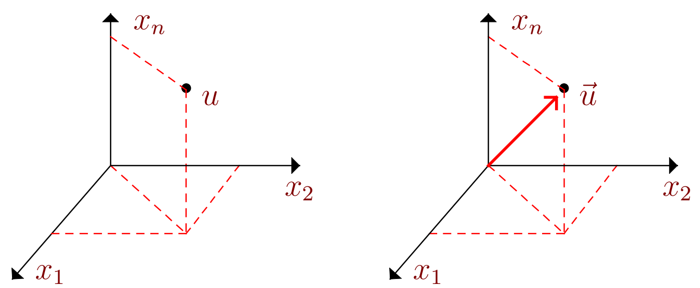
Appendix

1 Linear Algebra

1.1 $\mathbb{R}^n$ as a vector space

As a set, $\mathbb{R}^n$ is the collection of all n-tuple $(x_1, \ldots, x_n)$, $x_k \in \mathbb{R}$. Equivalently, $\mathbb{R}^n$ can be defined as the Cartesian product of n copies of $\mathbb{R}$ as $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$. A tuple $u = (x_1, \ldots, x_n)$ has a double life, as a point in the Cartesian space with coordinates x_k, and as a *vector* $\vec{u}$ with the tail at the origin and head at u. For this reason, $\mathbb{R}^n$ can be considered as a set of points or a set of vectors.



We need two more structures on $\mathbb{R}^n$ to make it a *vector space*, i.e., the *scalar multiplication*, and the *vector addition*. Let $c \in \mathbb{R}$ be a scalar, and $\vec{u} = (x_1, \ldots x_n), \vec{v} = (y_1, \ldots, y_n)$ two arbitrary vectors. We define

$$c\,\vec{u} = (cx_1, \ldots, cx_n), \vec{u} + \vec{v} = (x_1 + y_1, \ldots, x_n + y_n).$$

The geometry of these two operations are shown in Fig.1.

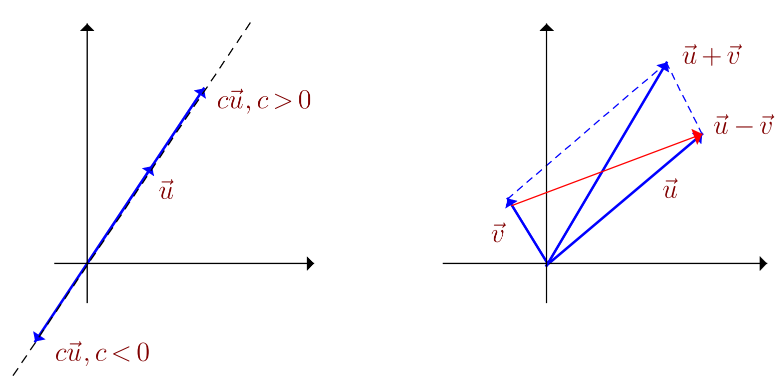


Figure 1.

It is seen that $\mathbb{R}^n$ is closed under the defined operations, that is, for any two vectors $\vec{u}$, $\vec{v}$ in $\mathbb{R}^n$, and $c_1, c_2 \in \mathbb{R}$, vector $c_1\vec{u} + c_2\vec{v}$ is in $\mathbb{R}^n$. Moreover, the following properties hold for any vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ and constant $c \in \mathbb{R}$:

i. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$

ii. $c(\vec{u} + \vec{v}) = c\,\vec{u} + c\,\vec{v}$

iii. $(c_1 + c_2)\vec{u} = c_1\,\vec{u} + c_2\,\vec{u}$

iv. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$

$\mathbb{R}^n$ has n standard *unit vectors* $\hat{e}_1, \ldots, \hat{e}_n$ defined below

$$\hat{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \hat{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \hat{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

The set $\{\hat{e}_k\}_{k=1}^n$ is a *basis* for $\mathbb{R}^n$ in the sense that any vector $\vec{u} = (x_1, \ldots, x_n)$ can be uniquely represented as

$$\vec{u} = x_1\,\hat{e}_1 + \cdots + x_n\,\hat{e}_n.$$

Remark. The magnitude or norm of a vector $\vec{u} = (x_1, \ldots, x_n)$ is defined as follows

$$\|\vec{u}\| = \sqrt{\sum_{k=1}^{n} x_k^2}.$$

A vector is called *unit* or a *direction vector* if its norm is equal 1. We usually denote unit vector by notation $\hat{}$.

1.2 Dot and Cross products

The *dot product* of two arbitrary vectors $\vec{u} = (x_1, \ldots, x_n), \vec{v} = (y_1, \ldots, y_n)$ is defined as

$$\vec{u} \cdot \vec{v} = \sum_{k=1}^{n} x_i\,y_i.$$

In particular, $\hat{e}_i \cdot \hat{e}_j = \delta_{i,j}$ where $\delta_{i,j} = \left\{ \begin{smallmatrix} 1 & i=j \\ 0 & i \neq j \end{smallmatrix} \right.$.

Problem 1. Show that the dot product enjoys the following properties for any vectors $\vec{u}, \vec{v}, \vec{w}$ and for arbitrary constant c

i. $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$,

ii. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$,

iii. $(c\,\vec{u}) \cdot \vec{v} = c\,(\vec{u} \cdot \vec{v})$

Problem 2. Show the following relations

a) $\|\vec{u}\|^2 = \vec{u} \cdot \vec{u}$

b) $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$

c) $\|\vec{u} + \vec{v}\|^2 - \|\vec{u} - \vec{v}\|^2 = 4\,\vec{u} \cdot \vec{v}$

Problem 3. The Cauchy inequality is as follows

$$\vec{u} \cdot \vec{v} \leq \|\vec{u}\|\,\|\vec{v}\|.$$

Try to prove the inequality by the following method: $\|\vec{u}+t\,\vec{v}\|^2 \geq 0$ for all $t\in\mathbb{R}$. Expand $\|\vec{u}+t\,\vec{v}\|^2$ in terms of t and conclude the inequality. By this inequality, one can *define* the angle between two nonzero vectors $\vec{u},\vec{v}$ as follows

$$\cos(\theta)=\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\,\|\vec{v}\|}.$$

By the above equality, one can write $(\vec{u},\vec{v})=\|\vec{u}\|\,\|\vec{v}\|\cos(\theta)$. Note that if $\vec{u}\cdot\vec{v}=0$ for non-zero vectors $\vec{u},\vec{v}$, then $\cos(\theta)=0$.

Problem 4. Show that if $\vec{u}_1,\ldots,\vec{u}_m$ are mutually orthogonal, that is, if $\vec{u}_i\cdot\vec{u}_j=0$ for $i\neq j$ then

$$\|\vec{u}_1+\cdots+\vec{u}_m\|^2=\|\vec{u}_1\|+\cdots+\|\vec{u}_m\|^2.$$

There is a standard product in $\mathbb{R}^3$ called *cross* or *external product*. For $\vec{u}=(x_1,y_1,z_1)$, $\vec{v}=(x_2,y_2,z_2)$, the cross product is defined as follows

$$\vec{u}\times\vec{v}=\begin{vmatrix}\hat{e}_1 & \hat{e}_2 & \hat{e}_3\\ x_1 & y_1 & z_1\\ x_2 & y_2 & z_2\end{vmatrix}=(y_1z_2-z_1y_2)\hat{e}_1+(z_1x_2-x_1z_2)\hat{e}_2+(x_1y_2-y_1x_2)\hat{e}_3.$$

Note that $\vec{u}\times\vec{v}$ is a vector, while their dot product is a scalar.

Problem 5. Show the relation $\vec{u}\times\vec{v}=-\vec{v}\times\vec{u}$.

Problem 6. Show that $\vec{u}\times\vec{v}$ is perpendicular to $\vec{u}$ and $\vec{v}$, that is,

$$(\vec{u}\times\vec{v})\cdot\vec{u}=(\vec{u}\times\vec{v})\cdot\vec{v}=0.$$

Problem 7. Show the identity

$$\|\vec{u}\times\vec{v}\|^2=\|\vec{u}\|^2\|\vec{u}\|^2-|\vec{u}\cdot\vec{v}|^2,$$

and conclude the relation

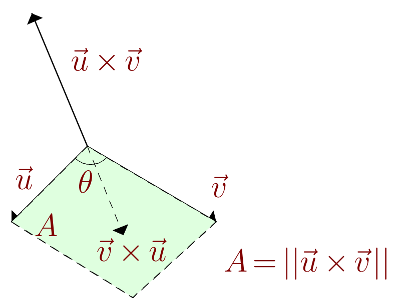
$$\|\vec{u}\times\vec{v}\|=\|\vec{u}\|\,\|\vec{v}\|\sin(\theta),$$

where θ is the angle between $\vec{u},\vec{v}$ in $[0,\pi]$.

By the above problem, we can write

$$\vec{u}\times\vec{v}=\|\vec{u}\|\,\|\vec{v}\|\sin(\theta)\,\hat{n},$$

where $\hat{n}$ is the unit vector perpendicular to the plane containing $\vec{u},\vec{v}$, that is, $\hat{n}=\frac{\vec{u}\times\vec{v}}{\|\vec{u}\times\vec{v}\|}$.



Problem 8. Show the following relation for any three vectors $\vec{u},\vec{v},\vec{w}$

$$\vec{u}\cdot(\vec{v}\times\vec{w})=\vec{v}\cdot(\vec{w}\times\vec{u}).$$

Problem 9. If $\vec{v},\vec{w}$ are orthogonal, show the following relation

$$\vec{u}\times(\vec{v}\times\vec{w})=(\vec{u}\cdot\vec{w})\vec{v}-(\vec{u}\cdot\vec{v})\vec{w}.$$

Use this result and relax the condition $\vec{v}$, $\vec{w}$ to be orthogonal. Use the formula and determine the conditions the following relation holds

$$\vec{u}\times(\vec{v}\times\vec{w})=(\vec{u}\times\vec{v})\times\vec{w}.$$

1.3 Subspaces and direct sum

Definition 1. *Tow vectors $\vec{u}$, $\vec{v}$ in $\mathbb{R}^n$ are called* linearly dependent *if there is a scalar c such that $\vec{u} = c\,\vec{v}$ or $\vec{v} = c\,\vec{u}$. A vector $\vec{u}$ is linearly dependent on vectors $\vec{v}_1, \ldots, \vec{v}_m$ if there are scalars $c_1, \ldots, c_m$ such that*

$$\vec{u} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m.$$

Vectors $\vec{v}_1, \ldots, \vec{v}_m$ in $\mathbb{R}^n$ are linearly independent if the linear combination

$$c_1\,\vec{v}_1 + \cdots + c_m\,\vec{v}_m = 0,$$

implies $c_1 = \cdots = c_m = 0$.

Problem 10. Vectors $\hat{e}_1, \ldots, \hat{e}_n$ are linearly independent in $\mathbb{R}^n$. Show that any $n+1$ vectors of $\mathbb{R}^n$ are linearly dependent.

Let $\{\vec{v}_1, \ldots, \vec{v}_d\}$ for $d \le n$ be a set of linearly independent vectors in $\mathbb{R}^n$. The span of vectors in the given set is the set of all possible linear combinations of $\vec{v}_1, \ldots, \vec{v}_d$, i.e.,

$$\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_d\} = \{c_1\vec{v}_1 + \cdots + c_d\,\vec{v}_d, c_k \in \mathbb{R}\}.$$

Note that $\mathbb{R}^n$ is itself equal to $\operatorname{span}\{\hat{e}_1, \ldots, \hat{e}_n\}$.

Proposition 1. *$\mathbb{V}_d := \operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_d\}$ is closed under the vector addition and scalar multiplication of $\mathbb{R}^n$. For this reason, $\mathbb{V}_d$ is called a* linear subspace *of $\mathbb{R}^n$.*

Definition 2. *Let $\mathbb{V}$ be a linear subspace of $\mathbb{R}^n$. The dimension of $\mathbb{V}$ is the maximum number of linearly independent vectors in $\mathbb{V}$.*

Example 1. Technically speaking, $\mathbb{R}^m$ is not a subspace of $\mathbb{R}^n$ for $m < n$, however, if we interpret $\mathbb{R}^m$ as $\operatorname{span}\{\hat{e}_1, \ldots, \hat{e}_m\}$ where each $\hat{e}_j$ is a vector in $\mathbb{R}^n$, then $\mathbb{R}^m$ is a linear subspace of $\mathbb{R}^n$.

If $\mathbb{V}$ is a linear subspaces of $\mathbb{R}^n$, then its orthogonal subspace $\mathbb{V}^\perp$ is defined as follows

$$\mathbb{V}^\perp = \{\vec{w} \in \mathbb{R}^n; \vec{w} \cdot \vec{v} = 0; \vec{v} \in \mathbb{V}\}.$$

Obviously, $\mathbb{V}^\perp$ is a linear subspace of $\mathbb{R}^n$ equipped with the vector addition and scalar multiplication operations.

Problem 11. Show that if $\mathbb{V} = \operatorname{span}\{(1,1,0),(0,1,1)\}$, then $\mathbb{V}^\perp$ is the one dimensional subspace spanned by $(1,-1,1)$.

Problem 12. Find the orthogonal subspace of $\mathbb{V} = \{(1,0,1)\}$ in $\mathbb{R}^3$.

Suppose $\mathbb{U}, \mathbb{V}$ are two subspaces of $\mathbb{R}^n$ and $\mathbb{U} \cap \mathbb{V} = \{0\}$. The *direct sum* $\mathbb{U} \oplus \mathbb{V}$ is defined as follows

$$\mathbb{U} \oplus \mathbb{V} = \{c_1\vec{u} + c_2\vec{v}; \quad \vec{u} \in \mathbb{U}, \vec{v} \in \mathbb{V}\}.$$

Problem 13. If $\mathbb{V}$ is a subspace of $\mathbb{R}^n$, show that $\mathbb{R}^n = \mathbb{V} \oplus \mathbb{V}^\perp$.

Problem 14. Let $\mathbb{V}$ be an arbitrary subspace of $\mathbb{R}^n$. Show that every vector $\vec{u}$ in $\mathbb{R}^n$ can be represented *uniquely* as $\vec{u} = c_1\vec{v} + c_2\,\vec{w}$ for $\vec{v} \in \mathbb{V}$ and $\vec{w} \in \mathbb{V}^\perp$.

1.4 Matrices and linear mappings

Definition 3. *A mapping $f: \mathbb{R}^n \to \mathbb{R}^m$ is called* linear *if for any constants c_1, c_2 and any vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, the following relation holds*

$$f(c_1 \vec{u} + c_2 \vec{v}) = c_1 f(\vec{u}) + c_2 f(\vec{v}).$$

Problem 15. If $f: \mathbb{R}^n \to \mathbb{R}^m$ is linear then $f(0) = 0$.

Proposition 2. *A linear mapping $f: \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a $m \times n$* matrix*.*

Proof. Remember that a matrix $A = [a_{ij}]_{m \times n}$ is a structure of n-columns of vectors belonging $\mathbb{R}^m$, i.e., $A = [A_1|\cdots|A_n]$, where $A_k \in \mathbb{R}^m$. The action of A to $\hat{e}_k$ is defined by the relation $A(\hat{e}_k) = A_k$. Now define A_f as

$$A_f = [f(\hat{e}_1)|f(\hat{e}_2)|\cdots|f(\hat{e}_n)].$$

It is simply seen that for arbitrary vector $\vec{u} \in \mathbb{R}^n$, the following relation holds $f(\vec{u}) = A_f(\vec{u})$. □

Problem 16. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ has the matrix representation $A_{2\times 2} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ in the standard basis. Find the matrix representation of T in the basis $\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Problem 17. Prove that a matrix 2×2 maps any parallelogram to a parallelogram.

Problem 18. Verify that the the matrix $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ rotates vectors in the plane by θ degree counter-clockwise. Verify that $R_{\theta_1} R_{\theta_2} = R_{\theta_1 + \theta_2}$ and conclude that $R_\theta R_{-\theta}$ is the identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Definition 4. *Let $f: \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. The kernel (or null space) of f, denoted by $\ker(f)$ (or just N_f) is a set of all vectors $\vec{n}$ of $\mathbb{R}^n$ such that $f(\vec{n}) = 0 \in \mathbb{R}^m$. The* image *of f denoted by $\operatorname{Im}(f)$ is the set of all vectors $\vec{w} \in \mathbb{R}^m$ such that $\vec{w} = f(\vec{v})$ for some $\vec{v} \in \mathbb{R}^n$.*

Proposition 3. *The kernel of a linear mapping $f: \mathbb{R}^n \to \mathbb{R}^m$ is a vector subspace of $\mathbb{R}^n$. The image of f is a vector subspace of $\mathbb{R}^m$.*

Problem 19. Prove the proposition.

Theorem 1. *Assume that $f: \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping. The following relation holds*

$$n = \dim \ker(f) + \dim \operatorname{Im}(f). \tag{1}$$

Problem 20. Let S denote the orthogonal subspace of $\ker(f)$. Show that $\dim S = \dim \operatorname{Im}(f)$ and conclude $f(S) = \operatorname{Im}(f)$.

1.5 Linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$

1.5.1 Determinant

Let $f: \mathbb{R}^n \to \mathbb{R}^n$ be a linear mapping, and let C be a unit cube constructed on $\hat{e}_k, k = 1, \ldots, n$. The image of C under f, that is $f(C)$, is a parallelogram. In fact, every vector $\vec{u} \in C$ is represented by the linear combination

$$\vec{u} = c_1 \hat{e}_1 + \cdots + c_n \hat{e}_n,$$

for $0 \leq c_k \leq 1$, and thus

$$f(\vec{u}) = c_1 \vec{f}_1 + \cdots + c_n \vec{f}_n,$$

where $f_k = f(\hat{e}_k)$. The set $\{c_1 \vec{f}_1 + \cdots + c_n \vec{f}_n\}$ for $0 \leq c_k \leq 1$ is a parallelogram constructed on $\vec{f}_1, \ldots, \vec{f}_n$; see Fig. 2 .

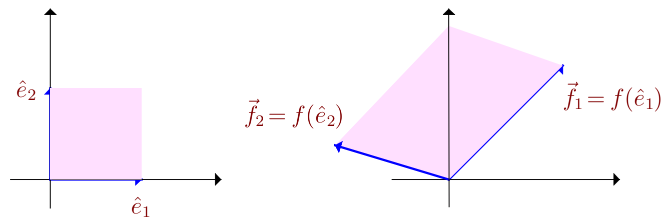


Figure 2.

Definition 5. *Let $f: \mathbb{R}^n \to \mathbb{R}^n$ be a linear mapping, and let C be the unit cube constructed on $\{\hat{e_k}\}_{k=1}^n$. The determinant of f denoted by $\det(f)$ is the algebraic volume of parallelogram $f(C)$. The algebraic volume is the signed volume with positive or negative signs.*

Example 2. In $\mathbb{R}^2$, the *determinant* of $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is defined by the following formula

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} a_{22} - a_{12} a_{21}. \tag{2}$$

It is simply verified that $|\det(A)| = ||A(\hat{e}_1)||\, ||A(\hat{e}_2)|| \sin(\theta)$, where θ is the angel between two columns of A.

