1 Function of two variables

Observations

1. A real-valued function is a box that accepts a number and produces another number.
2. We can describe a real-valued function in terms of its graph, which is a curve.
3. A function, in general, is a box that accepts something of one type, and produces something else of another type.
4. We study real-valued functions of two variables, which accept two numbers and produce another one.
5. We can describe such functions in terms of their graph, which is now a surface.
6. Formally, we have the following:

Definition 1.1. *(Real valued functions of two variables)*

1. *A function of* 2 *variables is a mapping* $(x, y) \to z = f(x, y)$.
2. *Its domain* $D \subseteq \mathbb{R}^2$ *is all the pairs* (x, y) *where the function is defined.*
3. *Its range* $R \subseteq \mathbb{R}$ *is* $R = \{f(x, y) : (x, y) \in D\}$.
4. *Its graph is* $G = \{(x, y, z) : (x, y) \in D,\ z = f(x, y)\} \subseteq \mathbb{R}^3$.
5. *The equipotential curve, or contour line at level* c*, is the set*

$$C(c) = \{(x, y, z) : (x, y) \in D,\ z = f(x, y) = c\} \subseteq \mathbb{R}^3.$$

INSERT FIRST DEMO HERE. TOPICS: Examples of functions and contour lines

Example 1.1. Let $f(x, y) = \frac{1}{3}\sqrt{36 - 9x^2 - 4y^2}$. Then the domain is the set

$$D = \{(x, y) : 36 - 9x^2 - 4y^2 \geq 0\},$$

while the contour lines satisfy the following equations:

$$c = \frac{1}{3}\sqrt{36 - 9x^2 - 4y^2} \Leftrightarrow 9x^2 + 4y^2 = 36 - 9c^2.$$

Therefore, contour lines are ellipses.

Definition 1.2. *(Planes) Linear functions of the form* $f(x, y) = ax + by + c$ *are called planes.*

Definition 1.3. *(Polynomials and rational functions)*

1. *A polynomial with variables* x, y *is a function of the form*

$$f(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x^i y^j.$$

2. *A rational function with variables* x, y *is a function of the form*

$$f(x, y) = \frac{p(x, y)}{q(x, y)},$$

where $p(x, y)$ *and* $q(x, y)$ *are polynomials.*

2 Limit and Continuity

Definition 2.1. *(Function limit) Function $f(x,y)$ has limit L at $(x,y)=(a,b)$, and we write*

$$\lim_{(x,y)\to(a,b)} f(x,y) = L \qquad \textit{or} \qquad f(x,y)\to L \quad \textit{for} \quad (x,y)\to(a,b),$$

if the following holds:

$$\forall\epsilon>0, \exists\delta>0 : 0<\sqrt{(x-a)^2+(y-b)^2}<\delta \Rightarrow |f(x,y)-L|<\epsilon.$$

Observations

1. We use the notation
$$\|(x,y)\| = \sqrt{x^2+y^2}$$
for the Euclidean length of a vector, therefore we also have
$$\|(x,y)-(a,b)\| = \sqrt{(x-a)^2+(y-b)^2}.$$

2. The basic intuition of the plain limit fully transfers: if we take concentric circles around (a,b), then, as their radius decreases, so does the maximum distance of the values of the function from limit L.

Theorem 2.1. *(Limits of polynomial and rational functions)*

1. *If $f(x,y)$ is a polynomial, then*
$$\lim_{(x,y)\to(a,b)} f(x,y) = f(a,b).$$

2. *If $f(x,y)=\frac{p(x,y)}{q(x,y)}$ where $p(x,y),\ q(x,y)$ are polynomials, then*

   (α′) If $q(a,b)\neq 0$, then
$$\lim_{(x,y)\to(a,b)} f(x,y) = \frac{p(a,b)}{q(a,b)}.$$

   (β′) If $q(a,b)=0$ and $p(a,b)\neq 0$, then the limit $\lim_{(x,y)\to(a,b)} f(x,y)$ does not exist.

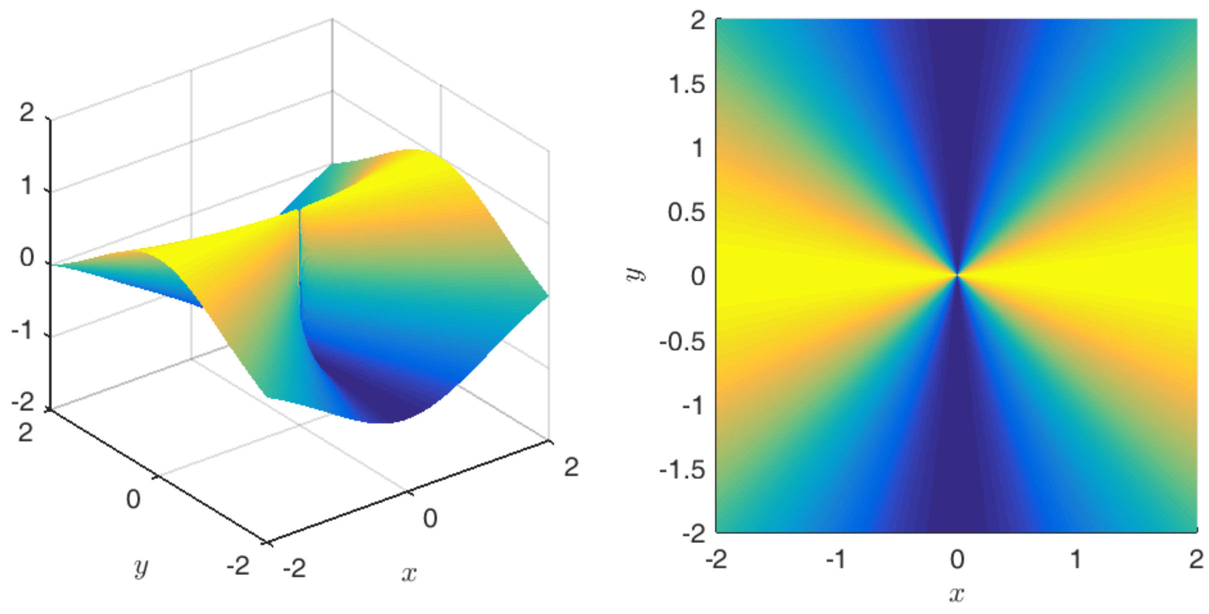
3. *If $q(a,b)=0$ and $p(a,b)=0$, anything goes: the limit might exist or it might not exist.*

Example 2.1. *(Limit calculations)*

1. $$\lim_{(x,y)\to(1,2)} (x^2y+3y) = 2+6 = 8.$$

2. The limit
$$\lim_{(x,y)\to(0,0)} \frac{x^2+y^2+1}{x^2-y^2}$$
does not exist. Why? Does the function go to $\pm\infty$?

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



Σχήμα 1: Example 2.1.

also does not exist. Indeed, observe that

$$\lim_{x\to 0} f(x,0) = \lim_{x\to 0} \frac{x^2}{x^2} = 1,$$
$$\lim_{y\to 0} f(0,y) = \lim_{y\to 0} -\frac{y^2}{y^2} = -1,$$

so the limit depends on the angle of approach! See Fig. 1.

Definition 2.2. *(Continuity) Function $f(x)$ is continuous in (a,b) if the limit $\lim_{(x,y)\to(a,b)} f(x,y)$ exists and we have*

$$\lim_{(x,y)\to(a,b)} f(x,y) = f(a,b).$$

Theorem 2.2. *(Basic properties of continuity)*

1. *If $f(x,y)$ and $g(x,y)$ are continuous at (a,b), then the functions $\lambda_1 f(x,y) + \lambda_2 g(x,y)$, where $\lambda_1, \lambda_2 \in \mathbb{R}$, and $f(x,y)g(x,y)$ are also continuous at (a,b).*

2. *If $f(x,y)$ and $g(x,y)$ are continuous at (a,b), and $g(a,b) \neq 0$, then $\frac{f(x,y)}{g(x,y)}$ is also continuous at (a,b).*

3. *If $g(x,y)$ is continuous at (a,b) and f is continuous at $g(a,b)$, then the composition $f(g(x,y))$ is continuous at (a,b).*

3 Partial derivatives

Definition 3.1. *(Partial Derivatives)*
The partial derivatives of $f(x,y)$ at location (x_0,y_0) are defined as follows:

$$\begin{aligned} f_x(x_0,y_0) &= \lim_{h\to 0}\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h}, \\ f_y(x_0,y_0) &= \lim_{h\to 0}\frac{f(x_0,y_0+h)-f(x_0,y_0)}{h}. \end{aligned}$$

Observations

1. We also use the notations

$$\begin{aligned} f_x(x_0,y_0) &= \left.\frac{\partial z}{\partial x}\right|_{(x_0,y_0)} = \left.\frac{\partial f(x,y)}{\partial x}\right|_{(x_0,y_0)} = \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)}, \\ f_y(x_0,y_0) &= \left.\frac{\partial z}{\partial y}\right|_{(x_0,y_0)} = \left.\frac{\partial f(x,y)}{\partial y}\right|_{(x_0,y_0)} = \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)}. \end{aligned}$$

2. Observe that the partial derivative with respect to x at point (x_0,y_0) is the usual derivative of the single-variable function $f(x,y_0)$ at $x=x_0$.

3. What is the physical interpretation of partial derivatives?

Definition 3.2. *(Second-order partial derivatives)*

$$\begin{aligned} f_{xx} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \\ f_{xy} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right), \\ f_{yx} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \\ f_{yy} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right). \end{aligned}$$

Observations

1. Higher-order derivatives are defined in a similar manner.

2. We use the following notation:

$$f_{xx}=\frac{\partial^2 f}{\partial x^2},\quad f_{yy}=\frac{\partial^2 f}{\partial y^2},\quad f_{xy}=\frac{\partial^2 f}{\partial y\partial x},\quad f_{yx}=\frac{\partial^2 f}{\partial x\partial y}.$$

Example 3.1. Let the function

$$f(x,y)=xe^y-\sin\left(\frac{x}{y}\right)+x^3y^2.$$

Its partial derivatives are as follows:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= e^y - \cos\left(\frac{x}{y}\right)\frac{1}{y} + 3x^2y^2, \\
\frac{\partial^2 f}{\partial x^2} &= \sin\left(\frac{x}{y}\right)\frac{1}{y^2} + 6xy^2, \\
\frac{\partial f}{\partial y} &= xe^y - \cos\left(\frac{x}{y}\right)\left(-\frac{x}{y^2}\right) + 2yx^3 \\
&= xe^y + \frac{x}{y^2}\cos\left(\frac{x}{y}\right) + 2yx^3, \\
\frac{\partial^2 f}{\partial y^2} &= xe^y - \frac{2x}{y^3}\cos\left(\frac{x}{y}\right) - \frac{x}{y^2}\sin\left(\frac{x}{y}\right)\left(-\frac{x}{y^2}\right) + 2x^3 \\
&= xe^y - \frac{2x}{y^3}\cos\left(\frac{x}{y}\right) + \frac{x^2}{y^4}\sin\left(\frac{x}{y}\right) + 2x^3, \\
\frac{\partial^2 f}{\partial x \partial y} &= e^y + \frac{1}{y^2}\cos\left(\frac{x}{y}\right) + \frac{x}{y^2}\left(-\sin\left(\frac{x}{y}\right)\right)\frac{1}{y} + 6x^2y \\
&= e^y + \frac{1}{y^2}\cos\left(\frac{x}{y}\right) - \frac{x}{y^3}\sin\left(\frac{x}{y}\right) + 6x^2y, \\
\frac{\partial^2 f}{\partial y \partial x} &= e^y + \sin\left(\frac{x}{y}\right)\left(-\frac{x}{y^2}\right)\frac{1}{y} - \cos\left(\frac{x}{y}\right)\left(-\frac{1}{y^2}\right) + 6x^2y \\
&= e^y - \frac{x}{y^3}\sin\left(\frac{x}{y}\right) + \frac{1}{y^2}\cos\left(\frac{x}{y}\right) + 6x^2y.
\end{aligned}$$

Observe that f_{xy} and f_{yx} are equal! This is not a coincidence.

Theorem 3.1. *(Equality of partial derivatives) If f_{xy} and f_{yx} are continuous on an open set S, then they are also equal.*

Observation: So, what does it mean for a two-dimensional set to be open? We discuss this in the next section.

4 Small detour on topology

Observation: The notation $A \subset B$ implies that the sets A, B cannot be equal. The notation $A \subseteq B$ implies that they might be equal.

Definition 4.1.

