

Structured Differentiation for Physicists

P. Reany

February 25, 2025

Abstract

Structured Differentiation is used to present the treatment of differentiation of maps of R^m to R^n . Along the way are partial derivatives, implicit functions, Jacobians, and the great bugaboo of total derivatives. Some rigor is left to the textbooks. I well remember how much trouble I had as an undergraduate math student in learning about partial derivatives and Jacobians and the like, so, if you are a math, chemistry, or an engineering student, you may also benefit from this paper and the rest of the series of papers I've written on Structured Differentiation.

1 Introduction

This paper was first published in the *AJNP* in October 1987 and then republished in it in January 1992.¹ Some corrections and additions have been made.

Anyone who has made a serious comparison of the various formalisms for differentiation used by 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. The student not yet versed in all these formalisms has his work cut out for him. 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 transformations from R^n to R^m , and it is most notably a matrix formulation of differentiation.

¹The impetus for this work began in 1985, when I was having a lot of trouble understanding the full ramifications of partial differentiation used in physics and advanced calculus.

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, cognitively, is the use of three distinct symbols representing the three semantically 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 many technical aspects of differentiability conditions which are completely ignored in this article. This article assumes that the functions used herein are nice differentiable functions. Specific functions encountered in the real world may have restrictions on their differentiability. Consult a text on advanced calculus to become familiar with them. Furthermore, familiarity with the Implicit Function Theorem would be helpful in some places.

2 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

$$f'(x) = \frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}, \quad (1)$$

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

$$dA = -kA(t)dt, \quad (2)$$

where $A(t)$ is the amount in the sample at time t , and k is a positive constant. Dividing (2) by dt gives

$$\frac{dA}{dt} = -kA(t). \quad (3)$$

Henceforth, we will 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

$$\frac{\delta \mathbf{f}}{\delta t} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{f}(t + \Delta t) - \mathbf{f}(t)}{\Delta t}, \quad (4)$$

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

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \partial f_1 / \partial x_1 & \partial f_1 / \partial x_2 & \dots & \partial f_1 / \partial x_m \\ \partial f_2 / \partial x_1 & \partial f_2 / \partial x_2 & \dots & \partial f_2 / \partial x_m \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_n / \partial x_1 & \partial f_n / \partial x_2 & \dots & \partial f_n / \partial x_m \end{bmatrix} \quad (5)$$

where $\mathbf{f} = (f_1, f_2, \dots, f_n)^t$ —the superscript t meaning transpose. The i th 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 i th component of \mathbf{f} , and represent it in indicial notation; using the chain rule we have

$$\delta f_i = \sum_j \frac{\partial f_i}{\partial x_j} \delta x_j \quad (i = 1, 2, \dots, n, j = 1, 2, \dots, m). \quad (6)$$

Equivalently, we can write this in matrix form as

$$\begin{bmatrix} \delta f_1 \\ \delta f_2 \\ \vdots \\ \delta f_n \end{bmatrix} = \begin{bmatrix} \partial f_1 / \partial x_1 & \partial f_1 / \partial x_2 & \dots & \partial f_1 / \partial x_m \\ \partial f_2 / \partial x_1 & \partial f_2 / \partial x_2 & \dots & \partial f_2 / \partial x_m \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_n / \partial x_1 & \partial f_n / \partial x_2 & \dots & \partial f_n / \partial x_m \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_m \end{bmatrix} \quad (7)$$

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

$$\delta \mathbf{f} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \delta \mathbf{x}. \quad (8)$$

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

$$\frac{\delta \mathbf{f}}{\delta \mathbf{x}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{\delta \mathbf{x}}{\delta \mathbf{x}}. \quad (9)$$

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

$$\frac{\delta \mathbf{f}}{\delta \mathbf{x}} = \begin{bmatrix} \delta f_1 / \delta x_1 & \delta f_1 / \delta x_2 & \dots & \delta f_1 / \delta x_m \\ \delta f_2 / \delta x_1 & \delta f_2 / \delta x_2 & \dots & \delta f_2 / \delta x_m \\ \vdots & \vdots & \ddots & \vdots \\ \delta f_n / \delta x_1 & \delta f_n / \delta x_2 & \dots & \delta f_n / \delta x_m \end{bmatrix}. \quad (10)$$

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 (9) 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 and only if f is functionally independent of x .

Similarly, 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. Conventional calculus has no equivalent to these definitions. They play, however, a dominant role in SD by helping students grasp the meaning of the different derivatives which correspond to different types of functional dependence.

