

### 0.0.1 一元函数的导数与微分（回顾）

我们在本科的高等数学课上已经学过一元函数的导数与微分。我们复述它们，一是为了明确符号的记法，二是为了与后面介绍的多维的情况作对比。

回顾高阶无穷小定义，简述为：若函数  $f(x)$  在点  $x_0$  处有极限  $\lim_{x \rightarrow x_0} f(x) = 0$ ，则称函数  $f(x)$  是当  $x \rightarrow x_0$  时的一个无穷小。若函数  $f(x), g(x)$  是当  $x \rightarrow x_0$  时的无穷小，且  $\lim_{x \rightarrow x_0} f(x)/g(x) = 0$ ，则称  $f(x)$  是当  $x \rightarrow x_0$  时  $g(x)$  的一个高阶无穷小。

**定义 0.1** (一元函数的导数). <sup>[1]定义 2.1.1</sup> 设函数  $y = f(x)$  在点  $x_0$  的某个邻域内有定义，当自变量  $x$  在点  $x_0$  处取得改变量  $\Delta x$ ，且点  $x_0 + \Delta x$  在上述邻域内时，相应地，函数的改变量为

$$\Delta y = f(x_0 + \Delta x) - f(x_0),$$

如果极限

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

存在，称函数  $f(x)$  在点  $x_0$  处可导（或存在导数），上述极限值称为函数  $f$  在点  $x_0$  处的导数，记为

$$f'(x_0), \quad y'|_{x=x_0}, \quad \left. \frac{df}{dx} \right|_{x=x_0}, \quad \text{或} \quad \left. \frac{dy}{dx} \right|_{x=x_0},$$

即

$$f'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

若该极限不存在，称函数  $f(x)$  在  $x_0$  处不可导。

**定义 0.2** (一元函数的微分). <sup>[1]定义 2.5.1</sup> 设函数  $y = f(x)$  在点  $x_0$  的某邻域内有定义，当  $x$  在点  $x_0$  处获得增量  $\Delta x$  时，如果相应的函数增量  $\Delta y = f(x_0 + \Delta x) - f(x_0)$  可以表示为

$$\Delta y = A\Delta x + o(\Delta x),$$

其中，常数  $A$  与  $\Delta x$  无关（仅与  $x_0$  有关），而  $o(\Delta x)$  是较  $\Delta x$  ( $\Delta x \rightarrow 0$ ) 的高阶无穷小，则称函数  $y = f(x)$  在点  $x_0$  处可微（differentiable at  $x_0$ ）。且称  $A\Delta x$  为函数  $y = f(x)$  在点  $x_0$  处对应于自变量增量  $\Delta x$  的微分（differential of  $y = f(x)$  at  $x_0$ ），记作  $dy$ ，即

$$dy = A\Delta x,$$

也常把  $A\Delta x$  ( $A \neq 0$ ) 称为函数增量  $\Delta y$  的线性主部。

在  $\Delta y$  的分解式中，由于其值主要取决于线性主部  $A\Delta x$ ，因此当  $\Delta x \rightarrow 0$  时，我们可以写

$$\Delta y \approx dy = A\Delta x.$$

**定理 0.1.** <sup>[1]定理 2.5.1</sup> 函数  $y = f(x)$  在点  $x_0$  处可微的充分必要条件是：函数  $y = f(x)$  在点  $x_0$  处可导，且当  $y = f(x)$  在点  $x_0$  处可微时，其微分是

$$dy = f'(x_0) \Delta x.$$

证明. 略<sup>[1]p. 103</sup>。 □

关于符号“d”的意义，这里给出一个比高等数学课本<sup>[1]p. 104</sup>更仔细の説明。

由函数微分的定义可知，函数  $f(x)$  在点  $x_0$  处对应于自变量增量  $\Delta x$  的微分  $dy = f'(x_0) \Delta x$ ，实际上是一个由点  $x_0$  引出的函数，记为  $df_{x_0} : \mathbb{R} \rightarrow \mathbb{R}, df_{x_0}(x) = f'(x_0)x, \forall x \in I$ ，其中区域  $I$  是所有满足  $x_0 + x$  在函数可微的  $x_0$  的邻域的所有实数  $x$  的集合。

若视  $x$  本身为一个恒等映射，即  $x(p) = p, \forall p \in \mathbb{R}$ ，则函数  $x$  在点  $p$  处的微分也是一个由点  $p$  引出的函数，且  $dx_p(u) = x'(p)u = 1 \times u, \forall u \in \mathbb{R}$ ，也是一个恒等映射。同时我们看到，函数  $dx_p(u)$  是  $u \rightarrow 0$  的无穷小。

考虑复合映射  $f(x(p))$  在点  $p_0$  处的微分，它是由点  $p_0$  引出的函数，

$$d(f \circ x)_{p_0}(u) = \left. \frac{d(f \circ x)}{dp} \right|_{p=p_0} u = f'(x(p_0))x'(p_0)u = f'(p_0)dx_{p_0}(u)$$