1. *Let $S \subseteq \mathbb{R}^2$. We define the complement of S as the set*

$$\overline{S} = \{(x, y) \in \mathbb{R}^2 : (x, y) \notin S\}.$$

2. *Let $S \subseteq \mathbb{R}^2$ and a point $P = (x_0, y_0)$. The ball of radius $\delta > 0$ centered at P is the set*

$$B(P, \delta) = \{Q = (x, y) : \|P - Q\| < \delta \Leftrightarrow \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta\}.$$

3. *A point $P \in S$ is called an interior point of S is there is a $B(P, \delta) \subset S$.*

4. *A point P is called a* <u>*boundary point*</u> *of S if for all $\delta > 0$, the set $B(P, \delta)$ contains both points in S and points in $\overline{S}$. The set of all boundary points of S is called the* <u>*boundary*</u> *∂S of S. Observe that each point in S is either a boundary point of S, or an interior point of S.*

5. *A set S is called* <u>*open*</u> *if all its points P are interior points, i.e., it contains no boundary points.*

6. *A set S is called closed if it contains all its boundary points. Equivalently, set S is closed if set $\overline{S}$ is open.*

7. *A set S is called* <u>*bounded*</u> *if there is a radius $R > 0$ such that $S \in B((0,0), R)$.*

Example 4.1. *(Examples of open and closed sets)*

1. The set $S = \{(x, y) : x^2 + y^2 < 1\}$ is open, and its boundary is $\partial S = \{(x, y)' : x^2 + y^2 = 1\}$.
2. The set $S = \{(x, y) : |y| < 1\}$ is open, and its boundary is $\partial S = \{(x, y) : y = 1\} \cup \{(x, y) : y = -1\}$.
3. The set $S = \{(x, y) : |y| < 1, \ x = 0\}$ is neither closed nor open! Its boundary is the set $\partial S = \{(x, y) : |y| \leq 1, x = 0\}$.
4. Both $\mathbb{R}^2$ and the empty set $\emptyset$ are both open and closed! Such sets are called <u>clopen</u>.

5 The gradient and differentiable functions

Definition 5.1. *Let a point $P = (x, y)$. The vector*

$$\nabla f(x, y) \triangleq (f_x(x, y), f_y(x, y))$$

is defined as the <u>*gradient*</u> *of $f(x, y)$ at the point (x, y).*

Theorem 5.1. *(Basic properties of the gradient)*

1.

$$\nabla(f(x, y) + g(x, y)) = \nabla f(x, y) + \nabla g(x, y).$$

2.

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

3.

$$\nabla(f(x, y)g(x, y) = g(x, y)\nabla f(x, y) + f(x, y)\nabla g(x, y).$$

Observations

1. The above properties have simple proofs, based the definition of the gradient.
2. The gradient points to the direction where the function increases the fastest. This will become clear later on.

Example 5.1. Let the functions $f(x, y) = x^2 y$ and $g(x, y) = 3xy$. We have

$$\begin{aligned} \nabla f(x, y) &= (2xy, x^2), \\ \nabla g(x, y) &= (3x, 3y), \\ \nabla(x^2 + 3xy) &= (2xy + 3y, x^2 + 3x). \end{aligned}$$

Definition 5.2. *(Differentiability) Function $f(x, y)$ is differentiable at the point $P = (x_0, y_0)$ if the following holds:*

$$\begin{aligned} f(x_0 + h_1, y_0 + h_2) &= f(x_0, y_0) + h_1 f_x(x_0, y_0) + h_2 f_y(x_0, y_0) + h_1\epsilon_1(h_1, h_2) + h_2\epsilon_2(h_1, h_2) \\ &= f(x_0) + (h_1, h_2) \cdot \nabla f(x_0, y_0) + (h_1, h_2) \cdot (\epsilon_1(h_1, h_2), \epsilon_2(h_1, h_2)), \end{aligned} \tag{1}$$

where

$$\lim_{(h_1,h_2)\to(0,0)} \epsilon_1(h_1, h_2) = 0, \qquad \lim_{(h_1,h_2)\to(0,0)} \epsilon_2(h_1, h_2) = 0.$$

Observations Setting $x = x_0 + h_1$ and $y = y_0 + h_2$, (1) becomes

$$f(x, y) = f(x_0, y_0) + (x - x_0)f_x(x_0, y_0) + (y - y_0)f_y(x_0, y_0) + h_1\epsilon_1(h_1, h_2) + h_2\epsilon_2(h_1, h_2).$$

On the other hand, observe that the function

$$T(x, y) = f(x_0, y_0) + (x - x_0)f_x(x_0, y_0) + (y - y_0)f_y(x_0, y_0) \tag{2}$$

describes a plane passing through the location $(x_0, y_0, f(x_0, y_0))$. Therefore, differentiability means that there is a plane, as prescribed above, that approximates the function really well close to that point.

Definition 5.3.

1. *The plane given by (2) is called the tangent plane of f at (x_0, y_0).*
2. *A function is called differentiable in an open set R if it is differentiable for all $(x, y) \in \mathbb{R}$.*

INSERT SECOND DEMO HERE. TOPICS: Tangent planes and partial derivatives

Theorem 5.2. *If $f(x, y)$ has continuous partial derivatives $f_x(x, y)$ and $f_y(x, y)$ everywhere on an open set that contains (a, b), then $f(x, y)$ is differentiable in (a, b).*

Observations

1. The inverse of the theorem does not hold!
2. Therefore, being differential at a point is different from having partial derivatives at a point. See Fig. 2. On the other hand, in one-variable functions, being differentiable means having a derivative!

Theorem 5.3. *If $f(x, y)$ is differentiable at (x_0, y_0), then it is continuous at (x_0, y_0).*

Απόδειξη. Observe that

$$f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0) = h_1 f_x(x_0, y_0) + h_2 f_y(x_0, y_0) + h_1\epsilon_1(h_1, h_2) + h_2\epsilon_2(h_1, h_2),$$

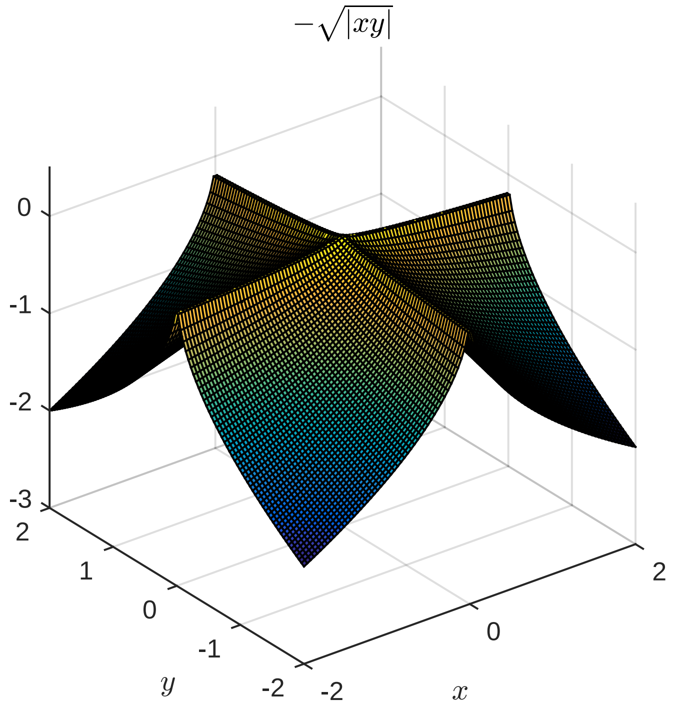
which does go to 0 as $(h_1, h_2) \to 0$. □

Example 5.2. For the function $f(x, y) = xe^y + x^2 y$, we have

$$\nabla f(x, y) = (e^y + 2xy, xe^y + x^2).$$

The partial derivatives are continuous everywhere, therefore f is differentiable everywhere. For example, the equation of the tangent plane that passes through the location

$$x = 2, \quad y = 0, \quad z = 2e^0 + 2^2 \times 0 = 2,$$



Σχήμα 2: A function with partial derivatives $(0,0)$ that is not differentiable there.

is the following:

$$\begin{aligned} z &= f(2,0) + \nabla f(2,0)(x-2, y-0) \\ &= 2 + (1,6)\cdot(x-2,y) \\ &= x + 6y. \end{aligned}$$

Example 5.3. Let the function

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

This function has partial derivatives everywhere outside of $(0,0)$, by known theorems. It also has partial derivatives at zero, as shown by definition:

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

However, it is not even continuous! To see this, set $y = x$, and the function becomes $f(x,y) = \frac{x^2}{2x^2} = \frac{1}{2}$ for all x!

6 Directional derivative

Observation: Observe that

$$\begin{aligned} f_x(x,y) &= \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h}, \\ f_y(x,y) &= \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h}. \end{aligned}$$

Let us set $\mathbf{p} = (x, y)$, $\mathbf{i} = (1, 0)$, and $\mathbf{j} = (0, 1)$. The above equations are then written as

$$\begin{aligned} f_x(\mathbf{p}) &= \lim_{h \to 0} \frac{f(\mathbf{p} + h\mathbf{i}) - f(\mathbf{p})}{h}, \\ f_y(\mathbf{p}) &= \lim_{h \to 0} \frac{f(\mathbf{p} + h\mathbf{j}) - f(\mathbf{p})}{h}. \end{aligned}$$

The first partial derivative describes the rate of change in the x direction, and the second describes the rate of change in the y direction. What about *other* directions? We use the following definition.

Definition 6.1. *(Directional derivative) The directional derivative of f at location $\mathbf{p} = (x, y)$ in the direction of the unit vector $\mathbf{u} = (u_1, u_2)$ is defined as*

$$D_{\mathbf{u}} f(\mathbf{p}) = \lim_{h \to 0} \frac{f(\mathbf{p} + h\mathbf{u}) - f(\mathbf{p})}{h}$$

Theorem 6.1. *If f is differentiable in $\mathbf{p} = (x, y)$, then it has a directional derivative in the direction of the unit vector $\mathbf{u} = (u_1, u_2)$ given by the formula*

$$D_{\mathbf{u}} f(\mathbf{p}) = \mathbf{u} \cdot \nabla f(\mathbf{p}) = u_1 f_x(x, y) + u_2 f_y(x, y).$$

Απόδειξη. Since $f(\mathbf{p})$ is differentiable, we have

$$\begin{aligned} & f(\mathbf{p} + h\mathbf{u}) = f(\mathbf{p}) + \nabla f(\mathbf{p}) \cdot (h\mathbf{u}) + \epsilon(h\mathbf{u}) \cdot (h\mathbf{u}) \\ & \qquad \Rightarrow \lim_{h \to 0} \frac{f(\mathbf{p} + h\mathbf{u}) - f(\mathbf{p})}{h} = \nabla f(\mathbf{p}) \cdot \mathbf{u} + \lim_{h \to 0} \epsilon(h\mathbf{u}) \cdot (h\mathbf{u}) = \nabla f(\mathbf{p}) \cdot \mathbf{u}. \end{aligned}$$

□

Theorem 6.2. *At location $\mathbf{p}$ a differentiable function is increasing the fastest in the direction of the gradient $\nabla f(\mathbf{p})$ at that location, and is decreasing the fastest in the opposite direction.*

Απόδειξη. Note that

$$D_{\mathbf{u}}(\mathbf{p}) = \mathbf{u} \cdot \nabla f(\mathbf{p}) = \|\mathbf{u}\| \|\nabla f(\mathbf{p})\| \cos\theta,$$

where θ is the angle formed between the two vectors $\mathbf{u}$ and $\nabla f(\mathbf{p})$. Therefore, the directional derivative is maximum when the two vectors point in the same direction, i.e., $\theta = 0$, in which case $D_{\mathbf{u}}(\mathbf{p}) = \|\nabla f(\mathbf{p})\|$, and minimum when they point in opposite directions, in which case $D_{\mathbf{u}}(\mathbf{p}) = -\|\nabla f(\mathbf{p})\|$. □

Observation: Observe that the two rates are equal, in absolute terms. Does this make sense to you?

Example 6.1. Let the function $f(x, y) = 4x^2 - xy + 3y^2$.

We will find the directional derivative of f at the location $\mathbf{p} = (2, -1)$ in the direction of the vector $\mathbf{a} = (4, 3)$. Firstly, observe that the given vector $\mathbf{a}$ is not of unit length. Can you understand what would have been the problem if we had used it? We will use the vector

$$\mathbf{u} = \frac{(4, 3)}{\sqrt{16 + 9}} = \left(\frac{4}{5}, \frac{3}{5} \right).$$

Also,

$$\nabla f = (8x - y, -x + 6y) \Rightarrow \nabla f(2, 1) = (17, -8),$$

therefore

$$D_{\mathbf{u}}f(2,-1) = \left(\frac{4}{5}, \frac{3}{5}\right) \cdot (17, -8) = \frac{44}{5}.$$

Secondly, observe that the fastest rate of increase is in the direction of the unit vector

$$\mathbf{u}' = \frac{(17,-8)}{\sqrt{17^2+8^2}}$$

and equals

$$D_{\mathbf{u}}(2,1) \cdot \mathbf{u}' = \frac{(17,-8)\cdot(17,-8)}{\sqrt{17^2+8^2}} = \sqrt{353}.$$

Finally, observe that the fastest rate of decrease is in the direction opposite to that of the unit vector, and so equals

$$D_{\mathbf{u}}(2,1) \cdot (-\mathbf{u}') = -\frac{(17,-8)\cdot(17,-8)}{\sqrt{17^2+8^2}} = -\sqrt{353}.$$

Theorem 6.3. *The gradient of f at the location* **p** *is perpendicular to the contour line going through P, i.e.,*

$$\nabla f(\mathbf{p}) \cdot \mathbf{u} = 0,$$

where **u** *is a unit vector tangent to the contour line.*

(Add figure here)

Example 6.2. Consider the function

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

The contour line passing through point $(2,1)$ is

$$\frac{x^2}{4} + y^2 = \frac{4}{4} + 1 = 2 \Leftrightarrow y = \pm\sqrt{2 - \frac{x^2}{4}},$$

whereas the gradient at that location is

$$\nabla f(x,y) = \left(\frac{x}{2}, 2y\right) = (1,2).$$

(Add figure here)

7 Unconstrained extrema of functions of 2 variables

Definition 7.1. *Let $f(x,y)$ be a function with domain S and a point* $\mathbf{p}_0$.

