

Chain Rules for Higher Derivatives

H.-N. HUANG, S. A. M. MARCANTOGNINI,
AND N. J. YOUNG

We define a notion of higher-order directional derivative of a smooth function and use it to establish three simple formulae for the n th derivative of the composition of two functions. These three “higher-order chain rules” are alternatives to the classical Faà di Bruno formula. They are less explicit than Faà di Bruno’s formula but are much simpler and avoid Diophantine or combinatorial complications.

In principle it is easy to calculate a higher derivative of the composition $f \circ g$ of two sufficiently differentiable functions f and g : one can simply apply the “chain rule”

$$(f \circ g)' = (f' \circ g)g'$$

as many times as needed. For example,

$$\begin{aligned} (f \circ g)'' &= ((f' \circ g)g')' = (f' \circ g)'g' + (f' \circ g)g'' \\ &= (f'' \circ g)(g')^2 + (f' \circ g)g''. \end{aligned}$$

By continuing in this vein we can readily obtain an expression for any particular higher-order derivative of $f \circ g$ in terms of derivatives of f and g . Formulae for $(f \circ g)^{(n)}$, for an arbitrary positive integer n , have been known since the 19th century: an admirable scholarly article by Warren P. Johnson [8] not only gives several such formulae but also describes the history of the problem. Of all the formulae the most popular is named after Francesco Faà di Bruno¹, who published it in 1855 [3,4], though according to Johnson [8, p. 228], the result had been anticipated by J. F. C. Tiburce Abadie in 1850. The formula states that

$$\frac{d^n}{dt^n} f(g(t)) = \sum \frac{n!}{k_1!k_2! \cdots k_n!} f^{(k)}(g(t)) \left(\frac{g'(t)}{1!} \right)^{k_1} \left(\frac{g''(t)}{2!} \right)^{k_2} \cdots \left(\frac{g^{(n)}(t)}{n!} \right)^{k_n}, \quad (1)$$

where the sum is over all nonnegative integer solutions of the Diophantine equation

$$k_1 + 2k_2 + \cdots + nk_n = n, \quad \text{and} \quad k := k_1 + k_2 + \cdots + k_n.$$

For example, in the case $n = 3$, we find that the nonnegative integer solutions of the equation

$$k_1 + 2k_2 + 3k_3 = 3$$

¹A remarkable man, who not only wrote an influential treatise on binary forms (“Théorie des formes binaires” (Paris, 1876), translated into German (Leipzig, 1881)) but was also declared a saint of the Catholic church for his charitable and religious works [5].

are $(k_1, k_2, k_3) = (0, 0, 1)$ or $(1, 1, 0)$ or $(3, 0, 0)$ with corresponding values of k being 1, 2, or 3, respectively. Faà di Bruno's formula (1) then gives

$$\begin{aligned} \frac{d^3}{dt^3} f(g(t)) &= \frac{3!}{0!0!1!} f'(g(t)) \left(\frac{g'''(t)}{3!} \right) + \frac{3!}{1!1!0!} f''(g(t)) \left(\frac{g'(t)}{1!} \right) \left(\frac{g''(t)}{2!} \right) + \frac{3!}{3!0!0!} f'''(g(t)) \left(\frac{g'(t)}{1!} \right)^3 \\ &= f'(g(t))g'''(t) + 3f''(g(t))g'(t)g''(t) + f'''(g(t))(g'(t))^3. \end{aligned}$$

Faà di Bruno's formula, and others given in [8], have a substantial combinatorial aspect; a less combinatorial and generally more transparent treatment is that of Spindler [11]. We shall present an alternative formula for $(f \circ g)^{(n)}$ that differs from Faà di Bruno's in two ways: it makes use of higher-order *directional* derivatives, and it avoids combinatorial or Diophantine questions. It would be surprising if anything new could be said about such a classical topic, but we have not found anything similar in [8] or other recent papers [2, 9, 11]. We needed the formula for the study of the *spectral Carathéodory-Fejér problem*: given $k \times k$ matrices V_0, V_1, \dots, V_n , determine whether there exists an analytic matricial function F in the unit disc \mathbb{D} such that

- the eigenvalues of $F(\lambda)$ lie in \mathbb{D} for all $\lambda \in \mathbb{D}$, and
- $F^{(j)}(0) = V_j$, $0 \leq j \leq n$.

We may derive a necessary condition with the aid of a higher chain rule. If such a function F exists then we may consider the function

$$f(z, \lambda) = \det(zI - F(\lambda)).$$

This f is analytic in λ on \mathbb{D} , and for fixed λ is a polynomial in z whose zeros lie in \mathbb{D} . Moreover, $\partial^j f / \partial \lambda^j|_{\lambda=0}$ for $j = 0, 1, \dots, n$ can be found explicitly in terms of the V_j by the chain rule in Theorem 2 below. One of the main results in [6] states that, subject to a genericity condition, the existence of a function $f(z, \lambda)$ with the appropriate properties is also sufficient for the solvability of a spectral Carathéodory-Fejér problem.

