Faà di Bruno's formula for variational calculus

Daniel E. Clark and Jeremie Houssineau

School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK.

Abstract

This paper determines the general formula for describing differentials of composite functions in terms of differentials of their factor functions. This generalises the formula commonly attributed to Faà di Bruno to functions in locally convex topological vector spaces. The result highlights the general structure of the higher-order chain rule in terms of partitions of the directions.

1. Introduction

Mathematicians have investigated formulae for expressing higher-order derivatives of composite functions in terms of derivatives of their factor functions for over 200 years. These formulae are often attributed to Faà di Bruno [4], though Craik [3] recently highlighted a number of researchers preceding his works, the earliest of which is thought to be by Arbogast [1]. Despite the fact that the idea of expressing these formulae in terms of derivatives of the factor functions is not new, a number of recent works have appeared on this topic, including those by Hardy [6] and Ma [8] on partial derivatives, and an alternative approach was presented by Huang *et al.* for Fréchet derivatives [7], though the general form for variational calculus was previously undetermined.

The paper is structured as follows. Section 2 describes the Gâteaux differential [5] and the chain differential [2]. Section 3 determines the general higher-order chain rule. The paper concludes with a short discussion in Section 4.

2. Gâteaux differentials and chain differentials

In this section we discuss two different forms of differential, the Gâteaux differential [5] and the chain differential [2].

We justify adopting the chain differential to determine Faà di Bruno's formula in Section 3, since it is possible to determine the chain rule whilst maintaining the general structure.

2.1. Gâteaux differentials

The following two definitions describe the Gâteaux differential and its n^{th}-order differential.

Definition 1 (Gâteaux differential). Let X and Y be locally convex topological vector spaces, and let Ω be an open subset of X and let $f : \Omega \to Y$. The Gâteaux differential at $x \in \Omega$ in the direction $\eta \in X$ is

$$\delta f(x;\eta) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(f(x + \epsilon\eta) - f(x) \right) \tag{1}$$

when the limit exists. If $\delta f(x;\eta)$ exists for all $\eta \in X$ then f is Gâteaux differentiable at x. The Gâteaux differential is homogeneous of degree one in η, so that for all real numbers α, $\delta f(x;\alpha\eta) = \alpha\delta f(x;\eta)$.

Definition 2 (n^{th}-order Gâteaux differential). The *n^{th}-order variation* of $f(x)$ in directions $\eta_1, \ldots, \eta_n \in X$ is defined recursively with

$$\delta^n f\left(x;\eta_1,\ldots,\eta_n\right) = \delta\left(\delta^{n-1} f\left(x;\eta_1,\ldots,\eta_{n-1}\right);\eta_n\right).$$

2.2. Chain differentials

Due to the lack of continuity properties of the Gâteaux differential, further constraints are required in order to derive a chain rule. Bernhard [2] proposed a new form of Gâteaux differential defined with sequences, which he called the chain differential, that is not as restrictive as the Fréchet derivative though it is still possible to find a chain rule that maintains the general structure.

Definition 3 (Chain differential). The function $f : X \to Y$, where X and Y are locally convex topological vector spaces, has a *chain differential* $\delta f(x;\eta)$ at x in the direction η if, for any sequence $\eta_n \to \eta \in X$, and any sequence of real numbers $\theta_n \to 0$, it holds that the following limit exists

$$\delta f(x;\eta) = \lim_{n \to \infty} \frac{1}{\theta_n} \left(f(x + \theta_n \eta_n) - f(x) \right).$$

Lemma 1 (Chain rule, from [2], Theorem 1). *Let X, Y and Z be locally convex topological vector spaces, $f : Y \to Z$, $g : X \to Y$ and g and f have chain differentials at x in the direction η and at $g(x)$ in the direction $\delta g(x;\eta)$ respectively. Let $h = f \circ g$, then h has a chain differential at x in the direction η, given by the chain rule*

$$\delta h(x;\eta) = \delta f(g(x);\delta g(x;\eta)).$$

In order to determine the general higher-order chain rule, it is useful to introduce the partial and total chain differentials as follows.

