Integration and change of variables

Introduction to dynamical systems #9

Hiroaki Sakamoto

§ Contents

1 Integrals: recap

1.1 Single dimensional case

- **Integral over intervals**
  - Consider a function $f : X \to \mathbb{R}$ defined over $X \subset \mathbb{R}$
  - For an interval $R := [a, b] \subset X$ and $n \in \mathbb{N}$, the number

$$\sum_{k=0}^{n-1} f(x_k)(x_{k+1} - x_k), \quad x_k := a + \frac{k}{n}(b - a)$$

    is called a (left) Riemann sum
  - We define the integral of f over R as the limit of this number for $n \to \infty$

$$\int_R f(x)dx := \lim_{n\to\infty} \sum_{k=0}^{n-1} f(x_k)(x_{k+1} - x_k)$$

  - We also write this integral as $\int_a^b f(x)dx$
- **Examples**
  - The integral of $f(x) := x$ over $R := [0, x]$ is

$$\int_0^x f(t)dt = \int_0^x tdt = \lim_{n\to\infty} \sum_{k=0}^{n-1} \left(k\frac{x}{n}\right)\frac{x}{n} = \lim_{n\to\infty} \frac{x^2}{n^2} \underbrace{\sum_{k=0}^{n-1} k}_{\frac{1}{2}n(n-1)} = \frac{1}{2}x^2$$

  - The integral of $f(x) := x^2$ over $R := [0, x]$ is

$$\int_0^x f(t)dt = \int_0^x t^2dt = \lim_{n\to\infty} \sum_{k=0}^{n-1} \left(k\frac{x}{n}\right)^2\frac{x}{n} = \lim_{n\to\infty} \frac{x^3}{n^3} \underbrace{\sum_{k=0}^{n-1} k^2}_{\frac{1}{3}n(n-1)\left(n-\frac{1}{2}\right)} = \frac{1}{3}x^3$$

- **Antiderivatives**
  - We say that a function $F : X \to \mathbb{R}$ is an *antiderivative* or *primitive function* of another function $f : X \to \mathbb{R}$ if $F'(x) = f(x)$ for all $x \in X$
  - If $F(x)$ is an antiderivative of $f(x)$, then the integral of $f(x)$ over $R = [a, b]$ is

$$\int_R f(x)dx = \lim_{n\to\infty} \sum_{k=0}^{n-1} \underbrace{f(x_k)}_{F'(x_k)}(x_{k+1} - x_k) = \lim_{n\to\infty} \sum_{k=0}^{n-1} (F(x_{k+1}) - F(x_k)) = F(b) - F(a)$$

- **Example**
  - Consider a function $f(x) := 1/x$
  - Since $F(x) := \ln(x)$ is an antiderivative of $f(x)$, the integral of $f(x)$ over $[1, x]$ is

$$\int_1^x \frac{1}{t}dt = \ln(x) - \ln(1) = \ln(x)$$

  - In fact, this can be seen as a definition of $\ln(x)$ when read from right to left
- **Measure of sets**
  - For any subset $D \subset \mathbb{R}$, we define

$$\mathbb{1}_D(x) := \begin{cases} 1 & \text{if } x \in D \\ 0 & \text{otherwise} \end{cases} \quad \forall x \in \mathbb{R},$$

    which is called the *indicator function* of D
  - Let $D \subset \mathbb{R}$ be an arbitrary set and $R = [a, b] \subset \mathbb{R}$ be an interval such that $D \subset R$
  - We define the (Lebesgue) measure of D as

$$|D| := \int_R \mathbb{1}_D(x)dx := \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{1}_D(x_k)(x_{k+1} - x_k)$$

  - The measure of an interval $R = [a, b]$ coincides with the length of the interval

$$|R| = \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{1}_R(x_k)(x_{k+1} - x_k) = \lim_{n\to\infty} \sum_{k=0}^{n-1} (x_{k+1} - x_k) = b - a$$

- **Integral over arbitrary sets**
  - Consider a function $f : X \to \mathbb{R}$ defined over $X \subset \mathbb{R}$
  - For any subset $D \subset X$, we define the integral of f over D as

$$\int_D f(x)dx := \int_R \mathbb{1}_D(x)f(x)dx := \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{1}_D(x_k)f(x_k)(x_{k+1} - x_k),$$

    where R is an interval that contains D, provided that the limit exists
  - Note that $\mathbb{1}_D(x)f(x)$ can be seen as a single function defined over X

$$\mathbb{1}_D(x)f(x) = \begin{cases} f(x) & \text{if } x \in D \\ 0 & \text{if } x \notin D \end{cases} \quad \forall x \in X$$

1.2 Two dimensional case

- **Integral over rectangles**
  - Consider a function $f : X \to \mathbb{R}$ defined over $X \subset \mathbb{R}^2$
  - For a rectangle $R := [a_1, b_1] \times [a_2, b_2] \subset X$ and $n_1, n_2 \in \mathbb{N}$, the Riemann sum is

$$\sum_{k_2=0}^{n_2-1} \sum_{k_1=0}^{n_1-1} f(x_{1,k_1}, x_{2,k_2})(x_{1,k_1+1} - x_{1,k_1})(x_{2,k_2+1} - x_{2,k_2}), \quad x_{i,k_i} := a_i + \frac{k_i}{n_i}(b_i - a_i)$$

  - We define the integral of f over R as

$$\begin{aligned} \int_R f(\boldsymbol{x}) d\boldsymbol{x} &:= \lim_{n_2\to\infty} \lim_{n_1\to\infty} \sum_{k_2=0}^{n_2-1} \sum_{k_1=0}^{n_1-1} f(x_{1,k_1}, x_{2,k_2})(x_{1,k_1+1} - x_{1,k_1})(x_{2,k_2+1} - x_{2,k_2}) \\ &= \lim_{n_2\to\infty} \sum_{k_2=0}^{n_2-1} \underbrace{\left(\lim_{n_1\to\infty} \sum_{k_1=0}^{n_1-1} f(x_{1,k_1}, x_{2,k_2})(x_{1,k_1+1} - x_{1,k_1}) \right)}_{\int_{a_1}^{b_1} f(x_1, x_{2,k_2}) dx_1} (x_{2,k_2+1} - x_{2,k_2}) \\ &= \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right) dx_2 \end{aligned}$$