而且函数  $d(f \circ x)_{p_0}(u)$  与  $dx_{p_0}(u)$  是  $u \rightarrow 0$  时的同阶无穷小， $f'(p_0)$  就是这两个无穷小的比值，该结论对所有  $p_0, u \in \mathbb{R}$  都成立，因此若去掉  $p_0$  和  $u$  简记：

$$df(x) = f'dx$$

则

$$f' = \frac{df}{dx}.$$

因此，当我们把导数当作分数来处理时，实际是在上述的意义上视导数为两个微分作为同阶无穷小的比值。

**定义 0.3** (一元向量函数的导数). 设函数  $\mathbf{f} : \mathbb{R} \supset D \rightarrow \mathbb{R}^n$  的极限

$$\lim_{h \rightarrow 0} \frac{\mathbf{f}(t+h) - \mathbf{f}(t)}{h}$$

存在，则称函数  $\mathbf{f}(x)$  在  $x = t$  处可导。该极限是函数  $\mathbf{f}(x)$  在  $x = t$  处的导数，记为

$$\left. \frac{d\mathbf{f}(x)}{dx} \right|_{x=t} \equiv \lim_{h \rightarrow 0} \frac{\mathbf{f}(t+h) - \mathbf{f}(t)}{h}$$

由一元标量函数求导的法则可证有以下一元向量函数求导法则：对任意函数  $\mathbf{f} : \mathbb{R} \supset D \rightarrow \mathbb{R}^n, \mathbf{g} : D \rightarrow \mathbb{R}^n$ ,

- $\frac{d\mathbf{c}}{dt} = 0, \mathbf{c} \in \mathbb{R}^n$  是常向量
- $\frac{d}{dt}(\alpha\mathbf{f} + \beta\mathbf{g}) = \alpha\frac{d\mathbf{f}}{dt} + \beta\frac{d\mathbf{g}}{dt}, \forall \alpha, \beta \in \mathbb{R}$
- $\frac{d}{dt}[u(t)\mathbf{f}(t)] = \frac{du}{dt}\mathbf{f} + u\frac{d\mathbf{f}}{dt}, \forall u: D \rightarrow \mathbb{R}$
- $\frac{d}{dt}(\mathbf{f} \cdot \mathbf{g}) = \frac{d\mathbf{f}}{dt} \cdot \mathbf{g} + \mathbf{f} \cdot \frac{d\mathbf{g}}{dt}$
- $\frac{d}{dt}(\mathbf{f} \times \mathbf{g}) = \frac{d\mathbf{f}}{dt} \times \mathbf{g} + \mathbf{f} \times \frac{d\mathbf{g}}{dt}$

**定义 0.4** (多元标量值函数的偏导数). 给定函数  $f: \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ , 若对某一  $i \in \{1, \dots, n\}$ , 极限

$$\lim_{t \rightarrow 0} \frac{f(\dots, x_i + h, \dots) - f(\dots, x_i, \dots)}{t}$$

当  $x_i = x_{i0}$  时存在, 则称该极限为函数  $f(\mathbf{x})$  对第  $i$  个变量  $x_i$  在  $x_i = x_{i0}$  处的偏导数 (partial derivative), 记为

$$\left. \frac{\partial f(\mathbf{x})}{\partial x_i} \right|_{x_i=x_{i0}} \equiv \lim_{t \rightarrow 0} \frac{f(\dots, x_i + h, \dots) - f(\dots, x_i, \dots)}{t}$$

**定理 0.2.** \* 如果函数  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  的一个二阶混合偏导数  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  在  $\mathbb{R}^2$  的一个开子集  $S$  上处处连续, 则二阶混合偏导数  $\frac{\partial^2 f}{\partial x_j \partial x_i}$  在  $S$  上处处存在, 且  $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ 。

证明. 略<sup>†</sup>

□

**定义 0.5** (向量函数的偏导数). 函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  对第  $i$  个变量在  $x_i = x_{i0}$  处的偏导数定义如下向量

$$\left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_i} \right|_{x_i=x_{i0}} \equiv \begin{pmatrix} \left. \frac{\partial f_1(\mathbf{x})}{\partial x_i} \right|_{x_i=x_{i0}} \\ \vdots \\ \left. \frac{\partial f_m(\mathbf{x})}{\partial x_i} \right|_{x_i=x_{i0}} \end{pmatrix} \in \mathbb{R}^m$$

其中,  $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ 。

## 0.0.2 向量函数的微分和导数

回顾多元标量函数的全微分的定义<sup>[2]</sup> “定义 7.3.1”, p. 19——

**定义 0.6** (多元标量值函数的微分). 若函数  $z = f(x, y)$  在点  $P_0(x_0, y_0)$  的全增量

$$\Delta z = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)$$

\*即高等数学<sup>[2]</sup>p. 16 定理 7.2.1 向  $n$  维的推广。

<sup>†</sup>进一步了解: [Symmetry of second derivatives](#)。

可表示为

$$\Delta z = A\Delta x + B\Delta y + o(\rho)$$

其中  $A, B$  只与点  $(x_0, y_0)$  有关, 而与  $\Delta x, \Delta y$  无关。又  $\rho = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ ,  $o(\rho)$  是当  $\rho \rightarrow 0$  时  $\rho$  的高阶无穷小, 则称函数  $z = f(x, y)$  在点  $P_0(x_0, y_0)$  可微分, 且把  $\Delta z$  的线性主部  $A\Delta x + B\Delta y$  称为函数  $z = f(x, y)$  在点  $P_0(x_0, y_0)$  的全微分 (differential), 记作

$$dz|_{x=x_0, y=y_0} = A\Delta x + B\Delta y \text{ 或 } df(x_0, y_0) = A\Delta x + B\Delta y$$

如果函数  $z = f(x, y)$  在区域  $D$  内每一点都可微, 则称这函数在  $D$  内可微 (differentiable in  $D$ )。

如果函数  $z = f(x, y)$  在点  $P_0(x_0, y_0)$  处可微, 则有

$$\begin{aligned} f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) &= A\Delta x + B\Delta y + o(\rho) \\ \Leftrightarrow o(\rho) &= f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) - (A\Delta x + B\Delta y) \\ \Leftrightarrow \lim_{\rho \rightarrow 0} \frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) - (A\Delta x + B\Delta y)}{\sqrt{\Delta x^2 + \Delta y^2}} &= 0 \\ \Leftrightarrow \lim_{\rho \rightarrow 0} \frac{\Delta z - dz}{\rho} &= 0 \end{aligned}$$