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

$$\delta f = \frac{\partial f}{\partial \mathbf{x}} \delta \mathbf{x} = \frac{\partial f}{\partial x_1} \delta x_1 + \frac{\partial f}{\partial x_2} \delta x_2 + \cdots + \frac{\partial f}{\partial x_m} \delta x_m. \quad (11)$$

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

$$\frac{\delta f}{\delta x_i} = \frac{\partial f}{\partial \mathbf{x}} \frac{\delta \mathbf{x}}{\delta x_i} = \sum_j \frac{\partial f}{\partial x_j} \frac{\delta x_j}{\delta x_i}. \quad (12)$$

This can be rewritten as

$$\frac{\delta f}{\delta x_i} = \frac{\partial f}{\partial x_i} + \sum_{j \neq i} \frac{\partial f}{\partial x_j} \frac{\delta x_j}{\delta x_i}. \quad (13)$$

The first term on the right of (13) is the partial derivative of f with respect to x_i ; the summation on the right we refer to as the *copartial derivative* of f with respect to x_i . A more compact notation for the copartial derivative is as follows:

$$\frac{\partial f}{\partial x_i} \equiv \sum_{j \neq i} \frac{\partial f}{\partial x_j} \frac{\delta x_j}{\delta x_i}. \quad (14)$$

Thus (13) becomes

$$\frac{\delta f}{\delta x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial x_i}. \quad (15)$$

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 delta 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 cognitive 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

$$\frac{\partial F}{\partial x} = \frac{\partial}{\partial x}(0) = 0. \quad (16)$$

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.

The deltal derivative can be treated as a derivative “operator”

$$\frac{\delta}{\delta \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} + \frac{\partial}{\partial \mathbf{x}}, \quad (17)$$

or, more simply as

$$\delta_{\mathbf{x}} = \partial_{\mathbf{x}} + \partial_{\mathbf{x}}. \quad (18)$$

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

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx} = 0. \quad (19)$$

Therefore,

$$\frac{dy}{dx} = - \left(\frac{\partial f}{\partial y} \right)^{-1} \frac{\partial f}{\partial x}, \quad (20)$$

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

$$\frac{\delta \mathbf{f}}{\delta \mathbf{x}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \frac{\delta \mathbf{y}}{\delta \mathbf{x}} = \mathbf{0}. \quad (21)$$

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

$$\frac{\delta \mathbf{y}}{\delta \mathbf{x}} = - \left(\frac{\partial \mathbf{f}}{\partial \mathbf{y}} \right)^{-1} \frac{\partial \mathbf{f}}{\partial \mathbf{x}}. \quad (22)$$

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = - \left(\frac{\partial \mathbf{f}}{\partial \mathbf{y}} \right)^{-1} \frac{\partial \mathbf{f}}{\partial \mathbf{x}}. \quad (23)$$

Equation (23) 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. 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.

3 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

$$\delta F = \frac{\partial F}{\partial \mathbf{x}} \delta \mathbf{x} = 0. \quad (24)$$

Of course, we can interpret

$$\frac{\partial F}{\partial \mathbf{x}} = \nabla F, \quad (25)$$

the gradient of F . Using the dot product, (24) becomes (dropping the δF):

$$\nabla F \cdot \delta \mathbf{x} = 0 \quad (26a)$$

or

$$\delta \mathbf{x} \cdot \nabla F = 0. \quad (26b)$$

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

$$\mathbf{a} \cdot \nabla F. \quad (27)$$

Dividing (26b) through by $|\delta\mathbf{x}|$, we get

$$\frac{\delta \mathbf{x}}{|\delta \mathbf{x}|} \cdot \nabla F = 0, \quad (28a)$$

or

$$\mathbf{a} \cdot \nabla F = 0, \quad (28b)$$

where

$$\mathbf{a} = \frac{\delta \mathbf{x}}{|\delta \mathbf{x}|}. \quad (28c)$$

What (28b) 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

$$F(\mathbf{x}) = c \quad (29)$$

in that region, with c an arbitrary constant. This constraint has the effect of reducing the dimension of the space on which (29) applies to be $(n - 1)$ -dimensional subregion, called a *level surface*.² On taking the differential across

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

(29), we re-obtain (24) and all its following arguments up to and including (28c). So, if \mathbf{b} is a unit vector lying in the tangent space of a level surface at a given point at which (29) 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 (29) 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 (29). The vector $\delta\mathbf{x}$ in (26b) and the vector \mathbf{a} in (28b) 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

$$G(x, y, z) = k, \quad (30)$$

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 ([5]) (192–198).³)

Since $\partial G/\partial z \neq 0$ and G is constrained according to (30), then by the *Implicit Function Theorem* we can write

$$z = z(x, y) = z(\boldsymbol{\eta}), \quad (31)$$

where z is in primitive form. Then

$$F(x, y, z) = F(\boldsymbol{\eta}, z(\boldsymbol{\eta})) = F(\mathbf{x}). \quad (32)$$

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

$$\delta F/\delta\boldsymbol{\eta} = \mathbf{0}, \quad (33)$$

which is a direct generalization from ordinary calculus. Therefore, applying the chain rule to expand (33), we get

$$\frac{\partial F}{\partial \boldsymbol{\eta}} + \frac{\partial F}{\partial z} \frac{\partial z}{\partial \boldsymbol{\eta}} = \mathbf{0}, \quad (34)$$

³It'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.