If $\det(f) = 0$, then the volume degenerates, that means vectors $\vec{f}_1, \ldots, \vec{f}_n$ are linearly dependent. If $\det(f) < 0$, then f changes the *standard orientation* of the basis $\{\hat{e}_k\}_{k=1}^n$ (remember the standard rotations in $\mathbb{R}^2$ and $\mathbb{R}^3$), for example $f(x,y) = (y,x)$ with the matrix representation

$$A = [f(\hat{e}_1) | f(\hat{e}_2)] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

changes the standard rotation. Let $A_f = [\vec{f}_1 | \cdots | \vec{f}_n]$ be the representation of $f: \mathbb{R}^n \to \mathbb{R}^n$ in the standard basis $\{\hat{e}_k\}_{k=1}^n$. The determinant $\det(A_f)$ satisfies the following properties:

i. If $\vec{f}_1, \cdots, \vec{f}_n$ are linearly dependent then $\det[A] = 0$

ii. $\det[\vec{f}_2 | \vec{f}_1 | \cdots | \vec{f}_n] = -\det[A_f]$. In general any switch between column i and j multiple the determinant by the factor $(-1)^{i+j}$.

iii. $\det[c\, \vec{f}_1 | \vec{f}_2 | \cdots | \vec{f}_n] = c \det(A_f)$

iv. $\det[c_1 \vec{f}_1 + c_2 \vec{f}_k | \vec{f}_2 | \cdots | \vec{f}_n] = c_1 \det[A_f]$ for any $k = 2, \ldots, n$.

By the above properties, it is seen that if $f, g: \mathbb{R}^n \to \mathbb{R}^n$ are two linear mappings then

$$\det(AB) = \det(A) \det(B)$$

Problem 21. Verify directly the above claim for 2×2 matrices.

1.5.2 Injective and surjective mappings

Definition 6. *A linear mapping* $f\colon \mathbb{R}^n \to \mathbb{R}^m$ *is called* one-to-one *or* injective *if equality* $f(\vec{u}) = f(\vec{v})$ *implies* $\vec{u} = \vec{v}$*. A linear mapping* $f\colon \mathbb{R}^n \to \mathbb{R}^m$ *is called* onto *or* surjective*, if* $\mathbb{R}^m = f(\mathbb{R}^n)$*.*

Problem 22. A linear mapping $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is one to one if and only if $\ker(f) = \emptyset$, and if and only if it is onto.

Problem 23. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Show that if $m > n$, then f can not be onto, if $m < n$ then f can not be one-to-one.

If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one (and then onto), the mapping $f^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$ is called the *inverse mapping* of f if the following relation holds

$$ff^{-1} = f^{-1}f = \mathrm{Id},$$

where Id is the identity mapping on $\mathbb{R}^n$. The identity mapping has the matrix representation $\mathrm{diag}(1,\ldots,1)$, where $\mathrm{diag}(1,\ldots,1)$ has 1 on the main diagonal and zero everywhere else. Note that $\mathrm{Id}(\vec{u}) = \vec{u}$ for any vector $\vec{u}$.

Problem 24. If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is a one to one linear map, show that f^{-1} is also a one to one linear map.

1.5.3 Eigenvalues and Eigenvectors

A vector $\vec{v} \in \mathbb{R}^n - \{0\}$ is called an *eigenvector* of a linear mapping $f\colon \mathbb{R}^n \to \mathbb{R}^n$ if there is a scalar λ such that $f(\vec{v}) = \lambda\vec{v}$. It is seen that if $\vec{v}$ is an eigenvector, then vector $\vec{w} = t\vec{v}$ for arbitrary scalar t is also an eigenvector. Accordingly, one can define an *eigendirection* of f that is $\mathrm{span}\{\vec{v}\} := \{t\,\vec{v}, t \in \mathbb{R}\}$; see Fig. 3.

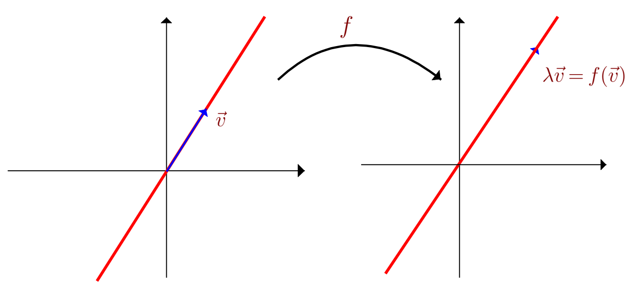


Figure 3.

Example 3. The vector $\vec{v} = (1, 1)$ is an eigenvector of the matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$ with the eigenvalue $\lambda = 2$, because $\begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$ has only one eigenvector and matrix $A = \begin{pmatrix} 2 & 3 \\ -4 & -5 \end{pmatrix}$ has two eigenvectors $\vec{v}_1 = (1, -1)$ and $\vec{v}_2 = (3, -4)$ with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = -2$ respectively. The rotation matrix $R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ has no (real) eigenvector for $\theta \neq 0, 2\pi$. Recall that R_θ rotates vectors counterclockwise by θ-angle. The identity matrix $\mathbb{I}_{2\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ has infinitely many eigenvectors. In fact, every vector in $\mathbb{R}^2$ is an eigenvector of $\mathrm{Id}_{2\times 2}$ with eigenvalue $\lambda = 1$.

Proposition 4. *If $f: \mathbb{R}^n \to \mathbb{R}^n$ has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then their associated eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ are linearly independent.*

Problem 25. Prove the proposition.

If $\vec{v}$ is an eigenvector of a linear mapping f with eigenvalue λ, then $(f - \lambda \mathrm{Id})\vec{v} = 0$, and since $\vec{v}$ is nonzero, $\vec{v}$ must belong to the kernel of $f - \lambda \mathrm{Id}$. Let A_f is a matrix representation of f, then the following relation holds

$$\det(A_f - \lambda \mathrm{Id}) = 0.$$

The above equation, which is an algebraic equation of λ, is called the *characteristic equation* of f. If $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, the *characteristic equation* is as follows

$$\lambda^2 - \mathrm{tr}(A)\,\lambda + \det(A) = 0, \tag{3}$$

where $\mathrm{tr}(A)$ (read trace A) is equal to $a_{11} + a_{22}$.

Problem 26. Show that if $A_{2\times 2}$ has a repeated eigenvalue λ with two linearly independent eigenvectors then all vectors of $\mathbb{R}^2$ is an eigenvector of A.

Problem 27. Let A be a 2×2 matrix. Prove that the following statements are equivalent

i. A is invertible.

ii. Two columns of A are linearly independent.

iii. The determinant of A is non-zero.

iv. No eigenvalue of A is zero.

Problem 28. If λ_1, λ_2 are two eigenvalues of $A_{2\times 2}$, show that $\det(A) = \lambda_1 \lambda_2$ and $\mathrm{tr}(A) = \lambda_1 + \lambda_2$.

Problem 29. If $Q_{2\times 2}$ is an invertible matrix, show the following relations

$$\mathrm{tr}(Q^{-1}AQ) = \mathrm{tr}(A), \quad \det(Q^{-1}AQ) = \det(A).$$

1.5.4 Symmetric mappings and Jordan forms

Definition 7. *A linear mapping $f: \mathbb{R}^n \to \mathbb{R}^n$ is called symmetric if the following equality holds for arbitrary vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$:*

$$f(\vec{u}) \cdot \vec{v} = \vec{u} \cdot f(\vec{v}).$$

Theorem 2. *If the linear mapping $f: \mathbb{R}^n \to \mathbb{R}^n$ is symmetric, then there are n mutually orthogonal eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ for f. Moreover, all eigenvalues of f are real.*

Problem 30. Assume $\vec{v}_1, \vec{v}_2$ are two eigenvectors of a symmetric mapping f. Show that $\langle \vec{v}_1, \vec{v}_2 \rangle = 0$.

If $f: \mathbb{R}^n \to \mathbb{R}^n$ has n linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$, then $\mathbb{R}^n$ can be decomposed by the direct sum $\mathbb{R}^n = \mathbb{V}_1 \oplus \cdots \oplus \mathbb{V}_n$ where $\mathbb{V}_k = \mathrm{span}\{\vec{v}_k\}$. The restriction of f to each $\mathbb{V}_k$ is a linear mapping $f_k: \mathbb{V}_k \to \mathbb{V}_k$, and thus we can decompose f as the direct sum $f = f_1 \oplus \cdots \oplus f_n$. With this interpretation, every vector $\vec{v} \in \mathbb{R}^n$ has a unique representation $\vec{v} = c_1 \vec{v}_1 + \cdots + c_n \vec{v}_n$, and thus $f(\vec{v})$ is

$$f(\vec{v}) = f_1(c_1 \vec{v}_1) + \cdots + f_n(c_n \vec{v}_n) = c_1 \lambda_1 \vec{v}_1 + \cdots + c_n \lambda_n \vec{v}_n$$

Definition 8. *A linear mapping $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is called positive definite if for any nonzero vector $\vec{v}$, the following inequality holds*

$$f(\vec{v}) \cdot \vec{v} > 0.$$

A negative definite linear mapping is defined similarly.

It is simply seen that if $A_f = [a_{ij}]$ is the matrix representation of the positive definite mapping f in the standard basis, then $a_{ii} > 0$ for $i = 1, \ldots, n$. Moreover, all real eigenvalues of A_f must be positive.

Problem 31. Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a symmetric mapping. A necessary and sufficient condition that f be positive definite is that its all eigenvalues are positive.

If $Q_{n\times n}$ is an *invertible* matrix, then two matrices $B = Q^{-1}AQ$ and A are called *similar*. It is seen that A, B have same characteristic polynomial as the following argument justifies it:

$$\det(Q^{-1}AQ - \lambda \mathrm{Id}) = \det Q^{-1}(A - \lambda \mathrm{Id})Q = \det(Q^{-1}) \det(A - \lambda \mathrm{Id}) \det Q = \det(A - \lambda \mathrm{Id}).$$

Proposition 5. *Suppose $A_{n\times n}$ has n linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$. Then the following relation holds*

$$Q^{-1}AQ = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

where for $Q = [\vec{v}_1 | \cdots | \vec{v}_n]$, and $\lambda_1, \ldots, \lambda_n$ are associated eigenvalues (not necessarily distinct). Matrix $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is called the JORDAN *form of A.*

Problem 32. Prove the above proposition.

Fig. 4 shows the relation between eigenvectors of A and $Q^{-1}AQ$.

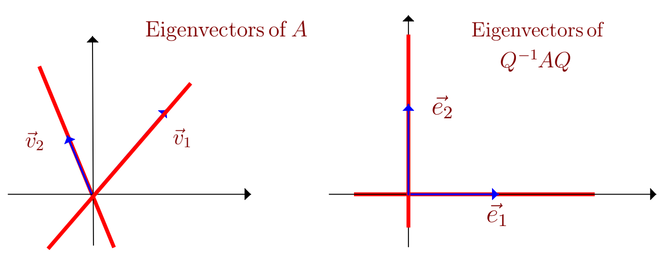


Figure 4.

Problem 33. Let $A = \begin{pmatrix} -4 & -3 \\ 3 & 2 \end{pmatrix}$. Find the matrix $Q^{-1}AQ$.

If a matrix $A_{n\times n}$ has n repeated eigenvalues λ with only one eigenvector, then the JORDAN form of $A_{n\times n}$ is a diagonal matrix with λ on the main diagonal and 1 on the upper diagonal. For example, for a 3×3 matrix with repeated eigenvalue λ, the Jordan form is $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$. If a matrix 2×2 has two complex eigenvalues $\lambda = \sigma \pm i\omega$, its Jordan form is $\begin{pmatrix} \sigma & -\omega \\ \omega & \sigma \end{pmatrix}$.

Problem 34. For a $n \times n$ matrix A show $\det(A) = \prod_{k=1}^{n} \lambda_k$, and conclude that A is one-to-one if and only if A does not have a zero eigenvalue.

1.6 Linear equations

Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping, and let $\vec{b} \in \mathbb{R}^m$ be an arbitrary vector. The linear equation $f(\vec{u}) = \vec{b}$ is *solvable* if and only if $\vec{b} \in \operatorname{Im}(f)$. If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear one-to-one mapping, then equation $f(\vec{u}) = \vec{b}$ is simply solved for $\vec{u} = f^{-1}(\vec{b})$. If $\dim \ker(f) > 0$ and $\vec{u}$ is a solution to the equation, then for any vector $\vec{n} \in \ker(f)$, $\vec{v} = \vec{n} + \vec{u}$ is also a solution. In this context, vectors in $\ker(f)$ are called the *homogeneous solutions* of $f(\vec{u}) = 0$.

Problem 35. Suppose $n \geq m$ and $f\colon R^n \to \mathbb{R}^m$ is a linear mapping. Show that if $\dim \ker(f) = n - m$, then the linear equation $f(\vec{u}) = \vec{b}$ is solvable for any $\vec{b} \in \mathbb{R}^m$. What if $\dim \ker(f) > n - m$? If $n > m$ and $\dim \ker(f) = n - m$, show that the equation $f(\vec{u}) = \vec{b}$ has infinitely many solutions.

Problem 36. Let $A = \begin{pmatrix} 2 & 6 \\ 1 & 3 \end{pmatrix}$. For what values of $\vec{b} \in \mathbb{R}^2$, the equation $A\vec{u} = \vec{b}$ is solvable? Verify that the solutions of the equation $A\vec{u} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ has the form $\vec{u} = t\begin{pmatrix} -3 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ for $t \in (-\infty, \infty)$.

Problem 37. Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear mapping and let $\vec{u}_1, \vec{u}_2$ be two solutions to equation $f(\vec{u}) = \vec{b}$. Show that $\vec{u}_1 - \vec{u}_2 \in \ker(f)$, and conclude every solution to the equation can be represented by $\vec{n} + \vec{u}$ where $\vec{n} \in \ker(f)$.

Problem 38. Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear mapping and suppose $\vec{u}_1$ is a solution to $f(\vec{u}) = \vec{b}_1$ and $\vec{u}_2$ is a solution to $f(\vec{u}) = \vec{b}_2$. Show that $\vec{u}_1 + \vec{u}_2$ is a solution to $f(\vec{u}) = \vec{b}_1 + \vec{b}_2$.

As we saw above, equation $f(\vec{u}) = \vec{b}$ is solvable if $\vec{b} \in \operatorname{Im}(f)$. The following problem answer the solvability of a linear equation $f(\vec{u}) = \vec{b}$ by the aid of the *transpose* of f. Remember that $f^t\colon \mathbb{R}^m \to \mathbb{R}^n$ is the transpose of $f\colon \mathbb{R}^n \to \mathbb{R}^m$ if the following equality holds for any $\vec{u} \in \mathbb{R}^n$, and $\vec{v} \in \mathbb{R}^m$

$$f(\vec{u}) \cdot \vec{v} = \vec{u} \cdot f^t(\vec{v}).$$

Problem 39. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Show $\ker(f^t) = [\operatorname{Im}(f)]^\perp$. Conclude that the linear equation $f(\vec{u}) = \vec{b}$ is solvable if $\langle \vec{b}, \vec{n} \rangle = 0$ for all $\vec{n} \in \ker(f^t)$. Also show

$$\dim \ker(f^t) - \dim \ker(f) = m - n.$$

Problem 40. Find $\ker(f^t)$ of the matrix $A = \begin{pmatrix} 2 & 6 \\ 1 & 3 \end{pmatrix}$ and verify that $\vec{b} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ is orthogonal to $\ker(f^t)$.

2 Functions of several variables

2.1 Topology of $\mathbb{R}^n$

For $p = (x_1, \cdots, x_n) \in \mathbb{R}^n$, the Euclidean *norm* $\|p\|$ is defined as

$$\|p\| = \sqrt{x_1^2 + \cdots + x_n^2},$$

and if $q = (y_1, \ldots, y_n)$, the Euclidean distance is defined as follows

$$\|p - q\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

An immediate result of the above definitions is the convergence of sequences in $\mathbb{R}^n$.

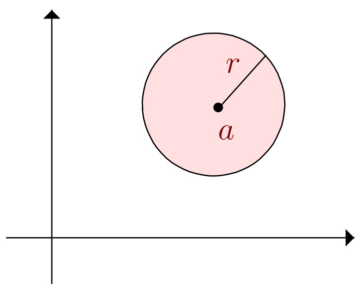
Definition 9. *A sequence $(p_m)_{m=1}^{\infty}$ is called convergent to a if*

$$\lim_{m\to\infty} \|p_m - a\| = 0.$$

Proposition 6. *A sequence p_m converges to a if and only if each coordinate of p_m converges to its associate coordinate of a.*

An open ball of radius r centered at $a \in \mathbb{R}^n$ is defined as

$$B_r(a) = \{p \in \mathbb{R}^n, \|p - a\| < r\}.$$



Problem 41. If $p_n \to a$ then any ball $B_r(a)$ contains infinitely many points of the sequence.

A set $D \subset \mathbb{R}^n$ is called *open* if for any point $a \in D$, there is $r > 0$ such that $B_r(a) \subset D$. A set $D \subset \mathbb{R}^n$ is called *bounded* if there is $r > 0$ such that $D \subset B_r$. If D is an open set then its *complement* D^c is closed. The complement set D^c is defined as

$$D^c = \{p \in \mathbb{R}^n; p \notin D\}.$$

If D is a set, its *closure*, $\mathrm{cl}(D)$ is the smallest closed set containing D, and $\mathrm{bnd}(D)$ denotes the boundary set of D. A point a is called a boundary point of a set D if for any $r > 0$, the following relation holds

$$B_r(a) \cap D \neq \emptyset, B_r(a) \cap D^c \neq \emptyset.$$

The above statements means that any ball centered at a crosses both D and D^c. We have $\mathrm{bnd}(B_r) = \{p, \|p\| = r\}$, and $\mathrm{cl}(B_r) = \{p; \|p\| \leq r\}$, and $B_r^c = \{q, \|q\| > r\}$.

Problem 42. A set $D \subset \mathbb{R}^n$ is closed if and only if any convergent sequence $(p_n); p_n \in D$ converges in D.

Problem 43. Consider the set $A = \left\{\frac{1}{n}, n = 1, 2, 3, \cdots\right\}$. Determine if A is open or closed.

Problem 44. Let $D \subset \mathbb{R}^n$ be any set, show that $\mathrm{bnd}(D)$ is closed.

Problem 45. Show that the set $A = \left\{y; 0 < y < \frac{1}{x}, x > 0\right\}$ is open. Find $\mathrm{bnd}(A)$.

Problem 46. If D_1, D_2 are open sets show that $D_1 \cup D_2$ and $D_1 \cap D_2$ are open. Repeat the argument if D_1, D_2 are closed.

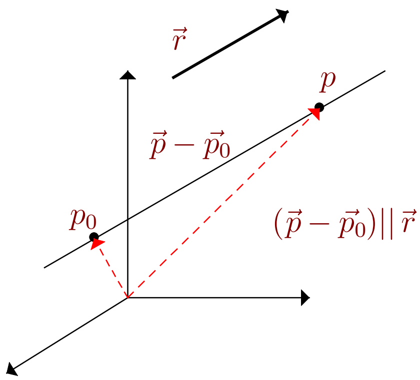
2.2 Straight lines and planes in $\mathbb{R}^3$

The parametric equation of a straight line passing through a point $p_0 = (x_0, y_0, z_0)$ and parallel to a vector $\vec{r} = (a, b, c)$ is $\vec{p_0} + t\,\vec{r}$, or equivalently

$$x(t) = x_0 + a\,t, \quad y(t) = y_0 + b\,t, \quad z(t) = z_0 + ct.$$

If $a \neq 0, b \neq 0, c \neq 0$, then we can rewrite the equation as follows

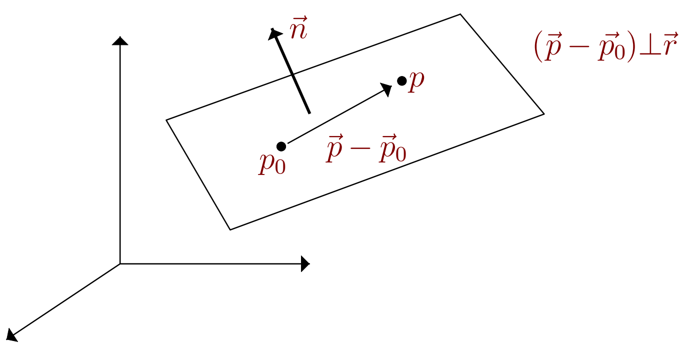
$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$



Similarly, the equation of a straight plane in $\mathbb{R}^3$ passing through a given point $p_0 = (x_0, y_0, z_0)$, and perpendicular to a given vector $\vec{n} = (a, b, c)$ is $\vec{n} \cdot (\vec{p} - \vec{p_0}) = 0$, or equivalently

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

or equivalently $a\,x + b\,y + c\,z = d$, for some constant d.