We give the formula in three slightly different versions, two for the case that g is a function of one variable and a third for more general g .

Directional Derivatives and Higher Chain Rules

Let X and Y be real or complex Banach spaces, let Ω be an open subset of X and let $f: \Omega \rightarrow Y$ be Fréchet-differentiable. (In the case that X and Y are Euclidean spaces, the notion of Fréchet differentiability coincides with the usual notion of differentiability from real analysis. The definition for the general case can be found in [7, Section 5.2] and [10, Chapter 29].) For any $x_0 \in \Omega$ and non-zero $x_1 \in X$, the *directional* or *Gateau derivative of f at x_0 along x_1* is defined to be

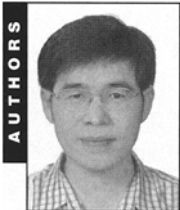
$$\lim_{z \rightarrow 0} \frac{1}{z} [f(x_0 + zx_1) - f(x_0)].$$

This is an element of Y , sometimes denoted by

$$D_{x_1} f(x_0) \quad \text{or} \quad \frac{d}{dx_1} f(x_0).$$

The limit exists, since f is supposed Fréchet differentiable: if $Df(x_0) \in \mathcal{L}(X, Y)$ is the Fréchet derivative of f at x_0 then, for any $x_1 \in X$,

$$\frac{d}{dx_1} f(x_0) = Df(x_0)x_1;$$



HUANG-NAN HUANG has been on the faculty at Tunghai University since 1995 and is now Head of the Department of Mathematics. He received his Ph.D. from National Chung-Hsing University in 1999. He mainly works on control and interpolation theory. In his spare time, he enjoys walking, running and Taichi Chuan.

Department of Mathematics
Tunghai University
Taichung 40704, Taiwan
e-mail: nhuang@thu.edu.tw



STEFANIA MARCANTOGNINI'S mathematical education was at the Universidad Central de Venezuela. At present, besides working at IVC, she is Professor at the Universidad Simón Bolívar. Her research area is operator theory. Her spare time, what little there is of it, she likes to spend with her husband and their twins.

Department of Mathematics
Instituto Venezolano de Investigaciones Científicas
Caracas 1020A, Venezuela
e-mail: smarcant@ivic.ve

here $\mathcal{L}(X, Y)$ denotes the space of bounded linear operators from X to Y . The reader is referred to [7, Section 5.2] and [10, Chapter 29] for further details.

We shall introduce notation for higher-order directional derivatives. As far as we know, this is a new notion.

DEFINITION 1 Let X, Y be real or complex Banach spaces, let Ω be an open set in X , and let $f: \Omega \rightarrow Y$ be n times Fréchet differentiable. We define, for $j = 0, 1, \dots, n-1$,

$$\Delta_j f : \Omega \times X^j \rightarrow Y$$

inductively for $x_0 \in \Omega, x_1, x_2, \dots, x_n \in X$ by

$$\begin{aligned} \Delta_0 f(x_0) &= f(x_0), \\ \Delta_{j+1} f(x_0, x_1, \dots, x_{j+1}) &= \frac{d}{d(x_1, x_2, \dots, x_{j+1})} \Delta_j f(x_0, x_1, \dots, x_j). \end{aligned} \quad (2)$$

Thus

$$\begin{aligned} \Delta_{j+1} f(x_0, x_1, \dots, x_{j+1}) &= \lim_{z \rightarrow 0} \frac{1}{z} \left[\Delta_j f(x_0 + zx_1, x_1 + zx_2, \dots, x_j + zx_{j+1}) - \Delta_j f(x_0, x_1, \dots, x_j) \right] \\ &= \frac{\partial}{\partial z} \Delta_j f(x_0 + zx_1, x_1 + zx_2, \dots, x_j + zx_{j+1}) \Big|_{z=0}, \end{aligned}$$

and we may write

$$\Delta_n f(x_0, x_1, \dots, x_n) = \frac{d}{d(x_1, \dots, x_n)} \frac{d}{d(x_1, \dots, x_{n-1})} \cdots \frac{d}{d(x_1, x_2)} \frac{d}{dx_1} f(x_0).$$

EXAMPLE 1: The exponential function.

