《机器学习》课程系列

雅可比矩阵*

Jacobian Matrix

武汉纺织大学数学与计算机学院

杜小勤

2020/07/11

Contents

1 定积分的换元法

在定积分中，可以使用换元法来简化定积分的求解[1]：

$$\int_a^b f(x)dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f\left(g(u)\right) g'(u)du \tag{1}$$

其中，$x = g(u)$ 和 $dx = g'(u)du$。

*本系列文档属于讲义性质，仅用于学习目的。Last updated on: July 17, 2020。

[1]关于换元法的定理证明，详见文献 [1]。

上述换元公式也可以反过来使用，为统一表达，将前述公式中的 x 与 u 对调：

$$\int_a^b f\left(g(x)\right) g'(x)dx = \int_{g(a)}^{g(b)} f(u)du \tag{2}$$

其中，$u = g(x)$ 和 $du = g'(x)dx$。

例 **1.1** 计算 $\int_0^{\frac{\pi}{2}} \cos^5 x \sin x dx$。

解：

1. 显式地使用换元法

   令 $u = \cos x$，则 $du = -\sin x dx$，且：

$$\begin{cases} x = 0 \quad \Rightarrow \quad u = 1 \\ x = \dfrac{\pi}{2} \quad \Rightarrow \quad u = 0 \end{cases} \tag{3}$$

   则：

$$\int_0^{\frac{\pi}{2}} \cos^5 x \sin x dx = -\int_1^0 u^5 du = \int_0^1 u^5 du = \left[\frac{u^6}{6}\right]_0^1 = \frac{1}{6} \tag{4}$$

2. 不显式地使用换元法

$$\int_0^{\frac{\pi}{2}} \cos^5 x \sin x dx = -\int_0^{\frac{\pi}{2}} \cos^5 x d\left(\cos x\right) = -\left[\frac{\cos^6 x}{6}\right]_0^{\frac{\pi}{2}} = \frac{1}{6} \tag{5}$$

□

2 极坐标换元法

有些二重积分，使用极坐标 (Polar Coordinates) 替换直角坐标 (Rectangular Coordinates) 或笛卡尔坐标 (Cartesian Coordinates)，积分区域与被积函数均会变得简便些。实际上，这是一种特殊的换元法，被称为极坐标换元法。后文将推导出下面的结论：

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ dA = rdrd\theta \end{cases} \tag{6}$$

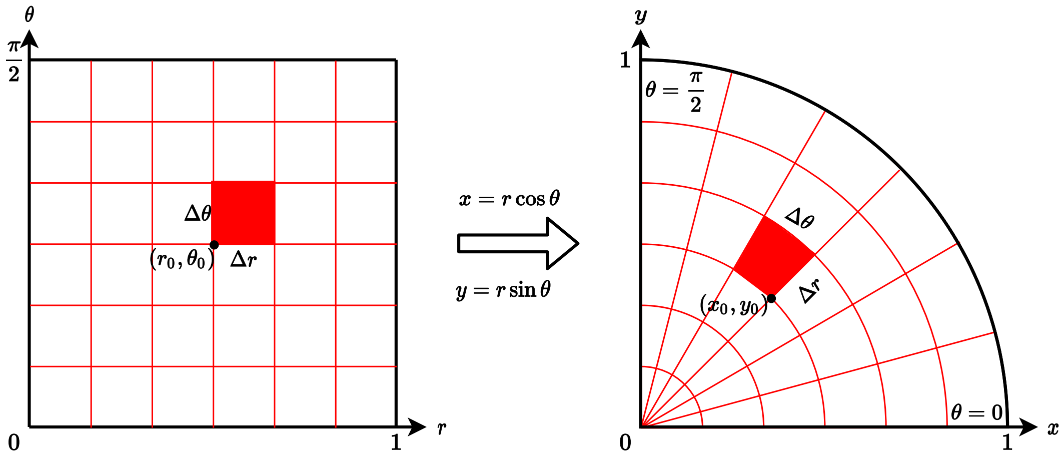

图 2-1: 极坐标换元法示意图

其中，r 与 θ 是极坐标系中的分量，dA 表示极坐标系中的面积微元。于是，得到：

$$\iint_R f(x,y)dxdy = \iint_S f(r\cos\theta, r\sin\theta)dA = \iint_S f(r\cos\theta, r\sin\theta)rdrd\theta \tag{7}$$