其中  $\rho = \sqrt{\Delta x^2 + \Delta y^2}$ 。上面的极限式是定义0.6的等价定义式。我们其实可以不引入某个高阶无穷小  $o(\rho)$ , 直接用使该极限式成立的  $A, B$  的存在性来定义函数的可微性。下面我们按照把函数微分的定义推广到一般的向量函数。

**定义 0.7** (向量值函数的微分). 若函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在其定义域内某点  $\mathbf{x}_0 \in D$  处, 存在一个线性变换  $\mathbf{L}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  使得对任意  $\mathbf{x}_0$  的邻域  $N(\mathbf{x}_0)$  中的点  $\mathbf{x} \in N(\mathbf{x}_0)$ ,

$$\lim_{\|\mathbf{x} - \mathbf{x}_0\| \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{L}(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \mathbf{0}$$

就称函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处可微分 (differentiable at  $\mathbf{x}_0$ )。自变量增量的线性主部  $\mathbf{L}(\mathbf{x} - \mathbf{x}_0)$  称为函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的微分 (differential of  $\mathbf{f}(\mathbf{x})$  at  $\mathbf{x}_0$ )。如果函数  $\mathbf{f}(\mathbf{x})$  在开集  $S \subseteq D$  内的每一点上都可微分, 则称函数是  $S$  上的可微函数 (differentiable function on  $S$ )。

我们可以根据向量函数的定义来看出, 定义0.7就是定义0.6的推广。延用定义0.7的设定, 设  $\mathbf{f} = (f_1, \dots, f_m)^T$ ,  $\mathbf{x} = (x_1, \dots, x_n)^T$ ,  $\mathbf{x}_0 = (x_{01}, \dots, x_{0n})^T$ , 则函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的全增量是:

$$\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) = \begin{pmatrix} f_1(x_1, \dots, x_n) - f_1(x_{01}, \dots, x_{0n}) \\ \vdots \\ f_m(x_1, \dots, x_n) - f_m(x_{01}, \dots, x_{0n}) \end{pmatrix}$$

可见我们实际考虑的是  $m$  个  $n$  元标量值函数分别在点  $x_{01}, \dots, x_{0n}$  处的全增量。如果在这些点上这些标量值函数分别都可微，则  $\mathbf{f}$  的每个坐标函数的全增量都可按定义0.6写成

$$f_i(x_1, \dots, x_n) - f_i(x_{01}, \dots, x_{0n}) = \sum_{j=1}^n L_{ji}(x_j - x_{0j}) + o(\rho), i = 1, \dots, m$$

其中， $L_{ij}$  是“线性主部”的系数，只与  $(x_{01}, \dots, x_{0n})$  有关， $\rho = \left(\sum_{j=1}^n (x_j - x_{0j})^2\right)^{1/2} = \|\mathbf{x} - \mathbf{x}_0\|$ ， $o(\rho)$  是当  $\rho \rightarrow 0$  时  $\rho$  的高阶无穷小。这等价于如下极限式

$$\lim_{\rho \rightarrow 0} \frac{1}{\|\mathbf{x} - \mathbf{x}_0\|} \begin{pmatrix} f_1(x_1, \dots, x_n) - f_1(x_{01}, \dots, x_{0n}) - \sum_{j=1}^n L_{1j}(x_j - x_{0j}) \\ \vdots \\ f_m(x_1, \dots, x_n) - f_m(x_{01}, \dots, x_{0n}) - \sum_{j=1}^n L_{mj}(x_j - x_{0j}) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

若线性变换  $\mathbf{L}$  在标准基下的矩阵坐标就是  $L_{ij}$ ，则上式等价于定义0.7的极限式。

我们以一个易于作图的例子说明函数微分的几何意义。函数  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  的图像是 3 维欧几里得空间中的一个曲面（如图0.1）。 $f$  在点  $\mathbf{x}_0$  处的微分  $(d_{\mathbf{x}=\mathbf{x}_0} f)(\mathbf{x})$  定义了该曲面在  $\mathbf{x} = \mathbf{x}_0$  处的切平面。随着  $\mathbf{x}$  接近  $\mathbf{x}_0$ ，函数  $f$  的全增量  $f(\mathbf{x}) - f(\mathbf{x}_0)$  将比  $\mathbf{x}$  接近  $\mathbf{x}_0$  更快地接近切平面。