Directly from Definition 1 we find, for $x_j \in \mathbb{C}$,

$$\begin{aligned} \Delta_1 \exp(x_0, x_1) &= \frac{\partial}{\partial z} e^{x_0 + zx_1} \Big|_{z=0} = e^{x_0} x_1, \\ \Delta_2 \exp(x_0, x_1, x_2) &= \frac{\partial}{\partial z} e^{x_0 + zx_1} (x_1 + zx_2) \Big|_{z=0} = e^{x_0} (x_1^2 + x_2), \\ \Delta_3 \exp(x_0, x_1, x_2, x_3) &= \frac{\partial}{\partial z} e^{x_0 + zx_1} [(x_1 + zx_2)^2 + x_2 + zx_3] \Big|_{z=0} \\ &= e^{x_0} (x_1^3 + 3x_1 x_2 + x_3). \end{aligned}$$

We regard $\Delta_n f(x_0, \dots, x_n)$ as the directional derivative of f of order n at x_0 in the direction (x_1, \dots, x_n) . There is another natural interpretation of the appellation “ n th-order directional derivative of f ”, to wit the function

$$\Delta^n f: \Omega \times X^{2^n-1} \rightarrow Y$$

defined inductively by

$$\Delta^1 f = \Delta_1 f, \quad \Delta^{n+1} f = \Delta_1(\Delta^n f).$$

There is a close connection between Δ_n and Δ^n .



NICHOLAS YOUNG works on operator theory and complex analysis, with a leaning toward questions arising from H^∞ control theory. He took his first and doctoral degrees in mathematics at Oxford University and has held posts at Glasgow, Lancaster, and Newcastle upon Tyne.

School of Mathematics
University of Leeds
Leeds, LS2 9JT, England
e-mail: N.J.Young@leeds.ac.uk

THEOREM 1 Let X, Y be Banach spaces, let Ω be an open subset of X , and let $f: \Omega \rightarrow Y$ be n times continuously Fréchet differentiable. For $m \in \mathbb{N}$, let

$$L_m : \Omega \times X^m \rightarrow \Omega \times X^{2^m-1}$$

be defined inductively by

$$L_1(x_0, x_1) = (x_0, x_1),$$

$$L_{m+1}(x_0, x_1, \dots, x_{m+1}) = (L_m(x_0, \dots, x_m); L_m(x_1, \dots, x_{m+1})),$$

for $x_0 \in \Omega$ and $x_j \in X$, $j \geq 1$. Then

$$\Delta_n f = (\Delta^n f) \circ L_n : \Omega \times X^n \rightarrow Y.$$

Thus

$$\Delta_2 f(x_0, x_1, x_2) = \Delta^2 f(x_0, x_1, x_1, x_2),$$

$$\Delta_3 f(x_0, x_1, x_2, x_3) = \Delta^3 f(x_0, x_1, x_1, x_2, x_1, x_2, x_2, x_3).$$

$\Delta_n f$ has the merit of a relative economy in the number of its arguments, and earns its keep by virtue of the following three higher-order chain rules. The first two concern functions $f \circ g$ where g is a function of a single real or complex variable. Let \mathbb{K} denote either \mathbb{R} or \mathbb{C} .

THEOREM 2 Let X, Y be Banach spaces over \mathbb{K} and let Ψ, Ω be open subsets of \mathbb{K}, X respectively. Let

$$g: \Psi \rightarrow \Omega, \quad f: \Omega \rightarrow Y$$

have n continuous Fréchet derivatives, for some $n \in \mathbb{N}$. For any $t \in \Psi$,

$$(f \circ g)^{(n)}(t) = \Delta_n f(g(t), g'(t), \dots, g^{(n)}(t)). \quad (3)$$

PROOF. In the case $n = 1$ we have

$$\begin{aligned} \Delta_1 f(g(t), g'(t)) &= \left. \frac{\partial}{\partial z} f(g(t) + zg'(t)) \right|_{z=0} \\ &= Df(g(t))(g'(t)) \\ &= (f \circ g)'(t) \end{aligned}$$

by the classical chain rule.

Suppose the assertion holds for some $j < n$, that is, for any $t \in \Psi$,

$$\frac{d^j}{dt^j} f \circ g(t) = \Delta_j f(g(t), g'(t), \dots, g^{(j)}(t)). \quad \square$$

Then, for $t \in \Psi$,

$$\begin{aligned} \frac{d^{j+1}}{dt^{j+1}} f \circ g(t) &= \frac{d^j}{dt^j} (f \circ g)'(t) \\ &= \frac{d^j}{dt^j} \Delta_1 f(g(t), g'(t)) \\ &= \frac{\partial^j}{\partial t^j} \left(\left. \frac{\partial}{\partial z} f(g(t) + zg'(t)) \right|_{z=0} \right) \\ &= \frac{\partial^j}{\partial t^j} \left. \frac{\partial}{\partial z} f(g(t) + zg'(t)) \right|_{z=0}. \end{aligned}$$