其中，R 是 $x-y$ 直角坐标系上的积分区域，S 是 $r-\theta$ 极坐标系上对应的积分区域。从上式可以看出，直角坐标系中的面积微元为 $dxdy$，而极坐标系中的面积微元为 $rdrd\theta$[2]。

为了获得对两个坐标系下积分区域形状变化的直观认识，我们来观察图2-1。可以看出，极坐标系中的大矩形映射成直角坐标系中的 $\frac{1}{4}$ 圆弧；极坐标系中的小矩形 (左子图的红色区域) 映射成直角坐标系中的截断扇形环 (右子图的红色区域)。此外，左子图中标记为 (r_0, θ_0) 的点，对应着右子图中标记为 (x_0, y_0) 的点，Δr 与 $\Delta\theta$ 在 2 个子图中均已标出。

下面来推导极坐标系中，dr 与 $d\theta$ 所形成的面积微元的计算公式。该过程可以借助于原有的直角坐标系——使用极坐标的划分方式，即以极点为中心的一簇同心圆来划分积分区域，参见图2-1的右子图。

为便于讨论，将右子图中的局部单独展示出来，如图2-2所示。红色区域的面积为：

$$\Delta A = \frac{1}{2}r_2^2\Delta\theta - \frac{1}{2}r_1^2\Delta\theta = \frac{r_2+r_1}{2}(r_2-r_1)\Delta\theta \tag{8}$$

其中，$\frac{r_2+r_1}{2}$ 是 2 个半径的平均长度。

[2]需要注意的是，在二重积分中，对积分区域的划分是任意的。例如，在直角坐标系中，如果以平行于坐标轴的网格来划分积分区域，则得到面积微元 $dxdy$，而如果以极点为中心的一簇同心圆来划分积分区域 (参见图2-1的右子图)，则得到面积微元 $rdrd\theta$，下文将表明这一点。

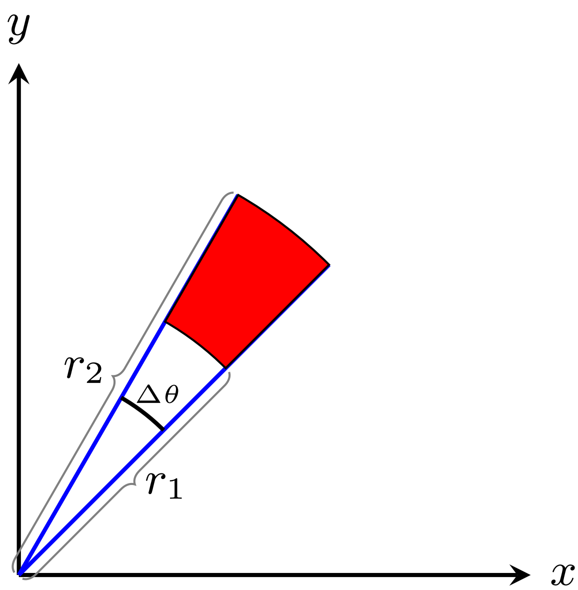

图 2-2: 截断扇形环

当 $r_2 - r_1 = \Delta r \to 0$ 时，$r_2 \approx r_1$，则 $\frac{r_2+r_1}{2} \approx r_i(i=1,2)$。因而，当 $\Delta r \to 0$ 和 $\Delta\theta \to 0$ 时：

$$\Delta A \approx r_i \Delta r \Delta\theta \quad \Rightarrow \quad dA = r dr d\theta \tag{9}$$

其中，$dA = rdrd\theta$ 就是极坐标系下的面积微元。于是，得到：

$$\iint_R f(x,y)dxdy = \iint_S f(r\cos\theta, r\sin\theta)dA = \iint_S f(r\cos\theta, r\sin\theta)rdrd\theta \tag{10}$$