where we used the fact that when we expanded $\delta F/\delta\boldsymbol{\eta}$ by the chain rule, we could relate $\delta z/\delta\boldsymbol{\eta}$ by $\partial z/\partial\boldsymbol{\eta}$ since z is in primitive form. Now, by differentiating (30) through by $\boldsymbol{\eta}$, we also have that

$$\frac{\delta G}{\delta\boldsymbol{\eta}} = \frac{\partial G}{\partial\boldsymbol{\eta}} + \frac{\partial G}{\partial z} \frac{\partial z}{\partial\boldsymbol{\eta}} = \mathbf{0}. \quad (35)$$

Solving this system for $\partial z/\partial\boldsymbol{\eta}$ and substituting that result into (34) we obtain

$$\frac{\partial F}{\partial\boldsymbol{\eta}} + \lambda \frac{\partial G}{\partial\boldsymbol{\eta}} = \mathbf{0} \quad (36)$$

where

$$\lambda = -\frac{\partial F/\partial z}{\partial G/\partial z} \quad (37)$$

On rewriting this we have

$$\frac{\partial F}{\partial z} + \lambda \frac{\partial G}{\partial z} = 0. \quad (38)$$

Combining this and (36) results in

$$\frac{\partial F}{\partial\mathbf{x}} + \lambda \frac{\partial G}{\partial\mathbf{x}} = \mathbf{0}. \quad (39)$$

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

$$\nabla F + \lambda \nabla G = \mathbf{0}. \quad (40)$$

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

So, if (30) and (40) 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 (30), there exists some nonzero scalar c such that the level surface defined by

$$F(x, y, z) = c \quad (41)$$

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

$$H(x, y, z) = F(x, y, z) + \lambda G(x, y, z) \quad (42)$$

such that the possible relative extrema of $F(x, y, z)$ subject to (30) can be found by solving the system of (30) and

$$\nabla H = \mathbf{0}. \quad (43)$$

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)

$$J(x, y, z, \lambda) = F(x, y, z) + \lambda G(x, y, z) \quad (44)$$

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

$$\square J = \mathbf{0}, \quad (45)$$

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

$$\square \equiv (\partial_x, \partial_y, \partial_z, \partial_\lambda). \quad (46)$$

Now, the first three components of (45) reproduce (40) and the fourth component reproduces the original constraint (30), 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 (45) 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?

There is perhaps no more difficult concept in this paper than that of convolution. A *convolution* is said to occur whenever a variable (or function) is functionally dependent on itself. (Our definition of convolution, however, has no direct connection to the convolution of the Laplace transform.) As we shall see later, the most efficient way to discovering the effects of making a change of fundamental variables in a given problem, is to make variables functionally dependent on themselves in a nontrivial way, i.e., through some intermediate variable.

4 Theory: ‘Change of Variables’

Up to this point we have encountered functions that are dependent on variable \mathbf{x} maybe both explicitly and implicitly, in which case by differentiation, we get

$$\frac{\delta \mathbf{f}}{\delta \mathbf{x}} = \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \frac{\delta \mathbf{y}}{\delta \mathbf{x}}, \quad (47)$$

where $\mathbf{f} = \mathbf{f}(\mathbf{y})$ and \mathbf{y} is the variant of \mathbf{f} , say. There are many situations that require us to extend our theory of SD to include problems in which the fundamental itself will change, that is, will be replaced by some other fundamental.

Now, this is not a new concept even to a calculus student. Say we have the integral

$$I = \int A(x) dx \quad (48)$$

and we prefer to integrate with respect to some other variable, such as u . Then we recast this integral as

$$I = \int A(u) \frac{dx}{du} du. \quad (49)$$

This expression $\frac{dx}{du}$ is called the *Jacobian* of the transformation of variables from the old variable x to the new variable u . We will soon encounter problems where the old variables and the new variables both are treated as vectors in SD.

I will now present a couple definitions of SD on how to deal with derivatives resulting from a change of fundamental, which I cannot promise that all authors will use:

Definition of Jacobian matrix

The total derivative of the old fundamental with respect to the new fundamental shall be called the *Jacobian matrix* of the transformation of variables.⁴ (Often, the total derivative can be reduced to the partial derivative.) The determinant of the Jacobian matrix is called the *Jacobian* of the transformation.

So, besides integration, where does this kind of change of variables come up? For one, when one wants to change coordinates, such from rectangular to polar or in 3D spherical polar coordinates, or back the other way. Another is in the transformation of variables in thermodynamics. We will soon see an example where a thermodynamic equation which is known for one pair of fundamental variables is to be converted to its equivalent for when a new couple of fundamental variables is chosen.

Now, a subtlety arises that hasn't to this point. We got by just fine with using $\frac{\partial \mathbf{f}}{\partial \mathbf{y}}$ in Eq. (47), but it may be safer to use a parallel to that which is done is ordinary chain rule. So, if $f(x) = f(u(x))$ then

$$\frac{df}{dx} = \frac{df}{du} \frac{du}{dx}. \quad (50)$$

Therefore, if $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{u}(\mathbf{x}))$, it may be safer (that is, more accurate) to write

$$\frac{\delta \mathbf{f}}{\delta \mathbf{x}} = \frac{\delta \mathbf{f}}{\delta \mathbf{u}} \frac{\delta \mathbf{u}}{\delta \mathbf{x}}, \quad (51)$$

and the utility of this convention should become obvious once we've done a couple problems in thermodynamics.