The function $(t, z) \mapsto f(g(t) + zg'(t))$ has $j+1$ continuous derivatives on an open set in \mathbb{K}^2 , and so the partial derivatives can be interchanged to give

$$\frac{d^{j+1}}{dt^{j+1}} f \circ g(t) = \left. \frac{\partial}{\partial z} \frac{\partial^j}{\partial t^j} f(g(t) + zg'(t)) \right|_{z=0}.$$

Hence, by the inductive hypothesis,

$$\begin{aligned} \frac{d^{j+1}}{dt^{j+1}} f \circ g(t) &= \left. \frac{\partial}{\partial z} \Delta_j f(g(t) + zg'(t), g'(t) + zg''(t), \dots, g^{(j)}(t) + zg^{(j+1)}(t)) \right|_{z=0} \\ &= \Delta_{j+1} f(g(t), g'(t), \dots, g^{(j+1)}(t)). \end{aligned}$$

The theorem follows by induction. □

We can give an alternative expression for $\Delta_n f$ and hence for the chain rule (3) in terms of the elementary symmetric functions σ_k . If $z = (z_1, \dots, z_n)$ then $\sigma_k(z)$ will denote the elementary symmetric function of degree k in z_1, \dots, z_n , so that, for all $\lambda \in \mathbb{C}$,

$$(\lambda + z_1) \cdots (\lambda + z_n) = \lambda^n + \sigma_1(z)\lambda^{n-1} + \cdots + \sigma_n(z).$$

LEMMA 1 *Let $f : \Omega \subset X \rightarrow Y$ have n continuous Fréchet derivatives, where X, Y are Banach spaces and Ω is an open set. For any $x_0 \in \Omega$ and $x_1, \dots, x_n \in X$,*

$$\Delta_n f(x_0, x_1, \dots, x_n) = \frac{\partial^n}{\partial z_n \partial z_{n-1} \cdots \partial z_1} f(x_0 + \sigma_1(z)x_1 + \cdots + \sigma_n(z)x_n) \Big|_{z=(0, \dots, 0)}.$$

PROOF. The case $n=1$ is simple. Suppose that $n>1$ and that the equality holds for $n-1$. If $z' = (z_1, \dots, z_{n-1})$ we have

$$\begin{aligned} \Delta_n f(x_0, x_1, \dots, x_n) &= \frac{\partial}{\partial z_n} \Delta_{n-1} f(x_0 + z_n x_1, \dots, x_{n-1} + z_n x_n) \Big|_{z_n=0} \\ &= \frac{\partial}{\partial z_n} \frac{\partial^{n-1}}{\partial z_{n-1} \cdots \partial z_1} f(x_0 + z_n x_1 + \sigma_1(z')(x_1 + z_n x_2) \\ &\quad + \cdots + \sigma_{n-1}(z')(x_{n-1} + z_n x_n)) \Big|_{z'=(0, \dots, 0), z_n=0}. \end{aligned}$$

Since, for $0 \leq j \leq n-1$,

$$\sigma_j(z_1, \dots, z_{n-1})z_n + \sigma_{j+1}(z_1, \dots, z_{n-1}) = \sigma_{j+1}(z_1, \dots, z_n),$$

we have

$$\Delta_n f(x_0, x_1, \dots, x_n) = \frac{\partial^n}{\partial z_n \partial z_{n-1} \cdots \partial z_1} f(x_0 + \sigma_1(z)x_1 + \cdots + \sigma_n(z)x_n) \Big|_{z=(0, \dots, 0)}$$

as required. □

COROLLARY 1 *If f, g are as in Theorem 2, then for any $t \in \Psi$,*

$$(f \circ g)^{(n)}(t) = \frac{\partial^n}{\partial z_n \partial z_{n-1} \cdots \partial z_1} f(g(t) + \sigma_1(z)g'(t) + \cdots + \sigma_n(z)g^{(n)}(t)) \Big|_{z=(0, \dots, 0)}.$$

The third chain rule applies to more general composite functions on Banach spaces.