3 雅可比矩阵

雅可比矩阵 (Jacobian Matrix) 是函数的一阶偏导数以一定方式排列而成的矩阵。假设函数 $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$，它将向量 $\boldsymbol{x} \in \mathbb{R}^n$ 映射到向量 $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^m$，那么函数 $\boldsymbol{f}$ 的雅可比矩阵被定义为 $m \times n$ 矩阵：

$$\boldsymbol{J} = \begin{bmatrix} \dfrac{\partial \boldsymbol{f}}{\partial x_1} & \cdots & \dfrac{\partial \boldsymbol{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{11}$$

当 $m = n$ 时，雅可比矩阵的行列式被称为雅可比行列式 (Jacobian Determinant)[3]。

[3] 在英文文献中，一般将雅可比矩阵与雅可比行列式统称为 Jacobian。

我们知道，从变换的角度看，矩阵表示向量空间之间的某个变换。雅可比矩阵是由偏导数组成的矩阵，而 (偏) 导数表示向量空间之间的线性变换。因此，雅可比矩阵表示的是，将向量空间 $\mathbb{R}^n$ 中的点 $(x_1, \cdots, x_n)$ 线性变换为向量空间 $\mathbb{R}^m$ 中的点 $(y_1, \cdots, y_m)$，即将一个 n 维欧式空间线性变换为另一个 m 维欧式空间。其意义在于，如果函数 $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$ 在点 $\boldsymbol{x}$ 可微，那么点 $\boldsymbol{x}$ 的雅可比矩阵就是函数 $\boldsymbol{f}$ 在该点的最佳线性近似[4]。

另外，可以将雅可比矩阵理解为，单值实数函数的微分向多值或向量实数函数微分的泛化或扩展，即雅可比矩阵是多值或向量函数 $\boldsymbol{f}$ 在点 $\boldsymbol{x}$ 的微分或者偏导数。因此，也可以将雅可比矩阵记为 $\frac{\partial(f_1,\cdots,f_m)}{\partial(x_1,\cdots,x_n)}$、$\frac{\partial(y_1,\cdots,y_m)}{\partial(x_1,\cdots,x_n)}$ 或 $\boldsymbol{J_f}(x_1, \cdots, x_n)$。

根据上述讨论，假设 $\boldsymbol{x}_0$ 是空间 $\mathbb{R}^n$ 中的点，且 $\boldsymbol{f}$ 在 $\boldsymbol{x}_0$ 处可微，那么 $\boldsymbol{J_f}(\boldsymbol{x}_0)$ 就是函数 $\boldsymbol{f}$ 在点 $\boldsymbol{x}_0$ 处的 (偏) 导数 (矩阵)。于是，当 $\boldsymbol{x}$ 充分靠近点 $\boldsymbol{x}_0$ 时，有：

$$\boldsymbol{f}(\boldsymbol{x}) \approx \boldsymbol{f}(\boldsymbol{x}_0) + \boldsymbol{J_f}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \tag{12}$$

或：

$$\Delta \boldsymbol{y} \approx \boldsymbol{J_f}(\boldsymbol{x}_0)\Delta \boldsymbol{x} \tag{13}$$

当 $m = 1$ 时，函数 $\boldsymbol{f}$ 退化为 $f: \mathbb{R}^n \to \mathbb{R}$，它就是常规的标量函数[5]，雅可比矩阵退化为由偏导数组成的行向量。特别地，当 $n = m = 1$ 时，$f: \mathbb{R} \to \mathbb{R}$ 是单变量标量函数，雅可比矩阵退化为一个导数标量。

4 雅可比行列式与通用换元法

前面，讨论了二重积分的极坐标换元法。下面，探讨适用于一般情形的通用换元法。

关于二重积分的通用换元法，首先给出如下结论[6]：

$$\iint\limits_R f(x, y)dxdy = \iint\limits_S f\left(x(u, v), y(u, v)\right) \left|\frac{\partial(x, y)}{\partial(u, v)}\right|_+ dudv \tag{14}$$