Definition 4 (Partial chain differential). Let $\{X_i\}_{i=1:n}$ and Y be locally convex topological vector spaces. The function $f : X_1 \times \ldots \times X_n \to Y$ has a *partial chain differential* with respect to the i^{th} variable $\delta_i f(x_1, \ldots, x_n;\eta)$ at $(x_1, \ldots, x_n)$ in the direction η if, for any sequence $\eta_m \to \eta \in X$, and any sequence of real numbers $\theta_m \to 0$, it holds that the following limit exists

$$\delta_i f(x_1,\ldots,x_n;\eta) = \lim_{m \to \infty} \frac{1}{\theta_m} \left(f(x_1,\ldots,x_i + \theta_m\eta_m,\ldots,x_n) - f(x_1,\ldots,x_n) \right).$$

Theorem 1 (Total chain differential). *Let $\{X_i\}_{i=1:n}$ and Y be locally convex topological vector spaces. The function $f : X_1 \times \ldots \times X_n \to Y$ has a total chain differential δf at $(x_1, \ldots, x_n)$ if*

1. *all the partial chain differentials exist in a neighbourhood $\Omega \subseteq X_1 \times \ldots \times X_n$ of $(x_1, \ldots, x_n)$ and in any direction, and*
2. *$\delta_i f$ is continuous over $\Omega \times X_i$.*

Then for $\underline{\eta} \in X_1 \times \ldots \times X_n$ such that $\underline{\eta} = (\eta_1, \ldots, \eta_n)$,

$$\delta f(x_1, \ldots, x_n; \underline{\eta}) = \sum_{i=1}^{n} \delta_i f(x_1, \ldots, x_n; \eta_i).$$

Proof. The result is proved in the case $n = 2$ from which the general case can be straightforwardly deduced:

$$\begin{aligned}
\delta f\,(x, y; (\eta, \xi)) &= \lim_{r\to\infty} \theta_r^{-1} \left[f(x + \theta_r \eta_r, y + \theta_r \xi_r) - f(x, y) \right] \\
&= \lim_{r\to\infty} \left(\theta_r^{-1} \left[g_1(y + \theta_r \eta_r) - g_1(y) \right] + \theta_r^{-1} \left[g_2(x + \theta_r \eta_r) - g_2(x) \right] \right),
\end{aligned}$$

where we define $g_1(y)$ and $g_2(x)$ as follows:

$$\begin{cases} g_1(y) = f(x + \theta_r \eta_r, y), \\ \\ g_2(x) = f(x, y). \end{cases}$$

Given $\theta_r \neq 0$, define $h : \mathbb{R} \to \mathbb{R}$ as $h(t) = g_1(y + t\xi_r)$. From the mean value theorem for real-valued functions, there exists $c_y \in [0, \theta_r]$ such that

$$\theta_r^{-1} \left[h(\theta_r) - h(0) \right] = \left. \frac{\mathrm{d}h}{\mathrm{d}t} \right|_{t=c_y} = \delta h(c_y; 1),$$

which, when replacing $h(t)$ by $g_1(y + t\xi_r)$, can be rewritten

$$\begin{aligned}
\theta_r^{-1} \left[g_1(y + \theta_r \xi_r) - g_1(y) \right] &= \delta \left(g_1(y + c_y \xi_r); 1 \right) \\
&= \delta g_1(y + c_y \xi_r; \xi_r),
\end{aligned}$$

where Lemma 1 has been used. Similarly for $g_2(x)$, there exists $c_x \in [0, \theta_r]$ such that

$$\theta_r^{-1} \left[g_2(x + \theta_r \eta_r) - g_2(x) \right] = \delta g_2(x + c_x \eta_r; \eta_r).$$

The last step in the proof is to demonstrate that the limit of the following term

$$\left| \delta g_2(x + c_x \eta_r; \eta_r) + \delta g_1(y + c_y \xi_r; \xi_r) - \delta_1 f\,(x, y; \eta_r) - \delta_2 f\,(x, y; \xi_r) \right|, \tag{2}$$

is equal to 0 when $r \to \infty$. By the triangle inequality, (2) is bounded above by the following summation

$$\left| \delta g_2(x + c_x \eta_r; \eta_r) - \delta_1 f\,(x, y; \eta_r) \right| + \left| \delta g_1(y + c_y \xi_r; \xi_r) - \delta_2 f\,(x, y; \xi_r) \right|.$$

Substituting g_1 and g_2 with f, we have

$$\left|\delta_1 f\left(x+c_x\eta_r,y;\eta_r\right)-\delta_1 f\left(x,y;\eta_r\right)\right|+\left|\delta_2 f\left(x,y+c_y\xi_r;\xi_r\right)-\delta_2 f\left(x,y;\xi_r\right)\right|,$$

which tends to 0 when $r\to\infty$ because of the continuity of $\delta_1 f$ and $\delta_2 f$.

The following result is then proved:

$$\delta f\left(x,y;(\eta,\xi)\right)=\delta_1 f\left(x,y;\eta\right)+\delta_2 f\left(x,y;\xi\right)$$

which is equivalent to the Proposition 3 in [2].