THEOREM 3 *Let W, X, Y be Banach spaces over \mathbb{K} and let Φ, Ω be open subsets of W, X , respectively. Let*

$$g : \Phi \rightarrow \Omega, \quad f : \Omega \rightarrow Y$$

have n continuous Fréchet derivatives, for some $n \in \mathbb{N}$. For any $w_0 \in \Phi$ and $w_1, w_2, \dots, w_n \in W$,

$$\Delta_n f \circ g(w_0, w_1, \dots, w_n) = \Delta_n f(g(w_0), \Delta_1 g(w_0, w_1), \dots, \Delta_n g(w_0, w_1, \dots, w_n)).$$

PROOF. Given any $w_0 \in \Phi$ and $w_1 \in W$, set $G(z) := g(w_0 + zw_1)$ for z in a neighbourhood Ψ of 0, so that $G : \Psi \subseteq \mathbb{K} \rightarrow \Omega$ is continuously differentiable at 0 with $G'(0) = \Delta_1 g(w_0, w_1)$. Then

$$\Delta_1 f \circ g(w_0, w_1) = \frac{\partial}{\partial z} f(g(w_0 + zw_1)) \Big|_{z=0} = (f \circ G)'(0).$$

By Theorem 2,

$$(f \circ G)'(0) = \Delta_1 f(G(0), G'(0)) = \Delta_1 f(g(w_0), \Delta_1 g(w_0, w_1)).$$

Hence

$$\Delta_1 f \circ g(w_0, w_1) = \Delta_1 f(g(w_0), \Delta_1 g(w_0, w_1)),$$

which proves the result in the case $n=1$.

Suppose the assertion holds for some $j < n$, that is, for any $w_0 \in \Phi$ and $w_1, \dots, w_j \in W$,

$$\Delta_j (f \circ g)(w_0, \dots, w_j) = \Delta_j f(g(w_0), \dots, \Delta_j g(w_0, \dots, w_j)).$$

Let $w_0 \in \Phi$, $w_1, \dots, w_{j+1} \in W$ be given and, for z in a neighbourhood Ψ of 0, set

$$Z_0(z) := g(w_0 + zw_1), \dots, Z_j(z) := \Delta_j g(w_0 + zw_1, \dots, w_j + zw_{j+1}).$$

By the inductive hypothesis,

$$\begin{aligned} \Delta_{j+1}(f \circ g)(w_0, \dots, w_{j+1}) &= \frac{\partial}{\partial z} \Delta_j(f \circ g)(w_0 + zw_1, \dots, w_j + zw_{j+1}) \Big|_{z=0} \\ &= \frac{\partial}{\partial z} \Delta_j f(Z_0(z), \dots, Z_j(z)) \Big|_{z=0}. \end{aligned}$$

Define $G: \Psi \subseteq \mathbb{K} \rightarrow \Omega \times X^{j+1}$ by

$$G(z) := (Z_0(z), \dots, Z_j(z)),$$

so that

$$\Delta_{j+1}(f \circ g)(w_0, \dots, w_{j+1}) = (\Delta_j f \circ G)'(0). \quad (4)$$

Write

$$\zeta_0 = Z_0(0), \dots, \zeta_j = Z_j(0)$$

and

$$\zeta_{j+1} = Z_j'(0).$$

Note that, for $1 \leq k \leq j+1$,

$$\zeta_k = Z_{k-1}'(0) = \Delta_k g(w_0, \dots, w_k),$$

whence G is differentiable at 0 with

$$G'(0) = (\zeta_1, \dots, \zeta_{j+1}).$$

By equation (4) and Theorem 2,

$$\begin{aligned} \Delta_{j+1}(f \circ g)(w_0, \dots, w_{j+1}) &= (\Delta_j f \circ G)'(0) = \Delta_1 \Delta_j f(G(0), G'(0)) \\ &= \Delta_1 \Delta_j f((\zeta_0, \dots, \zeta_j); (\zeta_1, \dots, \zeta_{j+1})) \\ &= \frac{\partial}{\partial t} \Delta_j f(\zeta_0 + t\zeta_1, \dots, \zeta_j + t\zeta_{j+1}) \Big|_{t=0} \\ &= \Delta_{j+1} f(\zeta_0, \dots, \zeta_{j+1}) \\ &= \Delta_{j+1} f(g(w_0), \dots, \Delta_{j+1} g(w_0, \dots, w_{j+1})). \end{aligned}$$

Therefore, by induction, the formula holds for $n \in \mathbb{N}$. □

Let us note some simple properties of higher directional derivatives.

THEOREM 4 *Let X, Y be Banach spaces over \mathbb{K} , let Ω be an open set in X , and let $f: \Omega \rightarrow Y$ have n continuous Fréchet derivatives. For any $x_0 \in \Omega$, $x_1, \dots, x_n, \xi_1, \xi_2 \in X$ and $\alpha \in \mathbb{K}$,*

$$\Delta_n f(x_0, \alpha x_1, \dots, \alpha^n x_n) = \alpha^n \Delta_n f(x_0, x_1, \dots, x_n)$$

and

$$\Delta_n f(x_0, \dots, x_{n-1}, \alpha \xi_1 + (1 - \alpha)\xi_2) = \alpha \Delta_n f(x_0, \dots, x_{n-1}, \xi_1) + (1 - \alpha) \Delta_n f(x_0, \dots, x_{n-1}, \xi_2).$$

These properties follow simply by induction. Note that $\Delta_n f(x_0, x_1, \dots, x_n)$ is affine in x_n . Second, a version of Leibniz's formula holds for Δ_n .