  - We may write this integral more explicitly as

$$\iint_R f(x_1, x_2) dx_1 dx_2, \quad \text{or} \quad \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2$$

- **Example**
  - The integral of $f(x_1, x_2) := x_1 x_2$ over $R := [0, x_1] \times [0, x_2]$ is

$$\int_0^{x_2} \int_0^{x_1} t_1 t_2 dt_1 dt_2 = \int_0^{x_2} \left(\int_0^{x_1} t_1 t_2 dt_1 \right) dt_2 = \int_0^{x_2} \left(\frac{1}{2} x_1^2 t_2 \right) dt_2 = \frac{1}{2}\frac{1}{2} x_1^2 x_2^2 = \frac{1}{4} x_1^2 x_2^2$$

- **Measure of sets**
  - For any subset $D \subset \mathbb{R}^2$, we define

$$\mathbb{1}_D(\boldsymbol{x}) := \begin{cases} 1 & \text{if } \boldsymbol{x} \in D \\ 0 & \text{otherwise} \end{cases} \quad \forall \boldsymbol{x} \in \mathbb{R}^2,$$

    which is called the *indicator function* of D
  - Let $R = [a_1, b_2] \times [a_2, b_2] \subset \mathbb{R}^2$ be a rectangle that contains $D \subset R^2$
  - We define the (Lebesgue) measure of D as

$$|D| := \int_R \mathbb{1}_D(\boldsymbol{x}) d\boldsymbol{x} := \lim_{n_2\to\infty} \lim_{n_1\to\infty} \sum_{k_2=0}^{n_2-1} \sum_{k_1=0}^{n_1-1} \mathbb{1}_D(x_{1,k_1}, x_{2,k_2})(x_{1,k_1+1} - x_{1,k_1})(x_{2,k_2+1} - x_{2,k_2})$$

    provided that the limit exists
  - The measure of a rectangle $R = [a_1, b_1] \times [a_2, b_2]$ coincides with the area of the rectangle

$$|R| = \lim_{n_2\to\infty} \lim_{n_1\to\infty} \sum_{k_1=0}^{n_1-1} (x_{1,k_1+1} - x_{1,k_1}) \sum_{k_2=0}^{n_2-1} (x_{2,k_2+1} - x_{2,k_2}) = (b_1 - a_1)(b_2 - a_2)$$

- **Integral over arbitrary sets**
  - Consider a function $f : X \to \mathbb{R}$ defined over $X \subset \mathbb{R}^2$
  - For any subset $D \subset X$, we define the integral of f over D as
$$\int_D f(\boldsymbol{x})d\boldsymbol{x} := \int_R \mathbb{1}_D(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}$$
    where R is a rectangle that contains D, provided that the limit exists

1.3 Higher dimensional case

- **Integral over hyperrectangles**
  - Consider a function $f : X \to \mathbb{R}$ defined over $X \subset \mathbb{R}^m$
  - For an m-dimensional hyperrectangle $R := \prod_{i=1}^m [a_i, b_i] \subset X$ and $n_1, n_2, \ldots, n_i \in \mathbb{N}$, the number
$$\sum_{k_m=0}^{n_m-1} \cdots \sum_{k_1=0}^{n_1-1} f(x_{1,k_1}, \ldots, x_{m,k_m}) \prod_{i=1}^m (x_{i,k_i+1} - x_{i,k_i}), \quad x_{i,k_i} := a_i + \frac{k_i}{n_i}(b_i - a_i)$$
    is called a (left) Riemann sum
  - We define the integral of f over R as
$$\int_R f(\boldsymbol{x})d\boldsymbol{x} := \lim_{n_m\to\infty} \cdots \lim_{n_1\to\infty} \left(\sum_{k_m=0}^{n_m-1} \cdots \sum_{k_1=0}^{n_1-1} f(x_{1,k_1}, \ldots, x_{m,k_m}) \prod_{i=1}^m (x_{i,k_i+1} - x_{i,k_i}) \right)$$
    provided that the limit exists
  - We may write this integral more explicitly as
$$\int \cdots \int_R f(x_1, \ldots, x_m)dx_1 \ldots dx_m \quad \text{or} \quad \int_{a_m}^{b_m} \cdots \int_{a_1}^{b_2} (x_1, \ldots, x_m)dx_1 \ldots dx_m$$
- **Measure of sets**
  - For any subset $D \subset \mathbb{R}^m$, we define
$$\mathbb{1}_D(\boldsymbol{x}) := \begin{cases} 1 & \text{if } \boldsymbol{x} \in D \\ 0 & \text{otherwise} \end{cases} \qquad \forall \boldsymbol{x} \in \mathbb{R}^m,$$
    which is called the *indicator function* of D
  - Let $R = \prod_{i=1}^m [a_i, b_i] \subset \mathbb{R}^m$ be a hyperrectangle that contains $D \subset R^m$
  - We define the (Lebesgue) measure of D as
$$|D| := \int_R \mathbb{1}_D(\boldsymbol{x})d\boldsymbol{x}$$
    provided that the limit on the right-hand side exists
  - The measure of a hyperrectangle $R = \prod_{i=1}^m [a_i, b_i]$ is
$$|R| = (b_1 - a_1)(b_2 - a_2) \cdots (b_m - a_m)$$
- **Integral over arbitrary sets**
  - Consider a function $f : X \to \mathbb{R}$ defined over $X \subset \mathbb{R}^m$
  - For any subset $D \subset X$, we define the integral of f over D as
$$\int_D f(\boldsymbol{x})d\boldsymbol{x} := \int_R \mathbb{1}_D(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}$$
    where R is a hyperrectangle that contains D, provided that the limit exists