其中，$|\frac{\partial(x,y)}{\partial(u,v)}|_+$ 表示雅可比行列式的绝对值。该换元公式的应用，需要满足如下几个条件：

[4]实际上，可以将微分理解为线性化，将 (偏) 导数理解为线性变换。

[5]与之对应的是，前述的 $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$ 被称为向量函数。

[6]具体的定理描述，请参考文献 [1]。

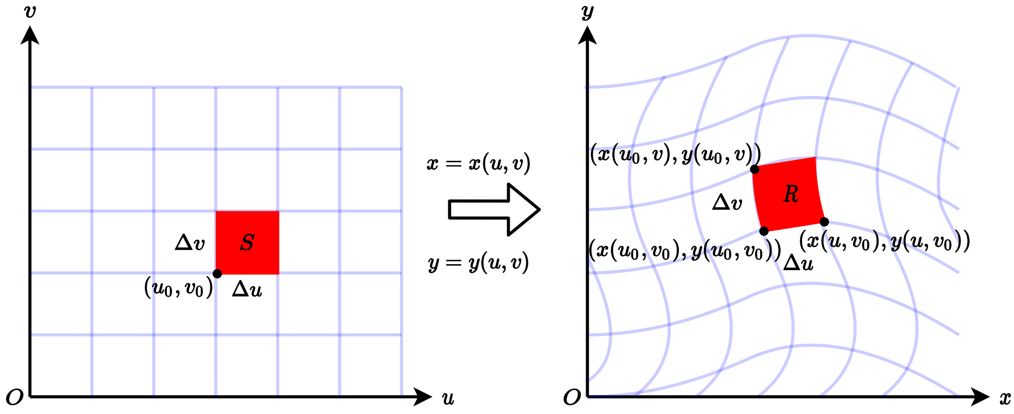

图 4-3: uOv 平面到 xOy 平面上的某个变换

- $x(u, v)$ 与 $y(u, v)$ 具有一阶连续偏导数；
- 雅可比行列式 $|\frac{\partial(x,y)}{\partial(u,v)}|$ 不等于 0；
- 在区域 R 与 S 上的变换是一对一的；

实际上，上述积分换元公式可以扩展到 n 重积分情形。

下面推导二重积分的通用换元公式。注意，下面的推导，均在直角坐标系的视角下进行[7]。图4-3展示了变换：

$$\begin{cases} x = x(u, v) \\ y = y(u, v) \end{cases} \tag{16}$$

它将直角坐标平面 uOv 上的点 (u, v)，通过上式，变换为直角坐标平面 xOy 上的点 (x, y)。

在图4-3中，标出了几个关键点及其坐标分量的对应关系，其中 $(x_0, y_0) = (x(u_0, v_0), y(u_0, v_0))$。为了便于理解直角坐标平面 uOv 下面积微元的求解过程，将图4-3的右子图单独放大显示，如图4-4所示。

根据变换公式 (16) 及“一对一”的约束条件，uOv 平面上的每一点 (u, v)，在 xOy 平面上，必然存在唯一的点 (x, y) 与之对应，反之亦然。为了表达方便，将

