Functions of Three or More Variables

Increments and differentials of functions of more than two variables are defined similarly. A function $w = f(x, y, z)$ has **increment**

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

and **differential**

$$dw = df = \frac{\partial f}{\partial x}\,\Delta x + \frac{\partial f}{\partial y}\,\Delta y + \frac{\partial f}{\partial z}\,\Delta z;$$

that is,

$$dw = \frac{\partial w}{\partial x}\,dx + \frac{\partial w}{\partial y}\,dy + \frac{\partial w}{\partial z}\,dz$$

if, as in Eq. (10), we write dx for Δx, dy for Δy, and dz for Δz.

EXAMPLE 5 You have constructed a metal cube that is supposed to have edge length 100 mm, but each of its three measured dimensions x, y, and z may be in error by as much as a millimeter. Use differentials to estimate the maximum resulting error in its calculated volume $V = xyz$.

Solution We need to approximate the increment

$$\Delta V = V(100 + dx, 100 + dy, 100 + dz) - V(100, 100, 100)$$

when the errors dx, dy, and dz in x, y, and z are maximal. The differential of $V = xyz$ is

$$dV = yz\,dx + xz\,dy + xy\,dz.$$

When we substitute $x = y = z = 100$ and $dx = \pm 1$, $dy = \pm 1$, and $dz = \pm 1$, we get

$$dV = 100 \cdot 100 \cdot (\pm 1) + 100 \cdot 100 \cdot (\pm 1) + 100 \cdot 100 \cdot (\pm 1) = \pm 30000.$$

It may surprise you to find that an error of only a millimeter in each dimension of a cube can result in an error of 30,000 mm^3 in its volume. (For a cube made of precious metal, an error of 30 cm^3 in its volume could correspond to a difference of hundreds or thousands of dollars in its cost.) ♦

Derivative Matrices

Matrix notation simplifies the description of differentials and linear approximation for functions of several variables. Let $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$ be a real-valued function of n variables. If

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^T \quad \text{and} \quad \mathbf{h} = [h_1 \quad h_2 \quad \cdots \quad h_n]^T,$$

then the linear approximation formula for f takes the form

$$f(\mathbf{x} + \mathbf{h}) \approx f(\mathbf{x}) + \frac{\partial f}{\partial x_1} h_1 + \frac{\partial f}{\partial x_2} h_2 + \cdots + \frac{\partial f}{\partial x_n} h_n \tag{12}$$

with one term for each independent variable. We introduce the **derivative matrix**

$$f'(\mathbf{x}) = [D_1 f(\mathbf{x}) \quad D_2 f(\mathbf{x}) \quad \cdots \quad D_n f(\mathbf{x})] = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_n}\right] \tag{13}$$

of the function $f(x_1, x_2, \ldots, x_n)$ of n variables; the elements of this *row* matrix are the n first-order partial derivatives of f (assuming that they exist).

Using this derivative matrix, the linear approximation formula in (12) takes the concise form

$$f(\mathbf{x} + \mathbf{h}) \approx f(\mathbf{x}) + f'(\mathbf{x})\mathbf{h}, \tag{14}$$

in pleasant analogy with the single-variable approximation $f(x+h) \approx f(x) + f'(x)h$ (writing h for Δx here). Note that, because $f'(\mathbf{x})$ is a $1 \times n$ row matrix and $\mathbf{h}$ is an $n \times 1$ column matrix, the matrix product on the right-hand side in (14) is defined and gives

$$f'(\mathbf{x})\mathbf{h} = D_1 f(\mathbf{x})h_1 + D_2 f(\mathbf{x})h_2 + \cdots + D_n f(\mathbf{x})h_n,$$

thus providing the linear terms on the right-hand side in (12). In analogy with the two-variable case in (5), the sum of these n linear terms is the **differential** $df = f'(\mathbf{x})\mathbf{h}$ of the function f of n real variables.

The derivative matrix $f'(\mathbf{x})$ is defined wherever all of the first-order partial derivatives of f exist. In Appendix K we give a proof of the linear approximation theorem stated next. This theorem assures us (in effect) that if the partial derivatives of f are also *continuous,* then the linear approximation in (14) is a *good* approximation when $|\mathbf{h}| = \sqrt{h_1^2 + h_2^2 + \cdots + h_n^2}$ is small.