2 Change of variables

2.1 Determinant revisited

- **Determinant of orthogonal matrices**
  - Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots \boldsymbol{a}_m \in \mathbb{R}^m$ be m-dimensional orthogonal vectors, i.e.,

$$\boldsymbol{a}_i^\top \boldsymbol{a}_j = 0 \quad \forall i \neq j$$

    and define $\boldsymbol{A} := \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_m \end{bmatrix}$
  - Then

$$\prod_{i=1}^{m} \|\boldsymbol{a}_i\| = |\det(\boldsymbol{A})|$$

    because

$$(|\boldsymbol{A}|)^2 = |\boldsymbol{A}||\boldsymbol{A}| = |\boldsymbol{A}^\top||\boldsymbol{A}| = |\boldsymbol{A}^\top \boldsymbol{A}| = \begin{vmatrix} \boldsymbol{a}_1^\top \boldsymbol{a}_1 & \boldsymbol{a}_1^\top \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1^\top \boldsymbol{a}_m \\ \boldsymbol{a}_2^\top \boldsymbol{a}_1 & \boldsymbol{a}_2^\top \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2^\top \boldsymbol{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_m^\top \boldsymbol{a}_1 & \boldsymbol{a}_m^\top \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_m^\top \boldsymbol{a}_m \end{vmatrix} = \prod_{i=1}^{n} \|\boldsymbol{a}_i\|^2 = \left(\prod_{i=1}^{n} \|\boldsymbol{a}_i\|\right)^2$$

  - In other words, if $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ are orthogonal,

$$\text{measure of } m\text{-dimensional hyperrectangle made by } \boldsymbol{a}_1, \dots, \boldsymbol{a}_m = \text{absolute value of } \det(\boldsymbol{A})$$

- **Skew translation and determinant**
  - Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots \boldsymbol{a}_m \in \mathbb{R}^m$ be orthogonal vectors
  - Now take the j-th column vector $\boldsymbol{a}_j$ and add $c_j \boldsymbol{a}_j$ with $c_j \neq 0$ to the i-th column to create another matrix

$$\boldsymbol{A}' := \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_i + c_j \boldsymbol{a}_j & \cdots & \boldsymbol{a}_j & \cdots & \boldsymbol{a}_m, \end{bmatrix} \quad i \neq j$$

    which geometrically represents an m-dimensional parallelepiped (i.e., an object you obtain by skewing the hyperrectangle made by $\boldsymbol{A}$ in the direction of $\boldsymbol{a}_j$)
  - Such a skew translation does not change the measure of the object, so we know that

$$\text{measure of object made by columns of } \boldsymbol{A}' = \text{measure of object made by columns of } \boldsymbol{A}$$

  - Since

$$\begin{aligned} |\boldsymbol{A}'| &= \begin{vmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_i + c_j \boldsymbol{a}_j & \cdots & \boldsymbol{a}_j & \cdots & \boldsymbol{a}_m \end{vmatrix} \\ &= \begin{vmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_i & \cdots & \boldsymbol{a}_j & \cdots & \boldsymbol{a}_m \end{vmatrix} + c_j \underbrace{\begin{vmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_j & \cdots & \boldsymbol{a}_j & \cdots & \boldsymbol{a}_m \end{vmatrix}}_{=0} = |\boldsymbol{A}| \end{aligned}$$

    we conclude that

$$\begin{aligned} &\text{measure of parallelepiped made by columns of } \boldsymbol{A}' \\ &\quad = \text{measure of hyperrectangle made by columns of } \boldsymbol{A} \\ &\quad = \text{absolute value of } \det(\boldsymbol{A}) \\ &\quad = \text{absolute value of } \det(\boldsymbol{A}') \end{aligned}$$

  - This observation should hold more generally because one can repeatedly apply skew translations without invalidating the argument

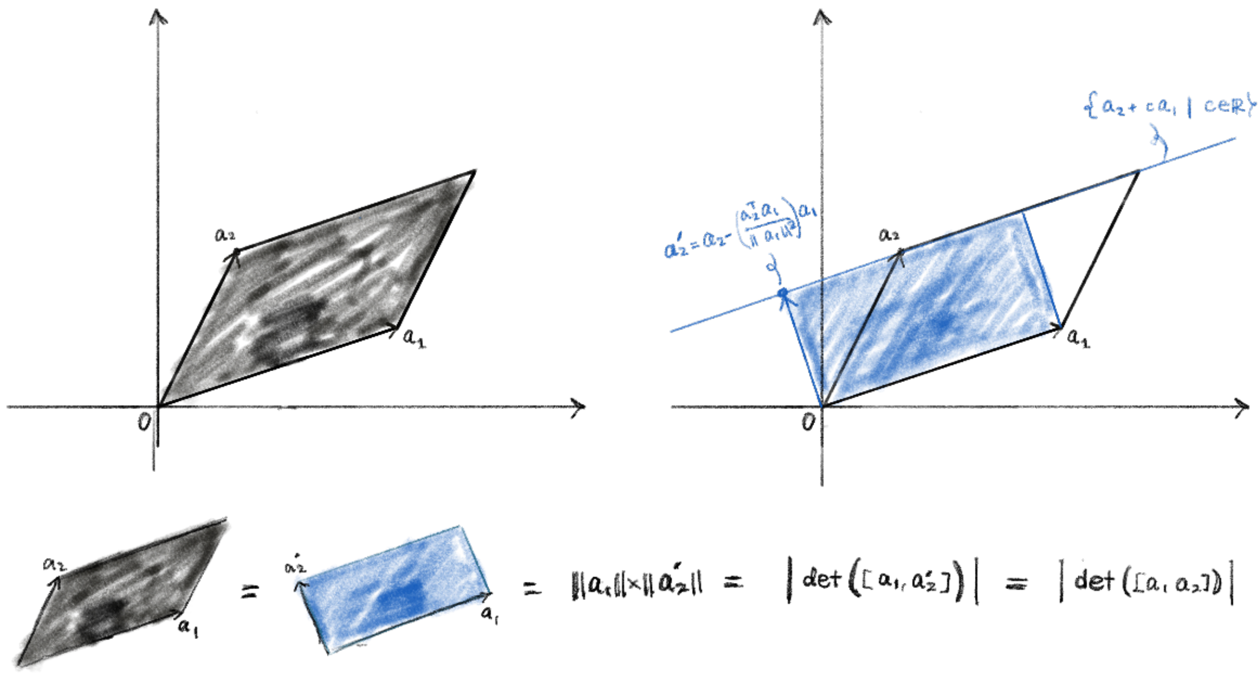

Figure 1: Skew translation and determinant

- **Determinant of general $m \times m$ matrices**
  - Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^m$ be a set of any vectors and define $\boldsymbol{A} := \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_m \end{bmatrix}$
  - We want to show that:

$$\text{measure of parallelepiped made by } \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^m = \text{absolute value of } \det(\boldsymbol{A})$$