Rule 1) When one fundamental variable η_i is totally differentiated by a cofundamental variable η_j , the result is

$$\frac{\delta \eta_i}{\delta \eta_j} = \delta_{ij}, \quad (52)$$

where δ_{ij} is the Kronecker delta, meaning that when a fundamental variable is totally differentiated by itself, the result is unity, but when it is totally differentiated by a different cofundamental variable, the result is zero; hence, the

⁴Of course, if both fundamentals are just scalars, then the Jacobian 'matrix' will be just a 1×1 matrix, or, a scalar.

notion of *variable independence* between different elements in a given fundamental is captured by this rule. Note: To remove ambiguity, any variable in a fundamental set is cofundamental to itself.⁵

Rule 2) In all other cases, i.e., when a new fundamental variable differentiates an old fundamental variable, or vice versa, the dotal derivative usually reduces to an explicit derivative. This is the standard thing to do in thermodynamics because, by default, we regard state variables as having no implicit dependence on other variables (at least this is my knowledge regarding them).⁶

5 Problem III. Example from Vector Analysis

Let \mathbf{u} be a vector field defined on the open subset \mathcal{D} of R^n , let \mathbf{x} be a vector field defined on the open subset \mathcal{D}' of R^n , and let T be a one-to-one, onto transformation which takes \mathbf{u} to \mathbf{x} by

$$\mathbf{x} = T(\mathbf{u}) = \mathbf{x}(\mathbf{u}). \quad (53)$$

Now T^{-1} exists on \mathcal{D}' so that

$$\mathbf{u} = T^{-1}(\mathbf{x}) = \mathbf{u}(\mathbf{x}). \quad (54)$$

The *Jacobian matrix* of the transformation to be

$$\mathbf{J}_T = \frac{\partial T}{\partial \mathbf{u}} = \frac{\partial \mathbf{x}}{\partial \mathbf{u}}, \quad (55)$$

where \mathbf{x} is the ‘old’ fundamental and \mathbf{u} is the ‘new’ fundamental. Thus, the *Jacobian* of the transformation to be

$$J_T \equiv |\mathbf{J}_T|, \quad (56)$$

where $|\mathbf{J}_T|$ is the determinant of the Jacobian matrix. We will show that

$$(\mathbf{J}_T)^{-1} = \mathbf{J}_{T^{-1}} \quad (57)$$

and

$$(J_T)^{-1} = J_{T^{-1}}. \quad (58)$$

To prove this we take the convolution of \mathbf{x} through \mathbf{u} and get

$$\mathbf{x} = \mathbf{x}(\mathbf{u}(\mathbf{x})). \quad (59)$$

On taking the total derivative of (59) by \mathbf{x} and simplifying we obtain

$$\mathbf{I} = \frac{\partial \mathbf{x}}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}. \quad (60)$$

But $\partial \mathbf{u} / \partial \mathbf{x} = \mathbf{J}_{T^{-1}}$; therefore, (60) establishes (57). To establish (58) we merely take the determinant of both sides of (60).

⁵We will consider it axiomatic in SD that the total derivative of any variable with respect to itself is unity.

⁶When a state variable **does** have an implicit dependence on some variable, then we cannot just drop the implicit derivative as we have done in this case.

6 Problem IV. Transformation on Coordinates

Let

$$\left. \begin{aligned} x &= r \sin \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta \end{aligned} \right\} \quad (61)$$

Find ∇r , $\nabla \theta$, and $\nabla \phi$ where $\nabla = \partial/\partial \mathbf{x}$. One way to solve this problem is to solve for r , θ , ϕ in terms of x , y , z , then differentiate, but we will take another path.

The gradients we want are just the row vectors of the matrix $\partial \mathbf{u}/\partial \mathbf{x}$ where $\mathbf{u} = (r, \theta, \phi)$. Now, treating \mathbf{x} as the old fundamental and \mathbf{u} as the new fundamental, the Jacobian matrix is

$$\frac{\partial \mathbf{x}}{\partial \mathbf{u}} = \begin{bmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{bmatrix} \quad (62)$$

and the Jacobian

$$\left| \frac{\partial \mathbf{x}}{\partial \mathbf{u}} \right| = r^2 \sin \theta. \quad (63)$$

Where the Jacobian is nonzero, we have from (60)

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{u}} \right)^{-1}. \quad (64)$$

Therefore,

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \begin{bmatrix} \nabla r \\ \nabla \theta \\ \nabla \phi \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \frac{1}{r} \cos \theta \cos \phi & \frac{1}{r} \cos \theta \sin \phi & -\frac{1}{r} \sin \theta \\ -\frac{\sin \phi}{r \sin \theta} & \frac{\cos \phi}{r \sin \theta} & 0 \end{bmatrix}. \quad (65)$$

7 Problem V. First Example from Thermodynamics

A certain thermodynamic state can be represented by the differential equation of state

$$\frac{\partial U}{\partial V} - T \frac{\partial P}{\partial T} + P = 0, \quad (66)$$

where V and T are the fundamental variables. What, then, is the appropriate form that (66) takes when U and P replace V and T as the fundamental variables? The current problem is from Buck [6], prob. 18, pg. 146.