[7]即便是前述的极坐标换元法，仍然可以认为是直角坐标系下执行的某种变换——把公式：

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \tag{15}$$

看成是从直角坐标平面 $rO\theta$ 到直角坐标平面 xOy 的某种变换。

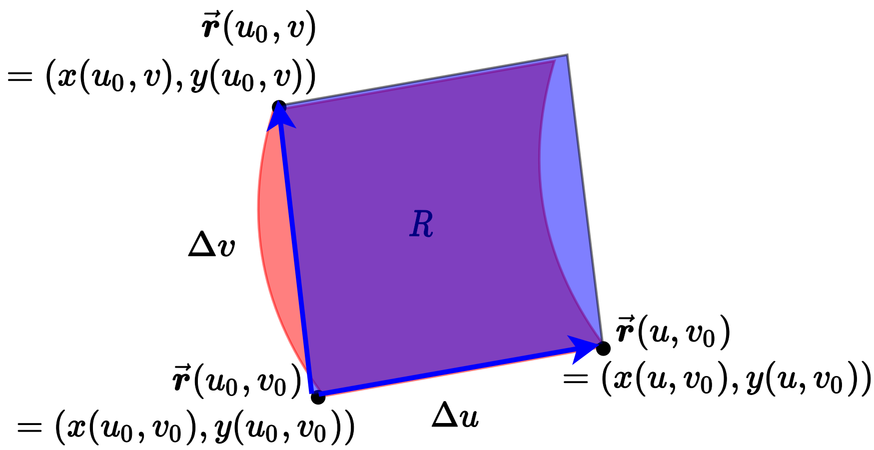

图 4-4: 面积微元示意图

(x, y) 表达成向量形式，于是，可令：

$$(x, y) = x\hat{\boldsymbol{i}} + y\hat{\boldsymbol{j}} = x(u, v)\hat{\boldsymbol{i}} + y(u, v)\hat{\boldsymbol{j}} = \vec{\boldsymbol{r}}(u, v) \tag{17}$$

即 $\vec{\boldsymbol{r}}(u, v)$ 表示 uOv 平面上的点 (u, v) 在 xOy 平面上形成的像 (Image) 点或位置向量 (Position Vector)。

于是，在图4-4中，点 $(x(u_0, v_0), y(u_0, v_0))$ 对应 $\vec{\boldsymbol{r}}(u_0, v_0)$，点 $(x(u, v_0), y(u, v_0))$ 对应 $\vec{\boldsymbol{r}}(u, v_0)$，点 $(x(u_0, v), y(u_0, v))$ 对应 $\vec{\boldsymbol{r}}(u_0, v)$。

当 $\Delta u \to 0$ 和 $\Delta v \to 0$ 时，图4-4中面积微元 (橙色)R 的面积，可近似为图4-4中平行四边形 (蓝色) 的面积，即[8]：

$$\Delta A \approx \|\, (\vec{\boldsymbol{r}}(u, v_0) - \vec{\boldsymbol{r}}(u_0, v_0)) \times (\vec{\boldsymbol{r}}(u_0, v) - \vec{\boldsymbol{r}}(u_0, v_0)) \,\| \tag{18}$$

设 $\vec{\boldsymbol{r}}(u_0, v_0)$ 处 u 方向的切向量 (Tangent Vector) 为 $\vec{\boldsymbol{r}}_u(u_0, v_0)$，则该切向量就是 $\vec{\boldsymbol{r}}(u, v)$ 关于 u 的偏导数在 (u_0, v_0) 处的值，即：

$$\vec{\boldsymbol{r}}_u(u_0, v_0) = \left.\frac{\partial x(u, v)}{\partial u}\hat{\boldsymbol{i}} + \frac{\partial y(u, v)}{\partial u}\hat{\boldsymbol{j}}\right|_{u=u_0, v=v_0} \tag{19}$$

类似地，设 $\vec{\boldsymbol{r}}(u_0, v_0)$ 处 v 方向的切向量为 $\vec{\boldsymbol{r}}_v(u_0, v_0)$，则该切向量就是 $\vec{\boldsymbol{r}}(u, v)$ 关于 v 的偏导数在 (u_0, v_0) 处的值，即：

$$\vec{\boldsymbol{r}}_v(u_0, v_0) = \left.\frac{\partial x(u, v)}{\partial v}\hat{\boldsymbol{i}} + \frac{\partial y(u, v)}{\partial v}\hat{\boldsymbol{j}}\right|_{u=u_0, v=v_0} \tag{20}$$

[8]公式的推导，请阅读《机器学习》课程系列之《叉积》一章。

而根据偏导数的定义有：

$$\vec{\boldsymbol{r}}_u(u_0,v_0)=\lim_{\Delta u\to 0}\frac{\vec{\boldsymbol{r}}(u_0+\Delta u,v_0)-\vec{\boldsymbol{r}}(u_0,v_0)}{\Delta u}=\lim_{\Delta u\to 0}\frac{\vec{\boldsymbol{r}}(u,v_0)-\vec{\boldsymbol{r}}(u_0,v_0)}{\Delta u} \tag{21}$$

因而，当 $\Delta u\to 0$ 时，有：

$$\vec{\boldsymbol{r}}(u,v_0)-\vec{\boldsymbol{r}}(u_0,v_0)\approx \Delta u\vec{\boldsymbol{r}}_u(u_0,v_0) \tag{22}$$

