Math Diversion Problem 434: Structured Differentiation: Lesson One

P. Reany

February 25, 2025

> Matters of notation play a role in
> connection with the chain rule. Wide
> varieties of usage exist where the
> chain rule is concerned.
> — Taylor & Mann [6]

1 Preface

This paper is meant to be the first of a series of articles on doing partial differentiation using Structured Differentiation (SD), which I have been working on for four decades. I've already written extensively on this topic, so I have a lot of material to borrow from to write this new series. I admit that the scope of my presentations may take the discussion outside of usual domain of math competitions, but then again, maybe not. Anyway, a serious mathematician needs to conquer this subject matter.

It seemed natural for me to look beyond the usual problems in complex numbers, algebra, and trig, to continue this series, or I'd probably soon run out of things to present. So, when I noticed that a significant number of downloads were made to my previous papers on partial differentiation, so it seemed natural for me to pursue it in depth.

2 Introduction

Anyone who has made a serious comparison of the various formalisms for differentiation used by mathematicians and physicists has surely noticed that two criteria dominate. The first is to symbolize distinction between explicit and implicit derivatives; the second is to represent a generalized total derivative. Add to this the fact that physics undergraduates usually learn differentiation from mathematicians who use still other formalisms which reflect their own design criteria.

The purpose of this paper is to present a structured, semantically unified formalism for differentiation to meet the needs of the undergraduate and graduate technical student. Much effort has been taken to incorporate the good ideas from the formalisms already in use, emphasizing the effect that making a change in the underlying 'independent' variables will have on the derivatives of variables of interest. Formally, we shall refer to this formalism as *structured differentiation* or SD. Mathematically, SD is not the most general formalism available for differentiation, nor is it meant to be. It does, however, address the theory and practice of maps from R^n to R^m, and it is most notably a matrix formulation of differentiation. I will not, however, be emphasizing the technical details of differentiability.

One of the major advantages of SD is the new terminology introduced to refer to some of the important features of differentiation left unnamed in conventional literature. Another advantage, semantically, is the use of three distinct symbols representing the three distinct derivatives: *total, explicit*, and *implicit*.

The student who has mastered SD can either use it directly, or else, because of the clarification that SD gives to the conventional approaches, he can use the conventional formalisms with greater ease and confidence.

Warning: There are a few technical aspects of differentiability that will be useful for the reader to know at certain places in the series. Case in point, familiarity with the Implicit Function Theorem will be helpful.

3 Standard Partial Differentiation

Let's now review the standard definition of a 'partial derivative': Let $f : R^n \to R$ be a function of the n variables $\{x_1, ..., x_n\}$. The partial derivative of f with respect to x_i (if it exists) at the point $(x_1, \ldots, x_n)$ is

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, x_2, \ldots, x_i + h, \ldots, x_n) - f(x_1, x_2, \ldots, x_i, \ldots, x_n)}{h}\,. \tag{1}$$

Before continuing, I need to define a term. The *variant vector* (or just *variant*) of a function is the ordered tuple of all the variables on which the function is explicitly dependent. Conventional calculus has no equivalent to this definition that I am aware of. It plays, however, a dominant role in SD by helping students grasp the meaning of the different derivatives which correspond to different types of functional dependence, that is, for example, explicit verses implicit. More about this later.

Let's look at some examples. In the function definition

$$f(x, y) = x + 3y\,, \tag{2}$$

where f is defined on a region of the x, y-plane, then the variant of f is $\mathbf{x} = (x, y)$.[1] The individual variables x and y are said to be *variants* of f.

---

[1] I'm cheating here by laying the vector on the line, but when we get to matrices, I will usually specify if the vector is a row vector or a column vector.

What then would be the variant of g defined this way:

$$g(x(t), y(t)) = x + 3y\,? \tag{3}$$

It would be $\mathbf{x} = (x, y)$. When deciding what variables make up the variant of a function, we don't care about the implicit dependencies of the function on variables. What about this case?

$$h(x, y, t) = x + 3yt\,. \tag{4}$$

Yes, $\mathbf{x} = (x, y, t)$. The individual variables x, y, t are said to be *variants* of h.

One more term. A function is said to be *primitive*, or in *primitive form*, if all of its variants are mutually independent of each other. Now, I'm ready to discuss the first thing that bothers me about standard partial differentiation. Let's see if you can catch it.

From Wikipedia:

> In mathematics, a partial derivative of a function of several variables is its derivative with respect to one of those variables, with the others held constant (as opposed to the total derivative, in which all variables are allowed to vary).