Let the old fundamental be represented by

$$\boldsymbol{\eta} = (V, T)^t, \quad (67)$$

and the new fundamental by

$$\boldsymbol{\eta}' = (U, P)^t. \quad (68)$$

For convenience, we also define the state vector

$$\boldsymbol{\psi} = (U, P, V, T)^t. \quad (69)$$

Then

$$\boldsymbol{\psi}'(\boldsymbol{\eta}') = \boldsymbol{\psi}(\boldsymbol{\eta}(\boldsymbol{\eta}')). \quad (70)$$

Differentiating this by $\boldsymbol{\eta}'$, we get

$$\frac{\delta \boldsymbol{\psi}'}{\delta \boldsymbol{\eta}'} = \frac{\delta \boldsymbol{\psi}}{\delta \boldsymbol{\eta}} \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'}, \quad (71)$$

and simplifying, we have

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \partial V/\partial U & \partial V/\partial P \\ \partial T/\partial U & \partial T/\partial P \end{bmatrix} = \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{J} \quad (72)$$

where $\mathbf{J} = \partial \boldsymbol{\eta}/\partial \boldsymbol{\eta}'$. (See Appendix 1 to see the important steps left out here.)

The relations we want can be extracted from (72) by inspection, yielding

$$\frac{\partial T}{\partial P} = \frac{\partial U}{\partial V} J \quad (73a)$$

and

$$\frac{\partial V}{\partial U} = \frac{\partial P}{\partial T} J. \quad (73b)$$

Solving these for $\partial U/\partial V$, $\partial P/\partial T$ and then substituting the results into (66) yields

$$\frac{\partial T}{\partial P} - T \frac{\partial V}{\partial U} + PJ = 0. \quad (74)$$

The technique used to derive (73a) and (73b) from (72) we refer to as *decomposition*, which we describe in brief. From the first and fourth rows of (72), we may write the 2×2 matrix equation⁷

$$\begin{bmatrix} 1 & 0 \\ \partial T/\partial U & \partial T/\partial P \end{bmatrix} = \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ 0 & 1 \end{bmatrix} \mathbf{J}, \quad (75)$$

and from the second and third rows, we may similarly write

$$\begin{bmatrix} 0 & 1 \\ \partial V/\partial U & \partial V/\partial P \end{bmatrix} = \begin{bmatrix} \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \end{bmatrix} \mathbf{J}. \quad (76)$$

Equations (73a) and (73b) follow by taking the determinant of these last two matrix equations, employing the fact that the determinant of a product is the product of the determinants.

⁷Proof given in Appendix 1.

8 Problem VI. Examples from Dynamics

Up to this point we have dealt almost exclusively with nonscalar fundamentals, for this is the realm in which SD shows its greatest power and elegance. It is not surprising, therefore, that SD does not much improve the study of dynamics, except possibly improving the clarity of interpretation of the equations. For this reason we will briefly study Hamiltonian dynamics and hydrodynamics. Since the functions here to be considered are dependent only on the scalar fundamental of time, we will use the d -differential instead of the δ -differential for the total derivative.

We will begin with Hamiltonian dynamics. Let $F = F(t, \mathbf{q}(t), \mathbf{p}(t))$, then

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \frac{\partial F}{\partial t} \quad (77)$$

$$= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial \mathbf{q}} \frac{d\mathbf{q}}{dt} + \frac{\partial F}{\partial \mathbf{p}} \frac{d\mathbf{p}}{dt} \quad (78)$$

$$= \frac{\partial F}{\partial t} + \frac{\partial F}{\partial \mathbf{q}} \dot{\mathbf{q}} + \frac{\partial F}{\partial \mathbf{p}} \dot{\mathbf{p}}. \quad (79)$$

Using Hamilton's equations this becomes

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \frac{\partial F}{\partial \mathbf{q}} \frac{\partial H}{\partial \mathbf{p}} - \frac{\partial F}{\partial \mathbf{p}} \frac{\partial H}{\partial \mathbf{q}} \quad (80)$$

$$= \frac{\partial F}{\partial t} + \{F, H\}, \quad (81)$$

where we have used the Poisson bracket of F and H . Therefore, we may write

$$\frac{\partial F}{\partial t} = \{F, H\}. \quad (82)$$

Moving now to hydrodynamics, consider the density function

$$\rho = \rho(t, \mathbf{x}(t)). \quad (83)$$

Differentiating this we obtain

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial \mathbf{x}} \dot{\mathbf{x}}. \quad (84)$$

Or, in the alternative notation

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho. \quad (85)$$

9 Conclusion

SD has been presented here as a semantically advanced language for differentiation. It stresses that differential operators not only have formal properties, they also have distinct meanings that people have to deal with. For example, mathematics texts use partial derivatives as total derivatives by placing function in primitive form (at least with respect to the variable of differentiation), thus making the implicit derivative identically zero. But this does not remove the distinction in meaning between partial and total derivatives. These same texts usually offer no motivation to the student to explain their preference for primitive functions—the lack of which I am sure adds unnecessarily to the perceived mysteriousness of partial differentiation and to its notorious reputation. Indeed, standard advanced calculus does not even have a sufficient vocabulary to give such an explanation. SD also obviates the need to use such cumbersome terms as “total” partial and “partial” partial derivatives, as I have seen in some books.