同理，当 $\Delta v\to 0$ 时，可得：

$$\vec{\boldsymbol{r}}(u_0,v)-\vec{\boldsymbol{r}}(u_0,v_0)\approx \Delta v\vec{\boldsymbol{r}}_v(u_0,v_0) \tag{23}$$

从而，从公式 (18) 出发，进一步得到：

$$\begin{aligned}\Delta A&\approx \|\left(\vec{\boldsymbol{r}}(u,v_0)-\vec{\boldsymbol{r}}(u_0,v_0)\right)\times\left(\vec{\boldsymbol{r}}(u_0,v)-\vec{\boldsymbol{r}}(u_0,v_0)\right)\|\\&\approx\|\Delta u\vec{\boldsymbol{r}}_u(u_0,v_0)\times\Delta v\vec{\boldsymbol{r}}_v(u_0,v_0)\|\\&=\Delta u\Delta v\|\vec{\boldsymbol{r}}_u(u_0,v_0)\times\vec{\boldsymbol{r}}_v(u_0,v_0)\|\end{aligned} \tag{24}$$

上式中，鉴于 (u_0,v_0) 的任意性，为表达简便，将其从公式中去掉，并将叉积写成行列式的形式：

$$\begin{aligned}\vec{\boldsymbol{r}}_u\times\vec{\boldsymbol{r}}_v&=\begin{vmatrix}\hat{\boldsymbol{i}}&\hat{\boldsymbol{j}}&\hat{\boldsymbol{k}}\\\frac{\partial x(u,v)}{\partial u}&\frac{\partial y(u,v)}{\partial u}&0\\\frac{\partial x(u,v)}{\partial v}&\frac{\partial y(u,v)}{\partial v}&0\end{vmatrix}\\&=\begin{vmatrix}\frac{\partial x(u,v)}{\partial u}&\frac{\partial y(u,v)}{\partial u}\\\frac{\partial x(u,v)}{\partial v}&\frac{\partial y(u,v)}{\partial v}\end{vmatrix}\hat{\boldsymbol{k}}\end{aligned} \tag{25}$$

得到：

$$\|\vec{\boldsymbol{r}}_u\times\vec{\boldsymbol{r}}_v\|=\begin{vmatrix}\frac{\partial x(u,v)}{\partial u}&\frac{\partial y(u,v)}{\partial u}\\\frac{\partial x(u,v)}{\partial v}&\frac{\partial y(u,v)}{\partial v}\end{vmatrix}_+=\left|\frac{\partial(x,y)}{\partial(u,v)}\right|_+ \tag{26}$$

上式就是雅可比行列式的绝对值。

因此，当 $\Delta u\to 0$ 和 $\Delta v\to 0$ 时，从公式 (24) 出发，得到：

$$\Delta A\approx\Delta u\Delta v\|\vec{\boldsymbol{r}}_u\times\vec{\boldsymbol{r}}_v\|\quad\Rightarrow\quad dA=\left|\frac{\partial(x,y)}{\partial(u,v)}\right|_+dudv \tag{27}$$

从而，得到：

$$\begin{aligned}\iint\limits_R f(x,y)dxdy&=\iint\limits_S f\left(x(u,v),y(u,v)\right)dA\\&=\iint\limits_S f\left(x(u,v),y(u,v)\right)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|_+dudv\end{aligned} \tag{28}$$

实际上，前述的极坐标换元法是通用换元法的特例：

$$\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right|_+ = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix}_+ = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix}_+ = r \tag{29}$$

于是，面积微元为：

$$dA = rdrd\theta \tag{30}$$

5 参考文献

1. 同济大学数学教研室。《高等数学》，1978 年 10 月第 1 版。
2. Jacobian matrix and determinant. https://en.wikipedia.org/wiki/Jacobian_matrix_and_determinant.
3. 雅可比矩阵. https://zh.wikipedia.org/wiki/雅可比矩阵.
4. Substitution. https://math.libretexts.org/.
5. Double Integration with Polar Coordinates. https://math.libretexts.org/.
6. Change of Variables in Multiple Integrals (Jacobians). https://math.libretexts.org/.