From Hubbard & Hubbard ([1]), under 'Partial Derivatives' (p. 127):

> One kind of a derivative of a function of several variables works just like a derivative of a function of one variable: take the derivative with respect to one variable, treating all the others as constants.

From Taylor-Mann, (p. 140):

> Let $f(x, y)$ be defined in a region R of the xy-plane. If we think of y as fixed and x as a variable, the derivative of $f(x, y)$ with respect to x is called the *partial derivative* with respect to x. Likewise the partial derivative with respect to y, $\dfrac{\partial f}{\partial y}$ or $\dfrac{\partial u}{\partial y}$, is the derivative of $f(x, y)$ with respect to y when x is regarded as a constant.

So here's my problem: Why are we 'treating' the other variants of the function as constants with respect to the variable of differentiation? Is it because all the other variables are functionally independent of the variant of differentiation (that is because the function is primitive), or is it because, even if functional dependencies might exist between them, we will simple by fiat 'treat' them as constant with respect to the variant of differentiation? This latter notion is what physicists frequently refer to as an 'explicit derivative', for obvious reasons. But the mathematicians I have corresponded with claimed not to know what an explicit derivative is, unless perhaps they had a good familiarity with physics literature.

If partial derivatives were only to be applied to primitive functions, then the subject would be trivial and uncomplicated, but it isn't. And how various authors generalize so-called 'partial differentiation' to non-primitive functions is where most of the complications, confusions, and ambiguities arise. The other main source of confusion is how to denote a generalized total derivative. We'll get to all of this in due time.

My objective is to sort this all out so that one can read the mathematics literature and the physics literature and be able to follow their reasoning on differentiation. To help prove this point, I will be presenting solved problems from the literature of both sides.

4 SD: Basic forms and Definitions

It is customary to introduce students to the concept of a derivative through the concept of a limit. Let $y = f(x)$, then the derivative of y with respect to x is

$$f'(x) = \frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}\,, \tag{5}$$

if the limit exists. We may think of the derivative of a scalar function as the ratio of two differentials—a differential being an infinitesimal quantity. To be precise, in SD we define a differential to be any scalar, vector, or matrix whose norm is an infinitesimal.

For example, suppose we want to solve the radioactive decay problem from basic principles; we know from laboratory measurements that in a small interval of time dt, the change in the amount of radioactive material (due only to decay) is given by

$$dA = -kA(t)dt\,, \tag{6}$$

where $A(t)$ is the amount in the sample at time t, and k is a positive constant. Dividing (6) by dt gives

$$\frac{dA}{dt} = -kA(t)\,. \tag{7}$$

Henceforth, we will usually represent a differential by a δ instead of a d, though the δ is equivalent to the d in derivatives that are "ordinary."

The derivative of a vector-valued function $\mathbf{f} = \mathbf{f}(t)$ with respect to t is given by

$$\frac{\delta \mathbf{f}}{\delta t} = \lim_{\Delta t \to 0} \frac{\mathbf{f}(t + \Delta t) - \mathbf{f}(t)}{\Delta t}\,, \tag{8}$$

if the limit exists. $\delta\mathbf{f}/\delta t$ is the *total derivative* of $\mathbf{f}$ with respect to t.

The *partial derivative*, or *gradient* of $\mathbf{f}$ with respect to the vector variable $\mathbf{x}$

is

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \dots & \partial f_1/\partial x_m \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \dots & \partial f_2/\partial x_m \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_n/\partial x_1 & \partial f_n/\partial x_2 & \dots & \partial f_n/\partial x_m \end{bmatrix} \tag{9}$$

where $\mathbf{f} = (f_1, f_2, \dots, f_n)^t$—the superscript t meaning transpose. The ith row of this matrix is the gradient of f_i.

In practice the operator $\partial/\partial\mathbf{x}$ is the most useful derivative for crafting formulas in their simplest form. Nevertheless, as you will see, often those clever equations made of partial derivatives, started from total derivatives. In SD a partial derivative will *always* mean an explicit derivative.