THEOREM Linear Approximation

Suppose that the function $f(\mathbf{x})$ of n variables has continuous first-order partial derivatives in a region that contains the neighborhood $|\mathbf{x} - \mathbf{a}| < r$ consisting of all points $\mathbf{x}$ at distance less than r from the fixed point $\mathbf{a}$. If $\mathbf{a} + \mathbf{h}$ lies in this neighborhood, then

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a})\mathbf{h} + \epsilon(\mathbf{h})\mathbf{h} \tag{15}$$

where $\epsilon(\mathbf{h}) = [\epsilon_1(\mathbf{h}) \;\; \epsilon_2(\mathbf{h}) \;\; \ldots \;\; \epsilon_n(\mathbf{h})]$ is a row matrix such that each element $\epsilon_i(\mathbf{h})$ approaches zero as $\mathbf{h} \to \mathbf{0}$.

REMARK 1 The multivariable function f is said to be **continuously differentiable** at a point provided that its first-order partial derivatives not only exist, but are continuous at the point. Thus the hypothesis of the linear approximation theorem is that the function f is continuously differentiable in the specified neighborhood of the point $\mathbf{a}$.

REMARK 2 The matrix product

$$\epsilon(\mathbf{h})\mathbf{h} = \epsilon_1(\mathbf{h})h_1 + \epsilon_2(\mathbf{h})h_2 + \cdots + \epsilon_n(\mathbf{h})h_n \tag{16}$$

in (15) is the **error** in the linear approximation—it measures the extent to which the *approximation* $f(\mathbf{a}+\mathbf{h}) \approx f(\mathbf{a}) + f'(\mathbf{a})\mathbf{h}$ fails to be an *equality*. We may regard the conclusion of the linear approximation theorem as saying that if $\mathbf{h}$ is "very small," then each element $\epsilon_i(\mathbf{h})$ is also "very small." In this event, each summand in (16) is a product of two very small terms, so we might say that the error $\epsilon(\mathbf{h})\mathbf{h}$ is "very very small."

Now let us divide by $|\mathbf{h}|$ in Eq. (16). Then we see that

$$\frac{\epsilon(\mathbf{h})\mathbf{h}}{|\mathbf{h}|} = \epsilon_1(\mathbf{h})\frac{h_1}{|\mathbf{h}|} + \epsilon_2(\mathbf{h})\frac{h_2}{|\mathbf{h}|} + \cdots + \epsilon_n(\mathbf{h})\frac{h_n}{|\mathbf{h}|} \to 0 \tag{17}$$

as $\mathbf{h} \to \mathbf{0}$. The reason is that, for each i $(1 \leqq i \leqq n)$,

$$\frac{h_i}{|\mathbf{h}|} \leqq 1 \quad \text{and} \quad \epsilon_i(\mathbf{h}) \to 0$$

as $\mathbf{h} \to \mathbf{0}$. Dividing both sides by $|\mathbf{h}|$ in Eq. (15) therefore gives the limit

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - f'(\mathbf{a})\mathbf{h}}{|\mathbf{h}|} = 0, \tag{18}$$

under the assumption that the function f is continuously differentiable near $\mathbf{a}$.

The Multivariable Newton's Method

Consider the problem of solving the two equations

$$f(x, y) = 0, \qquad g(x, y) = 0 \tag{21}$$

in the two unknowns x and y. The two scalar-valued functions $f : \boldsymbol{R}^2 \to \boldsymbol{R}$ and $g : \boldsymbol{R}^2 \to \boldsymbol{R}$ yield a single vector-valued function $F : \boldsymbol{R}^2 \to \boldsymbol{R}^2$ defined by $F(x, y) = [f(x, y) \;\; g(x, y)]^T$. That is, if $\mathbf{x} = [x \;\; y]^T$ and $\mathbf{u} = [u \;\; v]^T$, then the scalar components of $\mathbf{u} = F(\mathbf{x})$ are $u = f(x, y)$ and $v = g(x, y)$. Then the problem of solving simultaneously the two scalar equations in (21) amounts to solving the single vector equation

$$F(\mathbf{x}) = \mathbf{0}. \tag{22}$$

Just as the two scalar-valued differentiable functions f and g "combine" in a single vector-valued function F, their linear approximation formulas

$$f(\mathbf{x}) \approx f(\mathbf{a}) + f'(\mathbf{a})(\mathbf{x} - \mathbf{a}),$$
$$g(\mathbf{x}) \approx g(\mathbf{a}) + g'(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