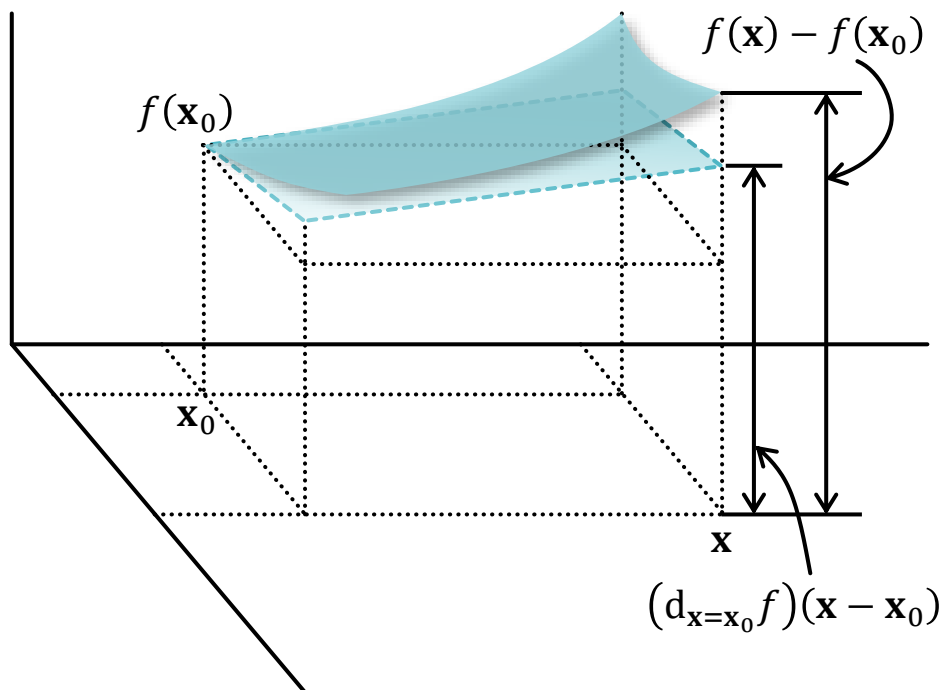


图 0.1: 向量函数的极限（定义??）的图像化理解。

我们接下来考察函数可微与可导的关系，即函数可微的充分和必要条件\*。首先，以下定

\*这部分内容可以与本科高等数学课上的相应内容对比学习，特别是关于必要非充份条件和充份非必要条件的例子（“二、全微分存在的条件”<sup>[2]p. 20。</sup>）

理是函数在某点处可微分的必要条件，它同时也给出了  $\mathbf{L}$  或其坐标矩阵  $L_{ij}$  的计算方法。

**定理 0.3.** 设函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在点  $\mathbf{x}_0 \in D$  处可微分，即存在线性变换  $\mathbf{L} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  满足

$$\lim_{\|\Delta \mathbf{x}\| \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + \Delta \mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{L}\Delta \mathbf{x}}{\|\Delta \mathbf{x}\|} = \mathbf{0}$$

则  $\mathbf{f}$  的每个坐标函数在  $\mathbf{x}_0$  处的每个偏导数

$$\left. \frac{\partial f_i(\mathbf{x})}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x}_0}, i = 1, \dots, m, j = 1, \dots, n$$

都存在。若  $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n\}, \{\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_m\}$  分别是  $\mathbb{R}^n, \mathbb{R}^m$  的标准基，则

$$\mathbf{L}\hat{\mathbf{e}}_j = \sum_{i=1}^m \left( \left. \frac{\partial f_i(\mathbf{x})}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x}_0} \right) \hat{\mathbf{u}}_i, j = 1, \dots, n$$

证明. 见附录。 □

注意到  $\mathbf{L}\hat{\mathbf{e}}_j$  其实就是线性变换的第  $j$  列，故上述定理说明，如果函数在某点处可微分，则其微分的线性变换  $\mathbf{L}$  就是函数的偏导数所形成的矩阵：

$$(\mathbf{L}) = \begin{pmatrix} \left. \frac{\partial f_1(\mathbf{x})}{\partial x_1} \right|_{\mathbf{x}=\mathbf{x}_0} & \dots & \left. \frac{\partial f_1(\mathbf{x})}{\partial x_n} \right|_{\mathbf{x}=\mathbf{x}_0} \\ \vdots & \ddots & \vdots \\ \left. \frac{\partial f_m(\mathbf{x})}{\partial x_1} \right|_{\mathbf{x}=\mathbf{x}_0} & \dots & \left. \frac{\partial f_m(\mathbf{x})}{\partial x_n} \right|_{\mathbf{x}=\mathbf{x}_0} \end{pmatrix}$$

以下定理解决了函数微分的线性变换  $\mathbf{L}$  的唯一性。

**定理 0.4.** 设函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在点  $\mathbf{x}_0 \in D$  处可微分，即存在线性变换  $\mathbf{L} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  满足

$$\lim_{\|\Delta \mathbf{x}\| \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + \Delta \mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{L}\Delta \mathbf{x}}{\|\Delta \mathbf{x}\|} = \mathbf{0}$$

则  $\mathbf{L}$  是唯一的。

证明. 见附录。 □

定理0.3和0.4共同构成了函数可微分的必要非充份条件。也就是说，并非每当函数在某点处的所有偏导数都存在，该函数就一定在该点可微<sup>[2]</sup>“例 2” .p. 21。不过，有了唯一性，我们至少可以把函数微分的线性变换  $\mathbf{L}$  定义为函数的导数，具体地——

**定义 0.8** (向量函数的导数). 设函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在点  $\mathbf{x}_0 \in D$  处可微分, 即存在线性变换  $\mathbf{L} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  满足

$$\lim_{\|\Delta \mathbf{x}\| \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + \Delta \mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{L}\Delta \mathbf{x}}{\|\Delta \mathbf{x}\|} = \mathbf{0}$$

则记  $\mathbf{L}$  为

$$\mathbf{L} \equiv \mathbf{Df}(\mathbf{x}_0)$$

并称为函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的导数 (*derivative of  $\mathbf{f}(\mathbf{x})$  at  $\mathbf{x}_0$* )。  $\mathbf{Df}(\mathbf{x}_0)$  标准基下的坐标矩阵称为函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的雅可比矩阵 (*Jacobian matrix*); 它的行列式称为函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的雅可比行列式 (*Jacobian determinant*)。

下面我们给出一个函数在某点处可微的充分非必要条件 (即未必一定要满足该条件函数才可微, 但满足该条件函数必可微)<sup>[2]</sup> “例 3”, p. 23。