1. *An open set N is called a* <u>*neighborhood*</u> *of* **p** *if* $\mathbf{p} \in N$.
2. *The value $f(\mathbf{p}_0)$ is a* <u>*(local) maximum*</u> *at the* <u>*(local) maximum point*</u> $\mathbf{p}_0$ *if there is a neighborhood N of* $\mathbf{p}_0$ *such that*
$$f(\mathbf{p}_0) \geq f(\mathbf{p}) \quad \forall \mathbf{p} \in N \cap S.$$
3. *The value $f(\mathbf{p}_0)$ is a* <u>*(local) minimum*</u> *at the* <u>*(local) minimum point*</u> $\mathbf{p}_0$ *if there is a neighborhood N of* $\mathbf{p}_0$ *such that*
$$f(\mathbf{p}_0) \leq f(\mathbf{p}) \quad \forall \mathbf{p} \in N \cap S.$$

4. *f has a (local) extremum at* **p** *if it has a local minimum or maximum at* **p**.

5. *The value $f(\mathbf{p}_0)$ is a global maximum at the global maximum point $\mathbf{p}_0$ if*

$$f(\mathbf{p}_0) \geq f(\mathbf{p}) \quad \forall \mathbf{p} \in S.$$

6. *The value $f(\mathbf{p}_0)$ is a global minimum at the global minimum point $\mathbf{p}_0$ if*

$$f(\mathbf{p}_0) \leq f(\mathbf{p}) \quad \forall \mathbf{p} \in S.$$

7. *f has a global extremum at* **p** *if it has a global minimum or maximum at* **p**.

Observations

1. Plurals of terms are minima, maxima, extrema.
2. Moving from one dimension to two dimensions, things can become much more complicated! See the next theorem:

Theorem 7.1. *(Necessary conditions for existence of local minimum) If f has an extremum at* **p**, *then one of the following holds:*

1. **p** *is a boundary point of S.*
2. **p** *is an interior point of S, but f is not differentiable there.*
3. **p** *is a stationary or critical point, i.e., all the following hold:*
   (α′) **p** *is an interior point of S.*
   (β′) *f is differentiable in p.*
   (γ′) $\nabla f(\mathbf{p}) = \mathbf{0}$.

INSERT THIRD DEMO HERE. TOPICS: extrema conditions

Theorem 7.2. *(Sufficient conditions for local extrema) Let the function $f(x, y)$ have continuous second-order partial derivatives in a neighborhood of (x_0, y_0) and $\nabla f(x_0, y_0) = \mathbf{0}$. Let*

$$D = D(x_0, y_0) = f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0).$$

The following hold:

1. *If $D > 0$ and $f_{xx}(x_0, y_0) < 0$, then $f(x_0, y_0)$ is a local maximum.*
2. *If $D > 0$ and $f_{xx}(x_0, y_0) > 0$, then $f(x_0, y_0)$ is a local minimum.*
3. *If $D < 0$ then $f(x_0, y_0)$ is not an extremum. We call it a saddle point.*
4. *If $D = 0$, then anything goes!*

Example 7.1. Let the function $f(x,y) = x^2 + y^2$ which, of course, we know has a minimum at $x = y = 0$ and no other extrema. Note that $\nabla f(x,y) = (2x, 2y)$, and $f_{xx} = f_{yy} = -2$, therefore the gradient is **0** only at the origin $x = y = 0$, where

$$D = 2 \times 2 - 2 \times 0 \times 2 \times 0 = 4,$$

so indeed Theorem 7.2 tells us we have a minimum at $(0,0)$.

Example 7.2. Now let the function $f(x,y) = -x^2 - y^2$ which, of course, we know has a maximum at $x = y = 0$ and no other extrema. Note that $\nabla f(x,y) = (-2x, -2y)$, and $f_{xx} = f_{yy} = -2$, therefore agai the gradient is **0** only at the origin $x = y = 0$, where

$$D = (-2) \times (-2) - 2 \times 0 \times 2 \times 0 = 4,$$

so indeed Theorem 7.2 tells us we have a maximum at $(0,0)$.

Example 7.3. Finally, let the function $f(x,y) = x^2 - y^2$. Note that $\nabla f(x,y) = (2x, -2y)$, and $f_{xx} = 2$, $f_{yy} = -2$. The gradient is **0** only at the origin $x = y = 0$, where

$$D = 2 \times (-2) - 2 \times 0 \times 2 \times 0 = -4,$$

so we have a saddle point.

Example 7.4. *(Monkey saddle)* Let the function $f(x,y) = x^3 - 3xy^2$. We have that

$$\nabla f = (3x^2 - 3y^2, -3y^2), \quad f_{xx} = 6x, \quad f_{yy} = 6y,$$

therefore $D = 0$ at $(x,y) = (0,0)$, and we cannot use Theorem 7.2.

Example 7.5. We will find the extrema of the function

$$f(x,y) = 3x^3 + y^2 - 9x + 4y$$

whose domain is $\mathbb{R}^2$.

Observe that the function is everywhere differentiable, and without a boundary, so, by Theorem 7.1 the only locations where there might be an extremum are the locations where the gradient is 0. Observe that

$$\nabla f(x,y) = (9x^2 - 9, 2y + 4) = 0 \Leftrightarrow y = -2, x = \pm 1.$$

Therefore, there are two candidates, $(1,-2)$ and $(-1,-2)$. We apply Theorem 7.2. Observe that

$$f_{xx}(x,y) = 18x, \quad f_{yy}(x,y) = 2, \quad f_{xy}(x,y) = 0,$$

therefore at $(x,y) = (1,-2)$ we have

$$D(1,-2) = 18 \times 2 - 0 = 36 > 0, f_{xx} = 18 > 0,$$

and therefore we have a local minimum. On the other hand, at $(x,y) = (-1,-2)$ we have

$$D = -18 \times 2 = -36,$$

and so we have a saddle point.

Example 7.6. *(2015 exam)* We will find the extrema of the function $f(x,y) = x^2y + xy + 1$ on $\mathbb{R}^2$. The function is everywhere differentiable and the domain has no boundary, so if there is a minimum it is on the locations where the gradient is **0**. Observe that

$$\begin{aligned}
\frac{\partial f}{\partial x} &= 2xy + y, \\
\frac{\partial f}{\partial y} &= x^2 + x, \\
\frac{\partial^2 f}{\partial x^2} &= 2y, \\
\frac{\partial^2 f}{\partial y^2} &= 0, \\
\frac{\partial^2 f}{\partial x \partial y} &= 2x + 1.
\end{aligned}$$

Requiring the gradient to be **0** gives

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (0,0) \Leftrightarrow (2xy + y, x^2 + x) = (0,0).$$

Therefore, we must have $x = 0$ or $x = -1$, and in both cases $y = 0$. Therefore, we have two candidates for extrema, $(0,0)$ and $(-1,0)$. Regarding the first point,

$$D(0,0) = f_{xx}(0,0)f_yy(0,0) - f_{xy}^2(0,0) = 0 \times 0 - 1^2 < 0,$$

therefore we have a saddle point. Regarding the second point, we have

$$D(-1,0) = f_{xx}(-1,0)f_{yy}(-1,0) - f_{xy}^2(-1,0) = 0 \times 0 - (-1)^2 < 0,$$

so again we have a saddle point. See Fig. 4.

Observations

1. The roadmap for solving such exercises is as follows:

   (α΄) First, we find candidate extrema, using Theorem 7.1.

   (β΄) Then, we investigate each of these candidates, using Theorem 7.2.

2. The following theorem explains nicely why Theorem 2 holds.

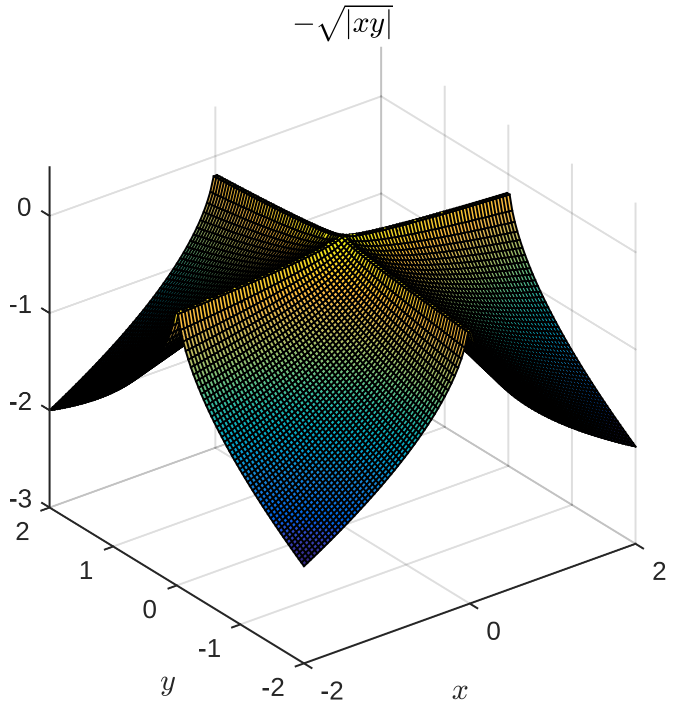
Theorem 7.3. *(Taylor expansion of two-variable functions) Let $f(x,y) = f(\mathbf{x})$ and a point* **a**. *The following holds, provided third-degree partial derivatives exist at an open sent containing point* **a**

$$f(\mathbf{x}) = f(\mathbf{a}) + (\mathbf{x} - \mathbf{a})^T \nabla f(\mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \nabla^2 f(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \ldots,$$

where the <u>*Hessian*</u> *$\nabla^2 f(\mathbf{a})$ is defined as*

$$\nabla^2 f(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$

Observations



Σχήμα 3: Exercise 7.6.

1. The omitted terms are higher-order terms that go to 0 much faster than the second one when $\mathbf{x} \to \mathbf{a}$.
2. When we are at a point where the gradient is not zero, the dominant term is the second one.
3. When we are a a point where the gradient is zero, the dominant term is the third one.

   Let λ_1, λ_2 be the eigenvalues of the Hessian. Then, we have the following cases:

   (α΄) If $\lambda_1, \lambda_2 > 0$, then $\frac{1}{2}(\mathbf{x}-\mathbf{a})^T\nabla^2 f(\mathbf{a})(\mathbf{x}-\mathbf{a})$ looks like a convex bowl and f has a minimum at $\mathbf{a}$.

   (β΄) If $\lambda_1, \lambda_2 < 0$, then $\frac{1}{2}(\mathbf{x}-\mathbf{a})^T\nabla^2 f(\mathbf{a})(\mathbf{x}-\mathbf{a})$ looks like an inverted bowl and f has a maximum at $\mathbf{a}$.

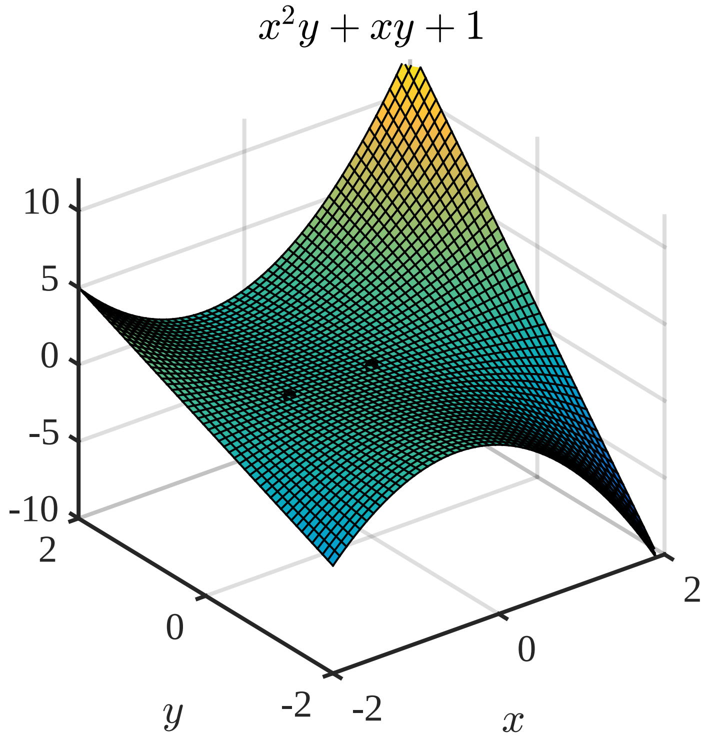
   (γ΄) If $\lambda_1\lambda_2 < 0$, then $\frac{1}{2}(\mathbf{x}-\mathbf{a})^T\nabla^2 f(\mathbf{a})(\mathbf{x}-\mathbf{a})$ looks like a saddle and f does not have an extremum at $\mathbf{a}$.