  - To show this, we apply a set of skew translations (i.e., the Gram-Schmidt orthogonalization) to transform the parallelepiped into a hyperrectangle (Figure 1):

$$\begin{aligned}
\boldsymbol{a}'_1 &:= \boldsymbol{a}_1 \\
\boldsymbol{a}'_2 &:= \boldsymbol{a}_2 - (\boldsymbol{a}_2^\top \boldsymbol{a}'_1)\frac{\boldsymbol{a}'_1}{\|\boldsymbol{a}'_1\|^2} \\
\boldsymbol{a}'_3 &:= \boldsymbol{a}_3 - (\boldsymbol{a}_3^\top \boldsymbol{a}'_1)\frac{\boldsymbol{a}'_1}{\|\boldsymbol{a}'_1\|^2} - (\boldsymbol{a}_3^\top \boldsymbol{a}'_2)\frac{\boldsymbol{a}'_2}{\|\boldsymbol{a}'_2\|^2} \\
&\vdots \\
\boldsymbol{a}'_m &:= \boldsymbol{a}_m - (\boldsymbol{a}_m^\top \boldsymbol{a}'_1)\frac{\boldsymbol{a}'_1}{\|\boldsymbol{a}'_1\|^2} - (\boldsymbol{a}_m^\top \boldsymbol{a}'_2)\frac{\boldsymbol{a}'_2}{\|\boldsymbol{a}'_2\|^2} \cdots - (\boldsymbol{a}_m^\top \boldsymbol{a}'_{m-1})\frac{\boldsymbol{a}'_{m-1}}{\|\boldsymbol{a}'_{m-1}\|^2}
\end{aligned}$$

    and let $\boldsymbol{A}' := \begin{bmatrix} \boldsymbol{a}'_1 & \boldsymbol{a}'_2 & \cdots & \boldsymbol{a}'_m \end{bmatrix}$
  - Observe:
    - the column vectors of $\boldsymbol{A}'$ are orthogonal to each other so:

$$\text{measure of hyperrectangle made by } \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_m = \text{absolute value of } \det(\boldsymbol{A}')$$

    - since the skew translations should not change the measure of the object:

$$\begin{aligned}
&\text{measure of hyperrectangle made by } \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_m \\
&\quad = \text{measure of parallelepiped made by } \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m
\end{aligned}$$

    - On the other hand, we know from the properties of determinant that $\det(\boldsymbol{A}') = \det(\boldsymbol{A})$

- Therefore

$$\begin{aligned}&\text{measure of parallelepiped made by } \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_m\\ &\quad = \text{measure of hyperrectangle made by } \boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_m\\ &\quad = \text{absolute value of } \det(\boldsymbol{A}')\\ &\quad = \text{absolute value of } \det(\boldsymbol{A})\end{aligned}$$

- **Example:** 2×2 **matrix**
  - Let $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^2$ be the column vectors of $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$
  - Then

$$\begin{aligned}(|\boldsymbol{A}|)^2 &= |\boldsymbol{A}||\boldsymbol{A}| = |\boldsymbol{A}^\top||\boldsymbol{A}| = |\boldsymbol{A}^\top\boldsymbol{A}| = \left|\begin{bmatrix}\boldsymbol{a}_1^\top\\ \boldsymbol{a}_2^\top\end{bmatrix}\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix}\right| = \left|\begin{bmatrix}\boldsymbol{a}_1^\top\boldsymbol{a}_1 & \boldsymbol{a}_1^\top\boldsymbol{a}_2\\ \boldsymbol{a}_2^\top\boldsymbol{a}_1 & \boldsymbol{a}_2^\top\boldsymbol{a}_2\end{bmatrix}\right|\\ &= \boldsymbol{a}_1^\top\boldsymbol{a}_1\boldsymbol{a}_2^\top\boldsymbol{a}_2 - \boldsymbol{a}_1^\top\boldsymbol{a}_2\boldsymbol{a}_2^\top\boldsymbol{a}_1 = \|\boldsymbol{a}_1\|^2\|\boldsymbol{a}_2\|^2 - (\boldsymbol{a}_1^\top\boldsymbol{a}_2)^2\\ &= \|\boldsymbol{a}_1\|^2\|\boldsymbol{a}_2\|^2\left(1 - \left(\frac{(\boldsymbol{a}_1^\top\boldsymbol{a}_2)}{\|\boldsymbol{a}_1\|\|\boldsymbol{a}_2\|}\right)^2\right)\\ &= \|\boldsymbol{a}_1\|^2\|\boldsymbol{a}_2\|^2\left(1 - \cos^2(\theta)\right) \quad \text{where} \quad \theta \text{ is the angle between } \boldsymbol{a}_1 \text{ and } \boldsymbol{a}_2\\ &= \|\boldsymbol{a}_1\|^2\|\boldsymbol{a}_2\|^2\sin^2(\theta) = (\|\boldsymbol{a}_1\|\|\boldsymbol{a}_2\|\sin(\theta))^2\\ &= (\text{area of parallelogram made by } \boldsymbol{a}_1 \text{ and } \boldsymbol{a}_2)^2,\end{aligned}$$

  which confirms

$$\text{area of parallelogram made by } \boldsymbol{a}_1 \text{ and } \boldsymbol{a}_2 = \text{absolute value of } \det(\boldsymbol{A})$$

2.2 Change of variable formula

- **What we want to do**
  - Let us say that we want to compute the integral

$$\int_Z f_Z(\boldsymbol{z})d\boldsymbol{z}$$

  for some function $f_Z : \mathbb{R}^m \to \mathbb{R}$ over some subset $Z \subset \mathbb{R}^m$
  - Suppose that the variable $\boldsymbol{z} \in Z$ can be transformed into another variable $\boldsymbol{x} \in \mathbb{R}^m$ through a bijective function $\phi : Z \to \phi(Z)$

$$\boldsymbol{x} = \phi(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in Z \subset \mathbb{R}^n$$

  or

$$\boldsymbol{z} = \phi^{-1}(\boldsymbol{x}) =: \psi(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in X := \phi(Z)$$