Suppose we take the differential of the ith component of $\mathbf{f}$, and represent it in indicial notation; using the chain rule we have

$$\delta f_i = \sum_j \frac{\partial f_i}{\partial x_j}\delta x_j \qquad (i = 1, 2, \dots, n\,, j = 1, 2, \dots, m)\,. \tag{10}$$

Equivalently, we can write this in matrix form as

$$\begin{bmatrix} \delta f_1 \\ \delta f_2 \\ \vdots \\ \delta f_n \end{bmatrix} = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \dots & \partial f_1/\partial x_m \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \dots & \partial f_2/\partial x_m \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_n/\partial x_1 & \partial f_n/\partial x_2 & \dots & \partial f_n/\partial x_m \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_m \end{bmatrix} \tag{11}$$

Both the indicial and matrix formalisms require the use of a basis which may be totally irrelevant to the problem, and will often be suppressed when dealing with the theory. Furthermore, they are less elegant than their SD equivalent:

$$\delta\mathbf{f} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\delta\mathbf{x}\,. \tag{12}$$

Dividing this by $\delta\mathbf{x}$ we have

$$\frac{\delta\mathbf{f}}{\delta\mathbf{x}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\frac{\delta\mathbf{x}}{\delta\mathbf{x}}\,. \tag{13}$$

We call $\delta\mathbf{f}/\delta\mathbf{x}$ the *total derivative* (or just *derivative*) of $\mathbf{f}$ with respect to $\mathbf{x}$; its matrix form is

$$\frac{\delta \mathbf{f}}{\delta \mathbf{x}} = \begin{bmatrix} \delta f_1/\delta x_1 & \delta f_1/\delta x_2 & \dots & \delta f_1/\delta x_m \\ \delta f_2/\delta x_1 & \delta f_2/\delta x_2 & \dots & f_2/\delta x_m \\ \vdots & \vdots & \ddots & \vdots \\ \delta f_n/\delta x_1 & \delta f_n/\delta x_2 & \dots & \delta f_n/\delta x_m \end{bmatrix}\,. \tag{14}$$

We define the *fundamental vector* (or just *fundamental*) of the function $\mathbf{f}$ to be the ordered tuple of all the "independent" variables of $\mathbf{f}$. (This is a naive introductory definition which will be made precise later by properly defining the notion of "variable independence.") The order of the variables in the fundamental, though arbitrary, must be adhered to throughout any given problem.

It should be clear that $\delta\mathbf{x}/\delta\mathbf{x} = \mathbf{I}$ in (13) is the identity matrix, if and only if $\mathbf{x}$ is the fundamental or a subset of it. In like manner, we say that for some function f, $\delta f/\delta x = 0$ if an only if f is functionally independent of x.

Once again stated, we define the *variant vector* (or just *variant*) of a function to be the ordered tuple of all the variables on which the function is explicitly dependent.

Let $f = f(\mathbf{x})$ where $\mathbf{x}$ is understood to be the variant of f; then the differential of f is

$$\delta f = \frac{\partial f}{\partial \mathbf{x}}\delta\mathbf{x} = \frac{\partial f}{\partial x_1}\delta x_1 + \frac{\partial f}{\partial x_2}\delta x_2 + \cdots + \frac{\partial f}{\partial x_n}\delta x_n \,. \tag{15}$$

On dividing this by the differential of the ith variant of f (usually context makes clear whether "variant" refers to the variant vector or just to one of its components) we have

$$\frac{\delta f}{\delta x_i} = \frac{\partial f}{\partial \mathbf{x}}\frac{\delta \mathbf{x}}{\delta x_i} = \sum_j \frac{\partial f}{\partial x_j}\frac{\delta x_j}{\delta x_i}\,. \tag{16}$$

This can be rewritten as

$$\frac{\delta f}{\delta x_i} = \frac{\partial f}{\partial x_i} + \sum_{j\neq i} \frac{\partial f}{\partial x_j}\frac{\delta x_j}{\delta x_i}\,. \tag{17}$$

The first term on the right of (17) is the partial (i.e., explicit) derivative of f with respect to x_i; the summation on the right we refer to as the *copartial* (implicit) derivative of f with respect to x_i. A more compact notation for the copartial derivative is as follows:

$$\frac{\not\partial f}{\not\partial x_i} \equiv \sum_{j\neq i} \frac{\partial f}{\partial x_j}\frac{\delta x_j}{\delta x_i}\,. \tag{18}$$

Thus (17) becomes

$$\frac{\delta f}{\delta x_i} = \frac{\partial f}{\partial x_i} + \frac{\not\partial f}{\not\partial x_i}\,. \tag{19}$$

We may interpret the partial derivative as an *explicit derivative*, and the copartial derivative as an *implicit derivative*. This leads to the simple result that the deltal derivative is always a total derivative, the partial derivative is always an explicit derivative, and the copartial derivative is always an implicit derivative. Neither mathematical nor scientific convention adhere as strictly to the meaning of the symbol "$\partial/\partial x$" as does SD. The use of a generalized total derivative is of semantic value that students will quickly appreciate. For example, consider the following problem. Given the equation $F(x, y, z) = 0$, what is $\partial F/\partial x$? Students may become confused when they write

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}(0) = 0\,. \tag{20}$$

This would be true if the derivative above were a generalized total derivative, but in mathematical convention it is often an explicit derivative. Often, there's just an ambiguity with its attending confusion.