3. Faà di Bruno's formula

The next result generalises the higher-order chain rule to variational calculus.

Theorem 2 (General higher-order chain rule). *Let X, Y and Z be locally convex topological vector spaces. Assume that $g: X\to Y$ has higher order chain differentials in any number of directions in the set $\{\eta_1,\dots,\eta_n\}\in X^n$ and that $f: Y\to Z$ has higher order chain differentials in any number of directions in the set $\{\delta^m g(x;S_m)\}_{m=1:n}$, $S_m\subseteq\{\eta_1,\dots,\eta_n\}$. Assume additionally that for all $1\le m\le n$, $\delta^m f(y;\xi_1,\dots,\xi_m)$ is continuous on an open set $\Omega\subseteq Y^{m+1}$ and linear with respect to the directions $\xi_1,\dots,\xi_m$, the n^{th}-order variation of composition $f\circ g$ in directions $\eta_1,\dots,\eta_n$ at point $x\in X$ is given by*

$$\delta^n(f\circ g)(x;\eta_1,\dots,\eta_n)=\sum_{\pi\in\Pi(\eta_{1:n})}\delta^{|\pi|}f\left(g(x);\xi_{\pi_1}(x),\dots,\xi_{\pi_{|\pi|}}(x)\right),$$

where $\xi_\omega(x)=\delta^{|\omega|}g\left(x;\omega_1,\dots,\omega_{|\omega|}\right)$ is the $|\omega|^{th}$-order chain differential of g in directions $\{\omega_1,\dots,\omega_{|\omega|}\}\subseteq\{\eta_1,\dots,\eta_n\}$. $\Pi(\eta_{1:n})$ represents the set of partitions of the set $\{\eta_1,\dots,\eta_n\}$ and $|\pi|$ denotes the cardinality of the set π.

Proof. Lemma 1 gives the base case $n=1$. For the induction step, we apply the differential operator to the case n to give the case $n+1$ and show that it involves a summation over partitions of elements $\eta_1,\dots,\eta_{n+1}$ in the following way

$$\delta^{n+1}(f\circ g)(x;\eta_1,\dots,\eta_{n+1})=\sum_{\pi\in\Pi(\eta_{1:n})}\delta\left(\delta^{|\pi|}f\left(g(x);\xi_{\pi_1}(x),\dots,\xi_{\pi_{|\pi|}}(x)\right);\eta_{n+1}\right).\tag{3}$$

The main objective in this proof is to calculate the term

$$\delta\left(\delta^k f\left(g(x);h_1(x),\dots,h_k(x)\right);\eta\right).\tag{4}$$

The additional differentiation with respect to η applies to every function on X, i.e. to g and to the h_i, where $1\le i\le k$. To highlight the structure of this result, we can define a multi-variate function F such that

$$\begin{aligned}F: Y^{k+1}&\to Z\\(y_0,\dots,y_k)&\mapsto\delta^k f(y_0;y_1,\dots,y_k),\end{aligned}$$

so that (5) can be rewritten $\delta\left(F(g(x),h_1(x),\ldots,h_k(x));\underline{\eta}^{(k+1)}\right)$, when denoting $\underline{\eta}^{(k)}=(\eta,\ldots,\eta)\in X^k$. Using Theorem 1 we find

$$\delta\left(F(g(x),h_1(x),\ldots,h_k(x));\underline{\eta}^{(k+1)}\right)=\sum_{i=1}^{k+1}\delta_i\left(F(g(x),h_1(x),\ldots,h_k(x));\eta\right). \tag{5}$$

From this point, differentiation with respect to $g(x)$ has to be dealt with separately due to the different properties of F with respect to its arguments.

- Consider the first term of the summation in (6):

  $$\delta_1\left(F(g(x),h_1(x),\ldots,h_k(x));\eta\right). \tag{6}$$

  Let $\theta_m\to 0$, $\eta^{(m)}\to\eta$, and let ϕ_m be defined as

  $$\phi_m(x)=\theta_m^{-1}(g(x+\theta_m\eta^{(m)})-g(x)).$$

  Following Theorem 1 in Bernhard [2], we have

  $$\begin{cases}\phi_m(x)\to\delta g(x;\eta),\\ g(x+\theta_m\eta^{(m)})=g(x)+\theta_m\phi_m(x).\end{cases} \tag{7}$$