(from Eq. (14) with $\mathbf{x} = \mathbf{a}$ and $\mathbf{h} = \mathbf{x} - \mathbf{a}$) combine in a single vector formula. That is, because $f'(\mathbf{a}) = [D_1 f(\mathbf{a}) \;\; D_2 f(\mathbf{a})]$ and $g'(\mathbf{a}) = [D_1 g(\mathbf{a}) \;\; D_2 g(\mathbf{a})]$, the two linear approximation formulas may be written as the single formula

$$\begin{bmatrix} f(\mathbf{x}) \\ g(\mathbf{x}) \end{bmatrix} \approx \begin{bmatrix} f(\mathbf{a}) \\ g(\mathbf{a}) \end{bmatrix} + \begin{bmatrix} D_1 f(\mathbf{a}) & D_2 f(\mathbf{a}) \\ D_1 g(\mathbf{a}) & D_2 g(\mathbf{a}) \end{bmatrix} (\mathbf{x} - \mathbf{a}).$$

Because $F(\mathbf{x}) = [f(\mathbf{x}) \;\; g(\mathbf{x})]^T$, this observation provides the *linear approximation formula*

$$F(\mathbf{x}) \approx F(\mathbf{a}) + F'(\mathbf{a})(\mathbf{x} - \mathbf{a}) \tag{23}$$

for the function $F : \boldsymbol{R}^2 \to \boldsymbol{R}^2$, where

$$F'(\mathbf{a}) = \begin{bmatrix} D_1 f(\mathbf{a}) & D_2 f(\mathbf{a}) \\ D_1 g(\mathbf{a}) & D_2 g(\mathbf{a}) \end{bmatrix} \tag{24}$$

is the 2×2 **derivative matrix** of the function F of two variables; its elements are the partial derivatives of the two component functions f and g of F.

We now use the vector linear approximation formula in (23) to derive Newton's method for the solution of the vector equation $F(\mathbf{x}) = \mathbf{0}$. Suppose that $\mathbf{r}$ is a solution of this equation and that $\mathbf{x}_0$ is an initial guess at the value of $\mathbf{r}$. Substituting $\mathbf{x} = \mathbf{r}$ and $\mathbf{x} = \mathbf{x}_0$ in (23) then yields the approximation

$$F(\mathbf{r}) \approx F(\mathbf{x}_0) + F'(\mathbf{x}_0)(\mathbf{r} - \mathbf{x}_0).$$

Because $F(\mathbf{r}) = \mathbf{0}$, this means that $F'(\mathbf{x}_0)(\mathbf{r} - \mathbf{x}_0) \approx -F(\mathbf{x}_0)$. If the 2×2 matrix $F'(\mathbf{x}_0)$ is invertible, then we can multiply by its inverse matrix $F'(\mathbf{x}_0)^{-1}$ to obtain the presumably improved estimate

$$\mathbf{r} \approx \mathbf{x}_0 - F'(\mathbf{x}_0)^{-1} F(\mathbf{x}_0)$$

of the solution $\mathbf{r}$. Thus having begun with the initial guess $\mathbf{x}_0$, we have derived the new approximation $\mathbf{x}_1$ given by

$$\mathbf{x}_1 = \mathbf{x}_0 - F'(\mathbf{x}_0)^{-1} F(\mathbf{x}_0).$$

In exactly the same way, having reached an nth approximation $\mathbf{x}_n$ to the actual solution $\mathbf{r}$, we thereby derive the next approximation

$$\mathbf{x}_{n+1} = \mathbf{x}_n - F'(\mathbf{x}_n)^{-1} F(\mathbf{x}_n). \tag{25}$$

14.7 THE MULTIVARIABLE CHAIN RULE

The single-variable chain rule of Section 3.3 says that the composition $h(x) = f(g(x))$ of two differentiable single-variable functions f and g is differentiable, and that its derivative is given in terms of their derivatives by $h'(x) = f'(g(x)) \cdot g'(x)$. An appropriate generalization to vector-valued functions is the multivariable chain rule of this section.