The intersection of two planes in $\mathbb{R}^3$ can be empty or a line depending on their position to each other. For example, two planes $P_1\colon a_1x + b_1y + c_1z = d_1$ and $P_2\colon a_2x + b_2y + c_2z = d_2$ intersect if their normal vectors $\vec{n}_1 = (a_1, b_1, c_1)$ and $\vec{n}_2 = (a_2, b_2, c_2)$ are not parallel to each other. In this case, the intersection line will be parallel to $\vec{n}_1 \times \vec{n}_2$

$$\vec{n}_1 \times \vec{n}_2 = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = (b_1c_2 - c_1b_2, c_1a_2 - a_1c_2, a_1b_2 - b_1a_2),$$

Hence, the equation of the intersection line is

$$\frac{x-x_0}{b_1c_2-c_1b_2}=\frac{y-y_0}{c_1a_2-a_1c_2}=\frac{z-z_0}{a_1b_2-b_1a_2},$$

where (x_0, y_0, z_0) is a point on the intersection of P_1, P_2.

Example 4. Find the intersection of two plane $P_1: 2x+y=1$, $P_2: y-z=-1$.

Solution. The associated normal vectors of two plains are $\vec{n}_1=(2,1,0)$, $\vec{n}_2=(0,1,-1)$, and they are not parallel. The intersection line of two plains are in the direction of

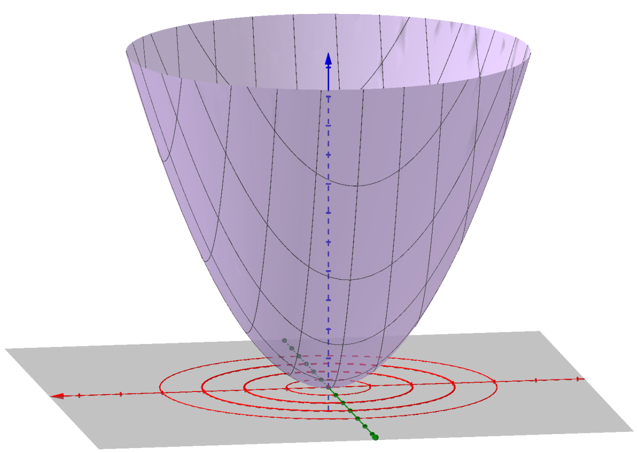
$$\vec{n}_1\times\vec{n}_2=(-1,2,2).$$

Obviously, the point $p_0=(0,1,2)$ lies on both plains, and thus the intersection line equation is

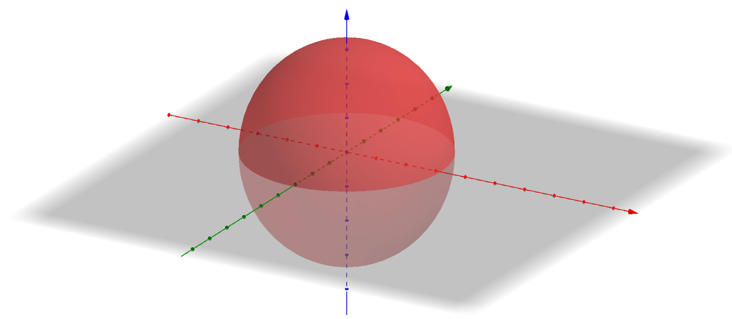
$$\frac{x}{-1}=\frac{y-1}{2}=\frac{z-2}{2}.$$

2.3 Scalar functions

A function f with the domain $D_f\subset\mathbb{R}^n$ is called a *scalar function* if $\mathrm{Im}_f\subset\mathbb{R}$. For example, a mapping that measures the temperature of each point of a room is a scalar function. The graph of a scalar function $y=f(x_1,\ ...x_n)$ is the set$\{(x_1,\ ...x_n,\ y)\}\subset\mathbb{R}^{n+1}$ where $(x_1,\ ...,\ x_n)\in D_f$, the domain of f. The graph of a function $z=f(x,\ y)$ is the surface $\{(x,y,f(x,y))\}$. For example, function $z=x^2+y^2$ is a *paraboloid* in the (x,y,z)-space



The set $\{f(x,y)=c\}$ for a fixed c is called the level set of f with value c. For example, the level sets of $f(x,y)=x^2+y^2$, is the set of circles of radius $\sqrt{c}$ centered at the origin. A level set is also called an *implicit function*, for example $x^2+y^2+z^2=c^2$, that is a sphere of radius c in $\mathbb{R}^3$.



Definition 10. *A scalar function $f: D \to \mathbb{R}$ has a limit L at $a \in D$ if for any sequence $p_m \in D$, $p_m \neq a$, the convergence $p_m \xrightarrow{\mathbb{R}^n} a$ implies $f(p_m) \xrightarrow{\mathbb{R}} L$, that is, for any $\varepsilon > 0$, there is $\delta > 0$ such that if $0 < \|p - a\| < \delta$ then $|f(p) - L| < \varepsilon$.*

It is simply seen that function $f(x, y) = \frac{xy}{x^2 + y^2}$, does not have a limit at $(0, 0)$. In fact, for sequence $(x_m, y_m) = \left(\frac{1}{m}, 0\right)$, the limit is 0. For sequence $(x_m, y_m) = \left(0, \frac{1}{m}\right)$, the limit is again 0, but the limit is $\frac{1}{2}$ for sequence $\left(\frac{1}{m}, \frac{1}{m}\right)$.

Problem 47. Determine if the following functions have a limit at $(0, 0)$

a) $f = \frac{x^2 y}{x^2 + y^2}$

b) $f = \frac{\sin(x) + \sin(y)}{x + y}$

Definition 11. *A scalar function $f: D \to \mathbb{R}$ is called* continuous *at $a \in D$ if for any sequence (p_m), $p_m \to a$, the sequence $f(p_m)$ converges to $f(a)$. The statement is equivalent to the following: for any $\varepsilon > 0$, there is $\delta > 0$ such that $|f(p) - f(a)| < \varepsilon$ for all $p \in B_\delta(a)$.*

Problem 48. Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is continuous. Show that for any open interval J, the set $f^{-1}(J) = \{p; f(p) \in J\}$ is open.

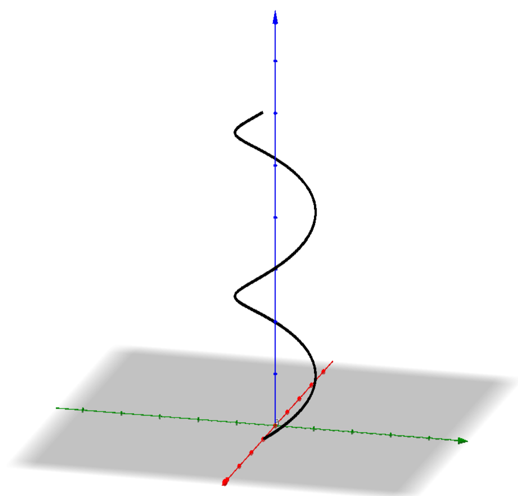
Problem 49. Show that the following function is not continuous at $(0, 0)$.

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}.$$

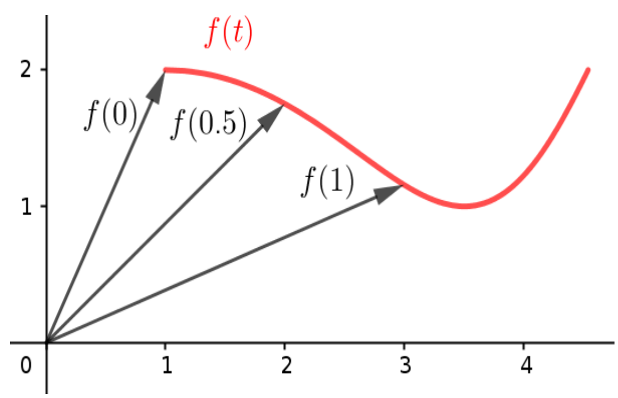
2.4 Vector functions

Let $I \subset \mathbb{R}$ be an interval. A mapping $f: I \to \mathbb{R}^n$ is called a vector valued function. A vector valued function f is usually denoted by $f(t) = (f_1(t), \ldots, f_n(t))$ for parameter t in I.

Example 5. The image of mapping $f(t) = (\cos(t), \sin(t))$ is a unit circle in the plane (x, y) as it satisfies the relation $[x(t)]^2 + [y(t)]^2 = 1$. The image of mapping $f(t) = (\cos t, \sin t, t)$ is a helix in $\mathbb{R}^3$ as shown below.



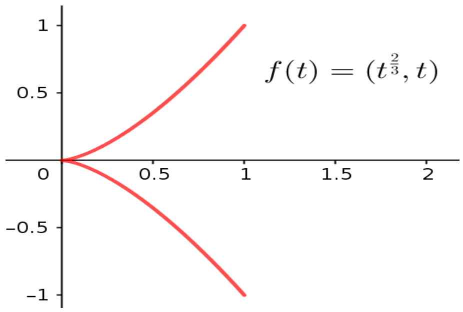
A vector function $f(t)$ has a limit at t_0 if and only if all coordinate functions $f_k(t)$ has a limit at t_0. Similarly, f is continuous if its all coordinate functions are continuous. We write $\lim\limits_{t \to t_0} f(t) = \vec{L}$, if $\lim\limits_{t \to t_0} f_k(t) = l_k$ for all $k = 1, \ldots, n$, and $L = (l_1, \ldots, l_n)$. A vector function is also called a one dimensional parametric function.



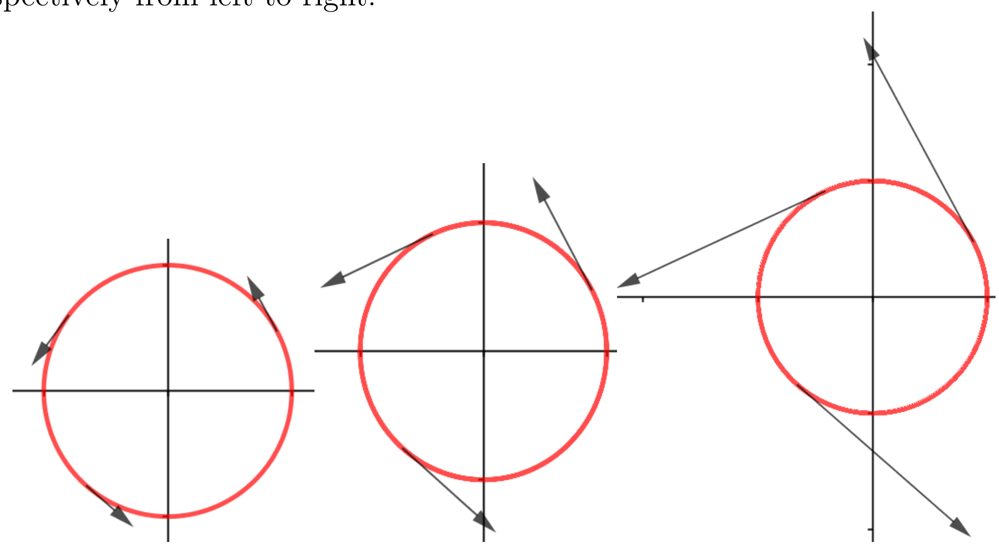
The derivative of vector function is defined coordinate-wise, i.e.,

$$f'(t) = (f_1'(t), \ldots, f_n'(t)).$$

For a fixed $t_0 \in I$, $f'(t_0)$ is the tangent vector on the curve $f(I)$ at t_0 as long as $f'(t_0)$ exists. If $f'(t_0)$ does not exist, we say f is singular at t_0. For example, function $f(t) = \left(t^{\frac{2}{3}}, t\right)$ is singular at $t = 0$.

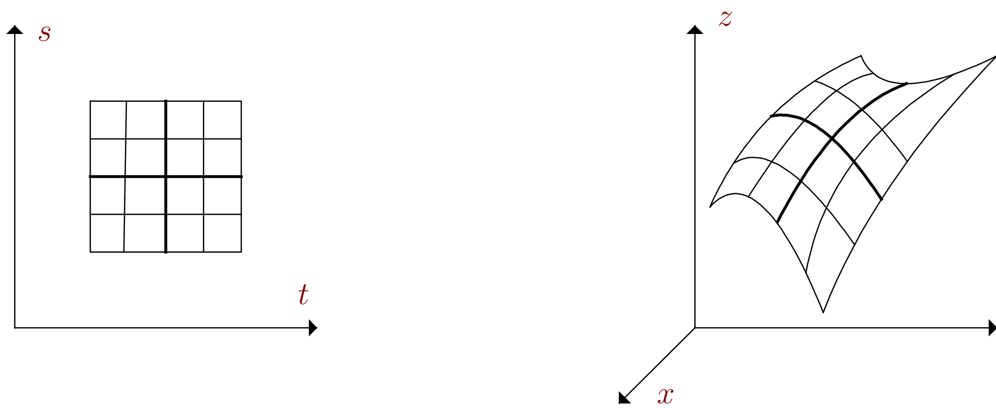


If $f(t)$ denotes the trajectory of a particle in $\mathbb{R}^n$, $f'(t_0)$ is called the velocity vector of that particle, and $\|f(t_0)\|$ is equal to its speed. The parametric representation of a curve provides us with more information than its image. For example, the image of mapping $f(t) = (\cos(\omega t), \sin(\omega t))$ is a unit circle for all values of $\omega \neq 0$, however, if $f(t)$ represents the trajectory of a particle, the speed of the particle would be a function of ω, as the relation $\|f'(t)\| = \omega$ justifies it. The following figure shows the velocity vectors for $\omega = 0.5,\ 1,\ 2$ respectively from left to right.



2.5 Parametric mappings

A mapping $f\colon D_f \subset \mathbb{R}^n \to \mathbb{R}^m$ is called a parametric mapping if $m > 1$. For example if $n = 2$ and $m = 3$, mapping $f(t, s) = (f_1(t, s), f_2(t, s), f_3(t, s))$ defines a parametric surface in $\mathbb{R}^3$.



Parametric mappings are very common for representing complex surfaces. For example, a torus is represented by the following parametric mapping

$$f(\theta,\phi)=((c+a\cos(\theta))\cos(\phi),(c+a\cos(\theta))\sin(\phi),a\sin(\theta))$$

where $c>a$ are some constants. The shape is shown below.

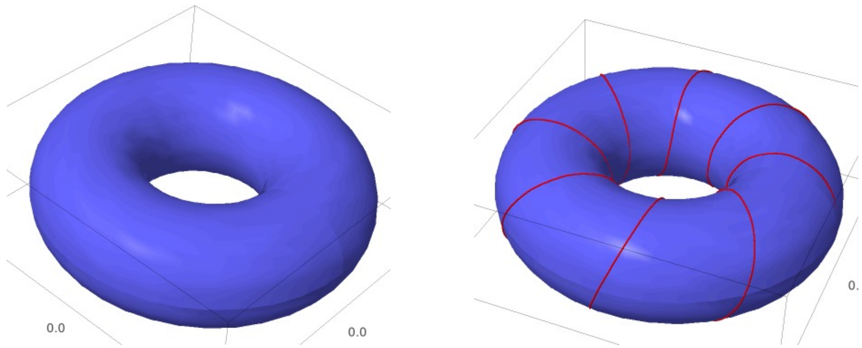


Figure 5.

The restriction of $f(t,s)$ to a curve in the (t,s) plane is mapped to a curve in the surface $f(t,s)$. For example, $\gamma\colon\phi=\theta/6$ is a straight line in (ϕ,θ) plane, and it is mapped under f to the following space curve

$$\Gamma(\theta)=((c+a\cos(\theta))\cos(\theta/6),(c+a\cos(\theta))\sin(\theta/6),a\sin(\theta)).$$

The following parametric mappings are respectively representations of a sphere and a cylinder

$$S\colon(\sin(\theta)\cos(\phi),\sin(\theta)\sin(\phi),\cos(\theta)),\quad C\colon(\cos(\theta),\sin(\theta),z)$$

The concept of limit and continuity is coordinate-wise as well, that is, f has a limit at $p\in D_f$ if each coordinate function has a limit at $p\in D_f$.

3 Derivatives

3.1 Derivatives of scalar functions

3.1.1 Partial derivatives.

Let D_f be an open set in $\mathbb{R}^n$. For a scalar function $f\colon D_f\to\mathbb{R}$, the *partial derivative* $\partial_k f$ at $p\in D_f$ is defined by the following limit

$$\partial_k f(p)=\lim_{t\to 0}\frac{f(p+t\hat{e}_k)-f(p)}{t}, \tag{4}$$

as long as the limit exists. For a two variable function $f(x,y)$, the partial derivatives $\partial_x f$, $\partial_y f$ at $p=(a,b)$ are defined respectively as follows

$$\partial_x f(a,b)=\lim_{t\to 0}\frac{f(a+t,b)-f(a,b)}{t},$$

$$\partial_y f(a,b)=\lim_{t\to 0}\frac{f(a,b+t)-f(a,b)}{t},$$

as long as the limits exist.

Remark 1. For the sake of simplicity, we use the flat notations ∂_x, ∂_y in this book instead of the standard ones $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$. Another notation for the partial derivative is f_x, f_y for $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$.

Remark 2. Similarly, we can define partial derivative *functions* $\partial_x f, \partial_y f$ in the open set D_f, the domain of f as

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

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

for $(x, y) \in D_f$.

Remark 3. The existence of partial derivatives of a function at a point does not guarantees the continuity of the function at that point. Consider the following function

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

Even though, $\partial_x f(0, 0), \partial_y f(0, 0)$ exist and are equal zero, the function is not continuous at the origin. However, $f(x, y)$ must be continuous and differentiable *with respect to* x at (a, b) in order that $\partial_x f(a, b)$ exists. Similarly, $f(x, y)$ must be continuous and differentiable *with respect to* y at (a, b) in order that $\partial_y f(a, b)$ exists.