  - We want to find a function $f_X : X \to \mathbb{R}$ such that

$$\int_B f_Z(\boldsymbol{z})d\boldsymbol{z} = \int_{\phi(B)} f_X(\boldsymbol{x})d\boldsymbol{x} \quad \forall B \subset Z \tag{1}$$

  or equivalently

$$\int_A f_X(\boldsymbol{x})d\boldsymbol{x} = \int_{\psi(A)} f_Z(\boldsymbol{z})d\boldsymbol{z} \quad \forall A \subset X \tag{2}$$

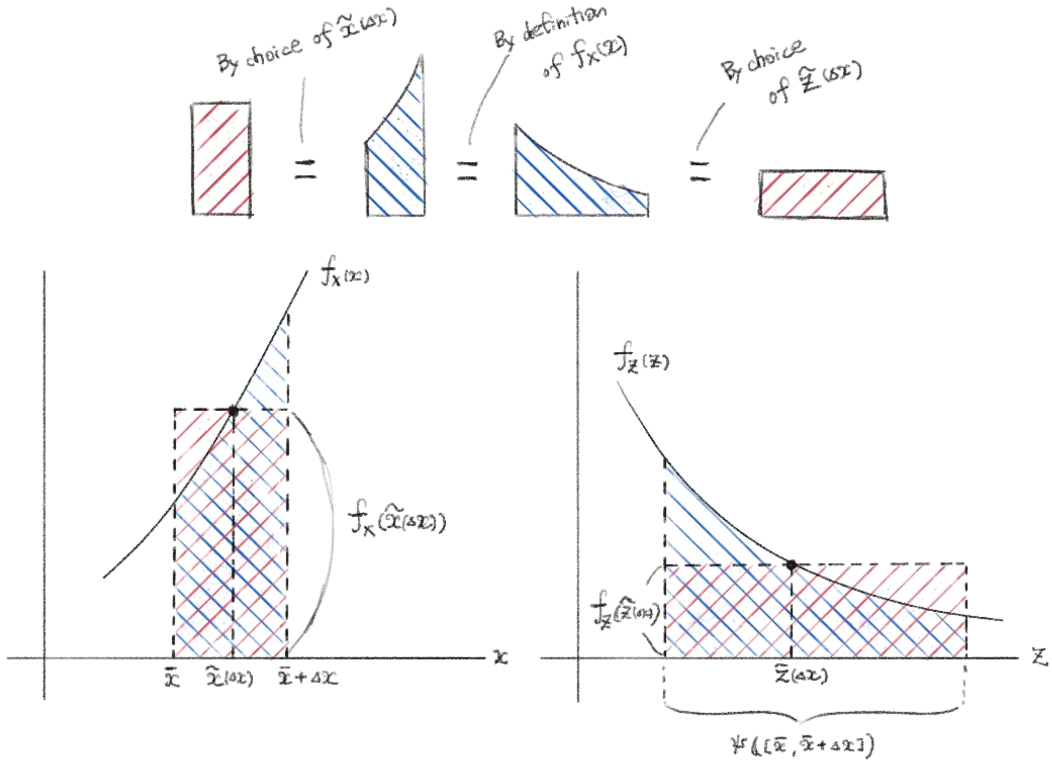

Figure 2: Choice of $\tilde{x}(\Delta x) \in [\bar{x}, \bar{x}+\Delta x]$ and $\tilde{z}(\Delta x) \in \psi([\bar{x}, \bar{x}+\Delta x])$ that satisfies (5) and (6)

- **Case with $m = 1$**
  - The function f_X that satisfies (1) or (2) is

$$f_X(x) = f_Z(\psi(x)) \left| \frac{d\psi(x)}{dx} \right| \quad \forall x \in X \tag{3}$$

  - To see this, fix $\bar{x} \in X$
  - If f_X satisfies (2), we must have

$$\int_{[\bar{x}, \bar{x}+\Delta x]} f_X(x)dx = \int_{\psi([\bar{x}, \bar{x}+\Delta x])} f_Z(z)dz \quad \forall \Delta x \geq 0 \tag{4}$$

  - Notice that for each $\Delta x > 0$,
    - there exists $\tilde{x}(\Delta x) \in [\bar{x}, \bar{x}+\Delta x]$ such that

$$\int_{[\bar{x}, \bar{x}+\Delta x]} f_X(x)dx = \int_{[\bar{x}, \bar{x}+\Delta x]} f_X(\tilde{x}(\Delta x))dx = f_X(\tilde{x}(\Delta x)) \underbrace{|[\bar{x}, \bar{x}+\Delta x]|}_{\Delta x}, \tag{5}$$

      which is illustrated in Figure 2
    - there exists $\tilde{z}(\Delta x) \in \psi([\bar{x}, \bar{x}+\Delta x])$ such that

$$\int_{\psi([\bar{x}, \bar{x}+\Delta x])} f_Z(z)dz = \int_{\psi([\bar{x}, \bar{x}+\Delta x])} f_Z(\tilde{z}(\Delta x))dz = f_Z(\tilde{z}(\Delta x))|\psi([\bar{x}, \bar{x}+\Delta x])| \tag{6}$$

      where $|\psi([\bar{x}, \bar{x}+\Delta x])|$ is the measure of the set $\psi([\bar{x}, \bar{x}+\Delta x])$
  - It follows from (4), (5), and (6) that

$$f_X(\tilde{x}(\Delta x)) = f_Z(\tilde{z}(\Delta x)) \frac{|\psi([\bar{x}, \bar{x}+\Delta x])|}{\Delta x}, \quad \forall \Delta x > 0$$