   (δ΄) If one eigenvalue is positive and the other is 0, then $\frac{1}{2}(\mathbf{x}-\mathbf{a})^T\nabla^2 f(\mathbf{a})(\mathbf{x}-\mathbf{a})$ looks like a drain, and higher order terms become important in some directions.

   (ε΄) If one eigenvalue is negative and the other is 0, then $\frac{1}{2}(\mathbf{x}-\mathbf{a})^T\nabla^2 f(\mathbf{a})(\mathbf{x}-\mathbf{a})$ looks like an inverted drain, and higher order terms become important in some directions.

   (ϛ΄) If both eigenvalues are 0, then $\frac{1}{2}(\mathbf{x}-\mathbf{a})^T\nabla^2 f(\mathbf{a})(\mathbf{x}-\mathbf{a}) = 0$, and higher order terms become important in all directions.

INSERT FOURTH DEMO HERE. TOPICS: eigenvalue cases



Σχήμα 4: Exercise 7.6.

8 Constrained Optimization

Theorem 8.1. *(Necessary condition) Consider the problem of maximizing or minimizing $f(x, y)$ subject to $g(x, y) = 0$ where $f(x, y)$ and $g(x, y)$ are differentiable. Let (x_0, y_0) be an extremum, where $\nabla g \neq \mathbf{0}$. Then, we must have $\nabla f = \lambda \nabla g$ at the extremum.*

Observations

1. There is a simple geometric explanation for this: if the two gradients are not parallel, then we can move along the curve ∇g and increase/decrease the value of the function. *(Insert figure here)*.

2. The method has a blind spot, i.e., locations where $\nabla g = 0$. These must be treated separately.

3. If we want $g(x, y) \geq 0$, then we must also have $\lambda \leq 0$. This can also be explained geometrically.

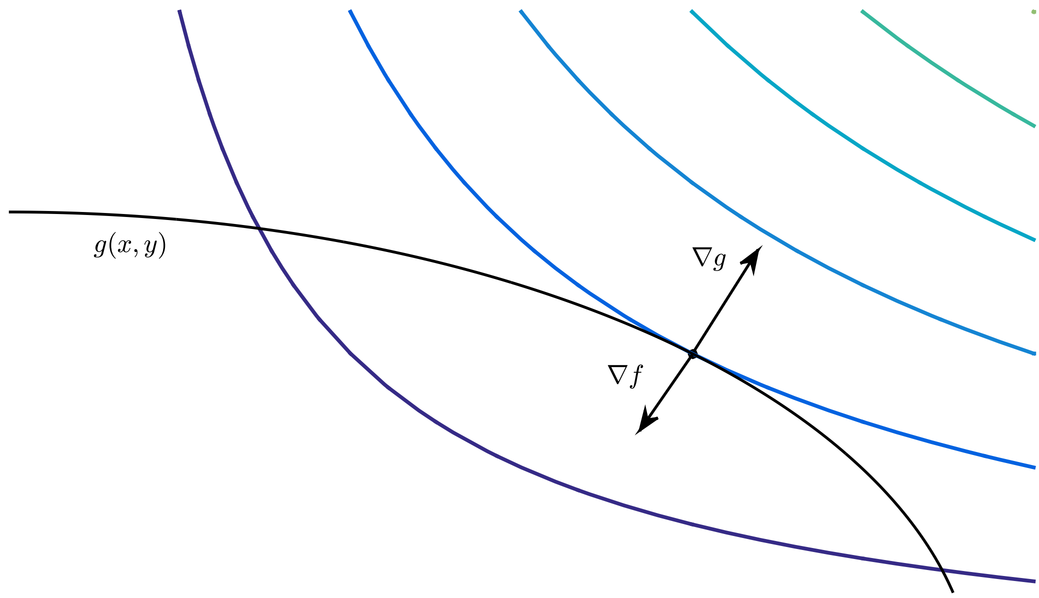
4. This condition is necessary, not sufficient.

5. In practice, we try to solve the system of 3 equations $\nabla f = \lambda \nabla g$ and $g(x, y) = 0$ to find λ, x, y. So, the new condition gives *two* equations and *one* unknown.

6. Another approach is to solve $g(x, y) = 0$ for x or y, substitute it in $f(x, y)$, and arrive at a one-dimensional function, which can be treated as usual. This method does not readily generalize to more dimensions, whereas the method of the theorem does.

Example 8.1. We want to minimize $f(x, y) = xy$ such that $\frac{x^2}{8} + \frac{y^2}{2} = 1$. The condition $\nabla f = \lambda \nabla g$ gives:

$$(y, x) = \lambda \left(\frac{x}{4}, y \right),$$



Σχήμα 5: Theorem 8.1

therefore we have the following three equations:

$$y = \frac{1}{4}\lambda x, \quad x = \lambda y, \quad \frac{x^2}{8} + \frac{y^2}{2} = 1.$$

It follows that $y = \lambda^2 y/4$. If $y = 0$, then $x = 0$, which cannot be a solution. If $\lambda = \pm 2$, then

$$\frac{x^2}{8} + \frac{x^2}{8} = 1 \Rightarrow \frac{x^2}{4} = 1 \Rightarrow x = \pm 2,\ y = \pm 1.$$

See Fig. 6. Finally, also observe that $\nabla g = \mathbf{0}$ only at the origin, which does not satisfy the constraint, so the origin cannot have an extremum.

Example 8.2. We will find extrema for the function $f(x,y) = 3x + 4y$ subject to the condition $x^2 + y^2 = 1$. Note that

$$\nabla f = \lambda \nabla g \Rightarrow (3,4) = \lambda(2x, 2y),$$

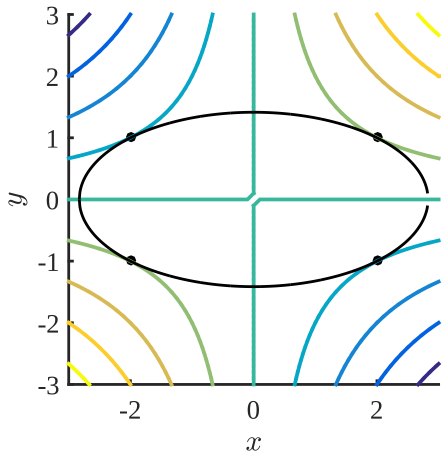
so we have the set of equations

$$2x\lambda = 3, \quad 2\lambda y = 4, \quad x^2 + y^2 = 1.$$

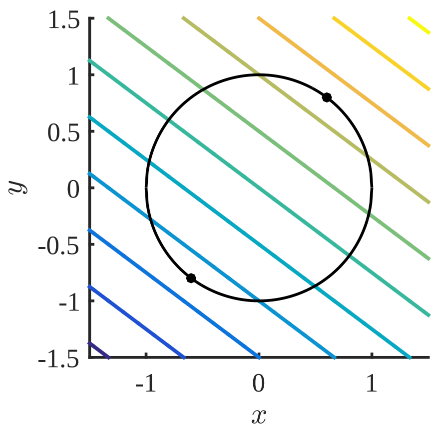
Observe that $\lambda \neq 0$, therefore

$$x = \frac{3}{2\lambda}, \quad y = \frac{2}{\lambda} \Rightarrow \frac{9}{4\lambda^2} + \frac{4}{\lambda^2} = 1 \Rightarrow \lambda^2 = \frac{9}{4} + \frac{16}{4} = \frac{25}{4} \Rightarrow \lambda = \pm\frac{5}{2} \Rightarrow (x,y) = \pm\left(\frac{3}{5}, \frac{4}{5}\right).$$

Again, also observe that $\nabla g = \mathbf{0}$ only at the origin, which does not satisfy the constraint, so the origin cannot have an extremum.



Σχήμα 6: Example 8.1.



Σχήμα 7: Example 8.2.

Example 8.3. We will now consider a problem case. We will minimize $f(x, y) = x + y$ subject to $g(x, y) = x^2 + y^2 = 0$. Obviously, the minimum is located at $(x, y) = (0, 0)$. However, using the theorem gives

$$\nabla f = \lambda \nabla g \Rightarrow (1, 1) = \lambda(2x, 2y),$$

which, together with the condition $x^2 + y^2 = 0$, has no solution! The problem was that $\nabla g = \mathbf{0}$ at the extremum.

Observation: The above example shows that we need to also examine, separately, the critical points where $g(x, y)$ has a vanishing gradient.

Theorem 8.2. *Consider the problem of finding extrema for the function $f(x, y, z)$ subject to two conditions*

$$g_1(x, y, z) = 0, \quad g_2(x, y, z) = 0.$$

Consider an extremum where the two gradients ∇g_1, ∇g_2 are linearly independent, i.e., they are not parallel. Then the following condition holds:

$$\nabla f = \lambda \nabla g_1 + \mu \nabla g_2.$$

Observation: The physical interpretation is as follows: the equations $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$ represent surfaces with the gradients ∇g_1 and ∇g_2 perpendicular to them. Jointly, the two conditions represent a curve, with both ∇g_1 and ∇g_2 vertical to it. We want ∇f to also be perpendicular to the curve, therefore we need it to be on the linear space spanned by the two gradients ∇g_1 and ∇g_2.

Example 8.4. We would like to minimize $f(x, y, z) = x^2 + y^2 + z^2$ subject to $x^2 + y^2 - 1 = 0$, which represents a cylinder, and $x + y + z - 1$=0, which represents a plane. According to the theorem,

$$\nabla f = \lambda \nabla g_1 + \mu \nabla g_2 \Leftrightarrow (2x, 2y, 2z) = \lambda(2x, 2y, 0) + \mu(1, 1, 1),$$

therefore we have the conditions

$$2x = 2x\lambda + \mu, \quad 2y = 2\lambda y + \mu, \quad 2z = \mu, \quad x^2 + y^2 = 1, \quad x + y + z = 1.$$

Removing μ, we have

$$x = x\lambda + z, \quad y = \lambda y + z, \quad x^2 + y^1 = 1, \quad x + y + z = 1.$$

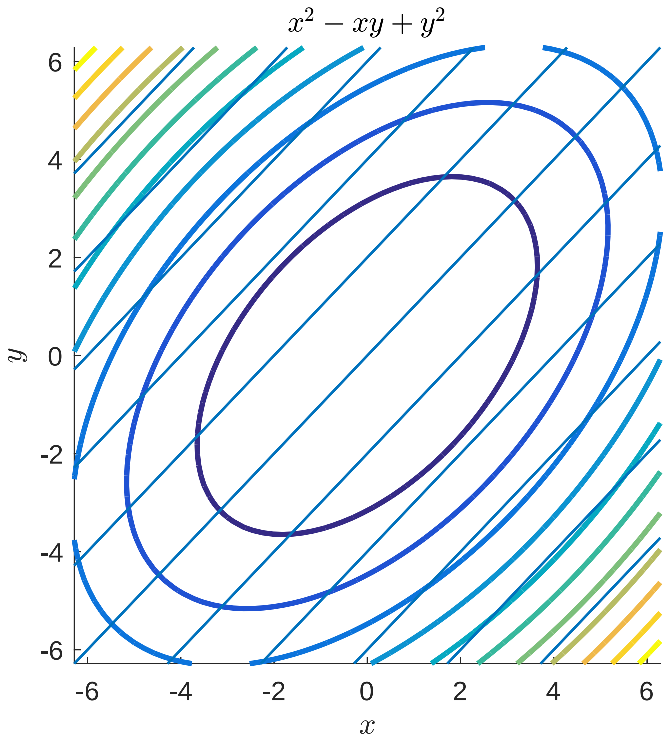
We now take cases:

1. If $\lambda = 1$, then $z = 0$, and we have

$$x^2 + y^2 = 1, \quad x + y = 1 \Rightarrow x^2 + 1 + x^2 - 2x = 1 \Rightarrow 2x^2 = 2x$$
$$\Rightarrow (x, y) = (0, 1) \text{ or } (x, y) = (1, 0).$$

2. If $\lambda \neq 1$, then

$$x = y = \frac{z}{1 - \lambda} \Rightarrow 2x^2 = 1, \quad 2x + z = 1$$
$$\Rightarrow (x, y, z) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 1 - \sqrt{2}\right) \quad \text{or} \quad (x, y, z) = \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 1 + \sqrt{2}\right).$$



Σχήμα 8: Exercise 8.5.

Example 8.5. (2015 exam) Find all the extrema of the function $f(x, y) = x^2 - xy + y^2$ along the line $g(x, y) = 0 \Leftrightarrow y - x - c = 0$, where c is a parameter of the problem not subject to optimization.