3.1.2 Interpretations of partial derivatives

Like a single variable function, there are two related interpretations of partial derivatives. Consider a two variable function $f(x, y)$ defined on an open set D_f. Fix a point $(a, b) \in D_f$, and consider the horizontal line parallel to x-axis passing through (a, b). The partial derivative $\partial_x f(a, b)$ measures the rate of change of f at (a, b) *along* the horizontal line, and similarly, $\partial_y f(a, b)$ measures the rate of change of f at (a, b) *along* the vertical line passing through (a, b). For example, the rate of change of function $f(x, y) = \sqrt{x^2+y^2}$ at $(1, 1)$ along x-axis is

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

The slope of tangent lines to the surface of $f(x, y)$ are expressed in terms of partial derivatives. The projection of line $(a+t, b)$, for $t \in (-c, c)$ for some $c > 0$ on the graph of $z = f(x, y)$ is a curve of the following form

$$\Gamma_1(t) = (a+t, b, f(a+t, b)).$$

This space curve passes through $(a, b, f(a, b))$ at $t = 0$. It is simply seen that $\partial_x f(a, b)$ is equal to the slope of tangent line to $\Gamma_1(t)$ in the (x, z)-plane at $t = 0$:

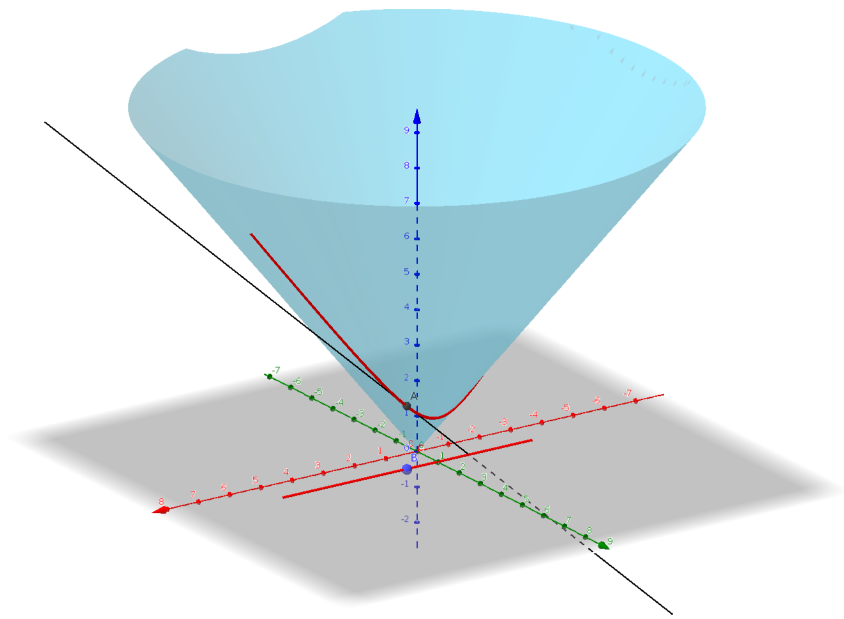
$$\left.\frac{d\Gamma_1}{dt}\right|_{t=0} = (1, 0, \partial_x f(a, b))$$

Similarly, $\partial_y f(a,\, b)$ is equal to the slope of tangent line to the curve $\Gamma_2(t) = (a,\, b + t,\, f(a,\, b + t))$ at $t = 0$ in the $(y,\, z)$-plane. The following figure shows the graph of function $f(x, y) = \sqrt{x^2 + y^2}$, and the projection of $\gamma_1(t) = (1 + t, 1)$ on it:

$$\Gamma_1(t) = \left(1 + t, 1, \sqrt{2 + t^2 + 2t}\right).$$

The black line is the tangent to the space curve Γ_1 at $t = 0$. The slope of the tangent line, is the tangent of the angle the line makes with the horizontal line γ_1, that is,

$$m = \frac{d}{dt}\sqrt{2 + t^2 + 2t}|_{t=0} = \frac{1}{\sqrt{2}}.$$



Note that $\frac{d\Gamma_1}{dt}(0)$ is the tangent vector to the curve Γ_1 at time 0

$$\frac{d\Gamma_1}{dt}(0) = \left(1, 0, \frac{1}{\sqrt{2}}\right).$$

A similar argument holds for $\partial_y f$, that is, if Γ_2 is the projection of $\gamma_2(t) = (1, 1 + t)$ on the graph of f, then

$$\frac{d\Gamma_2}{dt}(0) = \left(0, 1, \frac{1}{\sqrt{2}}\right).$$

Vectors $\frac{d\Gamma_1}{dt}(0)$, $\frac{d\Gamma_2}{dt}(0)$ are both tangent to the graph of f at $(1,\, 1,\, \sqrt{2})$ and thus the plane $\text{span}\{\frac{d\Gamma_1}{dt}(0),\, \frac{d\Gamma_2}{dt}(0)\}$ is the tangent plane to the surface of f at $(1,\, 1,\, \sqrt{2})$. The algebraic equation of the tangent plane is derived by the air of $\vec{n}$

$$\vec{n} = \vec{v}_1 \times \vec{v}_2 = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 1\right),$$

and thus the algebraic of the tangent plane is derived as

$$-\frac{1}{\sqrt{2}}(x-1)-\frac{1}{\sqrt{2}}(y-1)+z-\sqrt{2}=0.$$

Remark 4. For a general function $z=f(x,y)$, two principal tangent lines at $p=(a,b)$ are

$$\vec{v}_1=(1,0,\partial_x f(a,b)),\vec{v}_2=(0,1,\partial_y f(a,b)),$$

and thus $\vec{n}$, the *normal vector* to the graph of f at p is

$$\vec{n}=(-\partial_x f(p),-\partial_y f(p),1).$$

Accordingly, the algebraic equation on the tangent plane at p is

$$-\partial_x f(a,b)\,(x-a)-\partial_y f(a,b)\,(y-b)+z-f(a,b)=0.$$

3.1.3 Chain rule

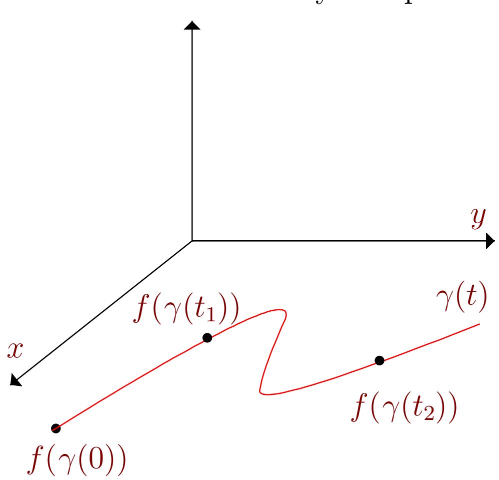
Let f be a scalar function on $D_f\subset\mathbb{R}^2$, and assume that x,y are functions of another variable, say t, i.e., $x=x(t)$, $y=y(t)$. In the final analysis, $f(x,\,y)$ is a function of t and thus the ordinary derivative $\frac{df}{dt}$ can exist.

$$\frac{df}{dt}(t)=\lim_{h\to 0}\frac{f(x(t+h),y(t+h))-f(x(t),y(t))}{h}.$$

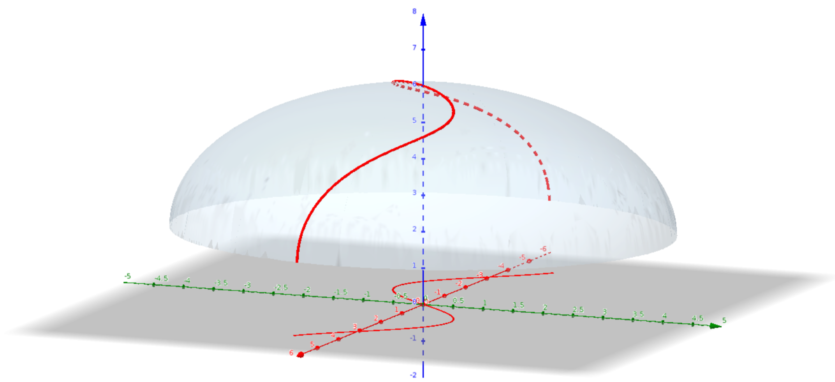
Proposition 7. *Assume that f is differentiable with respect to x and y, and moreover the partial derivatives are continuous. If $x(t)$, $y(t)$ are differentiable functions of t, then the following equality that is called* chain rule *holds*

$$\frac{df}{dt}=\partial_x f(x(t),y(t))\frac{dx}{dt}+\partial_y f(x(t),y(t))\frac{dy}{dt}. \tag{5}$$

There is an important interpretation for the above formula. First, note that $\gamma\colon(x(t),y(t))$ defines a parametric curve in the (x,y)-plane, and thus $f(x(t),y(t))$ can be considered as the *restriction* of f to γ. Also, we can consider γ as the path of a particle moving in the (x,y)-plane. Therefore, relation (5) states the rate of change of the value of that particle *along* γ. For example, if f is the density distribution function in the plane, then relation (5) states how fast or slow the density of a particle changes when it moves along path γ.



On the other hand, since the graph of f is a surface in $\mathbb{R}^3$, $f(\gamma(t))$ is the projection of $\gamma(t)$ in the surface as shown in the following figure. With this interpretation, relation (5) defines the slope of tangent to the curve at any instance of time.



Let us consider again equality (5) and rewrite it as follows

$$\partial_x f\,\frac{dx}{dt} + \partial_y f\,\frac{dy}{dt} = \begin{pmatrix} \partial_x f \\ \partial_y f \end{pmatrix} \cdot \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix}.$$

Note that $\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix}$ is just the tangent vector of the curve $\gamma = (x(t), y(t))$, i.e., $\gamma'(t)$. Vector $\begin{pmatrix} \partial_x f \\ \partial_y f \end{pmatrix}$ is called the gradient of f and is denoted by $\mathrm{grad}(f)$ or ∇f. Therefore, equality (5) can be rewritten as

$$\frac{df}{dt} = \nabla f \cdot \gamma'(t),$$

and for this reason, $\frac{df(\gamma(t))}{dt}$ is called also the derivative of f *along* $\gamma(t)$. The chain rule can be extended to higher dimensions. For example, if $f(x, y)$ is a differentiable function with respect to x, y and $x = (x(t, s)), y = y(t, s)$ are differentiable functions then

$$\partial_t f = \partial_x f \partial_t x + \partial_y f \partial_t y,$$

$$\partial_s f = \partial_x f \partial_s x + \partial_y f \partial_s y.$$

Problem 50. If $u = u(t, x)$, and $x = x(t)$, find $\frac{du}{dt}$.

Problem 51. If $u = f(x - 2t)$ find $\partial_t u$ and $\partial_x u$.

3.1.4 Directional derivative

Partial derivatives are just special cases of a more general derivative called the directional derivative. Assume that a direction vector $\hat{v} = (v_1, v_2)$ is given (a direction vector is a unit vector), and $f\colon D_f \to \mathbb{R}$ is a given continuous function. The directional derivative of f at $(a, b) \in D_f$ along $\hat{v}$ is defined by the following limit

$$\partial_{\hat{v}} f(a, b) = \lim_{t \to 0} \frac{f(a + t v_1, b + t v_2) - f(a, b)}{t},$$

as long as the limit exists. If so, then $\partial_{\hat{v}} f(a,b)$ measures the rate of change of f at (a,b) along $\hat{v}$. Obviously if $\hat{v}=(1,0)$ then $\partial_{\hat{v}} f(a,b)=\partial_x f(a,b)$ and if $\hat{v}=(0,1)$, it would be equal to $\partial_y f(a,b)$.

Proposition 8. *If $\partial_x f, \partial_y f$ are continuous at (a,b), that is,*

$$\lim_{(x,y)\to(a,b)} \partial_x f(x,y)=\partial_x f(a,b), \qquad \lim_{(x,y)\to(a,b)} \partial_y f(x,y)=\partial_y f(a,b),$$

then

$$\partial_{\hat{v}} f(a,b)=\nabla f(a,b)\cdot\hat{v}$$

The continuity in the above proposition is crucial, for example, consider the following function

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

If $\hat{v}=\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)$, then

$$\partial_{\hat{v}}(0,0)=\lim_{t\to 0}\frac{\frac{t^3}{2\sqrt{2}\,t^2}}{t}=\frac{1}{2\sqrt{2}},$$

however, $\nabla f(0,0)=\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$, and thus $\nabla f(0,0)\cdot\hat{v}=0$. The reason is that $\partial_x f$, $\partial_y f$ are not continuous at $(0,0)$. To see this, let us find $\partial_y f$ for $(x,y)\neq 0$ as

$$\partial_y f(x,y)=\frac{x^2(x^2-y^2)}{(x^2+y^2)^2},$$

and observe that $\partial_y f$ does not have even a limit at $(0,0)$. Note also that the directional derivative of a function f is a special case of the chain rule $\nabla f\cdot\gamma'(t)$.

Problem 52. For the following function

$$f(x,y)=\begin{cases} \frac{x^2 y}{x^4+y^2} & (x,y)\neq(0,0) \\ 0 & (x,y)=(0,0) \end{cases}$$

show that $\partial_{\hat{r}} f(0,0)$ exists for any direction $\hat{r}$ but partial derivatives are not continuous at $(0,0)$.

3.1.5 Gradient

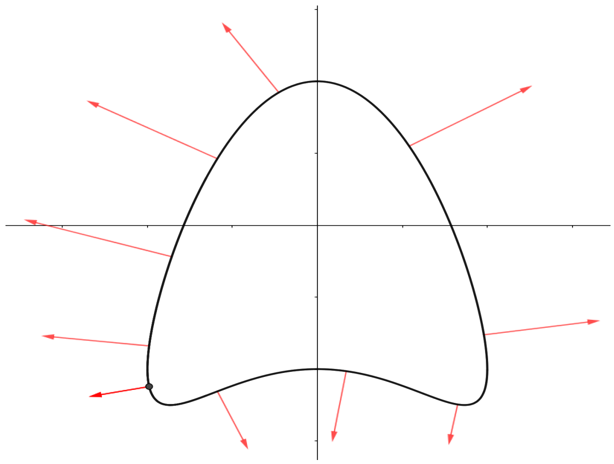
Consider the level curve defined by γ: $f(x,y)=c$. If $\gamma(t)=(x(t),y(t))$ is a parametrization of this curve then $\nabla f\cdot\gamma'(t)=0$ for any t as long as ∇f is a continuous vector function. This relation means that $\nabla f(\gamma(t))$ is always perpendicular to $\gamma(t)$. For example, the level curve

$$x^4+2x^2y+x^2+y^2=1,$$

has the gradient

$$\nabla f = \begin{pmatrix} 4x^3 + 4xy + 2x \\ 2x^2 + 2y \end{pmatrix}.$$

The following figure shows a few of gradient vectors Δf on the level curve



As it is observed, gradient vectors are perpendicular to level curves. On the other hand, if $\hat{n} = (n_1, n_2)$ is the direction vector at a point on the level curve, the the directional derivative $\partial_{\vec{n}} f$ of f is $\nabla f \cdot \hat{n}$, and since $\hat{n} = \frac{\nabla f}{\|\nabla f\|}$ (as long as the $\nabla f \neq 0$), we obtain

$$\partial_{\hat{n}} f = \|\nabla f\|.$$

Therefore, the magnitude of Δf at a point measures the rate of change of f along the normal direction on the level curve. This result is extremely useful to maximize (or minimize) a scalar function. The procedure is as follows. To maximize $f(x, y)$, we fix an initial point $p_0 = (x_0, y_0)$. The next point p_1 is obtained by the following relation

$$p_1 = p_0 + \lambda \frac{\nabla f(p_0)}{\|\nabla f(p_0)\|},$$

where $\lambda > 0$ is a small value. Geometrically, that mean we take a step of length λ along the direction $\nabla f(p_0)$. Iterating this procedure, that is,

$$p_{n+1} = p_n + \lambda \frac{\nabla f(p_n)}{\|\nabla f(p_n)\|},$$

converges the maximum point of f as long such a local or global maximum exists, and if f satisfies some other verifiable conditions.

In above, we used frequently the operator *nabla* $\nabla = \begin{pmatrix} \partial_1 \\ \vdots \\ \partial_n \end{pmatrix}$. It is applied to differentiable function as $\nabla f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_n f \end{pmatrix}$. We study this operator in more detail later in this appendix.

Problem 53. Show the following relations

a) $\nabla(f+g) = \nabla f + \nabla g$.

b) $\nabla(kf) = k\nabla f$, $k \in \mathbb{R}$.

c) $\nabla(fg) = f\nabla g + g\nabla f$.

3.1.6 Derivative and differential

Let $f\colon D_f \subset \mathbb{R}^n \to \mathbb{R}$ be a scalar function defined on an open set D_f. The derivative of f at p_0 is the linear mapping $D_{p_0}f\colon \mathbb{R}^n \to \mathbb{R}$ such that the following relation holds for any $\vec{h} \in \mathbb{R}^n$

$$\lim_{\vec{h}\to 0} \frac{f(p_0+\vec{h}) - f(p_0) - D_{p_0}f(\vec{h})}{\|\vec{h}\|} = 0 \tag{6}$$

Obviously if f is differentiable at p_0 then it is continuous at that point. Moreover, such a linear mapping must be unique. Note that if a function f is differentiable at a point a then directional derivatives along any direction at a exist. The reverse holds only if partial derivatives are continuous at a.

Problem 54. Verify that the above definition is compatible with the usual definition for single variable functions.

Problem 55. Show that if f is differentiable at p_0 it must be continuous at that point. Show also that its derivative (the linear mapping) is unique.

Proposition 9. *Assume that f has continuous partial derivatives at p_0, that is,*

$$\lim_{p\to p_0} \partial_k f(p) = \partial_k f(p_0),$$

for $k = 1, \ldots, n$, then $D_{p_0}f$ exists and has the matrix representation

$$D_{p_0}f = [\partial_1 f(p_0), \cdots, \partial_n f(p_0)].$$

According to the above proposition, if f has continuous partial derivatives at p_0 then for any $\vec{h} \in \mathbb{R}^n$, the following relation holds.

$$D_{p_0}f(\vec{h}) = \nabla f(p_0) \cdot \vec{h}.$$

For this reason, some texts write $D_{p_0}f = \nabla f(p_0)$ that we should keep in mind that $D_{p_0}f$ is a $1 \times n$ matrix, while $\nabla f(p_0)$ is a $n \times 1$ vector.

Example 6. For example, function

$$f(x,y)=\begin{cases} \frac{x^2y}{x^4+y^2} & (x,y)\neq(0,0) \\ 0 & (x,y)=(0,0) \end{cases}$$

has directional derivatives in all directions at the origin, however, the function is not differentiable at this point since it is not even continuous at the origin (why?). On the other hand, function $f(x,y)=x^2+y^2$ has continuous partial derivatives $\partial_x f=2x, \partial_y f=2y$ and for any $p_0=(x_0,y_0)$, we have

$$D_{p_0}f=2[x_0,y_0].$$

Let us verify definition (6) for f at $p_0=(1,-1)$. For arbitrary $\vec{h}=(h_1,h_2)$, we have

$$\lim_{\vec{h}\to 0}\frac{(1+h_1)^2+(-1+h_2)^2-2-2(h_1-h_2)}{\sqrt{h_1^2+h_2^2}}=\lim_{\vec{h}\to 0}\frac{h_1^2+h_2^2}{\sqrt{h_1^2+h_2^2}}=\lim_{\vec{h}\to 0}\|\vec{h}\|=0.$$

An immediate result of definition (6) is the *linear approximation* formula. If $f\colon D_f\to\mathbb{R}$ is continuously differentiable at $p_0\in D_f$, that is

$$\lim_{p\to p_0}D_pf=D_{p_0}f,$$

then

$$f(p)\approx f(p_0)+D_{p_0}f(\vec{p}-\vec{p}_0),$$

or equivalently

$$f(p)\approx f(p_0)+\nabla f(p_0)\cdot(\vec{p}-\vec{p}_0).$$

For functions of two variables, the above formula reads

$$f(x,y)=f(x_0,y_0)+\partial_x f\,(x-x_0)+\partial_y f\,(y-y_0).$$

Note that the right hand side is the equation of tangent plane at $(x_0,y_0,f(x_0,y_0))$:

$$T(x,y)=(x_0,y_0)+\partial_x f\,(x-x_0)+\partial_y f\,(y-y_0).$$

Theorem 3. (Mean Value Theorem) *Assume that $f\colon D_f\to\mathbb{R}$ is continuously differentiable everywhere in D_f. Fix a point $p_0\in D_f$. Then for any point $p\in D_f$, there is $\xi\in tp+(1-t)p_0$, $t\in(0,1)$ such that*

$$f(p)=f(p)+\nabla f(\xi)\cdot(\vec{p}-\vec{p}_0).$$

Problem 56. Prove the theorem. Hint: define $g(t)=f(tp+(1-t)p_0)$ and apply the mean value theorem for the single variable function $g(t), t\in[0,1]$.