One clarification needs to be made here: I said before that SD was developed to be a matrix algebraic approach to differentiation, but that matrices (or indices) introduce a basis, which may be irrelevant to the development of your mathematical line of reasoning. The resolution to this mystery is that the high-level, steps to your proofs can be performed in the bold-face notation, which is meant to suppress a basis, yet when one is ready to solve an actual problem, the basis can be made more or less explicit to perform calculations (at the very least, to set up the dimensions of the matrices involved).

Supplementary Problems

Problem 1

Given the system

$$\begin{aligned}x &= u \cos v, \\y &= u \sin v\end{aligned}\tag{86}$$

find $\partial u/\partial x$, $\partial u/\partial y$, $\partial v/\partial x$, $\partial v/\partial y$.

First, we rewrite the last system as

$$\begin{aligned}u \cos v - x &= 0, \\u \sin v - y &= 0\end{aligned}\tag{87}$$

and then make the definitions

$$\mathbf{F} \equiv (F_1, F_2)^t = (u \cos v - x, u \sin v - y)^t\tag{88}$$

$$\mathbf{u} = (u, v)^t, \quad \mathbf{x} = (x, y)^t\tag{89}$$

Since $\mathbf{F} = \mathbf{0}$ then $\delta\mathbf{F}/\delta\mathbf{x} = \mathbf{0}$, or

$$\frac{\partial\mathbf{F}}{\partial\mathbf{x}} + \frac{\partial\mathbf{F}}{\partial\mathbf{u}} \frac{\delta\mathbf{u}}{\delta\mathbf{x}} = \frac{\partial\mathbf{F}}{\partial\mathbf{x}} + \frac{\partial\mathbf{F}}{\partial\mathbf{u}} \frac{\partial\mathbf{u}}{\partial\mathbf{x}} = \mathbf{0}.\tag{90}$$

Solving for $\partial\mathbf{u}/\partial\mathbf{x}$ we have

$$\frac{\partial\mathbf{u}}{\partial\mathbf{x}} = - \left(\frac{\partial\mathbf{F}}{\partial\mathbf{u}} \right)^{-1} \frac{\partial\mathbf{F}}{\partial\mathbf{x}}.\tag{91}$$

Now

$$\frac{\partial\mathbf{F}}{\partial\mathbf{u}} = \begin{bmatrix} \cos v & -u \sin v \\ \sin v & u \cos v \end{bmatrix} \quad \text{and} \quad \det \left(\frac{\partial\mathbf{F}}{\partial\mathbf{u}} \right) = u.\tag{92}$$

So,

$$\left(\frac{\partial\mathbf{F}}{\partial\mathbf{u}} \right)^{-1} = \frac{1}{u} \begin{bmatrix} u \cos v & u \sin v \\ -\sin v & \cos v \end{bmatrix} = \begin{bmatrix} \cos v & \sin v \\ -\frac{\sin v}{u} & \frac{\cos v}{u} \end{bmatrix}\tag{93}$$

and therefore

$$\begin{aligned}\frac{\partial\mathbf{u}}{\partial\mathbf{x}} &= \begin{bmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{bmatrix} \\ &= - \begin{bmatrix} \cos v & \sin v \\ -\frac{\sin v}{u} & \frac{\cos v}{u} \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \cos v & \sin v \\ -\frac{\sin v}{u} & \frac{\cos v}{u} \end{bmatrix}.\end{aligned}\tag{94}$$

Problem 2

If $G_1(x_1, x_2, y)$, $G_2(x_1, x_2, y)$, and $f(x_1, x_2)$ are given, and if

$$g_i(x_1, x_2) \equiv G_i(x_1, x_2, f(x_1, x_2)) \quad (i = 1, 2) \quad (95)$$

Show that

$$\left| \frac{\partial(g_1, g_2)}{\partial(x_1, x_2)} \right| = \left| \frac{\partial(G_1, G_2)}{\partial(x_1, x_2)} \right| + \frac{\partial f}{\partial x_1} \left| \frac{\partial(G_1, G_2)}{\partial(y, x_2)} \right| + \frac{\partial f}{\partial x_2} \left| \frac{\partial(G_1, G_2)}{\partial(x_1, y)} \right|. \quad (96)$$

First, note that g_i is the reduction of G_i to primitive form. This reduction is not necessary in SD, which would have replaced the LHS of the last equation with $\left| \frac{\delta(G_1, G_2)}{\delta(x_1, x_2)} \right|$.