Note that

$$\nabla f = (2x - y, -x + 2y), \quad \nabla g = (-1, 1),$$

therefore

$$\nabla f = \lambda \nabla g \Leftrightarrow 2x - y = -\lambda, \quad -x + 2y = \lambda.$$

From these equations, we have that

$$2x - y = x - 2y \Rightarrow x = -y,$$

and since we must also have $y = x + c$, we have that $y = \frac{c}{2}$. Concluding,

$$x = -\frac{c}{2}, \quad y = \frac{c}{2}, \quad \lambda = \frac{3c}{2}.$$

See Fig. 8.

Example 8.6. (2016 exam)

1. We will find all the extrema of the function

$$3x^2 + 2y^2 + 2xy - 10x - 10y + 15$$

   Are they minima or maxima?

2. Then, we will repeat the previous part with the additional constraint that

$$x + y = -1,$$

   using a Lagrange multiplier.

1. We calculate the partial derivatives of $f(x,y)$:

$$\begin{aligned} f_x &= 6x+2y-10, \\ f_y &= 4y+2x-10, \\ f_{xx} &= 6, \\ f_{yy} &= 4, \\ f_{xy} &= 2. \end{aligned}$$

   We calculate points where the gradient is zero:

$$\nabla f = \mathbf{0} \Leftrightarrow (6x+2y-10, 4y+2x-10) = (0,0) \Leftrightarrow 3x+y=5, \quad 2y+x=5 \Leftrightarrow x=1,\ y=2.$$

   Also,

$$D = f_{xx}f_{yy} - (f_{xy})^2 = 6\times 4 - 2^2 = 20 > 0,$$

   therefore we have a minimum at $(x,y) = (1,2)$.

2. We require that

$$\nabla f = \lambda \nabla g \Leftrightarrow 6x+2y-10 = \lambda, \quad 4y+2x-10=\lambda.$$

   Solving for λ, we find $y = 2x$, and using the constraint $x+y=-1$ we find that $x = -\frac{1}{3}$, $y = -\frac{2}{3}$. We note that the given function is a positive definite quadratic form, therefore this extremum must be a minimum.

9 Double Integrals

Refer to book

Example 9.1. We will calculate the double integral of the function $f(x,y) = x^2y^2$ in the triangle formed by the points $(0,0)$, $(0,1)$, and $(1,1)$.

We will use both forms of Fubini's theorem and arrive at the same result. In the first form, we have

$$\begin{aligned} \iint_A x^2y^2\,dA &= \int_0^1 \left(\int_x^1 x^2y^2\,dy\right)dx = \int_0^1 x^2 \left(\int_x^1 \left(\frac{y^3}{3}\right)'\,dy\right) \\ &= \int_0^1 x^2\left(\frac{1}{3} - \frac{x^3}{3}\right)dx = \int_0^1 \left(\frac{x^2}{3} - \frac{x^5}{3}\right)dx = \left[\frac{x^3}{9} - \frac{x^6}{18}\right]_0^1 = \frac{1}{9} - \frac{1}{18} = \frac{1}{18}. \end{aligned}$$

Using the second form, we have

$$\begin{aligned} \iint_A x^2y^2\,dA &= \int_0^1 \left(\int_0^y x^2y^2\,dx\right)dy = \int_0^1 y^2 \left(\int_0^y \left(\frac{x^3}{3}\right)'\,dx\right)dy \\ &= \int_0^1 \frac{y^5}{3}\,dy = \int_0^1 \left(\frac{y^6}{18}\right)' = \frac{1}{18}. \end{aligned}$$

10 Vectors

Definition 10.1. *(Vectors and their operations)*

1. *We define a* **vector** $\mathbf{x}$ *of dimension* $n \in \mathbb{N}$ *to be an ordered set of* n **components** $x_i \in \mathbb{R}$, $1 \leq i \leq n$*:*

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \quad \text{or} \quad \mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}.$$

   The first notation is a **column vector**, *the second notation a* **row vector**. *In the rest, we will mostly be using the row vector notation, for convenience, but all definitions and theory apply for both notations, mutatis mutandis.*

2. *We define the* **transpose** *of a row (column) vector to be the corresponding column (row) vector:*

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}, \quad \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}.$$

3. *We define vector addition as follows:*

$$\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} + \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 & x_2 + y_2 & \dots & x_n + y_n \end{bmatrix}.$$

4. *We define multiplication of a vector with a real number as*

$$a \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} ax_1 & ax_2 & \dots & ax_n \end{bmatrix}.$$

5. *We define the* **null vector** *or* **zero vector**

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & \dots & 0 \end{bmatrix}.$$

6. *We denote the space of all vectors as* $\mathbb{R}^n$, *or, more specifically* $\mathbb{R}^{n\times 1}$, *for column vectors, and* $\mathbb{R}^{1\times n}$, *for row vectors.*

Theorem 10.1. *(Vector operation properties) The following properties hold for all* $\mathbf{x},\ \mathbf{y},\ \mathbf{z} \in \mathbb{R}^n$*:*

1. *(Commutativity)* $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. *(Associativity)* $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$.
3. *(Identity element)* $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x}$.
4. *(Inverse element)* $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
5. $a(b\mathbf{x}) = (ab)\mathbf{x}$.
6. $1\mathbf{x} = \mathbf{x}$.
7. *(Distributive property)* $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$.
8. *(Distributive property)* $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$.

Observations

1. The proofs of these properties follow easily from the manner we have defined the operations of the vectors.
2. Vectors are not the only mathematical constructs to satisfy the properties of the above theorem. Other constructs exist as well. For example, real functions also satisfy the same properties, mutatis mutandis. For this reason, all sets that satisfy the above theorem are called **vector spaces**.

Definition 10.2. *(Vector length) We define the **length** or **norm** $\|\mathbf{x}\|$ of a vector $\mathbf{x}$ as*

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Theorem 10.2. *(Properties of the length) The length of a vector satisfies the following properties, for all $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$, and all $a \in \mathbb{R}$:*

1. $\|\mathbf{x}\| \geq 0$.
2. $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.
3. $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$.
4. *(Triangle inequality)* $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Observations

1. The proof of these properties follows easily from the manner we have defined the operations (except the triangle inequality, which can most easily be proved using the Cauchy-Schwarz inequality, shown later).
2. Again, vectors and their length, defined as above, are not the only mathematical constructs to satisfy the properties of the above theorem. Other constructs exist as well. For example, vectors also satisfy these properties if we define the length of a vector as

$$\|\mathbf{x}\| = \left(x_1^p + x_2^p + \cdots + x_n^p\right)^{\frac{1}{p}},$$

where p is *any* positive real number. Functions also satisfy the same properties if we define the length of a function as, for example,

$$\|f\| = \left(\int_0^1 |f(x)|^p \, dx\right)^{\frac{1}{p}},$$

for any p positive real number. All vector spaces equipped with a norm that satisfies the above theorem are called **normed spaces**.

Definition 10.3. *(Inner product) We define the **inner product** of two vectors $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$ as*

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

Theorem 10.3. *(Properties of the inner product) For all $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$, the following properties hold:*

1. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.

2. $(a\mathbf{x})\cdot\mathbf{y} = a(\mathbf{x}\cdot\mathbf{y})$.

3. $(\mathbf{x}+\mathbf{y})\cdot\mathbf{z} = \mathbf{x}\cdot\mathbf{z}+\mathbf{y}\cdot\mathbf{z}$.

4. $\mathbf{x}\cdot\mathbf{x}\geq 0$.

5. $\mathbf{x}\cdot\mathbf{x} = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$.

Observations

1. Observe that $\mathbf{x}\cdot\mathbf{x} = \|\mathbf{x}\|^2$.

2. The above thoerem is easily proved using the definition of the inner product.

3. Any vector space equipped with an inner product satisfying the above theorem is called an **inner product space**. For example, The following will serve as an inner product of two real functions:

$$f\cdot g = \int_a^b f(x)g(x)\,dx.$$

Theorem 10.4. *(Cauchy-Schwarz Inequality) For all $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, the following inequality holds:*

$$|\mathbf{x}\cdot\mathbf{y}|\leq\|\mathbf{x}\|\|\mathbf{y}\|.$$

Απόδειξη. Observe that, for any $\lambda\in\mathbb{R}$,

$$0\leq\|\mathbf{x}-\lambda\mathbf{y}\|^2 = (\mathbf{x}-\lambda\mathbf{y})\cdot(\mathbf{x}-\lambda\mathbf{y}) = \mathbf{x}\cdot\mathbf{x}-2\lambda\mathbf{y}\cdot\mathbf{x}+\lambda^2\mathbf{y}\cdot\mathbf{y} = \lambda^2\|\mathbf{y}\|^2-(2\mathbf{y}\cdot\mathbf{x})\lambda+\|\mathbf{x}\|^2.$$

The above trinomial is always positive, so the discriminant must be non-positive, i.e.,

$$4(\mathbf{y}\cdot\mathbf{x})^2-4\|\mathbf{x}\|^2\|\mathbf{y}\|^2\leq 0\Leftrightarrow|\mathbf{x}\cdot\mathbf{y}|\leq\|\mathbf{x}\|\|\mathbf{y}\|.$$

□

Definition 10.4. *(Angles between vectors)*

1. *When $\mathbf{x}\cdot\mathbf{y}=0$, the two vectors $\mathbf{x}$, $\mathbf{y}$ are called **orthogonal**.*

2. *We define the angle θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ as*

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

Observe that this definition is possible due to the Cauchy-Schwarz inequality, which ensures that $\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\in[-1,1]$

Definition 10.5. *(Linear Independence) A set of m vectors $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m$ are called **linearly independent** if*

$$a_1\mathbf{x}_1+a_2\mathbf{x}_2+\cdots+a_m\mathbf{x}_m=\mathbf{0}\Rightarrow a_1=a_2=\cdots=a_m=0.$$

*Otherwise, they are called **linearly dependent**.*

Example 10.1. The vectors $\begin{bmatrix}2 & 1 & 0\end{bmatrix}^T$, $\begin{bmatrix}1 & 2 & 3\end{bmatrix}^T$ and $\begin{bmatrix}4 & 5 & 6\end{bmatrix}^T$ are linearly dependent because

$$\begin{bmatrix}2\\1\\0\end{bmatrix} + 2\begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}4\\5\\6\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Observations

1. If a set of vectors are linearly dependent, then there is a set of numbers $a_1, a_2, \ldots, a_m$, not all zero, such that $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_m\mathbf{x}_m = \mathbf{0}$. This readily gives that there is one vector $\mathbf{x}_k$ such that

$$\mathbf{x}_k = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_{k-1}\mathbf{x}_{k-1} + b_{k+1}\mathbf{x}_{k+1} + \cdots + b_m\mathbf{x}_m,$$

i.e., that vector can be written as a linear combination of the rest.

Definition 10.6. *(Subspaces) A subspace S of $\mathbb{R}^n$ is a set closed under vector addition and multiplication with a real number, i.e., if $\mathbf{x}_1$ and $\mathbf{x}_2$ belong to S, then so does $k_1\mathbf{x}_1 + k_2\mathbf{x}_2$, for all $k_1, k_2 \in \mathbb{R}$.*

Observation: Are lines and planes subspaces of $\mathbb{R}^3$?

Example 10.2. Let $S = \{\mathbf{x} : a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0\}$. Then it is straightforward to show that S is a subspace. Indeed, let $\mathbf{x},\ \mathbf{y} \in S$ with

$$\mathbf{x} = \begin{bmatrix}x_1 & x_2 & \ldots & x_n\end{bmatrix}, \ \mathbf{y} = \begin{bmatrix}y_1 & y_2 & \ldots & y_n\end{bmatrix},$$
$$k_1\mathbf{x} + k_2\mathbf{y} = \begin{bmatrix}k_1x_1 + k_2y_1 & k_1x_2 + k_2y_2 & \ldots & k_1x_n + k_2y_n\end{bmatrix},$$

and observe that

$$a_1(k_1x_1 + k_2y_1) + a_2(k_1x_2 + k_2y_2) + \cdots + a_n(k_1x_n + k_2y_n)$$
$$= k_1(a_1x_1 + a_2x_2 + \cdots + a_nx_n) + k_2(a_1y_1 + a_2y_2 + \cdots + a_ny_n) = 0.$$

Definition 10.7. *(Orthogonal Subspaces)*

1. *Two subspaces S_1 and S_2 are called **orthogonal** if, for all $\mathbf{s}_1 \in S_1$ and all $\mathbf{s}_2 \in S_2$, we have $\mathbf{s}_1 \cdot \mathbf{s}_2 = 0$.*
2. *The **orthogonal complement** $S^\perp$ of a subspace S is the set of all vectors that are orthogonal to S (which is easy to see it is a subspace itself).*

Observations

1. If S_1 and S_2 are orthogonal, which vectors do they have in common?
2. What are the orthogonal complements of lines and planes in $\mathbb{R}^3$?