A function $F : \boldsymbol{R}^n \to \boldsymbol{R}^m$ (that is, from $\boldsymbol{R}^n$ to $\boldsymbol{R}^m$) associates with a vector $\mathbf{x}$ in $\boldsymbol{R}^n$ a vector $\mathbf{y} = F(\mathbf{x})$ in $\boldsymbol{R}^m$. Each component of the m-vector $\mathbf{y}$ is then a function of the n-vector $\mathbf{x}$, so we may write

$$F(\mathbf{x}) = [F_1(\mathbf{x}) \quad F_2(\mathbf{x}) \quad \cdots \quad F_m(\mathbf{x})]^T$$

where $F_1, F_2, \ldots, F_m$ are the **component functions** of F. Each component function $F_i : \boldsymbol{R}^n \to \boldsymbol{R}$ associates a real number $y_i = F_i(\mathbf{x})$ with the n-vector $\mathbf{x}$, and we may write $F_i(\mathbf{x}) = F_i(x_1, x_2, \ldots, x_n)$ in view of the correspondence between n-vectors and n-tuples. Thus the function $F : \boldsymbol{R}^n \to \boldsymbol{R}^m$ expresses each of the m **dependent variables** $y_1, y_2, \ldots, y_m$ as a function $y_i = F_i(x_1, x_2, \ldots, x_n)$ of the n **independent variables** $x_1, x_2, \ldots, x_n$. We may write $F: \boldsymbol{R}^n_{\mathbf{x}} \to \boldsymbol{R}^m_{\mathbf{y}}$ to signify the use of x-coordinates in the domain and y-coordinates in the range of the function F, but other notation for the independent and dependent variables may be more appropriate in a particular situation.

The notation $F : \boldsymbol{R}^n \to \boldsymbol{R}^m$ means that the domain of F lies in $\boldsymbol{R}^n$ and that its range lies in $\boldsymbol{R}^m$. But F need not be defined everywhere in $\boldsymbol{R}^n$. For instance, the function $F : \boldsymbol{R}^3 \to \boldsymbol{R}$ described by $F(x, y, z) = xy/z$ is not defined at points of the xy-plane at which $z = 0$.

EXAMPLE 1 Recall the familiar equations

$$x = r\cos\theta, \quad y = r\sin\theta$$

that express the rectangular coordinates (x, y) of a point in the plane in terms of its polar coordinates r and θ. These equations define the component functions $x = T_1(r, \theta)$ and $y = T_2(r, \theta)$ of the function $T : \boldsymbol{R}^2_{r\theta} \to \boldsymbol{R}^2_{xy}$ that is defined by $T(r, \theta) = (r\cos\theta, r\sin\theta)$. Note that T has independent variables r, θ and dependent variables x, y. ♦

REMARK The function $T : \boldsymbol{R}^2 \to \boldsymbol{R}^2$ of Example 1 is sometimes called the *polar coordinates transformation* of the plane. The words *function, mapping,* and *transformation* are synonyms whose usage depends somewhat on the context—the latter two terms appear less frequently in single-variable situations than in multivariable ones, and the term "transformation" is often (but not always) used when $m = n$.

The following definition of differentiability reads essentially the same as the definition given in Section 14.6 for the special case of real-valued functions. We require that the function F be defined at and near the point $\mathbf{a}$, meaning that it is defined for all $\mathbf{x}$ in some neighborhood $|\mathbf{x} - \mathbf{a}| < \delta$ of the point $\mathbf{a}$.

DEFINITION Differentiable Function

Suppose that the function $F : \boldsymbol{R}^n \to \boldsymbol{R}^m$ is defined at and near the point $\mathbf{a}$ in $\boldsymbol{R}^n$. Then F is said to be **differentiable** at $\mathbf{a}$ provided that there exists a constant $m \times n$ matrix $\mathbf{C}$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{F(\mathbf{a}+\mathbf{h}) - F(\mathbf{a}) - \mathbf{Ch}}{|\mathbf{h}|} = \mathbf{0}. \tag{1}$$

In effect, this definition means that F is differentiable at $\mathbf{a}$ if there exists a linear function $L(\mathbf{h}) = \mathbf{Ch}$ of the n-vector $\mathbf{h}$ that approximates the increment $F(\mathbf{a}+\mathbf{h}) - F(\mathbf{a})$

so closely that (when $\mathbf{h}$ is small) the error is small even in comparison with $|\mathbf{h}|$. In short, differentiable functions are those that have "good" linear approximations.

