قاعدهٔ زنجیره‌ای (۱)

یکی از اساسی‌ترین قضیه‌های حساب دیفرانسیل قاعدهٔ زنجیره‌ای است که مشتق ترکیب دو تابع را برحسب مشتق‌های دو تابع بیان می‌کند. در این جلسه این قضیه را بیان و ثابت می‌کنیم.

(۲۹—۱) قاعدهٔ زنجیره‌ای. فرض کنید $S \subset \mathbb{R}^n$، $T \subset \mathbb{R}^m$، a یک نقطهٔ درونی S و b یک نقطهٔ درونی T است. تابع‌های $f : S \longrightarrow \mathbb{R}^m$ و $g : T \longrightarrow \mathbb{R}^p$ داده شده‌اند به طوری که $f(a) = b$، f در a مشتق‌پذیر و g در b مشتق‌پذیر است. در این صورت $g \circ f$ در a مشتق‌پذیر است و

$$(D(g \circ f))(a) = Dg(f(a)) \circ Df(a) \qquad (۱)$$

تعدادی توضیح در مورد صورت قضیه مفید است.

۱) $Df(a)$ یک تابع خطی از $\mathbb{R}^n$ به $\mathbb{R}^m$، $Dg(f(a))$ یک تابع خطی از $\mathbb{R}^m$ به $\mathbb{R}^p$، و $(D(g \circ f))(a)$ یک تابع خطی از $\mathbb{R}^n$ به $\mathbb{R}^p$ است. در نمایش ماتریسی، $Df(a)$ یک ماتریس $m \times n$، $Dg(f(a))$ یک ماتریس $p \times m$، و $(D(g \circ f))(a)$ یک ماتریس $p \times n$ است، و ترکیب تابع‌های خطی به صورت حاصل‌ضرب ماتریس‌های مربوط ظاهر می‌شود.

۲) در حالت $m = n =$ ۱ که هر دو تابع یک متغیری و دارای مقدار حقیقی هستند هر یک از ماتریس‌های فوق یک ماتریس ۱ $\times$ ۱ است، $Df(a) : [f'(a)]$، $Dg(f(a)) : [g'(f(a))]$، و

$[(g \circ f)'(a)] = (D(g \circ f))(a)$، پس (۱) را می‌توان به صورت زیر نوشت:

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a) \qquad (۲)$$

که صورت مأنوس قاعدهٔ زنجیره‌ای یک متغیری است.

۳) در حالت کلی، نقاط $\mathbb{R}^n$ را به $x = (x_۱, \ldots, x_n)$، نقاط $\mathbb{R}^m$ را به $y = (y_۱, \ldots, y_m)$، نقاط $\mathbb{R}^p$ را به $z = (z_۱, \ldots, z_p)$ نمایش می‌دهیم. در این صورت می‌توان نوشت:

$$\left. \begin{array}{l} (y_۱, \ldots, y_m) = f(x_۱, \ldots, x_n) \\ (z_۱, \ldots, z_p) = g(y_۱, \ldots, y_m) \end{array} \right\} \Longrightarrow (z_۱, \ldots, z_p) = (g \circ f)(x_۱, \ldots, x_n)$$

$$Dg(f(a)) : [\frac{\partial z_i}{\partial y_j}(f(a))] \quad , \quad Df(a) : [\frac{\partial y_i}{\partial x_j}(a)] \quad , \quad (D(g \circ f))(a) : [\frac{\partial z_i}{\partial x_j}(a)]$$

پس (۱) به صورت زیر قابل بیان است:

$$\begin{bmatrix} \frac{\partial z_۱}{\partial x_۱}(a) & \cdots & \frac{\partial z_۱}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial z_p}{\partial x_۱}(a) & \cdots & \frac{\partial z_p}{\partial x_n}(a) \end{bmatrix} = \begin{bmatrix} \frac{\partial z_۱}{\partial y_۱}(b) & \cdots & \frac{\partial z_۱}{\partial y_m}(b) \\ \vdots & & \vdots \\ \frac{\partial z_p}{\partial y_۱}(b) & \cdots & \frac{\partial z_p}{\partial y_m}(b) \end{bmatrix} \begin{bmatrix} \frac{\partial y_۱}{\partial x_۱}(a) & \cdots & \frac{\partial y_۱}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_۱}(a) & \cdots & \frac{\partial y_m}{\partial x_n}(a) \end{bmatrix} \qquad (۳)$$