**定理 0.5.** 若函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  的定义域  $D$  是开集, 偏微分  $\frac{\partial f_i}{\partial x_j}, i = 1, \dots, n, j = 1, \dots, m$  在  $D$  内都连续, 则  $\mathbf{f}$  在  $D$  内均可微分。

证明. 见附录。 □

我们不把  $\mathbf{L}$  记为  $\frac{d\mathbf{f}}{d\mathbf{x}}$ , 因为线性变换并非“两个向量的商”。

### 0.0.3 向量函数的导函数、连续可微函数

**定义 0.9** (向量函数的导函数). 设函数  $\mathbf{f}: \mathbb{R}^n \subseteq D \rightarrow \mathbb{R}^m$  是开集  $S \subseteq D$  上的可微函数, 则函数  $\mathbf{f}$  在  $S$  上的导函数 (*derivative function*) 是一个线性变换值函数  $\mathbf{L}: S \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ,

$$\mathbf{L}(\mathbf{x}) \equiv \mathbf{Df}(\mathbf{x}), \quad \forall \mathbf{x} \in S$$

注意, 向量函数的导函数的函数值是线性变换。向量函数在不同点上的导数是不同的线性变换。在上述定义中的“ $\mathbf{L}(\mathbf{x})$ ”, 不是指一个线性变换作用于一个向量, 而是一个线性变换值函数及其自变量。此时  $\mathbf{L}(\mathbf{x})$  可以作用于一个向量  $\mathbf{u}$  得到另一个向量  $\mathbf{v} = \mathbf{L}(\mathbf{x})\mathbf{u}$ 。

接下来我们准备讨论导函数的连续性, 这需要先引入“线性变换的范”的概念。

由定理??, 对任一线性变换  $\mathbf{L}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  皆存在一个正实数  $k > 0$  满足  $\|\mathbf{L}\mathbf{x}\| \leq k\|\mathbf{x}\|, \forall \mathbf{x} \in \mathbb{R}^n$ 。于是我们定义——

**定义 0.10** (线性变换的范). 设  $\mathcal{V}, \mathcal{W}$  是数域  $\mathbb{F}$  上的赋范向量空间, 线性变换  $\mathbf{L}: \mathcal{V} \rightarrow \mathcal{W}$  的范  $\|\mathbf{L}\|$  为满足  $\|\mathbf{L}\mathbf{x}\| \leq k\|\mathbf{x}\|, \forall \mathbf{x} \in \mathbb{R}^n$  的正实数  $k$  的最大下界 (*infimum*), 即

$$\|\mathbf{L}\| \equiv \inf \{k | k > 0, \|\mathbf{L}\mathbf{x}\| \leq k\|\mathbf{x}\| \forall \mathbf{x}\}$$

这一最大下界的存在性和唯一性是显然的。我们无需知道它具体数值。可验证，以上定义的范符合范的一般要求。又由于不同的范的定义是等价的（定理??），故后续命题的证明过程每当需要线性变换的范的定义时都不妨通过上述这种范的定义来求证。由此定义我们可立即获得性质： $\|\mathbf{L}\mathbf{x}\| \leq \|\mathbf{L}\| \|\mathbf{x}\|$ 。

有了向量函数的导函数的定义以及线性变换的范的定义，我们可以很容易理解导函数的连续性是什么意思。按照函数连续性的定义，若函数  $\mathbf{L}: S \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$  在  $S$  上连续，则对任意  $\varepsilon > 0$  总存在  $\delta > 0$  使得只要  $\|\mathbf{x} - \mathbf{x}_0\| < \delta, \mathbf{x}, \mathbf{x}_0 \in S$  就有  $\|\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_0)\| < \varepsilon$ 。

下面介绍函数“连续可微”的概念，它大致上说的是：导函数连续，则函数连续可微。准确定义如下。

**定义 0.11** (连续可微函数). 若函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在开集  $S \subseteq D$  上有导函数  $\mathbf{L}: S \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ，且其为  $S$  上的连续函数，则称函数  $\mathbf{f}(\mathbf{x})$  是  $S$  上的连续可微（*continuously differentiable*）函数。若函数  $\mathbf{f}$  的前  $k$  阶导数在  $S$  上都存在并连续，则称函数  $\mathbf{f}$  属于  $D$  上的  $k$  阶连续可微函数的集合  $C^k(D)$ 。

**推论 0.5.1.** 函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在开集  $D$  上连续可微当且仅当函数  $\mathbf{f}$  在  $D$  上的偏导数

$$\left. \frac{\partial f_i(\mathbf{x})}{\partial x_j} \right|_{\mathbf{x} \in D}, i = 1, \dots, m, j = 1, \dots, n$$

都存在且连续。

结合这一推论和定理0.5可知函数连续可微是函数可微的充份非必要条件。

## 0.0.4 方向导数与梯度

刚才在介绍雅可比矩阵的时候，我们利用全微分的性质考虑过以下极限：

$$\lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_0)}{t}, i = 1, \dots, n$$

并知道它就是  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0 = (x_{01}, \dots, x_{0n})^\top$  处对  $x_{0i}$  的偏导数，因为上式  $\Leftrightarrow$

$$\lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\mathbf{e}_i) - \mathbf{f}(\mathbf{x}_0)}{t}, i = 1, \dots, n$$

现在我们把  $\mathbf{e}_j$  改为任意单位向量  $\hat{\mathbf{n}}$ ，引入方向导数的定义。

**定义 0.12** (方向导数). 设函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在点  $\mathbf{x}_0 \in D$  的某邻域  $N$  有定义，给定单位向量  $\hat{\mathbf{n}} \in \mathbb{R}^n$ ，若极限

$$\lim_{t \rightarrow 0^+} \frac{\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{n}}) - \mathbf{f}(\mathbf{x}_0)}{t}$$