- Obviously,

$$\lim_{\Delta x \to 0} \tilde{x}(\Delta x) = \bar{x}, \quad \lim_{\Delta x \to 0} \tilde{z}(\Delta x) = \psi(\bar{x}),$$
$$|\psi([\bar{x}, \bar{x} + \Delta x])| = |\psi(\bar{x} + \Delta x) - \psi(\bar{x})| \quad \text{for sufficiently small } \Delta x,$$

  and therefore

$$\begin{aligned} f_X(\bar{x}) &= \lim_{\Delta x \to 0} f_X(\tilde{x}(\Delta x)) \\ &= \lim_{\Delta x \to 0} f_Z(\tilde{z}(\Delta x)) \frac{|\psi([\bar{x}, \bar{x} + \Delta x])|}{\Delta x} \\ &= f_Z(\psi(\bar{x})) \lim_{\Delta x \to 0} \frac{|\psi([\bar{x}, \bar{x} + \Delta x])|}{\Delta x} \\ &= f_Z(\psi(\bar{x})) \lim_{\Delta x \to 0} \left| \frac{\psi(\bar{x} + \Delta x) - \psi(\bar{x})}{\Delta x} \right| \\ &= f_Z(\psi(\bar{x})) \left| \frac{d\psi(\bar{x})}{dx} \right|, \end{aligned}$$

  where the absolute value is necessary because $\psi(\bar{x} + \Delta x)$ can be smaller than $\psi(\bar{x})$

- Since the argument above does not depend on the choice of $\bar{x}$, we conclude that the function f_X must be given by (3)

- **Case with $m = 2$ and higher m**

- The function f_X that satisfies (1) or (2) is

$$f_X(\boldsymbol{x}) = f_Z(\psi(\boldsymbol{x})) \left\| \frac{d\psi(\boldsymbol{x})}{d\boldsymbol{x}} \right\| \quad \forall \boldsymbol{x} \in X \tag{7}$$

- To see this, fix $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2) \in X$
- If f_X satisfies (2), we must have

$$\int_{[\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}]} f_X(\boldsymbol{x}) d\boldsymbol{x} = \int_{\psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])} f_Z(\boldsymbol{z}) d\boldsymbol{z} \quad \forall \Delta \boldsymbol{x} = (\Delta x_1, \Delta x_2) \geq \boldsymbol{0} \tag{8}$$

  where $[\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}] := [\bar{x}_1, \bar{x}_1 + \Delta x_1] \times [\bar{x}_2, \bar{x}_2 + \Delta x_2]$

- Notice that for each $\Delta \boldsymbol{x} > \boldsymbol{0}$,
  - there exists $\tilde{\boldsymbol{x}}(\Delta \boldsymbol{x}) \in [\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}]$ such that

$$\int_{[\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}]} f_X(\boldsymbol{x}) d\boldsymbol{x} = \int_{[\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}]} f_X(\tilde{\boldsymbol{x}}(\Delta \boldsymbol{x})) d\boldsymbol{x} = f_X(\tilde{\boldsymbol{x}}(\Delta \boldsymbol{x})) \underbrace{|[\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}]|}_{\Delta x_1 \Delta x_2} \tag{9}$$

  - there exists $\tilde{\boldsymbol{z}}(\Delta \boldsymbol{x}) \in \psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])$ such that

$$\int_{\psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])} f_Z(\boldsymbol{z}) d\boldsymbol{z} = \int_{\psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])} f_Z(\tilde{\boldsymbol{z}}(\Delta \boldsymbol{x})) d\boldsymbol{z} = f_Z(\tilde{\boldsymbol{z}}(\Delta \boldsymbol{x})) |\psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])| \tag{10}$$

    where $|\psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])|$ is the measure of the set $\psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])$

- It follows from (8), (9), and (10) that

$$f_X(\tilde{\boldsymbol{x}}(\Delta \boldsymbol{x})) = f_Z(\tilde{\boldsymbol{z}}(\Delta \boldsymbol{x})) \frac{|\psi([\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}} + \Delta \boldsymbol{x}])|}{\Delta x_1 \Delta x_2}, \quad \forall \Delta \boldsymbol{x} > 0$$

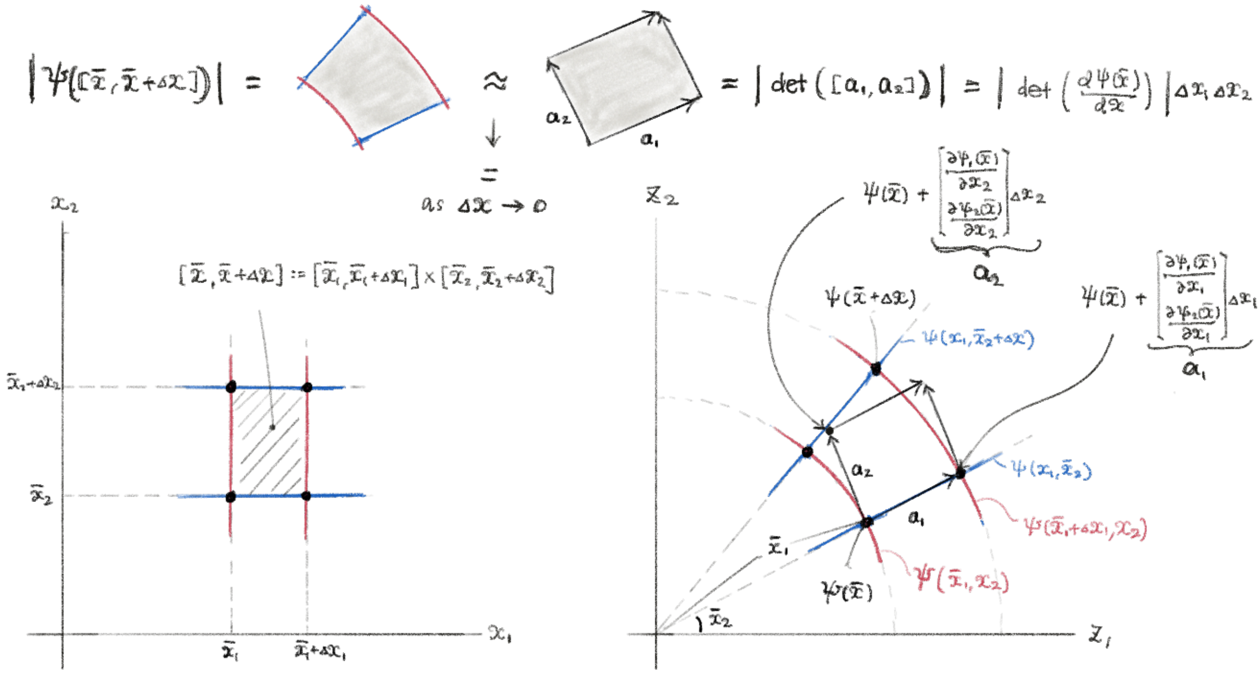

Figure 3: Derivation of $|\psi([\bar{x}, \bar{x} + \Delta x])|$ based on Example 2