که به طور نمادین و برای به یاد ماندن گاهی به صورت $[\frac{\partial z}{\partial x}] = [\frac{\partial z}{\partial y}][\frac{\partial y}{\partial x}]$ می‌نویسیم. وقتی $m = n = ۱$ فرمول (۳) تبدیل می‌شود به:

$$\frac{dz}{dx}(a) = \frac{dz}{dy}(b) \cdot \frac{dy}{dx}(a) \qquad (۴)$$

که صورت دیگری از (۲) است.

۴) با توجه به مفهوم حاصل‌ضرب دو ماتریس، درایهٔ نمونهٔ $\frac{\partial z_i}{\partial x_j}(a)$ از طرف چپ (۳) به صورت زیر محاسبه می‌شود:

$$\frac{\partial z_i}{\partial x_j}(a) = \frac{\partial z_i}{\partial y_۱}(b) \cdot \frac{\partial y_۱}{\partial x_j}(a) + \cdots + \frac{\partial z_i}{\partial y_m}(b) \cdot \frac{\partial y_m}{\partial x_j}(a) \qquad (۵)$$

گاهی به (۵) نیز قاعدهٔ زنجیره‌ای گفته می‌شود. مجدداً توجه کنید که در حالت $m = n = 1$، (۵) به (۴) تحویل می‌شود. طبق (۵)، برای محاسبهٔ مشتق z_i نسبت به x_j، همهٔ متغیرهای واسطه، یعنی $y_1, \dots, y_m$، می‌توانند دخیل شوند.

برای اثبات قاعدهٔ زنجیره‌ای نیاز به تخمینی در مورد نمو تابع برحسب مشتق داریم:

(۲–۲۹) تخمین نمو در یک نقطه. فرض کنید $S \subset \mathbb{R}^n$، a یک نقطهٔ درونی S است و $f : S \longrightarrow \mathbb{R}^m$ یک تابع که در نقطهٔ a مشتق‌پذیر است. اگر $K > 0$ طوری باشد که $K > |\frac{\partial f_i}{\partial x_j}(a)|$ برای هر $i = 1, \dots, m$ و هر $j = 1, \dots, n$، آنگاه $\rho > 0$ وجود دارد که:

$$|x - a| < \rho \Longrightarrow |f(x) - f(a)| \leq \sqrt{mn} K |x - a| \tag{٦}$$

اگر بنویسیم $y = f(x)$، $x - a$ را به Δx و $f(x) - f(a)$ را به Δy نمایش دهیم، آنگاه (٦) به صورت

$$|\Delta x| < \rho \Longrightarrow |\Delta y| \leq \sqrt{mn} K |\Delta x| \tag{۷}$$

نیز نوشته می‌شود. این نامساوی کرانی بالایی برای نسبت نمو مقدار تابع به نمو متغیر برحسب مشتقات پاره‌ای در نقطهٔ مورد نظر ارائه می‌کند. در حالت $m = n = 1$ به تعبیر شهودی زیر می‌رسیم. فرض کنید $K > |f'(a)|$. اگر دو خط راست گذرا از نقطهٔ $(a, f(a))$ با ضریب زاویهٔ K را در نظر بگیریم، ناحیهٔ مخروطی شکل بین این دو خط به وسیلهٔ نامساوی $|\frac{y - f(a)}{x - a}| \leq K$ مشخص می‌شود. طبق این گزاره، برای x به اندازهٔ کافی نزدیک a، نمودار f در داخل این مخروط قرار می‌گیرد.