存在，则称此极限值为函数  $\mathbf{f}(\mathbf{x})$  在点  $\mathbf{x}_0$  处沿方向  $\hat{\mathbf{n}}$  的方向导数 (*directional derivative*)，记作

$$\left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \hat{\mathbf{n}}} \right|_{\mathbf{x}=\mathbf{x}_0}$$

以下定理使得函数的方向导数可用该函数的导数来计算。

**定理 0.6.** 若函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在点  $\mathbf{x}_0 \in D$  的某邻域  $N$  可微分，则给定任一单位向量  $\hat{\mathbf{n}} \in \mathbb{R}^n$ ，就有

$$\left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \hat{\mathbf{n}}} \right|_{\mathbf{x}=\mathbf{x}_0} = (\mathbf{Df}(\mathbf{x}_0)) \hat{\mathbf{n}}$$

证明. 由函数  $\mathbf{f}$  的可微条件，自变量的增量  $t\hat{\mathbf{n}}$  造成的函数增量  $\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{n}}) - \mathbf{f}(\mathbf{x}_0)$  满足

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{n}}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{d}_{\mathbf{x}=\mathbf{x}_0} \mathbf{f}(\mathbf{x})(t\hat{\mathbf{n}})}{\|t\hat{\mathbf{n}}\|} &= \mathbf{0} \\ \Leftrightarrow \lim_{t \rightarrow 0} \frac{1}{\|t\hat{\mathbf{n}}\|} \left\| \frac{\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{n}}) - \mathbf{f}(\mathbf{x}_0)}{t} - \mathbf{d}_{\mathbf{x}=\mathbf{x}_0} \mathbf{f}(\mathbf{x}) \hat{\mathbf{n}} \right\| &= 0 \\ \Leftrightarrow \lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{n}}) - \mathbf{f}(\mathbf{x}_0)}{t} &= (\mathbf{Df}(\mathbf{x}_0)) \hat{\mathbf{n}} \end{aligned}$$

□

我们比较一个函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的微分及其在  $\mathbf{x}_0$  处关于  $\hat{\mathbf{n}}$  的方向导数可以发现，后者就是令前者中的  $\mathbf{x} - \mathbf{x}_0 = \hat{\mathbf{n}}$ 。注意到，函数在某点处关于标准基向量的方向导数就是函数关于该点相应分量的偏导数。

方向导数的几何意义是函数在某点处朝相应方向的变化率。定理0.6表明，拿一个函数在某点处的导数作用于一个单位向量，就可以得该函数在该点处朝该方向的变化率。这也是函数的导数的重要几何意义。

**定义 0.13** (向量函数的梯度). 若函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在点  $\mathbf{x}_0 \in D$  处的导数  $\mathbf{Df}(\mathbf{x}_0)$  存在，则记

$$\nabla \mathbf{f}(\mathbf{x}_0) \equiv (\mathbf{Df}(\mathbf{x}_0))^T$$

称为函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的梯度 (*gradient*)。

由线性变换的转置的定义，若视  $\mathbf{Df}(\mathbf{x}_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ，则  $\nabla \mathbf{f}(\mathbf{x}_0) \in \mathcal{L}(\mathbb{R}^{m*}, \mathbb{R}^{n*})$ 。由有限维里斯表示定理（定理??），必存在一组  $\mathbb{R}^m$  的规范正交基  $\{\hat{\mathbf{e}}_i\}_{i=1}^m$ ，使得函数  $\mathbf{f}(\mathbf{x})$  在点  $\mathbf{x}_0$  的导数和梯度满足关系

$$(\hat{\mathbf{e}}_i | (\mathbf{Df}(\mathbf{x}_0))(\mathbf{x} - \mathbf{x}_0)) = (\nabla \mathbf{f}(\mathbf{x}_0) \hat{\mathbf{e}}_i | \mathbf{x} - \mathbf{x}_0), \quad i = 1, \dots, m$$

其中  $\mathbf{x} \in N(\mathbf{x}_0)$ ,  $N(\mathbf{x}_0)$  是  $\mathbf{x}_0$  某邻域。

若用梯度来表示定理0.6, 则函数  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x}_0$  处的关于  $\hat{\mathbf{n}}$  的方向导数又可表示为

$$\left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \hat{\mathbf{n}}} \right|_{\mathbf{x}=\mathbf{x}_0} = (\nabla \mathbf{f}(\mathbf{x}_0))^T \hat{\mathbf{n}}$$

### 0.0.5 复合函数求导的链式法则、反函数定理、隐函数定理

**定理 0.7** (复合函数求导的链式法则). 如果函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  在  $\mathbf{x}_0 \in D$  处可微分; 函数  $\mathbf{g}: \mathbb{R}^m \supset E \rightarrow \mathbb{R}^p$  在  $\mathbf{f}(\mathbf{x}_0) \in E \cap \text{ran} \mathbf{f}$  处可微分, 则复合函数  $\mathbf{g} \circ \mathbf{f}$  在  $\mathbf{x}_0$  处可微分, 且其导数

$$D\mathbf{g} \circ \mathbf{f}(\mathbf{x}_0) = D\mathbf{g}(\mathbf{f}(\mathbf{x}_0)) D\mathbf{f}(\mathbf{x}_0)$$