- We know that for Δx close enough to $\mathbf{0}$

$$
\begin{aligned}
|\psi([\bar{x}, \bar{x} + \Delta x])| &\approx \text{area of parallelogram made by } \boldsymbol{a}_1 := \begin{bmatrix} \frac{\partial \psi_1(\bar{x})}{\partial x_1} \\ \frac{\partial \psi_2(\bar{x})}{\partial x_1} \end{bmatrix} \Delta x_1 \text{ and } \boldsymbol{a}_2 := \begin{bmatrix} \frac{\partial \psi_1(\bar{x})}{\partial x_2} \\ \frac{\partial \psi_2(\bar{x})}{\partial x_2} \end{bmatrix} \Delta x_2 \\
&= \left|\det\left(\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix}\right)\right| \\
&= \left| \begin{vmatrix} \frac{\partial \psi_1(\bar{x})}{\partial x_1} \Delta x_1 & \frac{\partial \psi_1(\bar{x})}{\partial x_2} \Delta x_2 \\ \frac{\partial \psi_2(\bar{x})}{\partial x_1} \Delta x_1 & \frac{\partial \psi_2(\bar{x})}{\partial x_2} \Delta x_2 \end{vmatrix} \right| = \left| \begin{vmatrix} \frac{\partial \psi_1(\bar{x})}{\partial x_1} & \frac{\partial \psi_1(\bar{x})}{\partial x_2} \\ \frac{\partial \psi_2(\bar{x})}{\partial x_1} & \frac{\partial \psi_2(\bar{x})}{\partial x_2} \end{vmatrix} \Delta x_1 \Delta x_2 \right| \\
&= \left| \left| \frac{d\psi(\bar{x})}{d\boldsymbol{x}} \right| \right| \Delta x_1 \Delta x_2
\end{aligned}
$$

  where $\left|\left|\frac{d\psi(\bar{x})}{dx}\right|\right|$ is the absolute value of the determinant of the Jacobian matrix $\frac{d\psi(\bar{x})}{dx}$ (See Figure 3 for an illustration)

- Obviously,

$$
\lim_{\Delta x \to \mathbf{0}} \tilde{x}(\Delta x) = \bar{x}, \quad \lim_{\Delta x \to \mathbf{0}} \tilde{z}(\Delta x) = \psi(\bar{x})
$$

  and therefore

$$
\begin{aligned}
f_X(\bar{x}) &= \lim_{\Delta x \to \mathbf{0}} f_X(\tilde{x}(\Delta x)) \\
&= \lim_{\Delta x \to \mathbf{0}} f_Z(\tilde{z}(\Delta x)) \frac{|\psi([\bar{x}, \bar{x} + \Delta x])|}{\Delta x_1 \Delta x_2} \\
&= f_Z(\psi(\bar{x})) \lim_{\Delta x \to \mathbf{0}} \frac{|\psi([\bar{x}, \bar{x} + \Delta x])|}{\Delta x_1 \Delta x_2} \\
&= f_Z(\psi(\bar{x})) \left| \left| \frac{d\psi(\bar{x})}{dx} \right| \right|
\end{aligned}
$$

- Since the argument above does not depend on the choice of $\bar{x}$, we conclude that the function f_X must be given by (7)
- The same is true of higher m

- **Example 1**
  - Consider the integral

$$\int_Z f_Z(z)dz, \quad f_Z(z) = \frac{(z^{-\alpha}+\gamma)^\beta}{z^{\alpha+1}} \quad Z := [a,b], \quad b > a > 0$$

    for some $\alpha, \beta, \gamma > 0$
  - Transform z into x by defining a function $\phi : Z \to \mathbb{R}$ as

$$x = \phi(z) := z^{-\alpha} + \gamma \quad \forall z \in Z$$

  - Notice that $\phi : Z \to X$ is a bijective (monotonically decreasing) function with

$$X := \phi(Z) = [\phi(b), \phi(a)] = [b^{-\alpha}+\gamma, a^{-\alpha}+\gamma]$$

    and the inverse function $\psi := \phi^{-1}$ is given by

$$z = \psi(x) := (x-\gamma)^{-\frac{1}{\alpha}} \quad \forall x \in X$$

  - Using the change of variable formula, one would obtain

$$\begin{aligned}
\int_Z f_Z(z)dz &= \int_X f_X(x)dx, \quad \text{where} \quad f_X(x) := f_Z(\psi(x)) \left|\frac{d\psi(x)}{dx}\right| \\
&= \int_X (x-\gamma)^{\frac{\alpha+1}{\alpha}} x^\beta \left| -\frac{1}{\alpha}(x-\gamma)^{-\frac{\alpha+1}{\alpha}} \right| dx \\
&= \int_X \frac{1}{\alpha} x^\beta dx \\
&= \int_{b^{-\alpha}+\gamma}^{a^{-\alpha}+\gamma} \frac{d}{dx}\left\{\frac{1}{\alpha\beta}x^\beta\right\} dx \\
&= \frac{1}{\alpha\beta}\left((a^{-\alpha}+\gamma)^\beta - (b^{-\alpha}+\gamma)^\beta\right)
\end{aligned}$$

- **Example 2**
  - Consider the integral

$$\int_Z f_Z(\boldsymbol{z})d\boldsymbol{z}, \quad \text{where} \quad f_Z(\boldsymbol{z}) := e^{-z_1^2-z_2^2}, \quad Z := \left\{(z_1,z_2) \in \mathbb{R}^2 \setminus \{\boldsymbol{0}\} \,\middle|\, z_1^2+z_2^2 \le a^2\right\}$$

  - Transform $\boldsymbol{z} = (z_1, z_2)$ into $\boldsymbol{x} = (x_1, x_2)$ by defining a function $\phi : Z \to \mathbb{R}^2$ as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \phi_1(z_1,z_2) \\ \phi_2(z_1,z_2) \end{bmatrix} := \begin{bmatrix} (z_1^2+z_2^2)^{1/2} \\ \arctan 2(z_2, z_1) \end{bmatrix} \quad \forall \boldsymbol{z} \in Z$$