اثبات ۲۹—۲. مشتق f در نقطهٔ a، یعنی $Df(a)$، به صورت یک ماتریس $m \times n$، $[\frac{\partial f_i}{\partial x_j}(a)]$ داده می‌شود. می‌نویسیم $x - a = h = (h_1, \dots, h_n)$، بنابراین

$$\begin{aligned} (Df(a))(h) &= (\sum_{j=1}^{n} \frac{\partial f_1}{\partial x_j}(a) \cdot h_j, \dots, \sum_{j=1}^{n} \frac{\partial f_m}{\partial x_j}(a) \cdot h_j) \\ |(Df(a))(h)|^2 &= \sum_{i=1}^{m} (\sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(a) \cdot h_j)^2 \\ &\leq m \cdot \max_{i=1,\dots,m} (\sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(a) \cdot h_j)^2 \end{aligned}$$

از طرفی دیگر توجه کنید که $\sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(a) \cdot h_j$ برابر حاصل‌ضرب داخلی $\nabla f_i(a) \cdot h$ است و طبق نامساوی کوشی‌ـ شوارتس داریم:

$$|\nabla f_i(a) \cdot h|^2 \leq |\nabla f_i(a)|^2 |h|^2$$

ولی:

$$\begin{aligned} |\nabla f_i(a)|^2 &= \sum_{j=1}^{n} (\frac{\partial f_i}{\partial x_j}(a))^2 \\ &\leq n \max_{j=1,\dots,n} |\frac{\partial f_i}{\partial x_j}(a)|^2 \end{aligned}$$

بنابراین نتیجه می‌شود که:

$$|(Df(a))(h)|^2 \leq mn|h|^2 \max_{i,j} |\frac{\partial f_i}{\partial x_j}(a)|^2 \tag{8}$$

حال طبق تعریف مشتق‌پذیری f در نقطهٔ a داریم:

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

کسر بالا را برای $h \neq \underline{0}$ برابر $\varphi(h)$ قرار می‌دهیم. چون حد $\varphi(h)$ وقتی $h \longrightarrow 0$ برابر صفر است، اگر تعریف کنیم $\varphi(\underline{0}) = \underline{0}$، تابع φ در دامنه خود (یعنی به‌ازای $|h|$ کوچک) پیوسته خواهد بود. داریم:

$$f(a+h) - f(a) - (Df(a))(h) = |h|\varphi(h)$$

$$f(a+h) - f(a) = (Df(a))(h) + |h|\varphi(h)$$

چون فرض کرده‌ایم $K > \max_{i,j}|\frac{\partial f_i}{\partial x_j}(a)|$، اگر قرار دهیم $e = \sqrt{mn}(K - \max_{i,j}|\frac{\partial f_i}{\partial x_j}(a)|)$، e عددی اکیداً مثبت است، بنابراین چون $\lim_{h\to ۰}\varphi(h) = \varphi(۰) = ۰$، $\rho > ۰$ وجود دارد که

$$|h| < \rho \Longrightarrow |\varphi(h)| < e$$

بدین ترتیب برای $|h| < \rho$ داریم:

$$\begin{aligned} |f(a+h) - f(a)| &\leq |(Df(a))(h)| + |h||\varphi(h)| \\ &\leq \sqrt{mn}|h|\max_{i,j}|\frac{\partial f_i}{\partial x_j}(a)| + |h|\sqrt{mn}(K - \max_{i,j}|\frac{\partial f_i}{\partial x_j}(a)|) \\ &\leq \sqrt{mn}|h|K \end{aligned}$$

و حکم به اثبات می‌رسد. ■

به کمک تخمین به‌دست آمده اکنون می‌توانیم قاعدهٔ زنجیره‌ای را ثابت کنیم.

اثبات (۲۹—۱). کافی است فرمول (۵) را ثابت کنیم. برای سهولت در تمیز دادن بیان اعداد و n تایی، n تایی $(h_۱, \ldots, h_n)$ را به $\vec{h}$ نمایش می‌دهیم. همچنان که در اثبات گزارهٔ بالا دیدیم، از مشتق‌پذیری f در a نتیجه که تابعی پیوسته φ وجود دارد با $\varphi(۰) = \underline{۰}$ که:

$$f(a+\vec{h}) - f(a) = (Df(a))(\vec{h}) + |\vec{h}|\varphi(\vec{h})$$

حال تابع $g = (g_۱, \ldots, g_m)$ را در نظر می‌گیریم. چون g (و معادلاً هر g_i) در نقطهٔ $b = f(a)$ مشتق‌پذیر است، تابع‌های پیوستهٔ $\sigma_۱, \ldots, \sigma_m$ وجود دارند که $\sigma_i(\underline{۰}) = \underline{۰}$ و:

$$g_i(b+\vec{k}) - g_i(b) = (Dg_i(b))(\vec{k}) + |\vec{k}|\sigma_i(\vec{k}) \quad , \quad \vec{k} = (k_۱, \ldots, k_m)$$

