suppose that $(\mathbf{x}_0, \mathbf{u}_0)$ is on the surface $\mathbf{g}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$. If $B(\mathbf{x}_0, \mathbf{u}_0)$ is an invertible matrix, then there is an open set V around $\mathbf{x}_0$ on which $\mathbf{u}$ is defined implicitly as a function of $\mathbf{x}$. That is, there exists a continuously differentiable function $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ such that for all $\mathbf{x} \in V$

$$\mathbf{g}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}.$$

Implicit Function Theorem

To find $D\mathbf{f}$ in terms of $D\mathbf{g}$ use the chain rule. Let $\mathbf{h} : \mathbb{R}^m \to \mathbb{R}^{m+n}$ be defined by

$$\mathbf{h}(\mathbf{x}) = \begin{pmatrix} \mathbf{x} \\ \mathbf{f}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ f_1(x_1, \ldots, x_m) \\ \vdots \\ f_n(x_1, \ldots, x_m) \end{pmatrix} \quad \Rightarrow \quad D_{\mathbf{x}}\mathbf{h} = \begin{pmatrix} I_m \\ D_{\mathbf{x}}\mathbf{f} \end{pmatrix}$$

where I_m is the $n \times n$ identity matrix. Differentiating the equation

$$\mathbf{g}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = (\mathbf{g} \circ \mathbf{h})(\mathbf{x}) = \mathbf{0}$$

gives

$$\mathbf{0} = D_{\mathbf{h}(\mathbf{x})}\mathbf{g}\, D_{\mathbf{x}}\mathbf{h} = \left(A(\mathbf{h}(\mathbf{x})) \mid B(\mathbf{h}(\mathbf{x}))\right) \begin{pmatrix} I_m \\ D_{\mathbf{x}}\mathbf{f} \end{pmatrix} = A(\mathbf{h}(\mathbf{x})) + B(\mathbf{h}(\mathbf{x}))D_{\mathbf{x}}\mathbf{f}.$$

Rearranging this gives $D_{\mathbf{x}}\mathbf{f} = -B(\mathbf{h}(\mathbf{x}))^{-1}A(\mathbf{h}(\mathbf{x}))$.

Implicit Function Theorem

Show that there are open sets $U \subset \mathbb{R}^2$ containing $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $V \subset \mathbb{R}^2$ containing $\begin{pmatrix} 1 \\ 5 \end{pmatrix}$ so that the equations

$$\begin{aligned} x^2 + xy + yu + u^2 - xv - 1 &= 0, \\ y^2 + xy - u^2 - v &= 0, \end{aligned}$$

define a differentiable function $\mathbf{f} : U \to V$ for which (x, y, u, v) satisfies the equations when $\begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{f}\begin{pmatrix} x \\ y \end{pmatrix}$.

Find the affine approximation to $\mathbf{f}$ near $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and hence find an approximate solution $\begin{pmatrix} u \\ v \end{pmatrix}$ to these equations when $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.2 \\ 1.9 \end{pmatrix}$.

Implicit Function Theorem

The equations can be written in the form $\mathbf{g}\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Then

$$D\mathbf{g} = \begin{pmatrix} 2x+y-v & x+u & y+2u & -x \\ y & x+2y & -2u & -1 \end{pmatrix}.$$

At the known point $\mathbf{x}_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\mathbf{u}_0 = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$, this gives

$$D\mathbf{g}(\mathbf{x}_0, \mathbf{u}_0) = \left(\begin{array}{cc|cc} -1 & 2 & 4 & -1 \\ 2 & 5 & -2 & -1 \end{array}\right) = [A \,|\, B].$$

B is invertible and so there is a C^1 function $\mathbf{f}\begin{pmatrix} x \\ y \end{pmatrix}$ defined on an open set around $\mathbf{x}_0$ so that $\mathbf{g}\left(\begin{pmatrix} x \\ y \end{pmatrix}, \mathbf{f}\begin{pmatrix} x \\ y \end{pmatrix}\right) = \mathbf{0}$, and

$$D\mathbf{f}(\mathbf{x}_0) = -B^{-1}A = -\frac{1}{-6}\begin{pmatrix} -1 & 1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} -1 & 2 \\ 2 & 5 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 3 & 3 \\ 6 & 24 \end{pmatrix}.$$

Implicit Function Theorem

Thus the affine approximation to $\mathbf{f}\begin{pmatrix} x \\ y \end{pmatrix}$ near $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \begin{pmatrix} 1 \\ 5 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 3 & 3 \\ 6 & 24 \end{pmatrix}\begin{pmatrix} x-1 \\ y-2 \end{pmatrix}.$$

In particular

$$\mathbf{f}\begin{pmatrix} 1.2 \\ 1.9 \end{pmatrix} \approx \begin{pmatrix} 1 \\ 5 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 3 & 3 \\ 6 & 24 \end{pmatrix}\begin{pmatrix} 0.2 \\ -0.1 \end{pmatrix} = \begin{pmatrix} 1.05 \\ 4.8 \end{pmatrix}.$$

You can check whether this is any good by calculating $\mathbf{g}\left(\begin{pmatrix} 1.2 \\ 1.9 \end{pmatrix}, \begin{pmatrix} 1.05 \\ 4.8 \end{pmatrix}\right)$.

Implicit Function Theorem

What we have done is to replace the original equations

$$\mathbf{g}(x, y, u, v) = \mathbf{0}$$

with the equations

$$\mathbf{T}(x, y, u, v) = \mathbf{g}(1, 2, 1, 5) + \begin{pmatrix} -1 & 2 & 4 & -1 \\ 2 & 5 & -2 & -1 \end{pmatrix} \begin{pmatrix} x-1 \\ y-2 \\ u-1 \\ v-5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where $\mathbf{T}$ is the best affine approximation to $\mathbf{g}$ near $(1, 2, 1, 5)$. That is, since $\mathbf{g}(1, 2, 1, 5) = \mathbf{0}$, the given equations are approximately

$$\begin{aligned} -(x-1) + 2(y-2) + 4(u-1) - (v-5) &= 0 \\ 2(x-1) + 5(y-2) - 2(u-1) - (v-5) &= 0 \end{aligned}$$

which simpifies to the pair of linear equations

$$\begin{aligned} -x + 2y + 4u - v &= -2 \\ 2x + 5y - 2u - v &= -5. \end{aligned}$$