The deltal derivative can be treated as a derivative "operator"

$$\frac{\delta}{\delta\mathbf{x}} = \frac{\partial}{\partial\mathbf{x}} + \frac{\not\partial}{\not\partial\mathbf{x}}\,, \tag{21}$$

or, more simply as

$$\delta_{\mathbf{x}} = \partial_{\mathbf{x}} + \not\partial_{\mathbf{x}}\,. \tag{22}$$

We shall refer to the separation of the deltal derivative into the sum of the partial and copartial derivatives as a *parametric split.*

With the exception of using δ-differentials instead of d-differentials, the form of SD is a trivial generalization of ordinary differentiation. The advantage in this is that the student's transition from the calculus of one variable to the calculus of many variables is made easier. Consider the problem of implicit differentiation in ordinary calculus. Let y be implicitly a function of x via the relation $f(x,y) = 0$, then what is dy/dx? Differentiating this equation by x yields

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0\,. \tag{23}$$

Therefore,

$$\frac{dy}{dx} = -\left(\frac{\partial f}{\partial y}\right)^{-1}\frac{\partial f}{\partial x}\,, \tag{24}$$

assuming that $\partial f/\partial y \neq 0$. Now consider the vector-valued function $\mathbf{f}(\mathbf{x},\mathbf{y}) = 0$, where $\mathbf{x}$ is the fundamental of $\mathbf{f}$, and $(\mathbf{x},\mathbf{y})$ is the variant of $\mathbf{f}$; then

$$\frac{\delta\mathbf{f}}{\delta\mathbf{x}} = \frac{\partial\mathbf{f}}{\partial\mathbf{x}} + \frac{\partial\mathbf{f}}{\partial\mathbf{y}}\frac{\delta\mathbf{y}}{\delta\mathbf{x}} = \mathbf{0}\,. \tag{25}$$

Assuming that $\dfrac{\partial\mathbf{f}}{\partial\mathbf{y}}$ is an invertible matrix, then

$$\frac{\delta\mathbf{y}}{\delta\mathbf{x}} = -\left(\frac{\partial\mathbf{f}}{\partial\mathbf{y}}\right)^{-1}\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\,. \tag{26}$$

We have not yet specified the type of dependence that $\mathbf{y}$ has on $\mathbf{x}$; if we stipulate that $\mathbf{y} = \mathbf{y}(\mathbf{x})$, or rather, that the variant of $\mathbf{y}$ is also the fundamental of $\mathbf{y}$, then

$$\frac{\partial\mathbf{y}}{\partial\mathbf{x}} = -\left(\frac{\partial\mathbf{f}}{\partial\mathbf{y}}\right)^{-1}\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\,. \tag{27}$$

Equation (27) is close to the form usually found in advanced calculus books.

We say that a function is *primitive* when the function's variant is the same as its fundamental, or madeup of a subset of it, or that the function's individual variants are functionally independent of each other. Conventional mathematics makes heavy use of primitive functions to reduce generalized total derivatives into explicit derivatives; this is because the total derivative of a primitive function reduces to an explicit derivative, since the implicit derivative is identically zero.

5 Problems from mathematical physics

Problem 1.) Level Surfaces of $F(\mathbf{x})$

Let F be a function defined on some n-dimensional space. We seek a point $\mathbf{x}_0$ such that for any allowable infinitesimal change $\delta\mathbf{x}$ from $\mathbf{x}_0$, the value of the function is unchanged, or $\delta F = 0$. But

$$\delta F = \frac{\partial F}{\partial \mathbf{x}}\delta\mathbf{x} = 0\,. \tag{28}$$