حال $z_i = (g_i \circ f)(x)$ را در نظر می‌گیریم. برای $|\vec{h}|$ کوچک که $a + \vec{h}$ در دامنهٔ f بماند، قرار می‌دهیم $\vec{k} = f(a+\vec{h}) - f(a)$، و داریم

$$\begin{aligned} (g_i \circ f)(a+\vec{h}) - (g_i \circ f)(a) &= g_i(b+\vec{k}) - g_i(b) \\ &= (Dg_i(b))(\vec{k}) + |\vec{k}|\sigma_i(\vec{k}) \end{aligned}$$

بنابراین

$$(g_i \circ f)(a+\vec{h}) - (g_i \circ f)(a) = \sum_{l=1}^{m} \frac{\partial g_i}{\partial y_l}(b) \cdot k_l + |\vec{k}|\sigma_i(\vec{k}) \tag{9}$$

حال می‌خواهیم $\frac{\partial z_i}{\partial x_j}(a)$، یعنی $\lim_{h\to 0} \frac{(g_i\circ f)(a+he_j)-(g_i\circ f)(a)}{h}$ را محاسبه کنیم. با قرار دادن $\vec{h} = he_j$، و با استفاده از این که $k_l = f_l(a+\vec{h}) - f_l(a)$، از (۹) داریم:

$$\frac{(g_i \circ f)(a+he_j) - (g_i \circ f)(a)}{h} = \sum_{l=1}^{m} \frac{\partial g_i}{\partial y_l}(b) \frac{f_l(a+he_j) - f_l(a)}{h} + \frac{|\vec{k}|}{h}\sigma_i(\vec{k}) \tag{10}$$

به حد طرف راست این رابطه وقتی $h \longrightarrow 0$ توجه می‌کنیم. از یک سو داریم:

$$\lim_{h\to 0} \frac{f_l(a+he_j) - f_l(a)}{h} = \frac{\partial y_l}{x_j}(a) \tag{11}$$

از طرفی دیگر، طبق تخمین نمو:

$$|\frac{|\vec{k}|}{h}| = \frac{|\vec{k}|}{|h|} = \frac{|\vec{k}|}{|\vec{h}|} \leq \sqrt{mn}K$$

که K یک عدد دلخواه بزرگتر از $\max_{i,j} |\frac{\partial f_i}{\partial x_j}(a)$ است. بنابراین

$$|\frac{|\vec{k}|}{h}\sigma_i(\vec{k})| \leq \sqrt{mn}K|\sigma_i(\vec{k})|$$

حال تابع f در نقطهٔ a پیوسته است (زیرا در آن نقطه مشتق‌پذیر است)، پس وقتی $\vec{h} \longrightarrow \underline{0}$، داریم $\vec{k} = f(a+\vec{h}) - f(a) \longrightarrow \underline{0}$ و چون $\sigma_i(\underline{0}) = \underline{0}$ و σ_i در $\underline{0}$ پیوسته است نتیجه می‌شود که $|\frac{|\vec{k}|}{h}\sigma_i(\vec{k})| \longrightarrow 0$ وقتی $h \longrightarrow 0$. با توجه به این مطلب، از (۱۰) و (۱۱) نتیجه می‌گیریم که

$$\frac{\partial z_i}{\partial x_j}(a) = \sum_{l=1}^{m} \frac{\partial z_i}{\partial y_l}(b) \frac{\partial y_l}{\partial x_j}(a)$$

و قاعدهٔ زنجیره‌ای به اثبات می‌رسد. ■

در جلسهٔ آینده مثال‌های متنوعی از کاربردهای قاعدهٔ زنجیره‌ای خواهیم دید.