Three important theorems in advanced calculus

Sean Fitzpatrick

June 13, 2018

Source: Marsden, J. E. and Tromba, A. J., *Vector Calculus*, 4th ed. W. H. Freeman and Company, New York, 1996.

We state three theorems of theoretical importance in multivariable calculus: the chain rule, the implicit function theorem, and the inverse function theorem. The first two are mentioned in most Stewart-style texts, but not in their most general form, and the last is not mentioned at all. We'll give a general proof of the chain rule, and state the implicit and inverse function theorems. (The proofs can be found in more advanced texts on real analysis.)

1 The Chain Rule

For a general function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, the derivative $\mathbf{D}f(\mathbf{x}_0)$ at a point $\mathbf{x}_0 \in U$ is the $m \times n$ matrix whose entries are given by the partial derivatives of f. That is, if

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)),$$

where each of the component functions f_i is a real-valued function of the variables $x_1, \ldots, x_n$, then the entry in the i^{th} row and j^{th} column of $\mathbf{D}f(\mathbf{x}_0)$ is $\dfrac{\partial f_i}{\partial x_j}(\mathbf{x}_0)$. The chain rule tells us that the derivative of a composition is given by the product of the derivatives, just as for the case of single-variable functions. First, we need a fact about linear functions:

Lemma 1.1. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function given by $T(\mathbf{x}) = A \cdot \mathbf{x}$, where $A = [a_{ij}]$ is an $m \times n$ matrix. Then T is continuous, and in particular, $\|T(\mathbf{x})\| \leq M\|\mathbf{x}\|$, where $M = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$.*

Proof. The components of T are given by $T_i(\mathbf{x}) = \sum_{j=1}^n a_{ij}x_j = \mathbf{a}_i \cdot \mathbf{x}$, where $\mathbf{a}_i = \langle a_{i1}, \ldots, a_{in}\rangle$. Thus,

$$\begin{aligned}
\|T(\mathbf{x})\| &= \sqrt{(T_1(\mathbf{x})^2 + \cdots + T_m(\mathbf{x})^2} \\
&= \sqrt{|\mathbf{a}_1 \cdot \mathbf{x}|^2 + \cdots + |\mathbf{a}_m \cdot \mathbf{x}|^2} \\
&\leq \sqrt{\|\mathbf{a}_1\|^2\|\mathbf{x}\|^2 + \cdots \|\mathbf{a}_m\|^2\|\mathbf{x}\|^2} \text{ (Cauchy-Schwartz inequality)} \\
&= \sqrt{(\|\mathbf{a}_1\|^2 + \cdots \|\mathbf{a}_m\|^2)\|\mathbf{x}\|^2} \\
&= M\|\mathbf{x}\|.
\end{aligned}$$

□

Theorem 1.2 (Chain Rule)**.** *Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open. Let $g : U \to \mathbb{R}^m$ and $f : V \to \mathbb{R}^p$ be given functions such that the range of g is contained in V, so that $f \circ g$ is defined. If g is differentiable at $\mathbf{x}_0 \in U$ and f is differentiable at $\mathbf{y}_0 = g(\mathbf{x}_0) \in V$, then $f \circ g$ is differentiable at $\mathbf{x}_0$ and*

$$\mathbf{D}(f \circ g)(\mathbf{x}_0) = \mathbf{D}f(g(\mathbf{x}_0))\mathbf{D}g(\mathbf{x}_0). \tag{1}$$

Proof. Using the definition of differentiability, we need to prove that the right-hand side of (1) defines a linear function from $\mathbb{R}^n$ to $\mathbb{R}^p$ such that

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\|f(g(\mathbf{x})) - f(g(\mathbf{x}_0)) - \mathbf{D}f(\mathbf{y}_0)\mathbf{D}g(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

The result then follows from the uniqueness of the derivative. By adding and subtracting $\mathbf{D}f(\mathbf{y}_0) \cdot (g(\mathbf{x}) - g(\mathbf{x}_0))$ in the numerator and applying the triangle inequality, we get, with $\mathbf{y} = g(\mathbf{x})$ and $\mathbf{y}_0 = g(\mathbf{x}_0)$,

$$\begin{aligned} \|f(\mathbf{y}) - f(\mathbf{y}_0) - \mathbf{D}f(\mathbf{y}_0)\mathbf{D}g(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)\| &\le \|f(\mathbf{y}) - f(\mathbf{y}_0) - \mathbf{D}f(\mathbf{y}_0) \cdot (\mathbf{y} - \mathbf{y}_0)\| \\ &\quad + \|\mathbf{D}f(\mathbf{y}_0) \cdot (g(\mathbf{x}) - g(\mathbf{x}_0) - \mathbf{D}g(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0))\|. \end{aligned}$$