Proof:

We begin by rewriting what we have in vector notation

$$\mathbf{g}(\mathbf{x}) = \mathbf{G}(\mathbf{x}, f(\mathbf{x})). \quad (97)$$

Then,

$$\frac{\delta \mathbf{g}}{\delta \mathbf{x}} = \frac{\delta \mathbf{G}}{\delta \mathbf{x}}. \quad (98a)$$

Furthermore, since \mathbf{g} is primitive in \mathbf{x} , then,

$$\frac{\delta \mathbf{g}}{\delta \mathbf{x}} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}. \quad (98b)$$

And then

$$\begin{aligned} \left| \frac{\partial(g_1, g_2)}{\partial(x_1, x_2)} \right| &= \left| \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right| = \left| \frac{\delta \mathbf{g}}{\delta \mathbf{x}} \right| = \left| \frac{\delta \mathbf{G}}{\delta \mathbf{x}} \right| = \left(\frac{\delta G_1}{\delta x_1} \frac{\delta G_2}{\delta x_2} - \frac{\delta G_1}{\delta x_2} \frac{\delta G_2}{\delta x_1} \right) \\ &= \left(\frac{\partial G_1}{\partial x_1} + \frac{\partial G_1}{\partial y} \frac{\partial f}{\partial x_1} \right) \left(\frac{\partial G_2}{\partial x_2} + \frac{\partial G_2}{\partial y} \frac{\partial f}{\partial x_2} \right) \\ &\quad - \left(\frac{\partial G_1}{\partial x_2} + \frac{\partial G_1}{\partial y} \frac{\partial f}{\partial x_2} \right) \left(\frac{\partial G_2}{\partial x_1} + \frac{\partial G_2}{\partial y} \frac{\partial f}{\partial x_1} \right) \\ &= \left(\frac{\partial G_1}{\partial x_1} \frac{\partial G_2}{\partial x_2} - \frac{\partial G_1}{\partial x_2} \frac{\partial G_2}{\partial x_1} \right) + \frac{\partial f}{\partial x_1} \left(\frac{\partial G_1}{\partial y} \frac{\partial G_2}{\partial x_2} - \frac{\partial G_1}{\partial x_2} \frac{\partial G_2}{\partial y} \right) \\ &\quad + \frac{\partial f}{\partial x_2} \left(\frac{\partial G_1}{\partial x_1} \frac{\partial G_2}{\partial y} - \frac{\partial G_1}{\partial y} \frac{\partial G_2}{\partial x_1} \right) \\ &= \left| \frac{\partial(G_1, G_2)}{\partial(x_1, x_2)} \right| + \frac{\partial f}{\partial x_1} \left| \frac{\partial(G_1, G_2)}{\partial(y, x_2)} \right| + \frac{\partial f}{\partial x_2} \left| \frac{\partial(G_1, G_2)}{\partial(x_1, y)} \right|. \end{aligned}$$

(Incidentally, this is the only problem I ever solved as a demonstration to fellow students that received a spontaneous ovation at its finish.)

Problem 3

A certain thermodynamic state can be represented by the differential equation of state

$$\frac{\partial U}{\partial V} - T \frac{\partial P}{\partial T} + P = 0 \quad (99)$$

where V and T are the fundamental variables. What, then, is the appropriate form that (99) takes when P and T replace V and T as the fundamental variables?

This problem is very similar to Problem V above, and it may well be advised to go over that solution and Appendix 1 before trying this problem.

As before, let the old fundamental be represented by

$$\boldsymbol{\eta} = (V, T)^t \quad (100)$$

and the new fundamental by

$$\boldsymbol{\eta}' = (P, T)^t. \quad (101)$$

The difference between this problem and Problem V is just that this time there is a common variable of T between the old and the new fundamental.

We again define the state vector

$$\boldsymbol{\psi} = (U, P, V, T)^t. \quad (102)$$

Then, as before

$$\boldsymbol{\psi}'(\boldsymbol{\eta}') = \boldsymbol{\psi}(\boldsymbol{\eta}'). \quad (103)$$

Differentiating this by $\boldsymbol{\eta}'$, we get

$$\frac{\delta \boldsymbol{\psi}'}{\delta \boldsymbol{\eta}'} = \frac{\delta \boldsymbol{\psi}}{\delta \boldsymbol{\eta}} \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'}, \quad (104)$$

which looks like this

$$\begin{bmatrix} \delta U / \delta P & \delta U / \delta T \\ \delta P / \delta P & \delta P / \delta T \\ \delta V / \delta P & \delta V / \delta T \\ \delta T / \delta P & \delta T / \delta T \end{bmatrix} = \begin{bmatrix} \delta U / \delta V & \delta U / \delta T \\ \delta P / \delta V & \delta P / \delta T \\ \delta V / \delta V & \delta V / \delta T \\ \delta T / \delta V & \delta T / \delta T \end{bmatrix} \mathbf{J}, \quad (105)$$

where $\mathbf{J} = \delta \boldsymbol{\eta} / \delta \boldsymbol{\eta}'$ is a 2×2 matrix (to be simplified later according to Rules 1 and 2 on page (11)). Now, let's analyze this last equation in depth.