In Problem 62 we ask you to verify that if F is differentiable at $\mathbf{a}$, then the component functions $F_1, F_2, \ldots, F_m$ of F all have first-order partial derivatives at $\mathbf{a}$, and that the matrix $\mathbf{C}$ in Eq. (1) is then given by

$$\mathbf{C} = \begin{bmatrix} D_1F_1(\mathbf{a}) & D_2F_1(\mathbf{a}) & \cdots & D_nF_1(\mathbf{a}) \\ D_1F_2(\mathbf{a}) & D_2F_2(\mathbf{a}) & \cdots & D_nF_2(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1F_m(\mathbf{a}) & D_2F_m(\mathbf{a}) & \cdots & D_nF_m(\mathbf{a}) \end{bmatrix}. \tag{2}$$

Note that the elements of the ith row of $\mathbf{C}$ are the n first-order partial derivatives of the ith component function F_i of F. And (just as in the case $m = 1$ discussed in Section 14.6) the function F is differentiable whenever all the partial derivatives appearing in Eq. (2) exist *and* are continuous—in which case F is said to be **continuously differentiable.** Thus:

- If F is continuously differentiable, then F is differentiable.
- If F is differentiable, then all its first-order partial derivatives exist.

Because the condition in Eq. (1) can be difficult to verify directly, the ordinary way of verifying differentiability of a multivariable function F is to calculate all first-order partial derivatives of its component functions, observe that they are all continuous, and hence conclude that the function F is differentiable because it is continuously differentiable.

The matrix in Eq. (2) is known as the *derivative matrix* of F at $\mathbf{a}$.

DEFINITION Derivative Matrix

The **derivative matrix** (sometimes called the **Jacobian matrix**) of the function $F: \mathbf{R}^n \to \mathbf{R}^m$ is the $m \times n$ matrix defined by

$$F'(\mathbf{x}) = [D_jF_i(\mathbf{x})] = \left[\frac{\partial F_i}{\partial x_j}\right] \tag{3}$$

wherever the indicated partial derivatives all exist.

Observe carefully the order of the subscripts in Eq. (3). The ith *row* of the derivative matrix consists of the partial derivatives of the ith component function F_i with respect to the n different independent variables $x_1, x_2, \ldots, x_n$ (first row, first function; second row, second function; ...). The jth *column* of $F'(\mathbf{x})$ consists of the partial derivatives of the m component functions $F_1, F_2, \ldots, F_m$ with respect to the jth independent variable x_j (first column, first variable; second column, second variable; ...).

EXAMPLE 2 The derivative matrix of the polar coordinate transformation $T(r, \theta) = (r\cos\theta, r\sin\theta)$ is a 2×2 matrix with the partial derivatives of $T_1(r, \theta) = r\cos\theta$ in its first row and the partial derivatives of $T_2(r, \theta) = r\sin\theta$ in its second row. Therefore

$$T'(r, \theta) = \begin{bmatrix} \dfrac{\partial T_1}{\partial r} & \dfrac{\partial T_1}{\partial \theta} \\ \dfrac{\partial T_2}{\partial r} & \dfrac{\partial T_2}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}.$$

All four partial derivatives we see here are obviously continuous everywhere, so we conclude that $T: \mathbf{R}^2 \to \mathbf{R}^2$ is continuously differentiable and, therefore, differentiable everywhere. ◆

The General Chain Rule

Given functions $G: \boldsymbol{R}^n \to \boldsymbol{R}^p$ and $F: \boldsymbol{R}^p \to \boldsymbol{R}^m$, their **composition**

$$H = F \circ G : \boldsymbol{R}^n \to \boldsymbol{R}^m$$

is defined by $H(\mathbf{x}) = F(G(\mathbf{x}))$ whenever $G(\mathbf{x})$ lies in the domain of F.