Let $\epsilon > 0$ be given. According to the lemma above $\|\mathbf{D}f(\mathbf{y}_0) \cdot \mathbf{v}\| \le M\|\mathbf{v}\|$ for any $\mathbf{v} \in \mathbb{R}^m$, for some constant $M > 0$. We will apply this for $\mathbf{v} = g(\mathbf{x}) - g(\mathbf{x}_0) - \mathbf{D}g(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$. Since g is differentiable at $\mathbf{x}_0$, we can find a $\delta_1 > 0$ such that $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta_1$ implies

$$\frac{\|g(\mathbf{x}) - g(\mathbf{x}_0) - \mathbf{D}g(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} < \frac{\epsilon}{2M}.$$

Also since g is differentiable at x_0, we can find a $\delta_2 > 0$ and a constant N such that $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta_2$ implies $\|g(\mathbf{x}) - g(\mathbf{x}_0)\| \le N\|\mathbf{x} - \mathbf{x}_0\|$. Since f is differentiable at $\mathbf{y}_0 = g(\mathbf{x}_0)$, we can find a $\delta_3 > 0$ such that $0 < \|\mathbf{y} - \mathbf{y}_0\| < \delta_3$ implies that

$$\|f(\mathbf{y}) - f(\mathbf{y}_0) - \mathbf{D}f(\mathbf{y}_0) \cdot (\mathbf{y} - \mathbf{y}_0)\| \le \frac{\epsilon}{2N}\|\mathbf{y} - \mathbf{y}_0\| = \frac{\epsilon}{2N}\|g(\mathbf{x}) - g(\mathbf{x}_0)\| < \frac{\epsilon}{2}\|\mathbf{x} - \mathbf{x}_0\|,$$

provided that $\|\mathbf{x} - \mathbf{x}_0\| < \min\{\delta_2, \delta_3/N\}$. Thus if we let $\delta = \min\{\delta_1, \delta_2, \delta_3/N\}$, we have

$$\frac{\|f(g(\mathbf{x})) - f(g(\mathbf{x}_0)) - \mathbf{D}f(g(\mathbf{x}_0))\mathbf{D}g(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} < \frac{\epsilon}{2} + M\frac{\epsilon}{2M} = \epsilon.$$

□

2 The Implicit and Inverse Function Theorems

Recall that for a level curve $g(x, y) = c$, we can solve for y as a function of x *locally* near a given point (x_0, y_0) on the curve, provided that the curve has a well-defined tangent line at that point, and that tangent line is not vertical. Notice that finding $y' = dy/dx$ by implicit differentiation is the same as finding y' via the relationship

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\frac{dy}{dx} = 0.$$

Thus, we can solve for y' provided $\dfrac{\partial g}{\partial y}(x_0, y_0) \neq 0$.

We will first state a special case that will be useful for dealing with level surfaces in $\mathbb{R}^n$ before stating the general result.

Theorem 2.1. *Suppose $F : \mathbb{R}^{n+1} \to \mathbb{R}$ is continuously differentiable. Denote points in $\mathbb{R}^{n+1}$ by $(\mathbf{x}, z)$, where $\mathbf{x} \in \mathbb{R}^n$ and $z \in \mathbb{R}$. If at a point $(\mathbf{x}_0, z_0) \in \mathbb{R}^{n+1}$ we have*

$$F(\mathbf{x}_0, z) = 0 \text{ and } \frac{\partial F}{\partial z}(\mathbf{x}_0, z_0) \neq 0,$$

then there is a ball U containing $\mathbf{x}_0$ in $\mathbb{R}^n$ and a interval (a, b) containing z in $\mathbb{R}$ such that there is a unique function $z = g(\mathbf{x})$ defined for $\mathbf{x} \in U$ and $z \in (a, b)$ that satisfies $F(\mathbf{x}, g(\mathbf{x})) = 0$. Moreover, if $\mathbf{x} \in U$ and $z \in (a, b)$ satisfy $F(\mathbf{x}, z) = 0$, then $z = g(\mathbf{x})$. Finally, $z = g(\mathbf{x})$ is continuously differentiable, with the derivative given by

$$\mathbf{D}g(\mathbf{x}) = -\frac{1}{\dfrac{\partial F}{\partial z}(\mathbf{x}, z)} \mathbf{D}_{\mathbf{x}} F(\mathbf{x}, z)|_{z=g(\mathbf{x})},$$

where $\mathbf{D}_{\mathbf{x}} F$ denotes the matrix of partial derivatives of F with respect to the variables $x_1, \ldots, x_n$. Equivalently, we have

$$\frac{\partial g}{\partial x_i} = -\frac{\partial F/\partial x_i}{\partial F/\partial z}, \quad i = 1, \ldots n.$$