Definition 10.8. *(Bases) Let a subspace S and a set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$.*

1. *If, for any $\mathbf{y} \in S$ we can write $\mathbf{y} = k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \cdots + k_m\mathbf{x}_m$, for some $k_1, k_2, \ldots, k_m \in \mathbb{R}$, then we say that the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ **span** S.*
2. *If, in addition, the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ are linearly independent, then we call them a **basis** of S.*
3. *If, in addition, all basis vectors are orthogonal with each other, the basis is called **orthogonal**.*
4. *If, in addition, the vectors have unit length, the basis is called **orthonormal**.*

Theorem 10.5. *(Dimension) For each subspace S, there is a number k called its* ***dimension*** *such that*

1. *Any $p > k$ vectors are linearly dependent.*
2. *Any $p = k$ linearly independent vectors span S.*
3. *No $p < k$ vectors can span S.*

Theorem 10.6. *(Orthonormal basis for $\mathbb{R}^n$) The space $\mathbb{R}^n$ has dimension n. One orthonormal basis is the following:*

$$\begin{aligned} e_1 &= [1\,0\,0\,\ldots,0],\\ e_2 &= [0\,1\,0\,\ldots,0],\\ e_3 &= [0\,0\,1\,\ldots,0],\\ \ldots &\ldots \ldots\\ e_n &= [0\,0\,0\,\ldots,1]. \end{aligned}$$

Observation: Given any two vectors **x** and **y**, we can break **y** in two components, one component $\mathbf{y}_{\|}$ parallel to **x** (its **projection** on **x**) and one component $\mathbf{y}_{\perp}$ perpendicular to **x**:

$$\mathbf{y}_{\|} = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|^2}\mathbf{x}, \quad \mathbf{y}_{\perp} = \mathbf{y} - \mathbf{y}_{\|}.$$

Indeed, $\mathbf{y}_{\|}$ is obviously parallel to **x**, and, furthermore, observe that

$$\mathbf{x}\cdot\mathbf{y}_{\perp} = \mathbf{x}\cdot\mathbf{y} - \frac{\mathbf{x}\cdot\mathbf{x}}{\|\mathbf{x}\|^2}(\mathbf{x}\cdot\mathbf{y}) = 0.$$

This observation explains the formulas used in the Gram-Schmidt process.

Definition 10.9. *(Gram-Schmidt process) Given a set $\mathbf{v}_1,\ \mathbf{v}_2,\ldots,\mathbf{v}_k$ of vectors, the following process creates a set of linearly independent vectors of unit length $\mathbf{e}_1,\ \mathbf{e}_2,\ \ldots,\mathbf{e}_p$, orthogonal to each other, that span the same subspace as the original set $\mathbf{v}_1,\ \mathbf{v}_2,\ldots,\mathbf{v}_k$, i.e., an orthonormal basis of that space:*

$$\begin{aligned} \mathbf{q}_1 = \mathbf{v}_1, &\quad \mathbf{e}_1 = \tfrac{\mathbf{q}_1}{\|\mathbf{q}_1\|},\\ \mathbf{q}_2 = \mathbf{v}_2 - (\mathbf{v}_2\cdot\mathbf{e}_1)\mathbf{e}_1, &\quad \mathbf{e}_2 = \tfrac{\mathbf{q}_2}{\|\mathbf{q}_2\|},\\ \mathbf{q}_3 = \mathbf{v}_3 - (\mathbf{v}_3\cdot\mathbf{e}_1)\mathbf{e}_1 - (\mathbf{v}_3\cdot\mathbf{e}_2)\mathbf{e}_2, &\quad \mathbf{e}_3 = \tfrac{\mathbf{q}_3}{\|\mathbf{q}_3\|},\\ \ldots\ldots\ldots &\quad \ldots\ldots\ldots \end{aligned}$$

Observations

1. If at some stage $\mathbf{q}_i = \mathbf{0}$, then we just skip it and move to creating the next vector $\mathbf{q}_{i+1}$. Therefore, p might be smaller than k.
2. Proof goes by induction, but is length and so is omitted. The basic idea is that, at each step, we subtract from the new vector its part that can be written as a linear combination of the vectors already in the set, its projections on those vectors already added to the set.

Example 10.3. (2016 exam) We will find an orthonormal basis for the subspace spanned by the vectors

$$\mathbf{v}_1 = [1 \quad 1 \quad 1 \quad 1], \quad \mathbf{v}_2 = [1 \quad 2 \quad 1 \quad 0], \quad \mathbf{v}_3 = [1 \quad 1 \quad 2 \quad 1]$$

using the Gram-Schmidt process.

The first vector in the basis is

$$\mathbf{e}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Let

$$\begin{aligned}
\mathbf{q}_2 &= \mathbf{v}_2 - (\mathbf{e}_1 \cdot \mathbf{v}_2)\mathbf{e}_1 \\
&= [1 \quad 2 \quad 1 \quad 0] - \left(\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \cdot [1 \quad 2 \quad 1 \quad 0]\right) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\
&= [1 \quad 2 \quad 1 \quad 0] - [1 \quad 1 \quad 1 \quad 1] \\
&= [0 \quad 1 \quad 0 \quad -1],
\end{aligned}$$

therefore the second vector in the basis is

$$\mathbf{e}_2 = \frac{\mathbf{q}_2}{\|\mathbf{q}_2\|} = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

Finally, let

$$\begin{aligned}
\mathbf{q}_3 &= \mathbf{v}_3 - (\mathbf{e}_1 \cdot \mathbf{v}_3)\mathbf{e}_1 - (\mathbf{e}_2 \cdot \mathbf{v}_3)\mathbf{e}_2 \\
&= [1 \quad 1 \quad 2 \quad 1] - \left(\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \cdot [1 \quad 1 \quad 2 \quad 1]\right) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\
&\quad - \left(\left[0 \quad \frac{\sqrt{2}}{2} \quad 0 \quad -\frac{\sqrt{2}}{2}\right) \cdot [1 \quad 1 \quad 2 \quad 1]\right) \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix} \\
&= [1 \quad 1 \quad 2 \quad 1] - \frac{5}{2} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} - 0 \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \end{bmatrix},
\end{aligned}$$

with

$$\|\mathbf{q}_3\| = \sqrt{\frac{1}{16} + \frac{1}{16} + \frac{9}{16} + \frac{1}{16}} = \frac{\sqrt{3}}{2}.$$

Therefore,

$$\mathbf{e}_3 = \frac{\mathbf{q}_3}{\|\mathbf{q}_3\|} = \begin{bmatrix} -\frac{1}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} & \frac{3}{\sqrt{3}} & -\frac{1}{2\sqrt{3}} \end{bmatrix}.$$

11 Matrices

Definition 11.1. *(Matrices and their operations)*

1. *We define a* ***matrix*** A *of size* $m \times n$*,* $m, n \in \mathbb{N}$*, to be a collection of* ***elements*** $a_{ij} \in \mathbb{R}$ *arranged in* m ***rows*** *and* n ***columns*** *as follows:*

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

2. *We define the* ***transpose*** A^T *of a matrix* A *as*

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}.$$

3. *We define the addition* $A + B$ *of two matrices* A *and* B *of the same size* $m \times n$ *as*

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{bmatrix}.$$

4. *We define the multiple of a matrix* A *with a real number* k *as*

$$k \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ ka_{21} & ka_{22} & \dots & ka_{2n} \\ \dots & \dots & \dots & \dots \\ ka_{m1} & ka_{m2} & \dots & ka_{mn} \end{bmatrix}.$$

5. *We define the* ***zero matrix***

$$0_{m\times n} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

6. *We define the product* AB *of a matrix* A *of size* $m \times k$ *with a matrix* B *of size* $k \times n$ *as the matrix* C *of size* $m \times n$ *for which*

$$c_{ij} = \sum_{1 \le p \le k} a_{ip} b_{pj}, \quad 1 \le i \le m,\ 1 \le j \le n.$$

7. *We denote the space of all matrices of size* $m \times n$ *as* $\mathbb{R}^{m\times n}$*.*

Observations

1. Observe that vectors may be thought of as special cases of matrices of size $n \times 1$ or $1 \times n$.

2. Among their many applications, matrices can be used to describe linear systems. Remember that the system

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2, \\ \dots & \dots & \dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m, \end{array}$$

may be written more succinctly as $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{11} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}.$$

Theorem 11.1. *(Matrices are a vector space) Let A, B, C be any matrices in $\mathbb{R}^{m\times n}$. The following properties hold.*

1. *(Commutativity)* $A + B = B + A$.
2. *(Associativity)* $A + (B + C) = (A + B) + C$.
3. *(Identity element)* $A + 0_{m\times n} = 0_{m\times n} + A$.
4. *(Inverse element)* $A + (-A) = 0_{m\times n}$.
5. $a(bA) = (ab)A$.
6. $1A = A$.
7. *(Distributive property)* $a(A + B) = aA + aB$.
8. *(Distributive property)* $(a + b)A = aA + bA$.

Observations

1. The proof of these properties follows easily from the manner we have defined the operations of the matrices. Therefore, $R^{m\times n}$ is a vector space, with all this entails (i.e., we can define bases, subspaces, etc.)

2. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad \mathbf{a}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \dots \\ a_{mi} \end{bmatrix}, \ 1 \le i \le n, \ \ \mathbf{b}_j = \begin{bmatrix} b_{j1} & b_{j2} & \dots & b_{jn} \end{bmatrix}, \ 1 \le j \le m.$$

i.e., the vectors $\mathbf{a}_i$ are the columns of the matrix and the vectors $\mathbf{b}_j$ are its rows. Let also vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 & z_2 & \dots & z_m \end{bmatrix}.$$

Then we can write

$$A\mathbf{x} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ \dots\\ x_n\end{bmatrix} = \sum_{i=1}^{n} x_i\mathbf{a}_i.$$

Therefore, we can think of a matrix as a collection of vectors, and multiplying it with another vector creates an element of the subspace spanned by these vectors.

Likewise, we write

$$\mathbf{z}A = \begin{bmatrix}z_1 & z_2 & \dots & z_m\end{bmatrix}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \dots\\ \mathbf{b}_m\end{bmatrix} = \sum_{j=1}^{m} z_j\mathbf{b}_j,$$

which has a similar interpretation.

Theorem 11.2. *(Properties of matrices) Let A a matrix of size $m \times k$ and B a matrix of size $k \times n$.*

1. $(AB)^T = B^T A^T$.
2. *However, $AB = BA$ does not hold in general, even if $m = n = k$, so that the dimensions of the two matrices allow both products to make sense.*

Definition 11.2. *(Fundamental subspaces) Let A be a $m \times n$ matrix.*

1. *We define the **range**, or **column space**, or **image** $R(A)$ of A to be the subspace*

$$R(A) = \{\mathbf{b} \in \mathbb{R}^{m\times 1} : \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^{n\times 1}\}.$$

*We call the dimension of this space the **rank** of the matrix, and we denote it as* $\operatorname{rank}(A)$.

2. *The **row space** of A is $R(A^T)$. Its dimension is also* $\operatorname{rank}(A)$.
3. *We define the **null space** or **kernel** $N(A)$ of A to be the subspace*

$$N(A) = \{\mathbf{x} \in \mathbb{R}^{n\times 1} : A\mathbf{x} = \mathbf{0}\}.$$

*The dimension of this space is called **nullity** and denoted by* $\operatorname{null}(A)$.

4. *The **left null space** is $N(A^T)$.*

Observations

1. It is straightforward to verify that these sets are indeed subspaces, i.e., they are closed under linear combinations. Indeed, let, for example, $\mathbf{b}_1, \mathbf{b}_2 \in R(A)$. Then, $\mathbf{b}_1 = A\mathbf{x}_1$, $\mathbf{b}_2 = A\mathbf{x}_2$, for some $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^m$. Then $k_1\mathbf{b}_1 + k_2\mathbf{b}_2$ also belongs to $R(A)$ because

$$A(k_1\mathbf{x}_1 + k_2\mathbf{x}_2) = k_1(A\mathbf{x}_1) + k_2(A\mathbf{x}_2) = k_1\mathbf{b}_1 + k_2\mathbf{b}_2.$$

Theorem 11.3. *(Fundamental Theorem of Linear Algebra)*

$$N(A) = (R(A^T))^{\perp}, \quad N(A^T) = (R(A))^{\perp}.$$

Definition 11.3. *(Full rank matrices) Let A be a matrix of size $m \times n$. If either the columns or the rows are linearly independent, we call the matrix **full rank**, otherwise we call it **rank deficient**.*

12 Square Matrices

Definition 12.1. *(Square matrices)*

1. *Matrices of size $n \times n$ are called **square** matrices.*
2. *The matrix*
$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$
*is called the **identity matrix**.*
3. *If A is square and $A^T = A$, then A is called **symmetric**.*
4. *If A is square and*
$$AB = BA = I,$$
*then B is called the **inverse** of A, and denoted by A^{-1}. It is simple to see that the inverse is unique. If the inverse of a matrix exists, we call that matrix **non-singular**, otherwise we call it **singular**.*
5. *Given the matrix A of size $m \times n$, let A_{ij} be the matrix of size $(m-1) \times (n-1)$ created from A if we remove row i and column j.*
6. *We define the **determinant** $\det(A)$ or $|A|$ of square matrix A by the recursive equation*
$$|A| = (-1)^{i+1} a_{i1} |A_{i1}| + (-1)^{i+2} a_{i2} |A_{i2}| + \cdots + (-1)^{i+n} a_{in} |A_{in}|,$$
which holds for any row i.