THEOREM 5 *Let W, X_1, X_2 , and Y be Banach spaces, let Ψ be an open subset of W , and let $B: X_1 \times X_2 \rightarrow Y$ be a continuous bilinear operator. Let $f: \Psi \rightarrow X_1$ and $g: \Psi \rightarrow X_2$ be n times continuously Fréchet differentiable. Then, for any $(w_0, w_1, \dots, w_n) \in \Psi \times W^n$,*

$$\Delta_n(B \circ (f \times g))(w_0, w_1, \dots, w_n) = \sum_{m=0}^n \binom{n}{m} B(\Delta_m f(w_0, \dots, w_m), \Delta_{n-m} g(w_0, \dots, w_{n-m})).$$

Examples

Three examples will suffice to illustrate the chain rule formulae in the preceding section.

EXAMPLE 2: Derivatives of e^g .

On combining Theorem 2 and Example 1 we find, for any \mathcal{C}^3 function g and $t \in \mathbb{C}$,

$$\frac{d^3}{dt^3} e^{g(t)} = \Delta_3 \exp(g(t), g'(t), g''(t), g'''(t)) = e^{g(t)} (g'(t)^3 + 3g'(t)g''(t) + g'''(t)).$$

By induction for any $n \in \mathbb{N}$ there is a polynomial Y_n of degree n in x_1, \dots, x_n such that

$$\Delta_n \exp(x_0, x_1, \dots, x_n) = e^{x_0} Y_n(x_1, \dots, x_n),$$

and so that

$$\frac{d^n}{dt^n} e^{g(t)} = e^{g(t)} Y_n(g'(t), \dots, g^{(n)}(t))$$

whenever g is n times continuously differentiable. The Y_n can be derived recursively: $Y_1(x_1) = x_1$, and the relation

$$e^{x_0} Y_{n+1}(x_0, \dots, x_{n+1}) = \frac{\partial}{\partial z} e^{x_0 + zx_1} Y_n(x_1 + zx_2, \dots, x_n + zx_{n+1}) \Big|_{z=0}$$

translates into the recursion

$$Y_{n+1}(x_0, \dots, x_{n+1}) = x_1 Y_n(x_1, \dots, x_n) + \sum_{j=1}^n x_{j+1} \frac{\partial Y_n}{\partial x_j}(x_1, \dots, x_n).$$

The polynomials Y_n have a long history. They are named after Eric T. Bell, who studied them in 1934 [1], but Johnson [8, Section 2] traces related polynomials to sundry 19th-century authors, including Faà di Bruno himself. Y_n can be defined combinatorially, or by means of the elegant determinantal formula (which apparently is essentially due to Faà di Bruno)

$$Y_n(x_1, \dots, x_n) = \begin{vmatrix} \binom{n-1}{0} x_1 & \binom{n-1}{1} x_2 & \binom{n-1}{2} x_3 & \dots & \binom{n-1}{n-2} x_{n-1} & \binom{n-1}{n-1} x_n \\ -1 & \binom{n-2}{0} x_1 & \binom{n-2}{1} x_2 & \dots & \binom{n-2}{n-3} x_{n-2} & \binom{n-2}{n-2} x_{n-1} \\ 0 & -1 & \binom{n-3}{0} x_1 & \dots & \binom{n-3}{n-4} x_{n-3} & \binom{n-3}{n-3} x_{n-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \binom{1}{0} x_1 & \binom{1}{1} x_2 \\ 0 & 0 & 0 & \dots & -1 & \binom{0}{0} x_1 \end{vmatrix}.$$

Faà di Bruno's formula can be succinctly expressed in terms of the coefficients of the Bell polynomials Y_n ; for proofs and much interesting history, see Johnson, *loc. cit.* The purpose of this note is rather to exhibit some simpler forms of higher chain rules.