THEOREM 1 The Chain Rule

If $G: \boldsymbol{R}^n \to \boldsymbol{R}^p$ is differentiable at $\mathbf{a}$ and $F: \boldsymbol{R}^p \to \boldsymbol{R}^m$ is differentiable at $G(\mathbf{a})$, then the composition $H = F \circ G: \boldsymbol{R}^n \to \boldsymbol{R}^m$ is differentiable at $\mathbf{a}$, and its derivative matrix is given by

$$H'(\mathbf{a}) = F'(G(\mathbf{a}))G'(\mathbf{a}). \tag{4}$$

Note that the $m \times n$ product matrix $H'(\mathbf{a}) = (F \circ G)'(\mathbf{a})$ on the right is defined because $F'(G(\mathbf{a}))$ is an $m \times p$ matrix and $G'(\mathbf{a})$ is a $p \times n$ matrix.

Observe that Eq. (4) is precisely the matrix analog of the single-variable scalar chain rule $h'(x) = f'(g(x)) \cdot g'(x)$ where $h = f \circ g$. At the conclusion of this section we outline a multivariable proof of Theorem 1 that generalizes the proof of the single-variable chain rule found in Section 4.2.

To see what the matrix product in Eq. (4) means, let's write each matrix in terms of partial derivatives of component functions (omitting arguments):

$$[D_j H_i] = \begin{bmatrix} D_1 F_1 & \cdots & D_r F_1 & \cdots & D_p F_1 \\ & & \vdots & & \\ D_1 F_i & \cdots & D_r F_i & \cdots & D_p F_i \\ & & \vdots & & \\ D_1 F_m & \cdots & D_r F_m & \cdots & D_p F_m \end{bmatrix} \begin{bmatrix} D_1 G_1 & \cdots & D_j G_1 & \cdots & D_n G_1 \\ \vdots & & \vdots & & \vdots \\ D_1 G_r & \cdots & D_j G_r & \cdots & D_n G_r \\ \vdots & & \vdots & & \vdots \\ D_1 G_p & \cdots & D_j G_p & \cdots & D_n G_p \end{bmatrix}.$$

Because the ijth element of $H'(\mathbf{a})$ is the product of the ith row of $F'(G(\mathbf{a}))$ and the jth column of $G'(\mathbf{a})$, we see that

$$D_j H_i(\mathbf{a}) = \sum_{r=1}^{p} D_r F_i(G(\mathbf{a})) \cdot D_j G_r(\mathbf{a}) \tag{5}$$

for each i $(1 \leqq i \leqq m)$ and each j $(1 \leqq j \leqq n)$.

To make this intricate-looking formula easier to remember, let's write $\mathbf{w} = F(\mathbf{x})$ and $\mathbf{x} = G(\mathbf{t})$, so that the notation

$$\boldsymbol{R}^n_{\mathbf{t}} \xrightarrow{\;\;G\;\;} \boldsymbol{R}^p_{\mathbf{x}} \xrightarrow{\;\;F\;\;} \boldsymbol{R}^m_{\mathbf{w}}$$

indicates domains and ranges. Thus F gives each of the **dependent variables** $w_1, w_2, \dots, w_m$ in terms of the **intermediate variables** $x_1, x_2, \dots, x_p$, and G gives each of these intermediate variables in terms of the **independent variables** $t_1, t_2, \dots, t_n$. Then, using partial differential rather than operator notation, Eq. (5) takes the more memorable form

$$\frac{\partial w_i}{\partial t_j} = \sum_{r=1}^{p} \frac{\partial w_i}{\partial x_r} \cdot \frac{\partial x_r}{\partial t_j} = \frac{\partial w_i}{\partial x_1} \cdot \frac{\partial x_1}{\partial t_j} + \frac{\partial w_i}{\partial x_2} \cdot \frac{\partial x_2}{\partial t_j} + \cdots + \frac{\partial w_i}{\partial x_p} \cdot \frac{\partial x_p}{\partial t_j} \tag{6}$$

for each i $(1 \leqq i \leqq m)$ and each j $(1 \leqq j \leqq n)$. In this formula for the partial derivative of the dependent variable w_i with respect to the independent variable t_j, we see one term for each of the p intermediate variables $x_1, x_2, \dots, x_p$. Moreover, each of these terms exhibits a characteristic "chain rule pattern"—as though the ∂x_r in the first factor $\partial w_i / \partial x_r$ somehow cancelled the ∂x_r in the second factor $\partial x_r / \partial t_j$, meaninglessly leaving $\partial w_i / \partial t_j$.

FIGUR
a cylir
(Exam