Multivariable Chain Rule, Total Differentials, and Implicit Differentiation

N-20260313

§1.1 The chain rule is composition of linear maps

The chain rule is not a collection of unrelated formulas. It says that first-order approximations compose.

Theorem 1.1.1 (Chain rule)

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at x, and let $G : \mathbb{R}^m \to \mathbb{R}^k$ be differentiable at $F(x)$. Then $G \circ F$ is differentiable at x, and

$$D(G \circ F)(x) = DG(F(x))\, DF(x).$$

Proof skeleton. Write

$$F(x+h) = F(x) + DF(x)h + r(h), \qquad \|r(h)\| = o(\|h\|).$$

Then apply the differentiability expansion for G at $F(x)$:

$$G(F(x)+u) = G(F(x)) + DG(F(x))u + s(u), \qquad \|s(u)\| = o(\|u\|).$$

Substitute $u = DF(x)h + r(h)$. Since $u = O(\|h\|)$, the remainder is still $o(\|h\|)$. The linear part is $DG(F(x))DF(x)h$. $\square$

This theorem is the source of all ordinary partial-derivative chain rules.

§1.2 One input variable, two intermediate variables

Suppose

$$z = f(u,v), \qquad u = u(x), \qquad v = v(x).$$

Then

$$\frac{dz}{dx} = f_u \frac{du}{dx} + f_v \frac{dv}{dx}.$$

In matrix form, this is

$$\frac{dz}{dx} = \begin{pmatrix} f_u & f_v \end{pmatrix} \begin{pmatrix} u_x \\ v_x \end{pmatrix}.$$

The row vector is Df, the column vector is the derivative of (u,v), and their product is the derivative of the composite.

Example 1.2.1

Let $z = u^2 + uv$, $u = x^2$, and $v = \sin x$. Then

$$z_u = 2u + v, \qquad z_v = u, \qquad u' = 2x, \qquad v' = \cos x.$$

Hence

$$\frac{dz}{dx} = (2u + v)2x + u \cos x.$$

Substituting $u = x^2$, $v = \sin x$ gives

$$\frac{dz}{dx} = 2x(2x^2 + \sin x) + x^2 \cos x.$$

§1.3 Two input variables

Suppose

$$z = f(u, v), \qquad u = u(x, y), \qquad v = v(x, y).$$

Then

$$\begin{pmatrix} z_x & z_y \end{pmatrix} = \begin{pmatrix} f_u & f_v \end{pmatrix} \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}.$$

Equivalently,

$$z_x = f_u u_x + f_v v_x, \qquad z_y = f_u u_y + f_v v_y.$$

Example 1.3.1

Let $z = e^u \cos v$, where $u = x^2 + y^2$ and $v = x - y$. Then

$$f_u = e^u \cos v, \qquad f_v = -e^u \sin v,$$

and

$$\begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} = \begin{pmatrix} 2x & 2y \\ 1 & -1 \end{pmatrix}.$$

Thus

$$z_x = e^u \cos v \cdot 2x - e^u \sin v, \qquad z_y = e^u \cos v \cdot 2y + e^u \sin v.$$

§1.4 Total differentials

The notation

$$dz = f_x\, dx + f_y\, dy$$

is best read as a compact way to write a linear approximation. If (x, y) changes by (dx, dy), then the first-order change in $z = f(x, y)$ is

$$dz \approx f_x\, dx + f_y\, dy.$$

For a vector-valued map F, the same statement is

$$dF = J_F\, dx.$$

The differential notation is safe when one remembers that it represents a linear map, not ordinary fraction arithmetic.

§1.5 Implicit differentiation in one equation

Suppose a relation

$$F(x, y) = 0$$

defines y locally as a differentiable function of x. Differentiating the identity $F(x, y(x)) = 0$ gives

$$F_x + F_y \frac{dy}{dx} = 0.$$

If $F_y \neq 0$, then

$$\frac{dy}{dx} = -\frac{F_x}{F_y}.$$

Example 1.5.1

For the circle $F(x, y) = x^2 + y^2 - 1 = 0$, where $y \neq 0$,

$$\frac{dy}{dx} = -\frac{2x}{2y} = -\frac{x}{y}.$$

This is the slope of the tangent line to the circle.

§1.6 Implicit differentiation for systems

The same idea becomes matrix algebra. Suppose

$$F(x, u) = 0, \qquad F : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^q,$$

and u is locally a function of x. Differentiating gives

$$F_x \, dx + F_u \, du = 0.$$

If F_u is invertible, then

$$du = -F_u^{-1} F_x \, dx,$$

so

$$Du(x) = -F_u^{-1} F_x.$$

Theorem 1.6.1 (Implicit function theorem, working form)

Let $F : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^q$ be C^1, and suppose

$$F(a, b) = 0, \qquad \det F_u(a, b) \neq 0.$$

Then near (a, b), the equation $F(x, u) = 0$ determines u as a C^1 function of x, and

$$Du(a) = -F_u(a, b)^{-1} F_x(a, b).$$

The theorem says when the formal linearized computation is legitimate. The condition $\det F_u \neq 0$ means the constraint can be solved in the u-directions.

§1.7 A shape table

Keeping track of shapes prevents many mistakes:

map	derivative shape	meaning
$f : \mathbb{R}^n \to \mathbb{R}$	$1 \times n$	row gradient
$F : \mathbb{R}^n \to \mathbb{R}^m$	$m \times n$	Jacobian
$G \circ F$	$k \times n$	$(k \times m)(m \times n)$
$F(x, u) = 0$	$q \times p,\ q \times q$	F_x, F_u

The formulas work because the linear maps have compatible domains and codomains.

§1.8 What to carry forward

The chain rule says that local linear approximations compose. Total differentials are the notation for those approximations. Implicit differentiation says that a constraint can be solved by inverting the linearized constraint in the unknown directions.