Definition 12. *The total differential of* $f\colon D_f \to \mathbb{R}$ *at* p_0 *is defined by the following formula*

$$df(p_0) = \partial_1 f(p_0)\, dx_1 + \cdots + \partial_n f(p_0)\, dx_n.$$

Remember that for a single variable function $y = y(t)$, the differential dy is defined by the relation $dy(t_0) = y'(t_0)\, dt$. Fig.(6) below shows this relation geometrically.

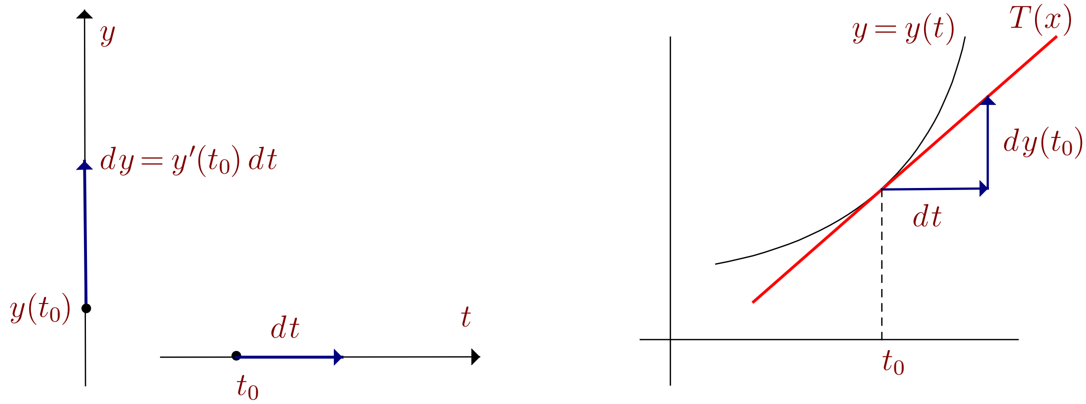


Figure 6.

Similarly for a 2-variable function $z = f(x, y)$, dz is defined by the following relation and its geometry is represented in Fig. 7.

$$dz = \partial_x f dx + \partial_y f dy.$$

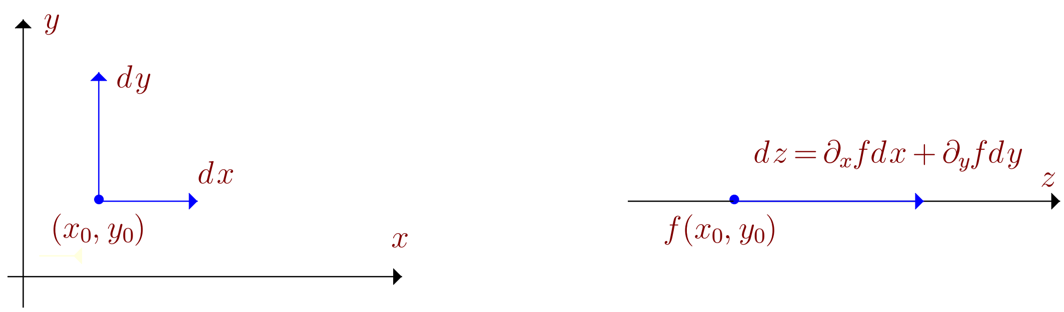


Figure 7.

3.1.7 Critical points and local max and min

Let $D_f \subset \mathbb{R}^n$ be an open set. A point $a \in D_f$ is called a local min (or max) of $f\colon D_f \to \mathbb{R}$ if there is a ball $B_\delta(a)$ such that $f(p) \leq f(a)$ (alternatively $f(p) \geq f(a)$) for all $p \in B_\delta(a)$. If f is differentiable at a, and if a is a local min or max, then $D_a f$ is a zero mapping. To see this, suppose a is a local min, choose an arbitrary direction vector $\vec{h}$, and write

$$0=\lim_{t\to 0}\frac{f(a+t\vec{h})-f(a)-D_af(t\vec{h})}{\|t\vec{h}\|}=\lim_{t\to 0}\frac{f(a+t\vec{h})-f(a)-t\,D_af(\vec{h})}{|t|}.$$

For $t>0$, we have

$$D_af(\vec{h})=\lim_{t\to 0}\frac{f(a+t\,\vec{h})-f(a)}{t}\geq 0$$

For $t<0$, we have

$$D_af(\vec{h})=\lim_{t\to 0}\frac{f(a+t\,\vec{h})-f(a)}{t}\leq 0,$$

and thus $D_af(\vec{h})=0$ for arbitrary $\vec{h}$, and thus D_af is a zero mapping (meaning $\nabla f(a)$ is a zero vector).

Definition 13. (Critical point) *A point a is called a critical point of a function f if either D_af does not exist or D_af is a zero mapping. If D_af is a zero mapping, a can be a local min, local max, a saddle point or non of them.*

In order to determine the type of a critical point in terms of min, max or saddle, we need the notion of second derivatives. Second order partial derivatives $\partial_{ij}f$ are defined as $\partial_{ij}f=\partial_i(\partial_j f)$. We have the following theorem.

Theorem 4. *Let $D_f\subset\mathbb{R}^n$ be an open set, and $f\colon D_f\to\mathbb{R}$. Furthermore assume that $\partial_{ij}f$ is continuous, then $\partial_{ij}f=\partial_{ji}f$.*

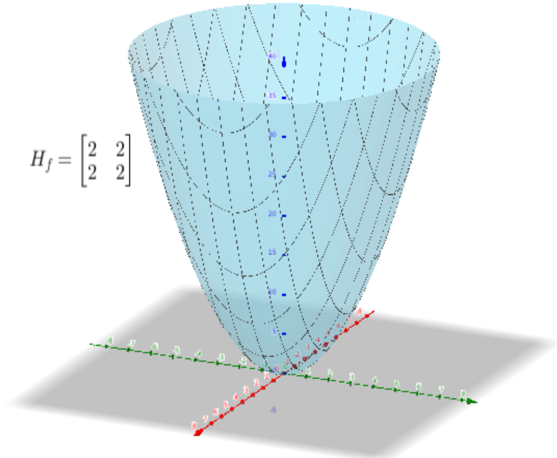
Theorem 5. *Assume that $f\colon D_f\to\mathbb{R}$ is continuously differentiable of order 2 on open set D_f. Fix $a\in D_f$, then there is $\xi\in tp+(1-t)a$ for some $t\in(0,1)$ such that the following relation holds for any for any $p\in D_f$*

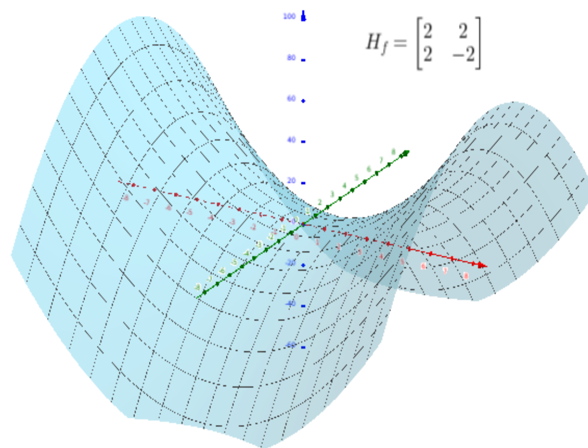
$$f(p)=f(a)+\nabla f(a)\cdot(\vec{p}-\vec{a})+(H_f(\xi)(\vec{p}-\vec{a}))\cdot(\vec{p}-\vec{a}),$$

where H_f is the Hessian matrix of f defined as $H_f=[\partial_{ij}f]_{i,j}$.

Corollary 1. *Assume that f is second order continuously differentiable function and $\nabla f(a)=0$, then a is a local min if $H_f(a)$ is a positive definite matrix, a is a local max if $H_f(a)$ is a negative definite matrix, and a saddle point if $H_f(a)$ has eigenvalues with opposite signs.*

The standard example of above three cases is $f=x^2+y^2$, $f=-x^2-y^2$, and $f=x^2-y^2$ as shown below





Note that $H_f(a)$ is a symmetric matrix, and thus it has n orthogonal eigenvectors, and the JORDAN form $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Therefore if all eigenvalues of $H_f(a)$ is positive then a is a local min, if its all eigenvalues are negative then a is local max, and if there are some positive and some negative eigenvalues, it is a saddle point. If there is at least one zero eigenvalue, or equivalently $\det(H_f(a)) = 0$, then a may not be any of these types.

Problem 57. Find all critical points of the function $f(x, y) = x^3 + y^3 + \frac{9}{2}x^2 - \frac{3}{2}y^2 + 6x$ and classify them.

3.2 Derivative of non-scalar mappings

3.2.1 Jacobi matrix

Let $D_f \subset \mathbb{R}^n$ be an open set and $f\colon D_f \to \mathbb{R}^m$ a continuous map. f is differentiable at $a \in D_f$ if there is a linear mapping $D_a f\colon \mathbb{R}^n \to \mathbb{R}^m$ that satisfies the following relation

$$\lim_{\vec{h} \to 0} \frac{\| f(a + \vec{h}) - f(a) - D_a f(\vec{h}) \|}{\| \vec{h} \|} = 0.$$

Equivalently, $f = (f_1, \ldots, f_m)$ is differentiable at a if and only if each coordinate function f_k is differentiable at a. On the other hand, the derivative of f_k at a is

$$D_a f_k = [\partial_1 f_k(a), \ldots, \partial_n f_k(a)],$$

and since each f_k is defined on $\mathbb{R}^n$, we obtain

$$D_a f = \begin{pmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \cdots & \partial_n f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) & \cdots & \partial_n f_2(a) \\ \vdots & \vdots & \vdots & \vdots \\ \partial_1 f_m(a) & \partial_2 f_m(a) & \ldots & \partial_n f_m(a) \end{pmatrix}.$$

The above matrix is called the Jacobi matrix of f at a, denoted also by $J_f(a)$. Let $\gamma(t)$ be a smooth curve passing through a at $t = 0$. This curve is mapped into $\mathbb{R}^m$ by f as $f(\gamma(t))$. In the final analysis, $f(\gamma(t))$ is a vector valued function and therefore, we have

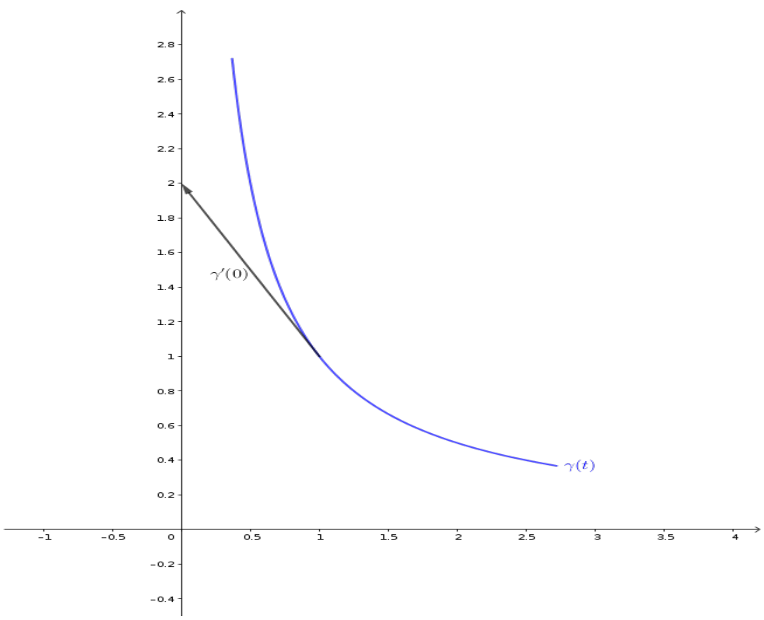
$$\frac{d f(\gamma(0))}{dt} = D_{\gamma(0)} f(\vec{\gamma}'(0)).$$

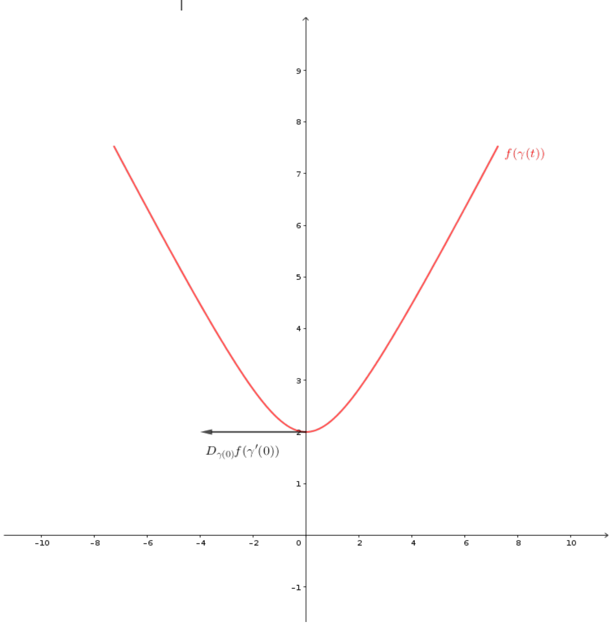
Since $\gamma'(0)$ is the tangent vector on $\gamma(t)$ at $t=0$, $D_{\gamma(0)}f(\vec{\gamma}'(0))$ is the tangent vector on $f(\gamma(t))$ at $t=0$. For example, for $f(x,y)=(x^2-y^2,x^2+y^2)$ we have

$$D_{(1,1)}f=\begin{pmatrix} 2 & -2 \\ 2 & 2 \end{pmatrix}.$$

If $\gamma(t)=(e^{-t},e^{t})$ (passing through $(0,0)$ at $t=0$), we have $\gamma'(0)=\left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right)$ and accordingly, $D_{\gamma(0)}f(\gamma'(0))=\left(\begin{smallmatrix} -4 \\ 0 \end{smallmatrix}\right)$. Note that

$$\frac{d}{dt}f(\gamma(t))|_{t=0}=\frac{d}{dt}(e^{-2t}-e^{2t},e^{-2t}+e^{-2t})|_{t=0}=\begin{pmatrix} -4 \\ 0 \end{pmatrix}.$$





Theorem 6. *Assume that the Jacobi matrix of a mapping $f: D_f \subset \mathbb{R}^n \to \mathbb{R}^n$ is invertible at a. Then there is a neighborhood $B_\delta(a)$ such that f is one to one on $B_\delta(a)$.*

Note that $\det(J_f(a))$ measures the volume of a parallelogram constructed on columns of $J_f(a)$. In other word, if C is a unit cube made at a, then $\det(J_f(a))$ is equal to the volume of parallelogram $J_f(a)(C)$. Hence, if $\det(J_f(a)) \neq 0$, there is a neighborhood $B_\delta(a)$ such that f is one to one on $B_\delta(a)$.

3.2.2 Smooth surfaces

Remember that a smooth space curve is represented by a curve map $\gamma(t)$ such that $\gamma'(t)$ is *nonzero*. Geometrically, this condition means that $\gamma(t)$ always admits a tangent vector that varies continuously along γ.

Definition 14. *A surface S in $\mathbb{R}^3$ is called smooth if S has a nonzero normal (perpendicular) vector at all points on S.*

Let $f(t, s) \in \mathbb{R}^3$ be parametric surface. Consider an arbitrary point $f(t_0, s_0)$ on S. Coordinate line $\gamma_1(t) = (t + t_0, s_0)$ is mapped on S as $\Gamma_1(t) = f(t + t_0, s_0)$. The tangent vector to this space curve is just $\Gamma_1'(t_0) = \partial_t f(t_0, s_0)$. Similarly, the coordinate line $\gamma_2(s) = (t_0, s + s_0)$ is mapped as $\Gamma_2(s) = f(t_0, s + s_0)$ and the tangent vector is $\Gamma_2'(s) = \partial_s f(t_0, s_0)$. Notice that both Γ_1', Γ_2' are tangent to S and therefore $\vec{n} := \Gamma_1' \times \Gamma_2'$ is perpendicular to S as long as it is *nonzero*

$$\vec{n} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \partial_t x\,(t_0, s_0) & \partial_t y(t_0, s_0) & \partial_t z(t_0, s_0) \\ \partial_s x\,(t_0, s_0) & \partial_s y(t_0, s_0) & \partial_s z(t_0, s_0) \end{vmatrix} \neq 0.$$

Let us see the result for a smooth function $z = f(x, y)$. The graph of function is

$$\phi(x, y) = (x, y, f(x, y)),$$

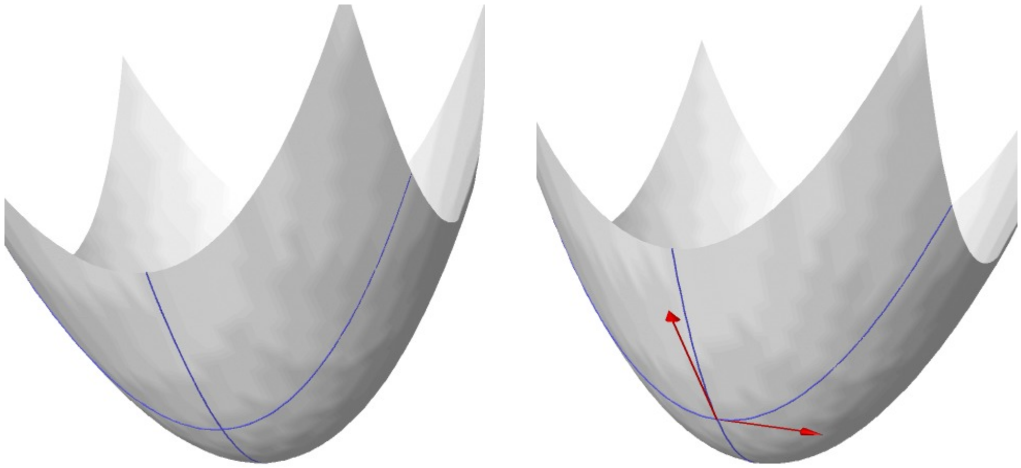
and thus $\partial_x \phi = (1, 0, \partial_x f)$ and $\partial_y \phi = (0, 1, \partial_y f)$. The normal vector $\vec{n}$ is

$$\vec{n} = \partial_x \phi \times \partial_y \phi = (-\partial_x f, -\partial_y f, 1).$$

Example 7. For $p_0 = (0, 1, 1)$ on the surface of $f(x, y) = x^2 + y^2$, consider coordinate lines $\gamma_1(x) = (x, 1), \gamma_2(y) = (0, y)$. We have $f(\gamma_1) = (x, 1, 1 + x^2)$, $f(\gamma_2) = (0, y, y^2)$, see the figure shown below. Respectively, the tangent vectors are as $T_1 = (1, 0, 0)$, $T_2 = (0, 1, 2)$. The normal vector $\vec{n}$ to the surface at p_0 is

$$\vec{n} = T_1 \times T_2 = (0, -2, 1),$$

and therefore the equation of tangent plane is $-2(y - 1) + z - 1 = 0$.