  One can then rewrite (7) as the limit when $m\to\infty$ of

  $$\theta_m^{-1}\left(F(g(x)+\theta_m\phi_m(x),h_1(x),\ldots,h_k(x))-F(g(x),h_1(x),\ldots,h_k(x))\right),$$

  which, when taking the limit and using (8), can be expressed as

  $$\delta_1\left(F(g(x),h_1(x),\ldots,h_k(x));\eta\right)=\delta_1 F(g(x),h_1(x),\ldots,h_k(x);\delta g(x,\eta)). \tag{8}$$

- Now consider all the other terms in (6):

  $$\delta_i\left(F(g(x),h_1(x),\ldots,h_k(x));\eta\right),\qquad 2<i\leq k+1. \tag{9}$$

  Let $\theta_m\to 0$ and $\eta^{(m)}\to\eta$. The terms in (10) can be expressed as the limit when $m\to\infty$ of

  $$\theta_m^{-1}\left(F(g(x),h_1(x),\ldots,h_i(x+\theta_m\eta^{(m)}),\ldots,h_k(x))-F(g(x),h_1(x),\ldots,h_k(x))\right).$$

  However, due to the linearity of F with respect to all its arguments except the first, the previous expression can be simplified and written as

  $$F(g(x),h_1(x),\ldots,\theta_m^{-1}\left(h_i(x+\theta_m\eta^{(m)})-h_i(x)\right),\ldots,h_k(x)).$$

  Taking the limit, this becomes

  $$\delta_i\left(F(g(x),h_1(x),\ldots,h_k(x));\eta\right)=F(g(x),h_1(x),\ldots,\delta h_i(x,\eta),\ldots,h_k(x)). \tag{10}$$

Considering $k = |\pi|$, $\eta = \eta_{n+1}$ and $h_i = \xi_{\pi_i}$ and replacing the results (9) and (11) into (4), we find

$$\delta^{n+1}(f \circ g)(x; \eta_1, \dots, \eta_{n+1}) = \\ \sum_{\pi \in \Pi(\eta_{1:n})} \delta^{|\pi|+1} f\left(g(x); \xi_{\pi_1}(x), \dots, \xi_{\pi_{|\pi|}}(x), \delta g(x, \eta_{n+1})\right) \\ + \sum_{\pi \in \Pi(\eta_{1:n})} \sum_{i=1}^{|\pi|} \delta^{|\pi|} f\left(g(x); \xi_{\pi_1}(x), \dots, \delta\xi_{\pi_i}(x; \eta_{n+1}), \dots, \xi_{\pi_{|\pi|}}(x)\right).$$

Following a similar argument used for the recursion of Stirling numbers of the second kind and their relation to Bell numbers [9, p74], the result above can be viewed as a means of generating all partitions of $n+1$ elements from all partitions of n elements: The first term corresponds to the creation of a new subset only containing η_{n+1}, and each term in the second summation appends η_{n+1} to one of the existing subset in $\pi \in \Pi(\eta_{1:n})$. This argument follows similar arguments previously used for ordinary and partial derivatives [8, 7, 6]. Hence the result is proved by induction.

4. Discussion

It is worth highlighting the structure of the result. In other forms of chain rule, Faà di Bruno's formula is a sum over partitions of products. However, in the general form for variational calculus, the outer functional has variations in directions that themselves are differentials of the inner functional.

Acknowledgements

Daniel Clark is a Royal Academy of Engineering/EPSRC Research Fellow. Jeremie Houssineau has a PhD scholarship sponsored by DCNS and Heriot-Watt University.

References

[1] Arbogast, L., 1800. Du Calcul des Derivations. Levrault, Strasbourg.

[2] Bernhard, P., 2005. Chain differentials with an application to the mathematical fear operator. Nonlinear Analysis 62, 1225–1233.

[3] Craik, A., 2005. Prehistory of Faa di Bruno's Formula. Amer. Math. Monthly 2, 119–130.

[4] F. Faà di Bruno, 1857. Note sur un nouvelle formule de calcul différentiel. Quarterly Journal of Pure and Applied Mathematics 1, 359–360.

[5] Gâteaux, R., 1919. Fonctions d'une infinité de variables indépendantes. Bulletin de la Société Mathématique de France 47, 70–96.

[6] Hardy, M., 2006. Combinatorics of partial derivatives. Electronic Journal of Combinatorics 13, R1.

[7] Huang, H.-N., Marcantognini, S. A. M., Young, N. J., 2006. Chain Rules for Higher Derivatives. The Mathematical Intelligencer 28(2), 61–69.

[8] Ma, T., 2009. Higher Chain Formula proved by Combinatorics. Electronic Journal of Combinatorics 16, N21.

[9] Stanley, R., 2012. Enumerative Combinatorics, 2nd Edition. Vol. 1. Cambridge University Press.