Of course, we can interpret

$$\frac{\partial F}{\partial \mathbf{x}} = \nabla F\,, \tag{29}$$

the gradient of F. Using the dot product, (28) becomes (dropping the δF):

$$\nabla F \cdot \delta\mathbf{x} = 0 \tag{30a}$$

or

$$\delta\mathbf{x} \cdot \nabla F = 0\,. \tag{30b}$$

Let $\mathbf{a}$ be an arbitrary unit vector in our full space. Then the rate of change of F in the direction of $\mathbf{a}$ is given by the so-called *directional derivative* of F in the direction $\mathbf{a}$ by

$$\mathbf{a} \cdot \nabla F\,. \tag{31}$$

Dividing (30b) through by $|\,\delta\mathbf{x}\,|$, we get

$$\frac{\delta\mathbf{x}}{|\,\delta\mathbf{x}\,|} \cdot \nabla F = 0\,, \tag{32a}$$

or

$$\mathbf{a} \cdot \nabla F = 0\,, \tag{32b}$$

where

$$\mathbf{a} = \frac{\delta\mathbf{x}}{|\,\delta\mathbf{x}\,|}\,. \tag{32c}$$

What (32b) is saying is that F is not changing in any direction that is tangent to the space defined by the $\delta\mathbf{x}$'s at point $\mathbf{x}_0$.

Let's reboot the problem at hand. Let $F = F(\mathbf{x})$ still be defined on some region, with $\mathbf{x}$ n-dimensional. Now, constrain F by the relation

$$F(\mathbf{x}) = c \tag{33}$$

in that region, with c an arbitrary constant. This constraint has the effect of reducing the dimension of the space on which (33) applies to be $(n-1)$-dimensional subregion, called a *level surface*.[2] On taking the differential across (33), we re-obtain (28) and all its following arguments up to and including (32c). So, if $\mathbf{b}$ is a unit vector lying in the tangent space of a level surface at a given point at which (33) applies, then the rate of change of F in the direction of $\mathbf{b}$ is zero, of course.

Let's recap: Let $\mathbf{x}_0$ be the point (or one point) at which (33) applies. Then the gradient vector $\nabla F(\mathbf{x}_0)$ is orthogonal to every tangent vector to the surface at $\mathbf{x}_0$ defined by (33). The vector $\delta\mathbf{x}$ in (30b) and the vector $\mathbf{a}$ in (32b) are vectors in the tangent space.

The gradient of F at every point on the level surface is orthogonal to every tangent vector $\delta\mathbf{x}$ at the same point in the surface. The gradient of F is defined on the entire n-dimensional space, which contains the level-surface points and that at a point on the surface the gradient is in the direction of maximal change away from the level surface at that point. Why is this? Because if it were not so, then the maximum rate of change of F at $\mathbf{x}_0$ would be in some other direction than the normal direction. This non-orthogonal direction would have a component in the orthogonal direction and a component (projection) onto the tangent space, which would imply a nonzero rate of change in F in that direction, contradicting our assumption that F is not changing value in that direction, at least not infinitesimally close to the point $\mathbf{x}_0$.

Problem II. Lagrange multipliers: a specific case

We wish to find the relative extrema of $F(x, y, z)$ subject to the constraint

$$G(x, y, z) = k\,, \tag{34}$$

where k is an arbitrary constant and $\partial G/\partial z \neq 0$. (If you're interested, compare the SD approach to that of Taylor-Mann ([6]) (192–198).[3]

Since $\partial G/\partial z \neq 0$ and G is constrained according to (34), then by the *Implicit Function Theorem* we can write

$$z = z(x, y) = z(\boldsymbol{\eta})\,, \tag{35}$$

where z is in primitive form. Then

$$F(x, y, z) = F(\boldsymbol{\eta}, z(\boldsymbol{\eta})) = F(\mathbf{x})\,. \tag{36}$$

Now, a necessary condition for an extremum point for F is that

$$\delta F/\delta\boldsymbol{\eta} = \mathbf{0}\,, \tag{37}$$

---

[2] In two-dimensional space, the level surface would be a *level curve*.