For a surface represented by the implicit function S: $f(x,y,z)=0$, the normal vector $\vec{n}$ is derived by the following procedure. Consider an arbitrary space curve $\gamma=(x(t),y(t),z(t))$ on S, that is, $f(x(t),y(t),z(t))=0$. The chain rule states

$$\frac{df}{dt}(\gamma)=\nabla f(\gamma(t))\cdot\gamma'(t)=0.$$

Since $\gamma'(t)$ is tangent to S, then ∇f is perpendicular to $\gamma(t)$ if ∇f is nonzero. On the other hand, since γ is arbitrary, then γ' belongs to the tangent plane on S, and thus ∇f is perpendicular to S. Therefore, we obtain $\vec{n}$ as

$$\vec{n}=(\partial_x f,\partial_y f,\partial_z f).$$

Problem 58. Write the equation of curve formed by the intersection of the unit sphere $x^2+y^2+z^2=4$ and the plane $x+y+z=1$.

3.3 Implicit function theorem

An implicit function $f(x,\,y)\,=\,0$ defines generally a planar curve in the $(x,\,y)$ plane. If $y=y(x)$, then by the chain rule, we can write

$$\frac{dy}{dx}=-\frac{\partial_x f}{\partial_y f}.$$

However, there is no guarantee in general that y could be solved in terms of x or x could be solved in terms of y. Question is this: is there any function $y=g(x)$ defined on an open interval I such that $f(x,g(x))=0$ for all $x\in I$. The following theorem answers the question.

Theorem 7. (implicit function theorem) *Suppose implicit function $f(x,y)=0$ satisfies the following conditions*

i. there is a point $p_0=(x_0,y_0)$ such that $f(x_0,y_0)=0$,

ii. there is an open ball $B_\varepsilon(p_0)$ such that f has continuous partial derivatives on it, *iii. and that* $\partial_y f(x_0, y_0) \neq 0$,

then there is an open interval $I = (x_0 - \delta, x_0 + \delta)$, *and a function* $y = g(x)$ *such that* $y_0 = g(x_0)$, *and* $f(x, g(x)) = 0$ *for all* $x \in I$.

Example 8. Consider the following function

$$e^{xy} + x + \sin(y) = 1.$$

The function defines a planar curve which is shown below in Fig.8.

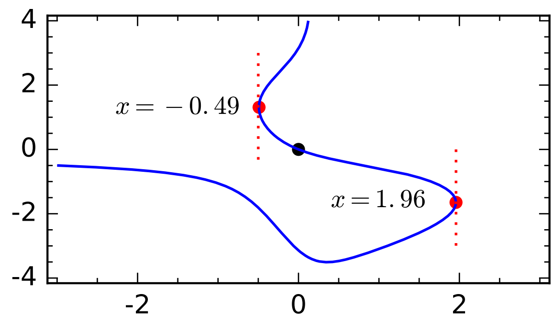


Figure 8.

The slopes at $x = -0.49$ and $x = 1.96$ are infinity, which implies $\partial_y f = 0$ at those points according to the formula $y' = -\frac{\partial_x f}{\partial_y f}$. Now fix the point $p_0 = (0, 0)$ on the curve. As it is seen from the figure, there is an explicit function $y = g(x)$ for $x \in (-0.49, 1.96)$ such that

$$e^{xg(x)} + x + \sin(g(x)) = 1. \tag{7}$$

The result can be generalized for functions $f: \mathbb{R}^n \to \mathbb{R}$ as follows.

Theorem 8. *Assume that implicit function* $f(x_1, x_2, \ldots, x_n) = 0$ *satisfies the following conditions*

i. there is a point $a = (a_1, a_2, \ldots, a_n)$ *such that* $f(a_1, \cdots, a_n) = 0$,

ii. There is an open ball $B_\varepsilon(a)$ *such that* f *has continuous partial derivatives on it,*

iii. and that $\partial_n f(a) \neq 0$,

then there exists a ball B_δ *at* $a' = (a_1, \cdots, a_{n-1})$ *and a function* g: $B_\delta(a') \to \mathbb{R}$ *such that* $a_n = g(a_1, \cdots, a_{n-1})$ *and* $f(x_1, x_2, \ldots, g(x_1, \ldots, x_{n-1})) = 0$ *for all* $(x_1, \ldots, x_{n-1}) \in B_\delta(a')$.

4 Integrals of mutivariable functions

4.1 Line integrals

Let $\gamma\colon (a,b)\to\mathbb{R}^n$ be a smooth curve map, that is $\gamma'(t)\neq 0$. The length of $\gamma(a,b)$ is *defined* by the following integral

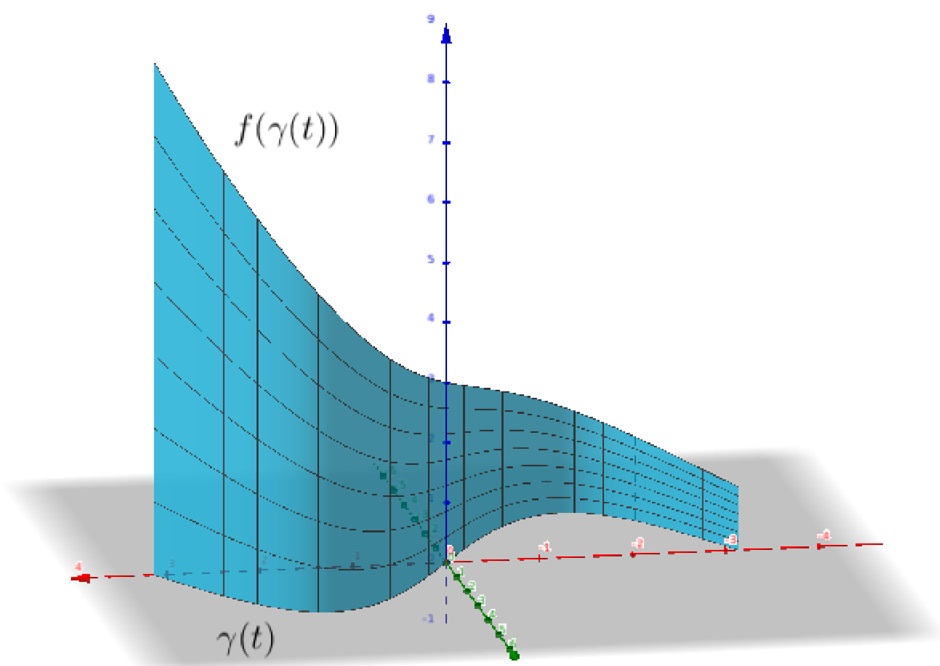
$$L=\int_a^b |\gamma'(t)|\,dt.$$

This definition coincides the intuitive notion of the length. For example, if the image of $\gamma(t)$ is a curved metal wire, and if we straightened the wired, we get the same length if we calculate the length by the above integration. In the above definition, $\gamma'(t)$ is the tangent vector on γ at t. The quantity $dl=|\gamma'(t)|\,\mathrm{d}t$ is called the differential arc length.

Now let $D\subset\mathbb{R}^n$ be an open set and let γ be a smooth curve in D. Assume that $f\colon D\to\mathbb{R}$ is a continuous function. The integral of f along γ is defined by the following integral

$$I=\int_a^b f(\gamma(t))dl=\int_a^b f(\gamma(t))\,|\gamma'(t)|\,dt.$$

Geometrically this integral is equal to the surface area constructed on the base $\gamma(t)$ an the height $f(\gamma(t))$. In particular if $f=1$, the above integral gives the arc length of γ. If f denote the density function of a metal wire represented by γ, the integral denote the total mass of the wire.



Problem 59. Consider the metal wire in the shape of the semi-circle $x^2+y^2=1$, $y\geq 0$. If the density of the wire is given by $\rho=k(1-y)$ for a constant k, find the center of mass of the wire.

4.2 Integrals over a bounded domain of $\mathbb{R}^2$

Now, let $R\colon[a,b]\times[c,d]$ be a rectangle and $f\colon R\subset\mathbb{R}^2\to\mathbb{R}$ be a continuous function. Similar to the integrals of single variable function, we can define the following double integral

$$I=\int\int_R f(x,y)dA,$$

by the aid of an infinite sum called the RIEMANN sum as

$$I=\lim_{n,m\to\infty}\sum_{j=1}^{m}\sum_{i=1}^{n} f(x_i,y_j)\,|\Delta_{ij}|.$$

Here Δ_{ij} is a partition of R, $|\Delta_{ij}|$ is the area of the rectangle ij and (x_i,y_j) is an arbitrary point in Δ_{ij}. We have the following theorem. Geometrically, I denote the volume constructed on the base R, and height f.

Theorem 9. (FUBINI) *Assume that the f is continuous (or piecewise continuous) in $R=[a,b]\times[c,d]$. Then we have*

$$\int_R f(x,y)\,dA=\int_a^b\left(\int_c^d f(x,y)\,\mathrm{d}y\right)dx=\int_c^d\left(\int_a^b f(x,y)\,\mathrm{d}x\right)dy.$$

Now, assume that $D\subset\mathbb{R}^2$ is a closed bounded domain and assume that $f\colon D\to\mathbb{R}$ is continuous. We can inscribe D inside a rectangle R and extend f on R as follows

$$\tilde{f}(x,y)=\begin{cases} f(x,y) & (x,y)\in D\\ 0 & (x,y)\notin D\end{cases}.$$

Even though $\tilde{f}$ may be discontinuous, we have the following fact

$$\int\int_D f(x,y)\,dA=\int\int_R \tilde{f}(x,y)\,dA.$$

Problem 60. Let $I=[a,b]$ and assume that $f(s,x)$ and $\frac{\partial f}{\partial x}(s,x)$ are continuous in $I\times I$. Use the fundamental theorem of calculus and prove the following formula

$$\frac{d}{dx}\int_{x_0}^{x} f(s,x)\,ds=f(x,x)+\int_{x_0}^{x}\frac{\partial}{\partial x}f(s,x)\,ds.$$

In the proof you may need to pass the limit inside the integral.

4.3 Change of variables in multiple integrals

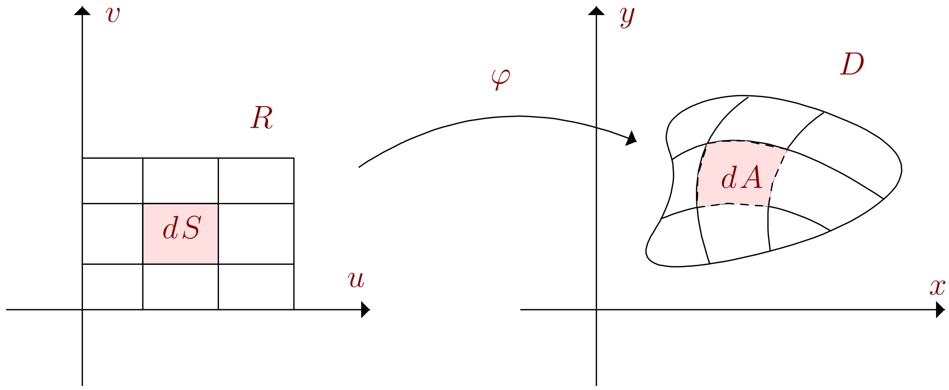
Of most important techniques to calculate a double integral (and also triple integrals) is the *change of variable* technique. Let $D\subset\mathbb{R}^2$ be a bounded domain, and assume $f\colon D\to\mathbb{R}$ is a continuous function. The goal is to calculate the integral

$$I=\int\int_D f(x,y)\,dA.$$

Now assume that there is a one to one mapping $\varphi: R\subset\mathbb{R}^2\rightarrow D$, where R is rectangle $[a, b]\times[c,d]$. If so, then we can calculate I in terms of an integral over R. The advantage is that the integration over rectangular domains are much more simpler than general domains. The procedure is as follows. Let $\varphi(u,v)=(x,y)$, where $(u,v)\in R$. See the following figure. The differential area dA in (x,y)-plane in terms of the differential area dS in the (u,v)-plane is

$$dA=|\det(J_\varphi)|\,dS,$$

where J_φ is the Jacobi matrix of the one to one transformation φ.



Accordingly, we have the following formula called the change of variable technique

$$\int\int_D f(x,y)\,dA=\int\int_R f(\varphi(u,v))\,|J_\varphi(u,v)|\,du\,dv.$$

Problem 61. Find the domain D formed by the transforming rectangle $[1,\ 2]\times[1,\ 2]$ under the transformation

$$\begin{cases} x=u/v \\ y=uv \end{cases}.$$

Calculate the following integral

$$\int\int_D e^{\sqrt{y/x}}\,e^{\sqrt{xy}}\,\mathrm{d}A.$$

4.4 Surface integrals over a surface in $\mathbb{R}^3$

Let S be a smooth surface in $\mathbb{R}^3$ with the representation

$$\varphi(u,v)=(x(u,v),y(u,v),z(u,v)),$$

where $(u,v)\in D\subset\mathbb{R}^2$, and assume that f is a continuous functions defined on S: $f: S\rightarrow\mathbb{R}$. We want to calculate the following integral

$$I=\int\int_S f(x,y,z)\,\mathrm{d}A,$$

where dA is the differential area of the surface S. Here we again transform the integral over S as an integral over D as follows. Note that

$$dA=||\partial_u\varphi\times\partial_v\varphi||\,dS,$$

where dS is a differential area in D. Remember that $\partial_u\varphi$ and $\partial_v\varphi$ are tangent vectors on S and $\|\partial_u\varphi \times \partial_v\varphi\|$ is equal to the area of parallelogram constructed on vectors $\partial_u\varphi, \partial_v\varphi$. Therefore, we can write

$$\int\int_S f(x,y,z)\mathrm{d}A = \int\int_D f(\varphi(u,v))\,||\partial_u\varphi \times \partial_v\varphi||dS.$$

Problem 62. If $f = f(x,y)$ is a smooth function defined on the bounded set $D \subset \mathbb{R}^2$, show that the area of the surface associated to u is

$$A = \int_D \sqrt{1 + |\partial_x f|^2 + |\partial_y f|^2}\,dx\,dy.$$

Problem 63. We show that the change of variable formula is independent of the transformation.

a) Assume $\varphi: D \to S$ and $\psi: D_1 \to S$ are two one to one transformations with image S. Define the map $\tilde{\varphi} := \varphi^{-1} \circ \psi: D_1 \to D$. Verify

$$||\partial_t\psi \times \partial_s\psi|| = ||\partial_u\varphi \times \partial_v\varphi||\,||\partial_t\tilde{\varphi} \times \partial_s\tilde{\varphi}||.$$

b) Now show

$$\int_D f(\varphi(u,v))\,||\partial_u\varphi \times \partial_v\varphi||du\,dv = \int_{D_1} f(\psi(s,t))|\partial_t\psi \times \partial_s\psi|\,ds\,dt.$$

We frequently use the polar and spherical coordinates for integrals. In polar coordinate, the transformation is defined by the relations $\varphi(r,\theta) = (r\cos\theta, r\sin\theta)$. The area differential dS in this case is

$$dS = \left|\begin{array}{cc} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{array}\right| dr\,d\theta = r\,dr\,d\theta.$$

If $f(x,y)$ is a function defined in disk B_a, a disk of radius a centered at the origin, its double integral in the polar coordinate is

$$\int_{B_a} f(x,y)dA = \int_0^{2\pi}\int_0^a f(r\cos\theta, r\sin\theta)\,r dr\,d\theta.$$

Problem 64. Show the following inequality

$$\frac{\pi}{4}(1 - e^{-a^2}) \le \int_0^a\int_0^a e^{-x^2-y^2}\mathrm{d}dx\,dy \le \frac{\pi}{4}(1 - e^{-2a^2}),$$

and conclude

$$\lim_{a\to\infty}\int_0^a\int_0^a e^{-x^2-y^2}\mathrm{d}A = \frac{\pi}{4}.$$

Use the above result and find

$$I = \int_0^\infty e^{-x^2}\mathrm{d}x.$$

In the spherical coordinate, the transformation is

$$\varphi(\rho,\phi,\theta) = (\rho\cos\phi\sin\theta, \rho\sin\phi\sin\theta, \rho\cos\theta).$$

The volume differential in this coordinate is

$$dV = \left|\begin{array}{ccc} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \rho\cos\theta\cos\phi & \rho\cos\theta\sin\phi & -\rho\sin\theta \\ -\rho\sin\theta\sin\phi & \rho\sin\theta\cos\phi & 0 \end{array}\right| d\rho d\phi d\theta = \rho^2\sin\theta\,d\rho\,d\phi\,d\theta.$$

If $f(x,y,z)$ is defined, for example, in a sphere of radius a, its integral on this sphere is equal to

$$\int_{S_a} f(x,y,z)dS = a^2\int_0^{2\pi}\int_0^{\pi} f(a\sin\theta\cos\phi, a\sin\theta\sin\phi, a\cos\theta)\sin\theta\, d\theta\, d\phi.$$

5 Calculus of vector fields

5.1 Vector field

Let $D\subset\mathbb{R}^n$ be an open set. A vector field is a mapping $f\colon D\to\mathbb{R}^n$. For each $p\in D$, $f(p)$ is a vector in $\mathbb{R}^n$, and thus we can interpret this mapping as an assignment a vector $f(p)$ to every point $p\in D$, that is, $p\mapsto f(p)$. This assignment is called a *vector field*.

A vector field $p\mapsto f(p)$ is continuous if the association varies continuously with respect to p, that is, the any change from p to an adjacent point q, the vector $f(p)$ continuously varies to $f(q)$. Mathematically speaking, this is equivalent to the continuity of f as a mapping on D. Remember that $f=(f_1, \ldots, f_n)$ is a continuous mapping if and only if each scalar function f_k is a continuous function. Similarly, a vector field $p\mapsto f(p)$ is called continuously differentiable if and only if all its coordinate functions are continuously differentiable.

A vector fields models many physical phenomena. For example, an electrical charge q located at the origin, generates an electrical field in the space as

$$E(r)=\frac{q}{4\pi\varepsilon_0}\frac{\vec{r}}{\|\vec{r}\|^3},$$

where ε_0 is the permittivity constant of the space. This field is a *force field* an is completely similar to the *gravitational field* generated by a mass M:

$$g(r)=GM\frac{\vec{r}}{\|\vec{r}\|^3},$$

where G is the universal constant.