EXAMPLE 3: Let us calculate the third derivative of $f(g_1(t), g_2(t))$ for suitably smooth functions f on \mathbb{R}^2 , g_j on \mathbb{R} . For $x_j, y_j \in \mathbb{R}$ we find

$$\begin{aligned} \Delta_1 f((x_0, y_0), (x_1, y_1)) &= \left(x_1 \frac{\partial}{\partial x} + y_1 \frac{\partial}{\partial y} \right) f(x_0, y_0), \\ \Delta_2 f((x_0, y_0), (x_1, y_1), (x_2, y_2)) &= \left(\left(x_2 \frac{\partial}{\partial x} + y_2 \frac{\partial}{\partial y} \right) f + \left(x_1 \frac{\partial}{\partial x} + y_1 \frac{\partial}{\partial y} \right)^2 f \right) (x_0, y_0), \\ \Delta_3 f((x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3)) &= \left(\left(x_3 \frac{\partial}{\partial x} + y_3 \frac{\partial}{\partial y} \right) f + 3 \left(x_2 \frac{\partial}{\partial x} + y_2 \frac{\partial}{\partial y} \right) \left(x_1 \frac{\partial}{\partial x} + y_1 \frac{\partial}{\partial y} \right) f + \left(x_1 \frac{\partial}{\partial x} + y_1 \frac{\partial}{\partial y} \right)^3 f \right) (x_0, y_0). \end{aligned}$$

Hence

$$\begin{aligned} \frac{d^3}{dt^3} f(g_1(t), g_2(t)) &= \Delta_3 f((g_1(t), g_2(t)), \dots, (g_1'''(t), g_2'''(t))) \\ &= \left(\left(g_1'''(t) \frac{\partial}{\partial x} + g_2'''(t) \frac{\partial}{\partial y} \right) f \right. \\ &\quad \left. + 3 \left(g_1''(t) \frac{\partial}{\partial x} + g_2''(t) \frac{\partial}{\partial y} \right) \left(g_1'(t) \frac{\partial}{\partial x} + g_2'(t) \frac{\partial}{\partial y} \right) f \right. \\ &\quad \left. + \left(g_1'(t) \frac{\partial}{\partial x} + g_2'(t) \frac{\partial}{\partial y} \right)^3 f \right) (g_1(t), g_2(t)). \end{aligned}$$

More generally, one can express $\Delta_n f$ in terms of higher-order Fréchet derivatives of f , e.g.,

$$\Delta_3 f(x_0, x_1, x_2, x_3) = Df(x_0)(x_3) + 3D^2f(x_0)(x_1, x_2) + D^3f(x_0)(x_1, x_1, x_1).$$

In view of Theorem 2, it is clear that a similarly explicit expression for $\Delta_n f$ would be equivalent to a version of the Faà di Bruno formula; indeed, from equation (1) (or its multivariate generalization) we must have

$$\Delta_n f(x_0, \dots, x_n) = \sum \frac{n!}{k_1! \dots k_n! (1!)^{k_1} \dots (n!)^{k_n}} D^k f(x_0)(x_1, \dots, x_1, \dots, x_n, \dots, x_n)$$

where, on the right-hand side, the argument x_j is repeated k_j times, and the sum is taken over all non-negative integer solutions for the k_j of the Diophantine equation $k_1 + 2k_2 + \dots + nk_n = n$, and $k := k_1 + \dots + k_n$. An independent proof of this formula, together with Theorem 2, would yield yet another proof of the Faà di Bruno formula.

Here is a multivariate example.

EXAMPLE 4: Let $f_m : \mathbb{C}^{k \times k} \rightarrow \mathbb{C}$ be defined by $f_m(X) = \text{tr}^2(X^m)$, and let us calculate $\Delta_2 f_m(X_0, X_1, X_2)$.

Write $p_m(X) = X^m$. By Theorem 3 we have

$$\begin{aligned} \Delta_2 f_m(X_0, X_1, X_2) &= \Delta_2(\text{tr}^2 \circ p_m)(X_0, X_1, X_2) \\ &= \Delta_2 \text{tr}^2(p_m(X_0), \Delta_1 p_m(X_0, X_1), \Delta_2 p_m(X_0, X_1, X_2)). \end{aligned}$$

On the one hand,

$$\begin{aligned} \Delta_1 \text{tr}^2(X_0, X_1) &= 2\text{tr} X_0 \text{tr} X_1, \\ \Delta_2 \text{tr}^2(X_0, X_1, X_2) &= 2\text{tr}^2 X_1 + 2\text{tr} X_0 \text{tr} X_2. \end{aligned}$$

On the other hand,

$$\begin{aligned} \Delta_1 p_m(X_0, X_1) &= \left. \frac{\partial}{\partial z} (X_0 + zX_1)^m \right|_{z=0} = \sum_{\substack{\alpha, \beta \geq 0 \\ \alpha + \beta = m-1}} X_0^\alpha X_1 X_0^\beta, \\ \Delta_2 p_m(X_0, X_1, X_2) &= \left. \frac{\partial}{\partial z} \sum_{\substack{\alpha, \beta \geq 0 \\ \alpha + \beta = m-1}} (X_0 + zX_1)^\alpha (X_1 + zX_2) (X_0 + zX_1)^\beta \right|_{z=0} \\ &= \sum_{\substack{\alpha, \beta \geq 0 \\ \alpha + \beta = m-1}} X_0^\alpha X_2 X_0^\beta + \sum_{\substack{\alpha, \beta, \gamma \geq 0 \\ \alpha + \beta + \gamma = m-2}} X_0^\alpha X_1 X_0^\beta X_1 X_0^\gamma. \end{aligned}$$