Note that the theorem essentially tells us when we can solve the equation $F(\mathbf{x}, z)$ for z in terms of $\mathbf{x}$. More generally, suppose we are given a system of equations of the form

$$\begin{aligned} F_1(x_1, \ldots x_n, z_1, \ldots, z_m) &= 0 \\ F_2(x_1, \ldots x_n, z_1, \ldots, z_m) &= 0 \\ \vdots \qquad\qquad & \quad \vdots \\ F_m(x_1, \ldots x_n, z_1, \ldots, z_m) &= 0. \end{aligned}$$

The general implicit function theorem tells us that we can solve the system for the z_i in terms of the x_j, giving $z_i = f_i(x_1, \ldots x_n)$ for unique smooth functions $f_1, \ldots, f_m$, provided that the *determinant* of the $m \times m$ matrix

$$\begin{bmatrix} \frac{\partial F_1}{\partial z_1} & \cdots & \frac{\partial F_1}{\partial z_m} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial z_1} & \cdots & \frac{\partial F_m}{\partial z_m} \end{bmatrix}$$

is non-zero. A special case of the general implicit function theorem is when $m = n$ and $F_i(y_1, \ldots y_n, x_1, \ldots, x_n) = y_i - f(x_1, \ldots x_n)$, (here the x_i above are now the y_i, and the z_i

above are now the x_i, just to keep you on your toes) so that we are trying to solve the system of equations

$$\begin{aligned} f_1(x_1, \dots x_n) &= y_1 \\ \vdots \quad & \quad \vdots \\ f_n(x_1, \dots x_n) &= y_n, \end{aligned}$$

which means we are trying to invert the system of equations to express the x_i as functions of the y_j. Note that this is equivalent to writing $\mathbf{y} = F(\mathbf{x})$ for the vector-valued function $F = \langle f_1, \dots f_n \rangle$ and asking for the inverse function such that $\mathbf{x} = F^{-1}(\mathbf{y})$. Given such a function F, we define the *Jacobian* $J(F)$ of F as the determinant of the derivative of F:

$$J(F)(\mathbf{x}) = \det \mathbf{D}f(\mathbf{x}).$$

Theorem 2.2. *Let $U \subseteq \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}^n$ be continuously differentiable. For any $\mathbf{x}_0 \in U$, if $J(F)(\mathbf{x}_0) \neq 0$, then there is a neighbourhood N of $\mathbf{x}_0$ contained in U and a unique function G that is also continuously differentiable, such that for each $\mathbf{x} \in N$ and each $\mathbf{y} = F(\mathbf{x})$, we have $\mathbf{x} = G(\mathbf{y})$. Moreover, we have*

$$\mathbf{D}G(\mathbf{y}) = (\mathbf{D}F(\mathbf{x}))^{-1}$$

for each $\mathbf{x} \in N$. (The -1 on the right-hand side denotes the matrix inverse.)

Note: The inverse function theorem only applies to maps $F : \mathbb{R}^n \to \mathbb{R}^n$ where the number of variables is equal to the number of components. Notice that even if F is not defined on all of $\mathbb{R}^n$, $\mathbf{D}F(\mathbf{x})$ *is* for each $\mathbf{x}$ in the domain of F: the domain of any linear function is all of $\mathbb{R}^n$. The condition that the determinant of $\mathbf{D}F(\mathbf{x})$ is nonzero is equivalent to requiring the linear function $\mathbf{D}F(\mathbf{x})$ to be both one-to-one and onto (and therefore invertible). Since this condition may hold at some points $\mathbf{x}$ in the domain of F and not at others, the invertibility of F only holds *locally* (e.g. on the neighbourhood N in the statement of the theorem).

If $F : \mathbb{R}^n \to \mathbb{R}^m$ with $m > n$, $\mathbf{D}F(\mathbf{x})$ is no longer a square matrix, and therefore cannot be invertible. In this case, the best we can ask for is that $\mathbf{D}F(\mathbf{x})$ is one-to-one. If this is true at each $\mathbf{x}$ in the domain of F, then F is called an *immersion*. An immersion is a map that "preserves structure" in a sense that is made precise in more advanced courses. For example, if $F : \mathbb{R}^2 \to \mathbb{R}^3$ is an immersion, and C is a curve contained in the domain of F, then the image of C under F will still be a curve in $\mathbb{R}^3$, and if D is a region in $\mathbb{R}^2$ (such as a disk or a rectangle), then the image of D will be a surface. (Roughly speaking, F "preserves the dimension" of these objects - it doesn't collapse a curve to a point or a region to a curve.

If $m < n$, then the strongest condition one can impose is that $\mathbf{D}F(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ be onto. When this is the case, F looks locally like a projection. Such maps are called *submersions*.