5.2 Vector fields and differential equations

The theory of ordinary differential equations can be formulated in terms of vector fields. In fact, the system

$$\begin{cases}\frac{dx_1}{dt}=f_1(x_1,\ldots,x_n)\\ \vdots\\ \frac{dx_n}{dt}=f_n(x_1,\ldots,x_n)\end{cases}.$$

defines a vector field $f=(f_1, \ldots, f_n)$ on some domain D, and a parametric curve $\gamma(t)=(x_1(t), \ldots, x_n(t))$ such that the tangent vector on γ at each instant of time t_0 that is $\gamma'(t_0)$ coincides with the vector assigned to point $\gamma(t_0)$. In other word, $f(\gamma(t))=\gamma'(t)$ for all t in an open interval. For example, the following system

$$\begin{cases}x'=-y\\ y'=x\end{cases},$$

defines vector field $f=(-y,x)$. As we know, the trajectory of a point p_0: (x_0, y_0) according to the above system is the line

$$\gamma(t)=(x_0\cos t - y_0 \sin t, x_0 \sin t + y_0 \cos t).$$

It is simply seen that

$$\gamma'(t)=(-x_0\sin t - y_0\cos t, x_0\cos t - y_0\sin t),$$

that is equal to $f(\gamma(t))$. Notice that $\gamma(t)$ is just the rotation mapping applied to p_o

$$\gamma(t)=\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}\begin{pmatrix} x_0 \\ y_0 \end{pmatrix},$$

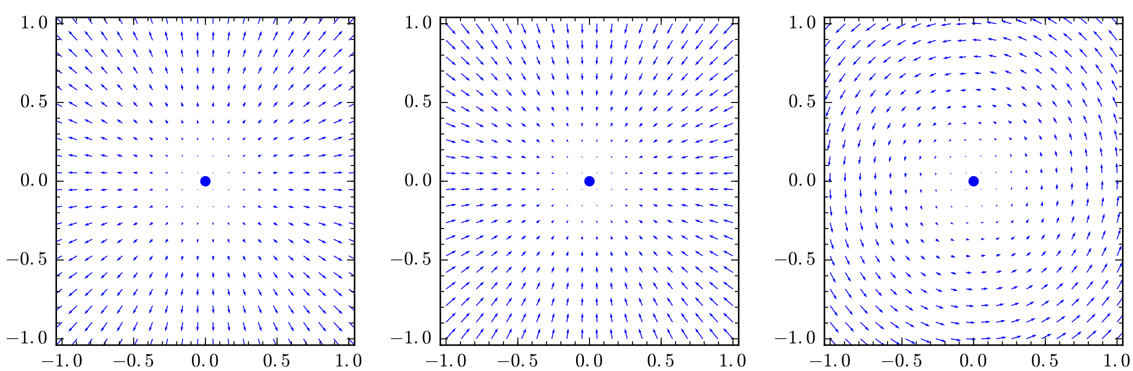
that coincides with the geometry of vector field f_3.

5.3 Gradient field

A vector fiend $p\mapsto f(p)$ is called a *potential, conservative* or just *gradient field* if there is a scalar function ϕ such that $f=-\nabla\phi$. The negative sign is just for historic reason. For example, the electric field $E(r)$ or the gravitational field $g(r)$ are potential. It is simply seen that $E(r)=-\nabla\phi$, where ϕ is the following scalar function:

$$\phi(r)=\frac{q}{4\pi\varepsilon_0}\frac{1}{\|\vec{r}\|}.$$

Potential fields satisfies very nice properties that we study in sequel. In the following figure, three vector fields are shown: $f_1=(x,y)$, $f_2=(-x,-y)$, and $f_3=(-y,x)$.



It is seen that f_1, f_2 are potential field with potentials $\phi_1=-\frac{1}{2}x^2-\frac{1}{2}y^2$, $\phi_2=\frac{1}{2}x^2\frac{1}{2}y^2$. The filed f_3 is not potential, i.e., there is no potential function ϕ such that $f_3=\nabla\phi$.

Problem 65. Prove the above claim, that is, show there is no scalar function ϕ such that $f_3=\nabla\phi$.

The force field generated by a potential is also called conservative. To see the reason, let us write the second NEWTON's law for a unit mass as

$$\begin{cases} \frac{d^2x}{dt^2}=p(x,y) \\ \frac{d^2y}{dt^2}=q(x,y) \end{cases},$$

where the field $f=(p,q)$ is potential generated by a potential function $\phi(x,y)$. Let $\gamma(t)$ be the trajectory of a particle initially located at (x_0,y_0). Then we have

$$\gamma''(t)=f(\gamma(t))=-\nabla\phi(\gamma(t)).$$

Let us define the energy along $\gamma(t)$ as follows

$$E(t)=\frac{1}{2}\|\gamma'(t)\|^2+\phi(\gamma(t)).$$

The derivative of E along $\gamma(t)$ is then

$$\frac{dE}{dt}=\gamma''(t)\cdot\gamma'(t)+\frac{d}{dt}\phi(\gamma(t)).$$

We have

$$\frac{d}{dt}\phi(\gamma(t))=\nabla\phi(\gamma(t))\cdot\gamma'(t),$$

and thus

$$\frac{dE}{dt}=\gamma''(t)\cdot\gamma'(t)+\nabla\phi\cdot\gamma'(t)=(\gamma''+\nabla\phi)\cdot\gamma'=0.$$

Therefore, the derivative of the energy function along a trajectory $\gamma(t)$ is zero, in other world, the energy is conserved along the trajectory of the particle.

5.4 Divergence, curl and Laplacian

Two important operations on smooth vector fields are divergence and curl. The divergence of a field $f=(f_1,\ldots,f_n)$ at a point p is defined by the following relation

$$\operatorname{div} f(p)=\partial_1 f_1(p)+\cdots+\partial_n f_n(p)$$

It is simply seen that the divergence of a vector field f is equal to the dot product of nabla ∇ and field f as $\operatorname{div} f=\nabla.f$. It is also equal to the trace of matrix D_pf. Intuitively speaking, $\operatorname{div} f(p)$ measure the flow of net *flux* passing through p. If $\operatorname{div} f(p)>0$, point p acts like a *source* that emits or generate flow. All points in filed $f_1=(x,y)$ are source points since $\operatorname{div} f(p)=2$. If $\operatorname{div} f(p)<0$, point p acts like a *sink* that absorbs or attracts flow to itself. All points in field $f_2=(-x,-y)$ are sink since $\operatorname{div} f(p)=-2$. If $\operatorname{div} f(p)=0$, then the net flow passing through p is zero, that the net amount of incoming flow is equal to outgoing flow. All points in field $f_3=(-y,x)$ are of this type.

Problem 66. If ϕ is a smooth scalar function and let f be a smooth vector field, show the following formula

$$\operatorname{div}(\phi f)=\phi\operatorname{div} f+f\cdot\nabla\phi \tag{8}$$

Another important operation regarding a vector field in $\mathbb{R}^3$ is the *curl* of the field at a point. The curl of a vector field $f=(f_1,f_2,f_3)$ at p is defined by the following relation

$$\operatorname{cur}(f)(p)=\begin{pmatrix}\hat{i} & \hat{j} & \hat{k}\\ \partial_x & \partial_y & \partial_z\\ f_1 & f_2 & f_3\end{pmatrix}(p)=(\partial_2 f_3-\partial_3 f_2,\partial_3 f_1-\partial_1 f_3,\partial_1 f_2-\partial_2 f_1)(p).$$

Symbolically, we can write the curl in terms of ∇ as $\nabla \times f$. Since $\operatorname{curl}(f)$ is a vector at every point, it also defines a new vector field in its domain. At each point p, $\operatorname{curl}(f)(p)$ measures the *rotation* of vector field f in three directions: 1) component $\partial_2 f_3(p) - \partial_3 f_2(p)$ that measures the rotation of f at p around x-axis, 2) component $\partial_3 f_1(p) - \partial_1 f_3(p)$ that measures the rotation of f at p around y-axis, 3) and $\partial_1 f_2(p) - \partial_2 f_1(p)$ that measure the rotation around z-axis . For example, $\operatorname{curl}(-y, x) = 2\,\hat{k}$ at all points, that means the filed rotates around z-axis with constant speed at all points. This rotation is evident from the figure of the field.

Assume that ϕ is a smooth scalar function defined on an open subset D of $\mathbb{R}^n$. The Laplacian of ϕ is defined by the following relation

$$\Delta\phi = \partial_{11}\phi + \cdots + + \partial_{nn}\phi.$$

It is simply seen that $\Delta\phi = \operatorname{div}(\operatorname{grad}\phi)$.

Problem 67. Consider field $f = (x^2 - y^2, y^2 - z^2, z^2 - x^2)$.

a) Find $\nabla \times f$ at the point $(1, 2, 3)$.

b) Verify that $(\nabla \times f).\,\hat{\imath}$ is equal to $\nabla \times (0, y^2 - z^2, z^2 - 1)$.

Problem 68. Assume that ϕ is a smooth scalar function defined in $\mathbb{R}^3$. Show $\nabla \times \nabla\phi = 0$.

Problem 69. Show the following relations for smooth fields f, g in $\mathbb{R}^3$ and smooth function ϕ

a) $\nabla \times (\phi f) = \phi \nabla \times f - f \times (\nabla\phi)$.

b) $\nabla.(f \times g) = g.(\nabla \times f) - f(\nabla \times g)$.

Problem 70. If f is a smooth vector field, what is $\nabla.\,(\nabla \times f)$?

Problem 71. Show the following relation

$$\nabla \times (\nabla \times f) = \nabla(\nabla.f) - \Delta f,$$

where $\Delta f = (\Delta f_1, \Delta f_2, \Delta f_3)$ for a smooth vector field $f = (f_1, f_2, f_3)$.

5.5 Line integrals in vector fields

Assume that the $f\colon D \subset \mathbb{R}^n \to \mathbb{R}^n$ is a continuous field, and C is a smooth curve in D. We define a method for the integral of f *along* curve C. For this, we take a parametrization of C as $\gamma\colon (a, b) \to D$, where $\gamma(a, b) = C$. The desired integral is defined as follows

$$\int_C f\,dc = \int_a^b f(\gamma(t)) \cdot \gamma'(t)\,dt. \tag{9}$$

It is seen that the integral is independent of parametrization, that is, if $\gamma_1\colon (c, d) \to D$ is another parametrization of C, then

$$I = \int_c^d f(\gamma_1(t)) \cdot \gamma_1'(t)\,dt.$$

Problem 72. Assume that $\gamma\colon (a, b) \to D$, $\gamma_1\colon (c, d) \to D$ are two smooth parametrizations of C. Show the relation

$$\int_a^b f(\gamma(t)) \cdot \gamma'(t) dt = \int_c^d f(\gamma_1(t)) \cdot \gamma_1'(t)\,dt.$$

If f is a force field, then the integral of f along a curve is called the *work* done by f. For example, let $\gamma(a) = p_0$, $\gamma(b) = p_1$, then I measure the total work done by f to move a particle from point p_0 to point p_1 *along* path γ. It is is interesting to note that if f is gradient, i.e., $f = \nabla\phi$ for some potential ϕ, then the integral is independent of path along a particle moves from p_0 to p_1. This important fact is shown below

$$\int_a^b \nabla\phi(\gamma(t)) \cdot \gamma'(t)\, dt = \int_a^b d\phi(\gamma(t)) = \phi(\gamma(b)) - \phi(\gamma(a)) = \phi(p_1) - \phi(p_0).$$

The path independence property of a field implies that the integral of field over any closed curve is zero, that is, if γ is a closed curve then

$$\oint_\gamma f(\gamma(t)) \cdot \gamma'(t)\, dt = 0. \tag{10}$$

We have the following theorem.

Theorem 10. *Assume $f: D \subset \mathbb{R}^n \to \mathbb{R}^n$ is a smooth vector field. If the integral of f over all closed curves in D is zero, then f is a conservative field on D.*

Problem 73. Let the condition of the above theorem holds. Fix $p_0 \in D$. For any $p \in D$ define ϕ as

$$\phi(p) = \int_0^t f(\gamma(t)) \cdot \gamma'(t)\, dt,$$

where $\gamma(t)$ is an arbitrary smooth curve in D such that $\gamma(0) = p_0$, and $\gamma(t) = p$.

a) Verify that ϕ is independent of path γ

b) Show that ϕ is a potential for f. For example, for two dimensional field $f = (p(x, y), q(x, y))$, verify the relation

$$\lim_{h\to 0} \frac{\phi(x+h, y) - \phi(x, y)}{h} = p(x, y), \lim_{h\to 0} \frac{\phi(x, y+h) - \phi(x, y)}{h} = q(x, y).$$

Problem 74. A smooth field $f = (p, q)$ in $\mathbb{R}^2$ is called *exact* if for all $(x, y) \in \mathbb{R}^2$ the following relation holds

$$\partial_y f(x, y) = \partial_x q(x, y).$$

a) Show that f is conservative.

b) Consider field $f = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$ defined everywhere in $\mathbb{R}^2$ except the origin. Show the following relation

$$\partial_y\left(\frac{-y}{x^2+y^2}\right) = \partial_x\left(\frac{x}{x^2+y^2}\right).$$

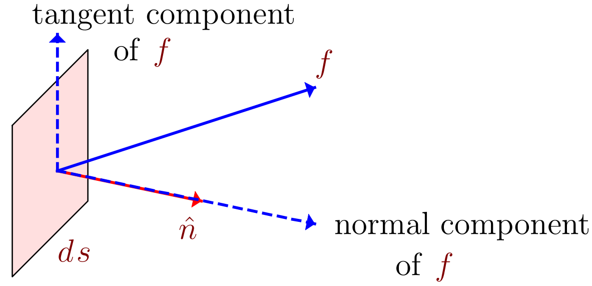
Now, consider closed curve $\gamma(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$, calculate the line integral of f along γ. The result is nonzero. Does it contradict the fact claimed in (a)?

5.6 Surface integrals in vector fields

Let S be a surface in $\mathbb{R}^3$ parameterized by a smooth map $\Gamma(t, s)$ where $(t, s) \in D$ for some open domain D. Note that Γ is smooth if $\partial_t\Gamma \times \partial_s\Gamma$ is nonzero for all points in D. Let smooth field $f = (f_1, f_2, f_3)$ given in $\mathbb{R}^3$. The integral of f on the surface S is defined by the following integral

$$I = \iint_S f \cdot \hat{n}\, dA,$$

where $\hat{n}$ is the unit normal vector on S. This integral measures the total flux passing outward through surface S. As it is shown in the following figure, the tangential component of f never leaves the surface (due to the fact $f\cdot\hat{n}=0$), and only the normal component of f contributes in the total flux. For example, the flow of water through a window measured in $\frac{m^3}{\sec}$ is a physical model for the flux.



Remark 5. In some branches of mathematics, the term *flux* is considered as a vector and not scalar. We will see this notion in the book for the mathematical modeling of heat flow through a conductive media.

For the parametrization $\Gamma(t,s)$, we have

$$\hat{n}\,dA=(\partial_t\Gamma\times\partial_s\Gamma)\,dt\,ds,$$

and therefore, the total flux is expressed in terms of double integral

$$I=\iint_D f(\Gamma(t,s))\cdot(\partial_t\Gamma\times\partial_s\Gamma)\,dt\,ds.$$

Theorem 11. (divergence) *Assume $D\subset\mathbb{R}^3$ is an open set with smooth (piecewise) boundary. If f is a smooth vector field in $\mathrm{cl}(D)$, the following relation holds*

$$\iiint_D \operatorname{div} f\,dv=\oiint_{\mathrm{bnd}(D)} f\cdot\hat{n}\,dA. \tag{11}$$

Note that If $\operatorname{div} f=0$ inside D, then the net amount of flow passing through $\mathrm{bnd}(D)$ is zero. Accordingly, the divergence of a field f in $\mathbb{R}^3$ at a point p can be defined by the following formula

$$\operatorname{div}(f)(p)=\lim_{r\to 0}\frac{1}{\mathrm{Vol}(B_r(p))}\oiint_{\mathrm{bnd}(B_r(p))} f\cdot\hat{n}\,dA. \tag{12}$$

The above formula coincides with our previous statement about the physical interpretation of divergence operator at a point, that $\operatorname{div}(f)(p)$ measures the net flux of f passing through point p.

Problem 75. Let $f=(x,y,z)$. Verify the formula (12) at the origin.

Example 9. Consider the identity field $f(x,y,z)=(x,y,z)$, and let B be the closed unit ball in $\mathbb{R}^3$ centered at the origin. The left hand side of formula (11) reads

$$\iiint_B \operatorname{div} f dV=3\iiint_B dV=4\pi. \tag{13}$$

The unit normal to $\mathrm{bnd}(B)$ is $\hat{n} = (x, y, z)$, and then the right hand side of formula (11) reads

$$\oiint_{\mathrm{bnd}(B)} (x^2 + y^2 + z^2)\, dA = \oiint_{\mathrm{bnd}(B)} dA = 4\pi. \tag{14}$$

Problem 76. Assume $f: D \subset \mathbb{R}^3 \to \mathbb{R}^3$ a smooth field. Use the divergence theorem and show

$$\iiint_D \Delta f\, dV = \oiint_{\mathrm{bnd}(D)} \partial_{\hat{n}} f\, ds, \tag{15}$$

where $\partial_{\hat{n}} f$ is the directional derivative of f in direction $\hat{n}$.

Problem 77. If D is a ball of radius R in $\mathbb{R}^n$, use the divergence theorem and show

$$\mathrm{Vol}(D) = \frac{R}{n} A(D),$$

where $\mathrm{Vol}(D)$ is the volume of D and $A(D)$ is the surface area of D. (Hint: consider the function $\phi = \sum_k x_k^2$).

Problem 78. Let f be a smooth field in $\mathbb{R}^n$ such that

$$|f(r)| \leq \frac{1}{(1 + ||r||)^{n+1}}.$$

Show that

$$\int_{\mathbb{R}^n} \mathrm{div}\,(f) = 0.$$

Proposition 10. (Integration by parts) *Let $D \subset \mathbb{R}^3$ be an open set and $f, g: D \to \mathbb{R}$ are smooth functions. We have*

$$\iiint_D (\partial_x f)\, g\, dV = \oiint_{\mathrm{bnd}(D)} f g n_1\, dA - \int_D f\, (\partial_x g)\, dV, \tag{16}$$

where n_1 is the first component of $\hat{n} = (n_1, n_2, n_3\rangle$ at $\mathrm{bnd}(D)$. Similar relations hold for the derivatives with respect to other components y, z.

The following proposition generalizes the above result.

Proposition 11. *Let D be a domain in $\mathbb{R}^3$ and $\mathrm{bnd}(D)$ is smooth. If f is a smooth vector field in D and g is a smooth function on D then*

$$\iiint_D g\, \mathrm{div}\,(f)\, dV = \oiint_{\mathrm{bnd}(D)} g\, (f \cdot \hat{n})\, dA - \iiint_D f \cdot \nabla g\, dV. \tag{17}$$

Problem 79. Prove the above proposition by the aid of divergence theorem.

Problem 80. By the above proposition show

$$\iiint_D g\, \Delta f\, dV = \oiint_{\mathrm{bnd}(D)} g\, \partial_{\hat{n}} f\, dA - \iiint_D \nabla f \cdot \nabla g\, dV,$$

and conclude the following relation called the Green's formula:

$$\iiint_D (g \Delta f - f \Delta g)\, dV = \oiint_{\mathrm{bnd}(D)} [g\, \partial_{\hat{n}} f - f \partial_{\hat{n}}\, g]\, dA.$$

Let S be a smooth surface in $\mathbb{R}^3$ with smooth boundary $\mathrm{bnd}(S)$. If f is a smooth field in $\mathbb{R}^3$, the following relation is called the Stoke's theorem:

$$\iint_S (\nabla \times f) \cdot \hat{n} \, dA = \oint_{\mathrm{bnd}(S)} f \cdot \hat{T} dl,$$

where $\hat{T}$ is the unit tangent vector on curve $\mathrm{bnd}(S)$ and $d\ell$ is the differential length of that curve.

Problem 81. Use the Stoke's theorem and show

$$(\nabla \times f(p)) \cdot \hat{i} = \lim_{r \to 0} \frac{1}{\pi r} \int_0^{2\pi} f(\gamma(t)) \cdot \gamma'(t) \, dt,$$

where $\gamma(t) = p + (r \cos t, r \sin t, 0)$. Derive similar formula for the second and third components, i.e., $(\nabla \times f(p)) \cdot \hat{j}$, and $(\nabla \times f(p)) \cdot \hat{k}$. Notice that $f \cdot \hat{T}$ measures the rotation of f along curve $\mathrm{bnd}(S)$.

Problem 82. Let $f = (-y + x, -x + z, -z + y)$. Verify the relations of the previous problem at the origin.