证明. 见附录。 □

**定理 0.8** (反函数定理). 设函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^n$  在  $D$  上是连续可微函数, 且其在  $\mathbf{x}_0 \in D$  处的导数  $D\mathbf{f}(\mathbf{x}_0)$  是可逆线性变换, 则在  $\mathbf{x}_0$  的某邻域  $N$  上, 函数  $\mathbf{f}$  有连续可微的反函数  $\mathbf{f}^{-1}$ ;  $N$  在函数  $\mathbf{f}$  下的像集  $\mathbf{f}(N)$  是开集; 且  $\mathbf{f}^{-1}(\mathbf{y})$  在  $\mathbf{y} = \mathbf{f}(\mathbf{x}_0)$  处的导数是  $\mathbf{f}(\mathbf{x})$  在  $\mathbf{x} = \mathbf{x}_0$  的导数的逆变换, 即

$$D\mathbf{f}^{-1}(\mathbf{f}(\mathbf{x}_0)) = (D\mathbf{f}(\mathbf{x}_0))^{-1}$$

证明. 见附录 □

定理0.8的证明过程依赖  $D\mathbf{f}(\mathbf{x}_0)$  是  $\mathbb{R}^n$  上的双射的事实。诚然, 在定理0.8中所讨论的函数  $\mathbf{f}$  的定义域和陪域是维数相同的。因此  $D\mathbf{f}(\mathbf{x}_0)$  若存在逆变换 (单射), 则其必为双射。若将定理0.8推广至关于函数  $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 则当  $n < m$  时定理需明确要求  $D\mathbf{f}(\mathbf{x}_0)$  是双射才能成立, 且仅有  $\mathbf{f}$  存在反函数  $\mathbf{f}^{-1}$  的结论;  $\mathbf{f}^{-1}$  未必连续可微 (读者可验证  $f(x) = x^3$  的情况)。作为推论正式表示如下。

**推论 0.8.1.** 设函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$  ( $n < m$ ) 在  $D$  上连续可微, 且其在  $\mathbf{x}_0 \in D$  处的导数  $D\mathbf{f}(\mathbf{x}_0)$  是双射线性变换, 则在  $\mathbf{x}_0$  某邻域  $N$  上函数  $\mathbf{f}$  存在反函数  $\mathbf{f}^{-1}$ 。

证明. 仅列出证明概要。设  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  是  $\mathbb{R}^m$  上的一组有线性无关向量组, 将其补全为  $\mathbb{R}^m$  的一组基  $\{\mathbf{e}_1, \dots, \mathbf{e}_n, \mathbf{e}_{n+1}, \dots, \mathbf{e}_m\}$ 。定义函数  $\mathbf{g}: \mathbb{R}^m \rightarrow \mathbb{R}^n$ ,

$$\mathbf{g}\left(\sum_{i=1}^m a_i \mathbf{e}_i\right) = (a_1, \dots, a_n)$$

即取  $\mathbb{R}^m$  上任一向量在有序基  $\{\mathbf{e}_i\}$  下的前  $n$  个坐标的函数。证明：复合函数  $\mathbf{g} \circ \mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^n$  满足定理0.8的条件。

□

当  $n > m$  时，由线性变换的维数定理  $D\mathbf{f}(\mathbf{x}_0)$  必不满秩，故  $\mathbf{f}$  不存在反函数。

反函数定理的逆命题一般不成立（仍可通过考虑  $f(x) = x^3$  理解）。以下推论成立。

**推论 0.8.2.** 设函数  $\mathbf{f}: \mathbb{R}^n \supset D \rightarrow \mathbb{R}^n$  在  $\mathbf{x}_0 \in D$  处的导数  $D\mathbf{f}(\mathbf{x}_0)$  存在。若函数  $\mathbf{f}$  在  $\mathbf{x}_0$  的某邻域  $N$  存在反函数  $\mathbf{f}^{-1}$  且其在  $N$  上可微，则  $D\mathbf{f}(\mathbf{x}_0)$  是可逆线性变换。

证明. 由于函数  $\mathbf{f}$  和  $\mathbf{f}^{-1}$  都可微，可应用定理0.7，由  $d_{\mathbf{x}=\mathbf{x}_0} \mathbf{f} \circ \mathbf{f}^{-1}(\mathbf{x}_0) = \mathbf{I}$  自然得出结论。 □

如果函数  $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  由函数  $\mathbf{F}: \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$  隐含定义，我们常常把  $\mathbf{F}$  写成这样一种映射： $\mathbf{F}: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ ， $\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}, \forall \mathbf{x} \in \mathbb{R}^n$ 。此时，我们记

$$\left. \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_0, \mathbf{y}=\mathbf{y}_0} = D\mathbf{F}(\mathbf{x}_0; \mathbf{y}_0)$$

引入这个记法有两个用处，一是为了以下例子，这个例子是后文引入物质导数的一个基础；二是为了引入隐函数定理。

**例 0.1.** 设函数  $\mathbf{f}: \mathbb{R}^{n+m} \rightarrow \mathbb{R}^p$  的定义域为  $D = \text{dom} \mathbf{f}$ ，若函数  $\mathbf{g}: \mathbb{R}^m \rightarrow \mathbb{R}^n$  满足  $\text{dom} \mathbf{g} = D$ ，则总可以构建函数  $\mathbf{h}: \mathbb{R}^m \rightarrow \mathbb{R}^{n+m}$  使得  $\text{dom} \mathbf{h} = D$  且

$$\begin{aligned} \mathbf{h}(\mathbf{x}) &= (\mathbf{g}(\mathbf{x}), \mathbf{x}) \forall \mathbf{x} \in D \\ h_i(\mathbf{x}) &= \begin{cases} g_i(\mathbf{x}), & i = 1, \dots, n \\ x_{i-n}, & i = n+1, \dots, n+m \end{cases} \end{aligned}$$