Therefore

$$\begin{aligned} \Delta_2 f_m(X_0, X_1, X_2) &= 2\text{tr}^2 \Delta_1 p_m(X_0, X_1) + 2\text{tr} p_m(X_0) \text{tr} \Delta_2 p_m(X_0, X_1, X_2) \\ &= 2m^2 \text{tr}^2(X_0^{m-1} X_1) + \\ &\quad 2\text{tr}(X_0^m) \times \left\{ m \text{tr}(X_0^{m-1} X_2) + \sum_{\beta=0}^{m-2} (m-1-\beta) \text{tr}(X_0^{m-2-\beta} X_1 X_0^\beta X_1) \right\}. \end{aligned}$$

ACKNOWLEDGEMENTS

The first and second named authors gratefully acknowledge the hospitality of the School of Mathematics and Statistics of the University of Newcastle upon Tyne. The third named author carried out this work while at Newcastle with partial support from EPSRC grant GR/T 20014/01.

REFERENCES

- [1] E. T. Bell, Exponential polynomials, *Ann. of Math. (2)* **35** (1934) 258–277.
- [2] G. M. Constantine and T. H. Savits, A multivariate Faà di Bruno formula with applications, *Transactions of the American Mathematical Society* **348**(2) (1996), 503–520.
- [3] Cav. F. Faà di Bruno, Sullo sviluppo delle funzioni, *Annali di Scienze Matematiche e Fisiche* **6** (1855), 479–480.
- [4] Cav. F. Faà di Bruno, Note sur une nouvelle formule de calcul différentiel, *Quarterly J. Pure Appl. Math.* **1** (1857) 359–360.
- [5] http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Faa_di_Bruno.html.
- [6] H.-N. Huang, S. A. M. Marcantognini, and N. J. Young, The spectral Carathéodory-Fejér problem, to appear in *Integr. Equ. Oper. Theory*.
- [7] I. L. Iribarren, *Calculo Diferencial en Espacios Normados*, Equinoccio, 1980.
- [8] W. P. Johnson, The curious history of Faà di Bruno's formula, *American Mathematical Monthly* **109** (2002), 217–234.
- [9] R. L. Mishkov, Generalization of the formula of Faà di Bruno for a composite function with a vector argument, *Internat. J. Math. Sci.* **24**(7) (2000), 481–491.
- [10] J. D. DePree, and C. W. Swartz, *Introduction to Real Analysis*, John Wiley & Sons, Inc., New York, 1998.
- [11] K. Spindler, A short proof of the formula of Faà di Bruno, *Elemente der Mathematik* **60** (2005), 33–35.

Solutions to "Ein Kleines Schach"

Problems on page 49

TIMOTHY CHOW

Solution to Problem 1

One first verifies that White is not in check from the Black queen, and so is not constrained by having to parry a check. 1.Ba5 is checkmate; all the flight squares of the king are covered, and the queen cannot capture the bishop. Trying to interpose the queen also does not work, because the bishop is checking from two opposite directions.

Closer analysis reveals that the bishop's moves enjoy a symmetry about the horizontal axis of the board; that is, barring obstructions, if the bishop can move to a certain square X on the board, then it can also move to the mirror image of X in the horizontal axis. It follows from this symmetry that 1.Ba4 is also checkmate.

Solution to Problem 2

Black plays Ka6, White plays Nf2, Black plays Kb7, and White checkmates with Nh1.

On an ordinary chess board, knight moves have parity; that is, if a knight can reach a certain square in an even number of moves, then it cannot also

reach the same square in an odd number of moves. Chess players say that a knight "cannot lose a tempo." However, on a Klein bottle, this is not true, because there are odd cycles (in the graph whose edges connect squares that are a knight's move apart), as this problem illustrates.

Note that pawns "reverse direction" when they cross the left or right edge of the board, and that the standard starting position in chess is illegal on a Klein bottle. This raises some questions about the proper rules for pawn promotion, double moves, and *en passant*, which we have glossed over in this problem, even though pawns appear. Fortunately, almost any reasonable way of settling these questions leaves the solution unchanged, as long as one stipulates (as we did) that Black's pawns are moving down the board and White's pawn is moving up the board in the set position.

Center for Communications Research
Princeton, NJ 08540
USA
e-mail: tchow@alum.mit.edu