Problem 83. Show the following relations

a)

$$\iint_S \phi \, \mathrm{curl}(f) \cdot \hat{n} \, dA = \iint_S (f \times \nabla \phi) \cdot \hat{n} \, dA + \oint_{\mathrm{bnd}(S)} \phi f \cdot \hat{T} dl.$$

b)

$$\iiint_D \phi \, \mathrm{div}\,(f) \, dV = \oiint_S \phi f \cdot \hat{n} \, dA - \iiint_D f \cdot \nabla \phi \, dV$$

c)

$$\iiint_D g \cdot \mathrm{curl}(f) \, dV = \oiint_S (f \times g) \cdot \hat{n} \, dA + \iiint_D f \cdot \mathrm{curl}(g) \, dV$$

We have the following theorem

Theorem 12. *If $\nabla \times f = 0$ everywhere, then f is gradient, i.e., there is a potential function ϕ such that $f = -\nabla \phi$. If $\mathrm{div}\,(f) = 0$ everywhere, then there is a vector field g such that $f = \mathrm{curl}(g)$.*

6 Orthogonal curvilinear coordinates

In many applications, it is convenient to use an orthogonal coordinate rather than the Cartesian one. The form of differential operators in a general orthogonal curvilinear coordinate is discussed below.

6.1 Unit vectors

Let (q_1, q_2, q_3) be a curvilinear coordinate system, and assume $(q_1, q_2, q_3) \xrightarrow{T} (x, y, z)$ is a one to one transformation to the Cartesian space. We first define the unit vectors $\hat{q}_1$, $\hat{q}_2$ and $\hat{q}_3$ in the directions of q_1, q_2 and q_3, and by the aid of that define the orthogonality notion. If we consider the restriction of T to q_1 as a one parameter map $\gamma(q_1)$, then the unit vector $\hat{q}_1$ can defined as

$$\hat{q}_1 := \frac{\gamma'(q_1)}{\|\gamma'(q_1)\|} = \frac{\partial_{q_1} T}{\|\partial_{q_1} T\|},$$

and since

$$T(q_1, q_2, q_3) = (x(q_1, q_2, q_3), y(q_1, q_2, q_3), z(q_1, q_2, q_3)),$$

we have

$$\partial_{q_1} T = (\partial_{q_1} x, \partial_{q_1} y, \partial_{q_1} z)$$

Similarly, we can define $\hat{q}_2$, $\hat{q}_3$ as

$$\hat{q}_2 = \frac{\partial_{q_2} T}{\|\partial_{q_2} T\|}, \hat{q}_3 = \frac{\partial_{q_3} T}{\|\partial_{q_3} T\|}.$$

A coordinate system (q_1, q_2, q_3) is called *orthogonal* if $\hat{q}_1$, $\hat{q}_2$ and $\hat{q}_3$ are mutually orthogonal, i.e., $\langle \hat{q}_i, \hat{q}_j \rangle = \delta_{ij}$.

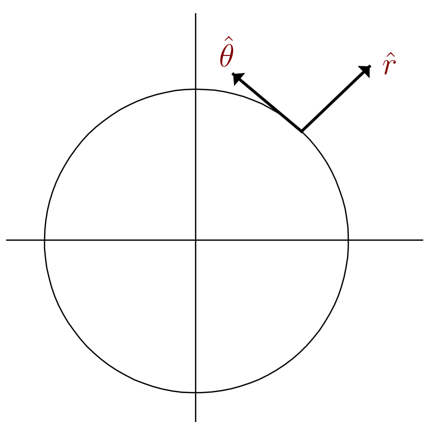
6.1.1 Polar, cylindrical and spherical coordinates

The polar coordinate (r, θ) is defined by the transformation $T(r, \theta) = (r\cos\theta, r\sin\theta)$ for $r = [0, \infty)$ and $\theta \in [0, 2\pi)$. The transformation is one to one everywhere in the domain except at the origin. Since $\partial_r T = \cos\theta\,\hat{\imath} + \sin\theta\,\hat{\jmath}$, we obtain

$$\hat{r} = \cos\theta\,\hat{\imath} + \sin\theta\,\hat{\jmath}.$$

Similarly, $\partial_\theta T = -r\sin\theta\,\hat{\imath} + r\cos\theta\,\hat{\jmath}$, and thus

$$\hat{\theta} = -\sin\theta\,\hat{\imath} + \cos\theta\,\hat{\jmath}.$$



The cylindrical transformation is defined by transformation

$$T(r, \theta, z) = (r\cos\theta . r\sin\theta, z)$$

The unit vectors are derived as

$$\hat{r} = \cos\theta\,\hat{\imath} + \sin\theta\,\hat{\jmath}, \hat{\theta} = -\sin\theta\,\hat{\imath} + \cos\theta\,\hat{\jmath}, \hat{z} = \hat{k}.$$

The spherical coordinate is defined by the transformations

$$T(\rho, \phi, \theta) = (\rho\cos\phi\sin\theta, \rho\sin\phi\sin\theta, \rho\cos\theta),$$

for $\rho \in [0, \infty), \phi \in [0, 2\pi), \theta = [0, \pi]$. The mapping is not one to one at $\rho = 0, \theta = 0, \pi$. The unit vectors are

$$\hat{\rho} = \sin\theta\cos\phi\,\hat{\imath} + \sin\theta\sin\phi\,\hat{\jmath} + \cos\theta\,\hat{k}, \tag{18}$$

$$\hat{\phi} = -\sin\phi\,\hat{\imath} + \cos\phi\,\hat{\jmath}, \tag{19}$$

$$\hat{\theta} = \cos\theta\cos\phi\,\hat{\imath} + \cos\theta\sin\phi\,\hat{\jmath} - \sin\theta\,\hat{k}. \tag{20}$$

6.2 Nabla ∇ in an orthogonal coordinate

Let f be a smooth scalar function given in an orthogonal coordinate system (q_1, q_2, q_3). We can write the gradient of f in this system as

$$\nabla f = f_1\,\hat{q}_1 + f_2\,\hat{q}_2 + f_3\,\hat{q}_3,$$

that implies $f_1 = \langle \nabla f, \hat{q}_1\rangle$, $f_2 = \langle \nabla f, \hat{q}_2\rangle$, and $f_3 = \langle \nabla f, \hat{q}_3\rangle$. Let us calculate f_1 for example. ∇f is coordinate free, and thus we can replace its form in the Cartesian coordinate, i.e., $\nabla f = \partial_x f\,\hat{i} + \partial_y f\,\hat{j} + \partial_z f\,\hat{k}$ in the associated dot product and derive

$$\langle \nabla f, \hat{q}_1\rangle = \frac{1}{\|\partial_{q_1} T\|}\langle \partial_x f\,\hat{i} + \partial_y f\,\hat{j} + \partial_z f\,\hat{k}, \partial_{q_1} x\,\hat{i} + \partial_{q_1} y\,\hat{j} + \partial_{q_1} z\,\hat{k}\rangle.$$

By the orthogonality, we obtain

$$\langle \nabla f, \hat{q}_1\rangle = \frac{1}{\|\partial_{q_1} T\|}(\partial_x f \partial_{q_1} x + \partial_y f \partial_{q_1} y + \partial_z f \partial_{q_1} z) = \frac{1}{\|\partial_{q_1} T\|}\,\partial_{q_1} f.$$

We obtain similar forms for f_2, f_3, that are,

$$\langle \nabla f, \hat{q}_2\rangle = \frac{1}{\|\partial_{q_2} T\|}\,\partial_{q_2} f, \quad \langle \nabla f, \hat{q}_3\rangle = \frac{1}{\|\partial_{q_3} T\|}\,\partial_{q_3} f$$

and therefore the operator ∇ in (q_1, q_2, q_3) is

$$\nabla = \frac{\hat{q}_1}{\|\partial_{q_1} T\|}\partial_{q_1} + \frac{\hat{q}_2}{\|\partial_{q_2} T\|}\partial_{q_2} + \frac{\hat{q}_3}{\|\partial_{q_3} T\|}\partial_{q_3}.$$

6.2.1 Polar, cylindrical and spherical coordinate

Applying the formula obtained above for polar, cylindrical and spherical systems gives respectively

$$\text{Polar:}\ \ \nabla = \hat{r}\,\partial_r + \frac{1}{r}\,\hat{\theta}\,\partial_\theta.$$

$$\text{Cylindrical:}\ \ \nabla = \hat{r}\,\partial_r + \frac{1}{r}\,\hat{\theta}\,\partial_\theta + \hat{k}\,\partial_z.$$

$$\text{Spherical:}\ \ \nabla = \hat{\rho}\,\partial_\rho + \frac{1}{\rho\sin\theta}\,\hat{\phi}\,\partial_\phi + \frac{1}{\rho}\,\hat{\theta}\,\partial_\theta.$$

To calculate the Laplacian operator $\Delta := \nabla.\nabla$ in (q_1, q_2, q_3) system, we do as follows:

$$\Delta := \nabla.\nabla = \left\langle \frac{\hat{q}_1}{\|\partial_{q_1} T\|}\partial_{q_1} + \frac{\hat{q}_2}{\|\partial_{q_2} T\|}\partial_{q_2} + \frac{\hat{q}_3}{\|\partial_{q_3} T\|}\partial_{q_3}, \frac{\hat{q}_1}{\|\partial_{q_1} T\|}\partial_{q_1} + \frac{\hat{q}_2}{\|\partial_{q_2} T\|}\partial_{q_2} + \frac{\hat{q}_3}{\|\partial_{q_3} T\|}\partial_{q_3} \right\rangle.$$

Here we need $\partial_i(\hat{q}_j)$ for $i, j = 1, 2, 3$. It turns out that the derivatives of $\hat{q}_k$ always lies in the normal plane to it, for example,

$$\partial_{q_2}(\hat{q}_1) = \alpha\,\hat{q}_2 + \beta\,\hat{q}_3.$$

The reason is for the relation $\langle \hat{q}_i, \hat{q}_i\rangle = 1$ and thus $\langle \partial_{q_j}(\hat{q}_i), \hat{q}_i\rangle = 0$.

Problem 84. In polar coordinate, show the following relations

$$\partial_\theta \hat{r} = \hat{\theta}, \quad \text{and} \quad \partial_\theta \hat{\theta} = -\hat{r}, \tag{21}$$

and conclude

$$\Delta f = \partial_{rr} f + \frac{1}{r}\partial_r f + \frac{1}{r^2}\partial_{\theta\theta} f. \tag{22}$$

Problem 85. In cylindrical coordinate show the relation

$$\Delta f = \partial_{rr} f + \frac{1}{r}\partial_r f + \frac{1}{r^2}\partial_{\theta\theta} f + \partial_{zz} f.$$

Problem 86. In spherical coordinate show the following relations

$$\partial_\phi \hat{\rho} = \sin(\theta)\,\hat{\phi} \quad \text{and} \quad \partial_\theta \hat{\rho} = \hat{\theta}, \tag{23}$$

$$\partial_\phi \hat{\phi} = -\sin(\theta)\,\hat{\rho} - \cos(\theta)\,\hat{\theta} \quad \text{and} \quad \partial_\theta \hat{\phi} = 0, \tag{24}$$

$$\partial_\phi \hat{\theta} = \cos(\theta)\,\hat{\phi} \quad \text{and} \quad \partial_\theta \hat{\theta} = -\hat{\rho}. \tag{25}$$

and conclude

$$\Delta f = \frac{1}{\rho^2}\partial_\rho(\rho^2\,\partial_\rho f) + \frac{1}{\rho^2\sin^2\theta}\,\partial_{\phi\phi} f + \frac{1}{\rho^2 \sin\theta}\partial_\theta(\sin\theta\,\partial_\theta f).$$

7 Function Series

Working in a function vector space where functions play the role of well-known vectors in $\mathbb{R}^n$, requires to study sequences whose elements are functions. These type of sequences are a natural generalization of numeric sequences that we suppose the reader is familiar with.

7.1 The different notions of convergence

We assume that the reader is familiar with numeric sequences. Here we consider sequences and series whose elements are functions. Let $f_n\colon [a,b] \to \mathbb{R}$ for $n = 1, 2, \ldots$ be a sequence of continuous functions. We say f_n converges *pointwise* in $[a,b]$ to f, and write $f_n \to f$ if

$$\lim_{n\to\infty} |f_n(x) - f(x)| = 0,$$

for all $a \in [a,b]$. In other word, if $a_n = f_n(x)$ for a fixed x, then (a_n) as a numeric sequence converges to value $b = f(x)$. This notion of convergence is equivalent to the following: for every $x \in [a,b]$ and every $\varepsilon > 0$, there is an integer $N_0 = N_0(x, \varepsilon) > 0$ such that

$$\forall n \geq N_0 \Rightarrow |f_n(x) - f(x)| < \varepsilon.$$

A function sequence f_n converges *uniformly* to f in $[a,b]$ if for for any $\varepsilon > 0$, there is and integer $N = N(\varepsilon)$ such that

$$\forall n \geq N_0 \Rightarrow \max_{x\in[a,b]} |f_n(x) - f(x)| < \varepsilon.$$

The norm of a continuous function in a closed interval $[a,b]$ is defined as

$$\|f\| = \max_{x\in[a,b]} |f(x)|.$$

For this reason, the uniform convergence is usually written as

$$\lim_{n\to\infty} \|f_n - f\| = 0.$$

Remark 6. The pointwise and uniform convergence are two different types of convergence and they may should not be considered as equivalent. In fact, if $f_n \to f$ uniformly, then $f_n \to f$ pointwise, while the converse is not generally true. For example, consider the sequence of functions $f_n(x) = x^n$ in $[0, 1]$. The sequence converges pointwise to $f(x) = \begin{cases} 1 & x=1 \\ 0 & \text{otherwise} \end{cases}$. However, it is seen

$$\max_{x\in[0,1]} |x^n - f(x)| > \frac{1}{2},$$

and thus f_n does not converges uniformly to f. In fact, f_n does not converge uniformly to any function.

Problem 87. If a function sequence (f_n) converges pointwise or uniformly, then its limit function is unique.

Problem 88. If a function sequence (f_n) converges uniformly to f in $[a, b]$ then it converges pointwise to f as well.

Problem 89. Assume that $f_n: [a, b] \to \mathbb{R}$ is a sequence of continuous functions converges uniformly to $f(x)$. Show that $f(x)$ is continuous.

There are other notions of convergence that we study in the book, for example, the convergence in norm as

$$\lim_{n\to\infty} \int_a^b |f_n(x) - f(x)| dx = 0.$$

For example, $f_n(x) = x^n$ in $[0, 1]$ converges to $f(x) = \begin{cases} 1 & x=1 \\ 0 & \text{otherwise} \end{cases}$ or $f(x) \equiv 0$ or $f(x) = \begin{cases} 1 & x=\frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$. As we see, the limit function is not unique in the usual sense, and we have to do something to remedy this situation.

Problem 90. If (f_n) converges uniformly to f in $[a, b]$, it converges to f in the following sense

$$\int_a^b |f_n(x) - f(x)| dx \to 0.$$

Problem 91. Assume that a function sequence (f_n) of continuous functions on $[a, b]$ converges in norm to f as

$$\int_a^b |f_n(x) - f(x)| dx \to 0$$

Show that

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx.$$

7.2 δ-sequence

A function sequence $f_n(x)$ is called a Dirac δ-sequence function at $x = 0$ if for any continuous bounded function $g(x)$ at $x = 0$, we have

$$\lim_{n\to\infty} \int g(x)\, f_n(x)\, dx = g(0).$$

For example, the sequence

$$f_n(x) = \begin{cases} \frac{n}{2} & -\frac{1}{n} < x < \frac{1}{n} \\ 0 & \text{otherwise} \end{cases},$$

is a δ-sequence. If $g(x)$ is any bounded continuous function at $x = 0$, then

$$\min_{x \in \left(-\frac{1}{n}, \frac{1}{n}\right)} g(x) \leq \int_{-\frac{1}{n}}^{\frac{1}{n}} g(x) \frac{n}{2} \, dx \leq \max_{x \in \left(-\frac{1}{n}, \frac{1}{n}\right)} g(x),$$

and thus

$$\lim_{n \to \infty} \int_{-\frac{1}{n}}^{\frac{1}{n}} g(x) \frac{n}{2} \, dx = g(0).$$

7.3 Differentiation and integration of function sequences

Let (f_n) be a function sequence in (a, b) or $[a, b]$. If $f_n \to f$ pointwise or uniformly, can we claim

$$\int_a^b f_n(x) dx \to \int_a^b f(x) dx$$

when $n \to \infty$? Let us see an important example:

$$f_n(x) = \begin{cases} 2^{2n} x & x \in \left[0, \frac{1}{2^n}\right] \\ 2^n (2 - 2^n x) & x \in \left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right] \\ 0 & \text{otherwise} \end{cases}.$$

The graph for some n are shown in Fig.9. Each function f_n is a triangle with the base $\left[0, \frac{1}{2^{n-1}}\right]$ and the height 2^n such that the area under each triangle is equal 1 independent of n.

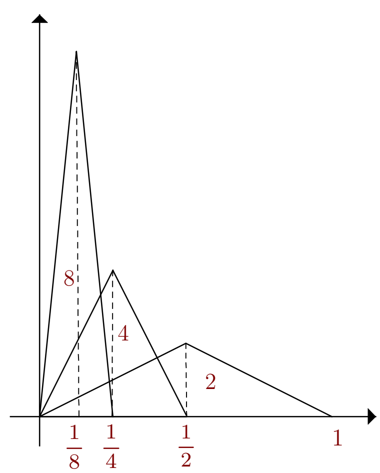


Figure 9.

Therefore, we have for all n:

$$\int_0^1 f_n(x) dx = 1.$$

On the other hand, it is seen f_n converges to $f\equiv 0$ on $[0,1]$ (why?) and thus

$$\int_0^1 f_n(x)dx \nrightarrow \int_0^1 f(x)dx.$$

In other word,

$$\lim_{n\to\infty}\int f_n(x)\,dx \neq \int \underbrace{\lim_{n\to\infty} f_n(x)}_{f(x)}\,dx.$$

Problem 92. Assume that (f_n), $f_n: I\to\mathbb{R}$ is a sequence of continuous functions converging uniformly to f. Show

$$\lim_{n\to\infty}\int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.$$

The following theorem gives a sufficient condition for passing the limit inside an integral.

Theorem 13. (dominant convergence) *Assume that (f_n) is a sequence of continuous functions converging pointwise to a function f in (a,b) (the interval may be finite or infinite). If there is a function $g(x)$ such that*

$$|f_n(x)|\leq g(x), \qquad \forall x\in(a,b),$$

and

$$\int_a^b g(x)\,dx<\infty,$$

then

$$\lim_{n\to\infty}\int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.$$

We also need the following form of the above theorem.

Corollary 2. *Assume that $f(t,x)$ is a continuous function defined on $(c,d)\times(a,b)$, and furthermore*

$$f(t,x)\leq g(x), \forall t\in(c,d)$$

such that

$$\int_a^b g(x)\,dx<\infty$$

Then function

$$F(t)=\int_a^b f(t,x)\,dx,$$

is continuous in (c,d).

Problem 93. Prove the corollary by the aid of Theorem (13).

The same story of differentiation is the same. If a sequence of differentiable functions (f_n) converges pointwise or even uniformly to f, then there is no guarantee that $f_n'\to f'$. As a simple example, sequence $f_n(x)=\frac{1}{n}\sin(nx)$ defined on $(0,\pi)$ converges uniformly to the constant function $f(x)=0$. However, $f_n'(x)=\cos(nx)$ which is not a convergent sequence.