Observation: We can find the inverse of a matrix A, if it exists, through the Gauss-Jordan method as shown in the following example.

Example 12.1. (2016 exam) We will find the inverse of the following matrix, using the Gauss-Jordan elimination method
$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 1 & 1 & 0 \end{bmatrix}.$$

We have:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 3 & 4 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & -1 & -3 & -1 & 0 & 1 \end{array}\right] \qquad (R2 = R2 - 2R1,\ R3 = R3 - R1)$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 \end{array}\right] \qquad (R3 = R3 - R2)$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & -1 & 0 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{array}\right] \qquad (R2 = -R2,\ R3 = -R3)$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 4 & -3 & 2 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{array}\right] \qquad (R2 = R2 - R3)$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 4 & -3 & 3 \\ 0 & 1 & 0 & 4 & -3 & 2 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{array}\right] \qquad (R1 = R1 - R3)$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -4 & 3 & -1 \\ 0 & 1 & 0 & 4 & -3 & 2 \\ 0 & 0 & 1 & -1 & 1 & -1 \end{array}\right] \qquad (R1 = R1 - 2R2)$$

Therefore, the inverse of the given matrix is

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

Theorem 12.1. *(Properties of square matrices)*

1. $(AB)^{-1} = B^{-1}A^{-1}$.
2. $(A^T)^{-1} = (A^{-1})^T$.
3. $$|A| = (-1)^{i+1}a_{1i}|A_{1i}| + (-1)^{i+2}a_{2i}|A_{2i}| + \cdots + (-1)^{i+n}a_{ni}|A_{ni}|,$$
   so we can select any column as well, when calculating the determinant.
4. $|A^T| = |A|$.
5. $|aA| = a^n|A|$.
6. $|AB| = |A||B|$.
7. *If B is created by exchanging two rows or two columns of matrix A, then $|B| = -|A|$.*
8. *If we add to any column or row the multiple of another column or row (respectively) the determinant does not change.*

9. *Let the k column vectors $\mathbf{a}_1, \ \mathbf{a}_2, \ \ldots, \mathbf{a}_k$ and the matrix*

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_k \end{bmatrix}$$

The vectors are linearly independent if we can delete some of the rows of A such that a matrix of size $k \times k$ of non-zero determinant is created.

Theorem 12.2. *(Non-singular matrices) Let A be a square matrix of size $n \times n$. The following statements are equivalent:*

1. *The matrix has an inverse A^{-1}.*
2. *The determinant $|A| \neq 0$.*
3. $\text{rank}(A) = n$.
4. *A has n linearly independent columns and n linearly independent rows.*
5. $N(A) = N(A^T) = \{\mathbf{0}\}$.
6. *All eigenvalues of A (to be defined shortly) are non-zero.*

Example 12.2. (2015 exam) We will answer the following questions:
What is the rank of the following matrices?

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix},$$

To which of the above matrices does the vector $[1 \quad -1 \quad 10]^T$ belong to the range?

Regarding the first matrix, observe that the first and the third rows are equal, so the rank cannot be 3. On the other hand, the determinant of the matrix that is created if we remove the last row and the last column is $1 \times 2 - 1 \times 1 = 1$, therefore the rank is 2. For the vector $[1 \quad -1 \quad 10]^T$ to belong to the range of that matrix, it is necessary for the following system to have a solution:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 10 \end{bmatrix}.$$

However, subtracting the first system from the last one shows that the system has no solution, so the given vector does not belong to the range of the matrix.

Regarding the second matrix, all rows are equal, so the rank must be 1. Again, for the vector $[1 \quad -1 \quad 10]^T$ to belong to the range of that matrix, it is necessary for the following system to have a solution:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 10 \end{bmatrix}.$$

The system clearly does not have a solution, so the given vector does not belong to the range of the matrix.

Regarding the final matrix, observe that the determinant

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1 \times (5 \times 9 - 6 \times 8) - 2 \times (4 \times 9 - 6 \times 7) + 3 \times (4 \times 8 - 5 \times 7) = 0,$$

so again the rank of that matrix cannot be 3. Indeed, if we try to solve the system

$$\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{bmatrix} \times \begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}0\\ 0\\ 0\end{bmatrix},$$

we find that the system becomes:

$$\begin{aligned}
&\left(\begin{array}{ccc|c}1 & 2 & 3 & 0\\ 4 & 5 & 6 & 0\\ 7 & 8 & 9 & 0\end{array}\right)\\
\Leftrightarrow &\left(\begin{array}{ccc|c}1 & 2 & 3 & 0\\ 0 & -3 & -6 & 0\\ 0 & -6 & -12 & 0\end{array}\right) \qquad (R_2 = R_1 - 4R_1, \quad R_3 = R_3 - 7R_1)\\
\Leftrightarrow &\left(\begin{array}{ccc|c}1 & 2 & 3 & 0\\ 0 & 1 & 2 & 0\\ 0 & 0 & 0 & 0\end{array}\right) \qquad (R_2 = -\frac{1}{3}R_2, \quad R_3 = R_3 - 2R_1)
\end{aligned}$$

therefore one solution is $x_3 = 1$, $x_1 = 1$, $x_2 = -2$. Regarding the given vector, we try to solve the system

$$\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{bmatrix} \begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}1\\ -1\\ 10\end{bmatrix}.$$

as follows:

$$\begin{aligned}
&\left(\begin{array}{ccc|c}1 & 2 & 3 & 1\\ 4 & 5 & 6 & -1\\ 7 & 8 & 9 & 10\end{array}\right)\\
\Leftrightarrow &\left(\begin{array}{ccc|c}1 & 2 & 3 & 1\\ 0 & -3 & -6 & -5\\ 0 & -6 & -12 & 3\end{array}\right) \qquad (R_2 = R_1 - 4R_1, \quad R_3 = R_3 - 7R_1)\\
\Leftrightarrow &\left(\begin{array}{ccc|c}1 & 2 & 3 & 1\\ 0 & 1 & 2 & \frac{5}{3}\\ 0 & 0 & 0 & 13\end{array}\right), \qquad (R_2 = -\frac{1}{3}R_2, \quad R_3 = R_3 - 2R_1)
\end{aligned}$$

which has no solution, therefore the vector $[1 \quad -1 \quad 10]^T$ does not belong to the range of that matrix as well

Definition 12.2. *(Characteristic polynomial) We define the* ***characteristic polynomial*** *of matrix A be the polynomial of λ of order n given by*

$$|A - \lambda I| = 0.$$

Its n roots are called eigenvalues. If λ is such a root, any vector x such that $A\mathbf{x} = \lambda\mathbf{x}$ is called a (right) eigenvector and any vector $\mathbf{x}$ such that $\mathbf{x}A = \lambda x$ is called a left eigenvector.

Example 12.3. We will find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix}1 & -1\\ 2 & 4\end{bmatrix}.$$

The characteristic polynomial is

$$det(A - \lambda I) = 0 \Leftrightarrow \begin{vmatrix} 1-\lambda & -1 \\ 2 & 4-\lambda \end{vmatrix} = 0 \Leftrightarrow (1-\lambda)(4-\lambda) + 2 = 0$$

$$\Leftrightarrow \lambda^2 - 5\lambda + 6 = 0 \Leftrightarrow \lambda_1 = 2, \quad \lambda_2 = 3.$$

Regarding the first eigenvector, we have

$$Ax = 2x \Leftrightarrow \begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow x_1 + x_2 = 0 \Leftrightarrow x_1 = -x_2.$$

Therefore, one eigenvector is $\left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]^T$.

Regarding the second eigenvector, we have

$$Ax = 3x \Leftrightarrow \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow 2x_1 + x_2 = 0 \Leftrightarrow x_2 = -2x_1.$$

Therefore, one eigenvector is $\left[\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right]^T$.

Observations

1. Eigenvalues are special, in that when λ is an eigenvalue, there are non-zero vectors **x** such that $A\mathbf{x} = \lambda\mathbf{x}$, i.e., multiplication with A does not change the direction of the vector **x**, only its magnitude.

2. The set of eigenvectors belonging to a specific eigenvalue form a subspace, as it is easy to show.

Theorem 12.3. *(Properties of eigenvalues)*

1. *Diagonal and triangular matrices have their eigenvalues on the main diagonal.*

2. *If $\lambda_1 \neq \lambda_2$ eigenvalues, then the eigenvectors of λ_1 cannot belong to the subspace spanned by the eigenvectors of λ_2.*

3. *Symmetric matrices have real eigenvalues and eigenvectors corresponding to different eigenvalues are orthogonal to each other.*

13 Diagonalization

Let us assume that a square matrix A has n linearly independent eigenvectors, $x_1, x_2, \ldots, x_n$. Let us create a matrix whose columns are these vectors:

$$S = \begin{bmatrix} x_1 \; x_2 \; \ldots \; x_n \end{bmatrix}.$$

Let also Λ the diagonal matrix with the corresponding eigenvalues in the diagonal, i.e.,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \ldots & \ldots \\ 0 & \lambda_2 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

We have

$$A = S\Lambda S^{-1}.$$

To prove this, note that

$$
\begin{aligned}
AS &= A\begin{bmatrix}x_1\; x_2\; \dots\; x_n\end{bmatrix} = \begin{bmatrix}Ax_1\; Ax_2\; \dots\; Ax_n\end{bmatrix} \\
&= \begin{bmatrix}\lambda_1 x_1\; \lambda_2 x_2\; \dots\; \lambda_n x_n\end{bmatrix} = \begin{bmatrix}x_1\; x_2 \dots\; x_n\end{bmatrix}\Lambda \\
&= S\Lambda,
\end{aligned}
$$

from which the result follows.

Diagonalization is useful for a number of reasons. For example, observe that

$$A^k = S\Lambda S^{-1} S\Lambda S^{-1} \dots S\Lambda S^{-1} = S\Lambda^k S^{-1},$$

where

$$\Lambda^k = \begin{bmatrix} \lambda_1^k & 0 & \dots & \dots \\ 0 & \lambda_2^k & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix}$$

Unfortunately, not all matrices can be diagonalized, as diagonalization requires n linearly independent eigenvectors, and not all matrices have them. See the next example

Example 13.1. We will calculate the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The characteristic polynomial is

$$\begin{vmatrix} -s & 1 \\ 0 & -s \end{vmatrix} = 0 \Leftrightarrow s^2 = 0 \Leftrightarrow s_1 = s_2 = 0.$$

Regarding the eigenvectors, we have

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Leftrightarrow x_2 = 0,$$

therefore the subspace of eigenvectors is the one-dimensional set $\{(x_1, 0) : x_1 \in \mathbb{R}\}$.

Example 13.2. We will find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix},$$

and then write it in the form $A = S\Lambda S^{-1}$ where Λ is a diagonal matrix.

To find the eigenvalues, we solve the equation

$$
\begin{aligned}
\begin{vmatrix} 3-s & 1 \\ 1 & 3-s \end{vmatrix} = 0 \Leftrightarrow (s-3)^2 - 1 = 0 \Leftrightarrow s^2 + 9 - 6s - 1 = 0 \Leftrightarrow s^2 - 6s + 8 = 0 \\
\Leftrightarrow s = \frac{6 \pm \sqrt{36-32}}{2} = 2, 4.
\end{aligned}
$$

To find the eigenvector corresponding to the eigenvalue $s_1 = 2$, we have

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Leftrightarrow 3x_1 + x_2 = 2x_1, \quad x_1 + 3x_2 = 2x_2 \Leftrightarrow x_1 + x_2 = 0.$$

Therefore, one choice for an eigenvector is $\left[\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\right]^T$.

Likewise, to find the eigenvector corresponding to the eigenvalue $s_1 = 4$, we have

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 4 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Leftrightarrow 3x_1 + x_2 = 4x_1, \quad x_1 + 3x_2 = 4x_2 \Leftrightarrow x_1 = x_2.$$

Therefore, one choice for an eigenvector is $\left[\frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}}\right]^T$.

It follows that we can write A as

$$A = S\Lambda S^{-1},$$

where

$$S = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix},$$

and

$$S^{-1} = S^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},$$

since S is orthonormal.