[3] It's important to understand that in this type of problem, the purpose of the relation $G(x, y, z) = k$ is to define a subset of 3-space on which the search for a relative extrema of $F(x, y, z)$ will be performed. The geometric meaning of $\partial G/\partial z \neq 0$ is that nowhere on this subset of points will the gradient on this function G have a z-component equal to zero.

which is a direct generalization from ordinary calculus. Therefore, applying the chain rule to expand (37), we get

$$\frac{\partial F}{\partial \boldsymbol{\eta}} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial \boldsymbol{\eta}} = \mathbf{0}\,, \tag{38}$$

where we used the fact that when we expanded $\delta F/\delta\boldsymbol{\eta}$ by the chain rule, we could replace $\delta z/\delta\boldsymbol{\eta}$ by $\partial z/\partial\boldsymbol{\eta}$ since z is in primitive form. Now, by differentiating (34) through by $\boldsymbol{\eta}$, we also have that

$$\frac{\delta G}{\delta \boldsymbol{\eta}} = \frac{\partial G}{\partial \boldsymbol{\eta}} + \frac{\partial G}{\partial z}\frac{\partial z}{\partial \boldsymbol{\eta}} = \mathbf{0}\,. \tag{39}$$

Solving this system for $\partial z/\partial\boldsymbol{\eta}$ and substituting that result into (38) we obtain

$$\frac{\partial F}{\partial \boldsymbol{\eta}} + \lambda\frac{\partial G}{\partial \boldsymbol{\eta}} = \mathbf{0} \tag{40}$$

where

$$\lambda = -\frac{\partial F/\partial z}{\partial G/\partial z} \tag{41}$$

On rewriting this we have

$$\frac{\partial F}{\partial z} + \lambda\frac{\partial G}{\partial z} = 0\,. \tag{42}$$

Combining this and (40) results in

$$\frac{\partial F}{\partial \mathbf{x}} + \lambda\frac{\partial G}{\partial \mathbf{x}} = \mathbf{0}\,. \tag{43}$$

Or, equivalently, for some scalar λ, $F(\mathbf{x})$ must satisfy the condition

$$\nabla F + \lambda \nabla G = \mathbf{0}\,. \tag{44}$$

The parameter λ is referred to as a *Lagrange multiplier.*

So, if (34) and (44) have a solution $\mathbf{x}_0$ for some nonzero λ, then $\mathbf{x}_0$ may be a point of relative extrema of F. Geometrically, this means that on the level surface defined by (34), there exists some nonzero scalar c such that the level surface defined by

$$F(x, y, z) = c \tag{45}$$

has a common point $\mathbf{x}_0$ such that both constrained functions share the same tangent spaces at $\mathbf{x}_0$, and therefore, at that point the gradients of F and G are nonzero scalar multiples of each other.

As a final note on this problem, it is possible to introduce a new function

$$H(x, y, z) = F(x, y, z) + \lambda G(x, y, z) \tag{46}$$

such that the possible relative extrema of $F(x, y, z)$ subject to (34) can be found by solving the system of (34) and

$$\nabla H = \mathbf{0}\,. \tag{47}$$

It appears that H is constant in some 3-dimensional neighborhood of $\mathbf{x}_0$. But some people go a step further, and introduce a new function (I use my own notation to exemplify what is done)

$$J(x, y, z, \lambda) = F(x, y, z) + \lambda G(x, y, z) \tag{48}$$

and then claim that the solution is found by solving the equation

$$\Box J = \mathbf{0}\,, \tag{49}$$

for $\mathbf{x}_0$, where $\Box$ is defined as

$$\Box \equiv (\partial_x, \partial_y, \partial_z, \partial_\lambda)\,. \tag{50}$$

Now, the first three components of (49) reproduce (44) and the fourth component reproduces the original constraint (34), though we are required to rewrite G such that $k = 0$ just to fulfill this unneeded mnemonic trick. My personal view of treating a Lagrange multiplier problem as the solution to an equation like found in (49) is at best a subtle mnemonic that adds nothing cogent to the real understanding of the problem, but does add to the mysteriousness and confusion already inherent in all aspects of partial differentiation. Why add more confusion than is necessary?

References

[1] J. Hubbard and B. Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms.* Prentice Hall NJ. (2002).

[2] K. Huang. *Statistical Mechanics.* John Wiley & Sons. New York. 126 (1963).

[3] S.N. Rasband. *Dynamics.* John Wiley & Sons. New York. 152 (1983).

[4] C. Page. *Physical Mathematics.* D. Van Nostand Co. Princeton, New Jersey. 44 (1955).

[5] M. Abbott and H. Van Ness. *Theory and Problems of Thermodynamics.* Mc-Graw Hill Book Co. 66-97 (1972).

[6] A. Taylor and W.R. Mann. *Advanced Calculus,* 2ed. John Wiley & Sons. New York (1972).

[7] R.C. Buck. *Advanced Calculus*, 3ed. McGraw-Hill Book Co. (1984).