这时， $\mathbf{f}(\mathbf{g}(\mathbf{x}), \mathbf{x}) = \mathbf{f}(\mathbf{h}(\mathbf{x}))$ 。若  $\mathbf{g}$  在某处可微分则  $\mathbf{h}$  在该处也可微分。由链式法则定理，在该处有

$$\begin{aligned} D\mathbf{f}(\mathbf{g}(\mathbf{x}), \mathbf{x}) &= D\mathbf{f}(\mathbf{h}(\mathbf{x})) D\mathbf{h}(\mathbf{x}) \\ &= \begin{pmatrix} \frac{\partial f_1}{\partial h_1} & \dots & \frac{\partial f_1}{\partial h_n} & \frac{\partial f_1}{\partial h_{n+1}} & \dots & \frac{\partial f_1}{\partial h_{n+m}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial h_1} & \dots & \frac{\partial f_p}{\partial h_n} & \frac{\partial f_p}{\partial h_{n+1}} & \dots & \frac{\partial f_p}{\partial h_{n+m}} \end{pmatrix} \begin{pmatrix} \frac{\partial h_1}{\partial y_1} & \dots & \frac{\partial h_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \dots & \frac{\partial h_n}{\partial y_m} \\ \frac{\partial h_{n+1}}{\partial y_1} & \dots & \frac{\partial h_{n+1}}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_{n+m}}{\partial y_1} & \dots & \frac{\partial h_{n+m}}{\partial y_m} \end{pmatrix} = C \end{aligned}$$

其中矩阵  $C$  的分量

$$C_{ij} = \sum_{k=1}^{n+m} \frac{\partial f_i}{\partial h_k} \frac{\partial h_k}{\partial x_j}, i = 1, \dots, p, j = 1, \dots, m$$

注意到

$$\frac{\partial f_i}{\partial h_k} \frac{\partial h_k}{\partial x_j} = \begin{cases} \frac{\partial f_i}{\partial g_k} \frac{\partial g_k}{\partial x_j}, & k = 1, \dots, n \\ \frac{\partial f_i}{\partial x_{k-n}} \delta_{k-n,j}, & k = n+1, \dots, n+m \end{cases}, i = 1, \dots, p, j = 1, \dots, m$$

故

$$\begin{aligned} C_{ij} &= \sum_{k=1}^n \frac{\partial f_i}{\partial g_k} \frac{\partial g_k}{\partial x_j} + \sum_{j=1}^m \frac{\partial f_i}{\partial x_j} \\ \Leftrightarrow \text{Df}(\mathbf{g}(\mathbf{x}), \mathbf{x}) &= \frac{\partial \mathbf{f}(\mathbf{y}, \mathbf{x})}{\partial \mathbf{y}} \Big|_{\mathbf{y}=\mathbf{g}(\mathbf{x})} \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \mathbf{f}(\mathbf{y}, \mathbf{x})}{\partial \mathbf{x}} \end{aligned}$$

**定理 0.9** (隐函数定理). 设  $\mathbf{F}: \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$  是连续可导函数, 且对  $\mathbf{x}_0 \in \mathbb{R}^n$  和  $\mathbf{y}_0 \in \mathbb{R}^m$  有

- $\mathbf{F}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ ;
- 导数  $\frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \Big|_{\mathbf{x}=\mathbf{x}_0, \mathbf{y}=\mathbf{y}_0}$  可逆;

则  $\mathbf{x}_0$  的某邻域  $N$  存在由  $\mathbf{F}$  隐函定义的连续可微函数  $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  满足  $\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}, \forall \mathbf{x} \in N$ , 且在  $N$  上有

$$\text{Df}(\mathbf{x}) = - \left[ \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \Big|_{\mathbf{x}=\mathbf{x}, \mathbf{y}=\mathbf{f}(\mathbf{x})} \right]^{-1} \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}, \mathbf{y}=\mathbf{f}(\mathbf{x})}$$

上式中的“ $-1$ ”是指线性变换的逆。

证明. 待补充<sup>[3]p. 593</sup>。

□

**例 0.2.** 设函数  $\mathbf{F}: \mathbb{R}^4 \rightarrow \mathbb{R}^2$ ,  $\mathbf{F}(u, v, x, y) = (F_1, F_2)$ ,  $\mathbf{F} = \mathbf{0}$  隐含定义了  $(x, y) = \mathbf{f}(u, v)$ ,  $\mathbf{f}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , 则

$$\begin{aligned} \frac{\partial \mathbf{F}}{\partial u} &= \mathbf{0}, \quad \frac{\partial \mathbf{F}}{\partial v} = \mathbf{0} \\ \Leftrightarrow \begin{cases} \frac{\partial F_1}{\partial u} + \frac{\partial F_1}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial F_1}{\partial y} \frac{\partial y}{\partial u} = 0 \\ \frac{\partial F_2}{\partial u} + \frac{\partial F_2}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial F_2}{\partial y} \frac{\partial y}{\partial u} = 0 \end{cases}, & \begin{cases} \frac{\partial F_1}{\partial v} + \frac{\partial F_1}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial F_1}{\partial y} \frac{\partial y}{\partial v} = 0 \\ \frac{\partial F_2}{\partial v} + \frac{\partial F_2}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial F_2}{\partial y} \frac{\partial y}{\partial v} = 0 \end{cases} \\ \Leftrightarrow \begin{pmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{pmatrix} + \begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} &= \mathbf{0} \\ \Leftrightarrow \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} &= - \begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{pmatrix} \end{aligned}$$