  - Notice that $\phi : Z \to X$ is bijective with

$$X := \phi(Z) = \{(x_1,x_2) \,|\, a \ge x_1 > 0, \pi \ge x_2 \ge -\pi\}$$

    and the inverse function $\psi := \phi^{-1}$ is given by

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \psi_1(x_1,x_2) \\ \psi_2(x_1,x_2) \end{bmatrix} := \begin{bmatrix} x_1\cos(x_2) \\ x_1\sin(x_2) \end{bmatrix} \quad \forall \boldsymbol{x} \in X$$

- Using the change of variable formula,

$$\begin{aligned}
\int_Z f_Z(\boldsymbol{z})d\boldsymbol{z} &= \int_X f_X(\boldsymbol{x})d\boldsymbol{x}, \quad \text{where} \quad f_X(\boldsymbol{x}) = f_Z(\psi(\boldsymbol{x})) \left|\left|\frac{d\psi(\boldsymbol{x})}{d\boldsymbol{x}}\right|\right| \\
&= \int_X f_Z(\psi(\boldsymbol{x})) \left|\left|\frac{d\psi(\boldsymbol{x})}{d\boldsymbol{x}}\right|\right| d\boldsymbol{x} \\
&= \int_{-\pi}^{\pi} \int_0^a e^{-x_1^2\cos^2(x_2)-x_1^2\sin^2(x_2)} \left|\left|\begin{matrix}\cos(x_2) & -x_1\sin(x_2) \\ \sin(x_2) & x_1\cos(x_2)\end{matrix}\right|\right| dx_1dx_2 \\
&= \int_{-\pi}^{\pi} \int_0^a e^{-x_1^2\cos^2(x_2)-x_1^2\sin^2(x_2)} \left|x_1\cos^2(x_2) + x_1\sin^2(x_2)\right| dx_1dx_2 \\
&= \int_{-\pi}^{\pi} \int_0^a e^{-x_1^2} x_1 dx_1dx_2 \\
&= \int_{-\pi}^{\pi} \frac{1}{2}\left(1 - e^{-a^2}\right) dx_2 \\
&= \pi\left(1 - e^{-a^2}\right)
\end{aligned}$$

- Since $\lim_{a\to\infty} Z = \mathbb{R}^2 \setminus \{\boldsymbol{0}\}$, we have

$$\int_{\mathbb{R}^2} e^{-z_1^2-z_2^2}d\boldsymbol{z} = \int_{\mathbb{R}^2\setminus\{\boldsymbol{0}\}} e^{-z_1^2-z_2^2}d\boldsymbol{z} = \lim_{a\to\infty} \pi\left(1 - e^{-a^2}\right) = \pi,$$

  which in turn allows us to compute the Gaussian integral:

$$\int_{\mathbb{R}} e^{-z^2}dz = \left(\left(\int_{\mathbb{R}} e^{-z_1^2}dz_1\right)\left(\int_{\mathbb{R}} e^{-z_2^2}dz_2\right)\right)^{1/2} = \left(\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-z_1^2}e^{-z_2^2}dz_1dz_2\right)^{1/2} = \sqrt{\pi} \quad (11)$$

- **Example 3**
- Consider the integral

$$\int_Z f_Z(\boldsymbol{z})d\boldsymbol{z}, \quad Z \subset \mathbb{R}^m$$

  for some function $f_Z : \mathbb{R}^m \to \mathbb{R}$
- Transform $\boldsymbol{z} \in \mathbb{R}^m$ into $\boldsymbol{x} \in \mathbb{R}^m$ by defining a function $\phi : \mathbb{R}^m \to \mathbb{R}^m$ as

$$\boldsymbol{x} = \phi(\boldsymbol{z}) := \boldsymbol{A}\boldsymbol{z} + \boldsymbol{b} \quad \forall \boldsymbol{z} \in \mathbb{R}^m$$

  for some $\boldsymbol{A} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{b} \in \mathbb{R}^m$
- Notice that, as long as $\boldsymbol{A}$ is non-singular, the function $\phi : Z \to X$ is bijective and the inverse function $\psi := \phi^{-1}$ is

$$\boldsymbol{z} = \psi(\boldsymbol{x}) := \boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{b}) \quad \forall \boldsymbol{x} \in \mathrm{X} := \phi(\mathrm{Z}) = \{\boldsymbol{x} \in \mathbb{R}^m \,|\, \boldsymbol{x} = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{b}\}$$

- Using the change of variable formula, we arrive at

$$\begin{aligned}
\int_Z f_Z(\boldsymbol{z})d\boldsymbol{z} &= \int_X f_X(\boldsymbol{x})d\boldsymbol{x}, \quad \text{where} \quad f_X(\boldsymbol{x}) = f_Z(\psi(\boldsymbol{x})) \left|\left|\frac{d\psi(\boldsymbol{x})}{d\boldsymbol{x}}\right|\right| \\
&= \int_X f_Z(\psi(\boldsymbol{x})) \left|\left|\frac{d\psi(\boldsymbol{x})}{d\boldsymbol{x}}\right|\right| d\boldsymbol{x} \\
&= \int_X f_Z(\boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{b})) \left|\left|\boldsymbol{A}^{-1}\right|\right| d\boldsymbol{x} \\
&= \int_X f_Z(\boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{b})) \frac{1}{||\boldsymbol{A}||} d\boldsymbol{x}
\end{aligned}$$

- **Example 4**
  - Let us say we want to compute the integral

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz$$

  - With the application of Gaussian integral (11) in mind, let us consider the following "change of variable"

$$x = \underbrace{\frac{1}{\sqrt{2}} z}_{\phi(z)} \qquad \text{or equivalently} \qquad z = \underbrace{\sqrt{2}x}_{\psi(x)}$$

  - Then the change of variable formula implies

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\psi(x))^2} \underbrace{|\psi'(x)|}_{\sqrt{2}} dx = \sqrt{2} \underbrace{\int_{-\infty}^{\infty} e^{-x^2} dx}_{\sqrt{\pi}} = \sqrt{2\pi}$$

  - Note that this is a special case of Example 3, where $\boldsymbol{A} = 1/\sqrt{2}$ and $\boldsymbol{b} = 0$