Example 13.3. We will compute the diagonalization of the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

To this effect, we calculate the roots of the characteristic polynomial:

$$det(A - \lambda I) = 0 \Leftrightarrow \begin{vmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{vmatrix} = 0$$
$$\Leftrightarrow (1-\lambda)\left[(1-\lambda)^2 - 1\right] - (1-\lambda-1) + (1-1+\lambda) = 0 \Leftrightarrow (1-\lambda)(\lambda^2 - 2\lambda) + 2\lambda = 0$$
$$\Leftrightarrow (1-\lambda)\lambda(\lambda-2) + 2\lambda = \lambda(2 + (1-\lambda)(\lambda-2)) = \lambda(2 + \lambda - 2 - \lambda^2 + 2\lambda) = \lambda^2(3-\lambda) = 0,$$

therefore the eigenvalues are

$$\lambda_1 = \lambda_2 = 0, \quad \lambda_3 = 3.$$

Therefore, matrix Λ is

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Regarding the eigenvectors of the first eigenvalue, we have:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow x_1 + x_2 + x_3 = 0.$$

Therefore, we have two degrees of freedom, and two eigenvectors are $x_1 = [1 \quad -1 \quad 0]^T$ and $x_2 = [1 \quad 0 \quad -1]^T$.

Regarding the eigenvectors of the second eigenvalue, we have:

$$\begin{bmatrix} 1-3 & 1 & 1 \\ 1 & 1-3 & 1 \\ 1 & 1 & 1-3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow -2x_1 + x_2 + x_3, \quad x_1 - 2x_2 + x_3 = 0, \quad x_1 + x_2 - 2x_3 = 0.$$

Two of these three equations are independent. We can then set the values of two components, and find the third. For example, one eigenvector is $[1\ 1\ 1]^T$.

Therefore, the required matrix S is

$$S = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}.$$

It remains to find the inverse of S. To this effect, we can use Gauss-Jordan elimination:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\Leftrightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \qquad (R_2 = R_2 + R1)$$

$$\Leftrightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 3 & 1 & 1 & 1 \end{array}\right] \qquad (R_3 = R_3 + R_2)$$

$$\Leftrightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right] \qquad (R_3 = R_3/3)$$

$$\Leftrightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right] \qquad (R_2 = R_2 - 2R_3)$$

$$\Leftrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right] \qquad (R_1 = R_1 - R_2 - R_3)$$

Therefore,

$$S^{-1} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Concluding, we have

$$\begin{aligned} A &= S\Lambda S^{-1} \\ &= \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}. \end{aligned}$$

As a means of verification, observe that if we get the product we have

$$\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 3 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Alternatively, we could have observed that A is symmetric, so we can make an orthonormal basis with its eigenvectors, and use this to construct S, in which case S^{-1} is simply equal to S^T. That would have made the Gauss-Jordan elimination unnecessary.

14 Symmetric matrices

Definition 14.1. *Let A be a symmetric (i.e. $A^T = A$) $n \times n$ matrix.*

1. *The expression*
$$\mathbf{x}^T A\mathbf{x}$$
is called a quadratic form.
2. *Matrix A is called positive definite if for all $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^T A\mathbf{x} > 0$.*
3. *Matrix A is called positive semidefinite if for all $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^T A\mathbf{x} \geq 0$.*
4. *Matrix A is called negative definite if for all $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^T A\mathbf{x} < 0$.*
5. *Matrix A is called negative semidefinite if for all $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^T A\mathbf{x} \leq 0$.*
6. *Matrix A is called indefinite if for some $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^T A\mathbf{x} > 0$ and for some $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^T A\mathbf{x} < 0$.*

Theorem 14.1.
1. *A matrix A is positive definite if all its eigenvalues are positive.*
2. *A matrix A is positive semidefinite if all its eigenvalues are nonnegative.*
3. *A matrix A is negative definite if all its eigenvalues are negative.*
4. *A matrix A is negative semidefinite if all its eigenvalues are nonpositive.*
5. *A matrix A is indefinite if it has both positive and negative eigenvalues*

Example 14.1. Which of the following matrices are positive (or negative) definite?

$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad \begin{bmatrix}1 & 0\\ 0 & -\end{bmatrix}, \quad \begin{bmatrix}10 & 1\\ 1 & 10\end{bmatrix}.$$

To answer this question, it suffices to calculate the eigenvalues of the matrices.

Regarding the first one, which is the identity matrix, we clearly have $\lambda_1,\ \lambda_2 = 1 > 0$, therefore that matrix is positive definite.

Regarding the second one, we have

$$det(A - \lambda I) = 0 \Leftrightarrow \begin{vmatrix} 1-\lambda & 0\\ 0 & -1-\lambda \end{vmatrix} = 0 \Leftrightarrow \lambda^2 - 1 = 0 \Leftrightarrow \lambda = \pm 1,$$

therefore that matrix is indefinite.

Finally, regarding the third matrix, we have

$$det(A - \lambda I) = 0 \Leftrightarrow \begin{vmatrix} 10-\lambda & 1\\ 1 & 10-\lambda \end{vmatrix} = 0 \Leftrightarrow (10-\lambda)^2 = 1 \Leftrightarrow \lambda_1 = 9,\ \lambda_2 = 11,$$

therefore this matrix is positive definite.

15 Singular Value Decomposition

Any $n \times m$ matrix A can be written as follows:

$$A = U\Sigma V^T.$$

This is the Singular Value Decomposition (SVD) of matrix A. In the above,

1. Σ is a diagonal matrix of size $n \times m$, whose diagonal elements are the square roots $\sigma_i = \sqrt{\lambda_i}$ of the non-zero eigenvalues of matrix AA^T (and also $A^T A$), in decreasing order.

2. U is an orthogonal matrix of size $n \times n$, comprised of columns that are the eigenvectors of AA^T.

3. V is an orthogonal matrix of size $m \times m$, comprised of columns that are the eigenvectors of $A^T A$.

Note that the eigenvectors appearing in the i-place of U and V must correspond to the i-larger eigenvalue. Note that $A^T A$ always has real non-negative eigenvalues.

The geometric interpretation of the SVD is as follows: a vector multiplied by A is first rotated in its own space, then its components are individually scaled, and then the resulting vector is rotated in the space of the output vector.

Observe that

$$A = \begin{bmatrix} u_1\, u_2\, \ldots u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & \ldots \\ 0 & \sigma_2 & 0\ldots & \\ 0 & 0 & \sigma_3 & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \ldots \\ v_n^T \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1\; \sigma_2 u_2\; \ldots \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \ldots \\ v_n^T \end{bmatrix} = \sum_{i=1}^{k} u_i \sigma_i v_i^T,$$

where k is the number of singular values. Based on this equation, we have

$$Av_i = \sigma_i u_i, \qquad A^T u_i = \sigma_i v_i$$

Example 15.1. We will find the SVD of the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

We have

$$AA^T = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

whose eigenvalues are given by the equation

$$det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (\lambda - 2)^2 - 1 = 0,$$

therefore $\lambda_1 = 1$ and $\lambda_2 = 3$.

Therefore,

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Regarding the first eigenvalue, we have

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} \Leftrightarrow x_1 = x_2,$$

therefore an eigenvector of unit length is $\begin{bmatrix} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \end{bmatrix}$.

Regarding the second eigenvector, we have

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} \Leftrightarrow x_1 + x_2 = 0,$$

therefore an eigenvector of unit length is $\begin{bmatrix} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \end{bmatrix}$.

It follows that

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Regarding the eigenvectors of V, we could repeat the above process, however since the eigenvectors have an arbitrary direction, and V is not uniquely determined it is not certain that we will arrive at a V that will satisfy the SVD equation. One solution is to use the method of the next exercises.

Alternatively, note that $v_i = \frac{1}{\sigma_i} A^T u_i$, therefore, for the first two vectors at least, we have

$$\begin{aligned} v_1 &= \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \\ v_2 &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}. \end{aligned}$$

We cannot find v_3 with this method, however we know that it must form an orthonormal basis together with v_1 andv_2 and so, using, for example, the Gram-Schmidt process, we find that $v_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}} \end{bmatrix}$. Therefore,

$$V = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{sqrt6} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

Note that we could also have found the eigenvectors of V first, and use them to find the two eigenvectors of U. Indeed,

$$A^T A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

whose characteristic polynomial is

$$\begin{aligned} det(A^T A - \lambda I) &= \begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 2-\lambda & 1 \\ 0 & 1 & 1-\lambda \end{vmatrix} = (1-\lambda)[(\lambda-1)(\lambda-2)-1)] - (1-\lambda) \\ &= (1-\lambda)(\lambda^2+1-\lambda-2\lambda)+\lambda-1 = (1-\lambda)(\lambda^2+1-3\lambda)+\lambda \\ &= \lambda^2+1-3\lambda-\lambda^3-\lambda+3\lambda^2+\lambda-1 \\ &= \lambda(\lambda-3-\lambda^2-1+3\lambda+1) = \lambda(\lambda-3-\lambda^2+3\lambda) = \lambda[(\lambda-3)-\lambda(\lambda-3)] \\ &= \lambda(\lambda-3)(1-\lambda), \end{aligned}$$

therefore the eigenvalues are

$$\lambda_1 = 3, \quad \lambda_2 = 1, \quad \lambda_3 = 0,$$

as expected.

Regarding the first eigenvector,

$$\begin{bmatrix} -2 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow -2x_1 + x_2 = 0, \quad x_2 - 2x_3 = 0,$$

therefore one eigenvector is $[1\ 2\ 1]^T$, which, normalizing to have unit length, becomes $\left[\frac{1}{\sqrt{6}}\ \frac{2}{\sqrt{6}}\ \frac{1}{\sqrt{6}}\right]^T$.

Regarding the second eigenvector,

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow x_2 = 0, \quad x_3 = -x_1,$$

therefore one eigenvector is $[1\ 0\ -1]^T$, which, normalizing to have unit length, becomes $\left[\frac{1}{\sqrt{2}}\ 0\ -\frac{1}{\sqrt{2}}\right]^T$.

Regarding the third eigenvector,

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow x_2 = -x_1, \quad x_2 = -x_3,$$

therefore one eigenvector is $[-1\ 1\ -1]^T$, which, normalizing, becomes $\left[-\frac{1}{\sqrt{3}}\ \frac{1}{\sqrt{3}}\ -\frac{1}{\sqrt{3}}\right]^T$.

Observe that the eigenvectors we found with this method coincide with those found with the previous one, but this is a coincidence.

Example 15.2. One method of finding the SVD is to solve the eigenvalue system

$$\begin{bmatrix} \mathbf{0} & A^T \\ A & \mathbf{0} \end{bmatrix} \begin{bmatrix} V & V \\ U & -U \end{bmatrix} = \begin{bmatrix} V & V \\ U & -U \end{bmatrix} \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & -\Sigma \end{bmatrix}$$
$$\Leftrightarrow A^T U = V\Sigma, \quad -A^T U = -V\Sigma, \quad AV = U\Sigma, \quad AV = U\Sigma \Leftrightarrow A = U\Sigma V^T$$

where $\mathbf{0}$ is a matrix with zeros and proper dimensions. Observe that the large block matrix is symmetric, therefore it is guaranteed to have an orthogonal set of eigenvectors.

Example 15.3. (2015 exam) We will find the SVD of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

We have

$$AA^T = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix},$$

whose characteristic polynomial is

$$det(AA^T - \lambda I) = \begin{vmatrix} 5-\lambda & 10 \\ 10 & 20-\lambda \end{vmatrix} = (\lambda-5)(\lambda-20) - 100 = 0 \Leftrightarrow \lambda^2 - 25\lambda + 100 - 100 = 0 \Leftrightarrow \lambda(\lambda-25) = 0.$$

Therefore, we have

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}$$

Next, we find the eigenvectors of AA^T. Regarding the eigenvector corresponding to $\lambda = 0$, we have

$$\begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow 5x_1 + 10x_2 = 0,$$

therefore one eigenvector is $[2 - 1]^T$. Normalizing it, gives the eigenvector of unit length $\left[\frac{2}{\sqrt{5}} - \frac{1}{\sqrt{5}}\right]$.

Regarding the eigenvector corresponding to $\lambda = 25$, we have

$$\begin{bmatrix} 5-25 & 10 \\ 10 & 20-25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow 10x_1 - 5x_2 = 0,$$

therefore one eigenvector is $[12]^T$. Normalizing it, gives the eigenvector of unit length $\left[\frac{1}{\sqrt{5}} \frac{2}{\sqrt{5}}\right]$.

Therefore, we found that

$$U = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}$$

Regarding V, we note that $v_i = \frac{1}{\sigma_i} A^T u_i$, therefore, regarding the first one,

$$v_1 \frac{1}{5} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$$

Regarding the second column of V, as the corresponding singular value is zero, we can use the method above. Requiring, however v_2 to form an orthonormal basis together with v_1, we readily find

$$V = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}$$

Wrapping everything up, we found that

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}^T$$

Alternatively, we could have notices that A is symmetric with an eigenvector space of dimension 2, therefore its SVD coincides with its diagonalization! The required calculations would be fewer.

16 Example Exercises

Multivariate Calculus

1. Find extrema when there are no constraints
2. Find extrema with constraints
3. Plot contour lines.

Linear Algebra

1. Find eigenvalues
2. Find SVD

3. Find if a set of vectors are linearly independent.
4. Find inverse matrix
5. What is a positive (negative) (semi)definite matrix, and how to determine if a given matrix belongs to any of these classes of matrices
6. Find the rank, range, and null space of a matrix.