First, since the total derivative of any variable by itself is unity; then for the derivatives in the 4×2 matrix on the LHS, we have that

$$\delta P / \delta P = 1, \quad \delta T / \delta T = 1, \quad (106)$$

and since P and T are independent of each other (again, on the LHS),

$$\delta P/\delta T = \delta T/\delta P = 0. \quad (107)$$

Therefore, upon substitution and simplification, we have

$$\begin{bmatrix} \delta U/\delta P & \delta U/\delta T \\ \delta P/\delta P & \delta P/\delta T \\ \delta V/\delta P & \delta V/\delta T \\ \delta T/\delta P & \delta T/\delta T \end{bmatrix} \longrightarrow \begin{bmatrix} \partial U/\partial P & \partial U/\partial T \\ 1 & 0 \\ \partial V/\partial P & \partial V/\partial T \\ 0 & 1 \end{bmatrix}. \quad (108)$$

Arguing similarly for the 4×2 matrix on the RHS, we have that

$$\delta V/\delta V = 1, \quad \delta T/\delta T = 1, \quad (109)$$

and since V and T are independent of each other,

$$\delta V/\delta T = \delta T/\delta V = 0. \quad (110)$$

Therefore, upon substitution and simplification, we have

$$\begin{bmatrix} \delta U/\delta V & \delta U/\delta T \\ \delta P/\delta V & \delta P/\delta T \\ \delta V/\delta V & \delta V/\delta T \\ \delta T/\delta V & \delta T/\delta T \end{bmatrix} \longrightarrow \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (111)$$

Now we substitute these simpler expressions back into (105) to get

$$\begin{bmatrix} \partial U/\partial P & \partial U/\partial T \\ 1 & 0 \\ \partial V/\partial P & \partial V/\partial T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{J}, \quad (112)$$

where \mathbf{J} is equal to

$$\mathbf{J} = \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'} = \begin{bmatrix} \delta V/\delta P & \delta V/\delta T \\ \delta T/\delta P & \delta T/\delta T \end{bmatrix} = \begin{bmatrix} \partial V/\partial P & 0 \\ \partial T/\partial P & 1 \end{bmatrix}, \quad (113)$$

where we used Rules 1 and 2 on page (11), and from which we get that

$$\begin{bmatrix} \partial U/\partial P & \partial U/\partial T \\ 1 & 0 \\ \partial V/\partial P & \partial V/\partial T \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \partial V/\partial P & 0 \\ \partial T/\partial P & 1 \end{bmatrix}. \quad (114)$$

The relations we want can be extracted from (114) by inspection, yielding

$$\frac{\partial U}{\partial P} = \frac{\partial U}{\partial V} \frac{\partial V}{\partial P}, \quad (115)$$

and

$$\frac{\partial V}{\partial T} = -\frac{\partial P}{\partial T} \frac{\partial V}{\partial P}. \quad (116)$$

Substituting these into (99) yields

$$\frac{\partial U}{\partial P} + T \frac{\partial V}{\partial T} + P \frac{\partial V}{\partial P} = 0. \quad (117)$$

10 Appendix 1

Here we fill in some of the missing steps from Problem V. The old fundamental is

$$\boldsymbol{\eta} = (V, T)^t, \quad (118)$$

and the new fundamental is

$$\boldsymbol{\eta}' = (U, P)^t. \quad (119)$$

We also define the state vector

$$\boldsymbol{\psi} = (U, P, V, T)^t. \quad (120)$$

Then

$$\boldsymbol{\psi}'(\boldsymbol{\eta}') = \boldsymbol{\psi}(\boldsymbol{\eta}'). \quad (121)$$

Differentiating this by $\boldsymbol{\eta}'$, we get

$$\frac{\delta \boldsymbol{\psi}'}{\delta \boldsymbol{\eta}'} = \frac{\delta \boldsymbol{\psi}}{\delta \boldsymbol{\eta}} \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'}, \quad (122)$$

which looks like this

$$\begin{bmatrix} \delta U / \delta U & \delta U / \delta P \\ \delta P / \delta U & \delta P / \delta P \\ \delta V / \delta U & \delta V / \delta P \\ \delta T / \delta U & \delta T / \delta P \end{bmatrix} = \begin{bmatrix} \delta U / \delta V & \delta U / \delta T \\ \delta P / \delta V & \delta P / \delta T \\ \delta V / \delta V & \delta V / \delta T \\ \delta T / \delta V & \delta T / \delta T \end{bmatrix} \mathbf{J}, \quad (123)$$

where $\mathbf{J} = \partial \boldsymbol{\eta} / \partial \boldsymbol{\eta}'$ is a 2×2 matrix. Now, let's analyze this last equation in depth.

First, since the total derivative of any variable by itself is unity; then for the derivatives in the 4×2 matrix on the LHS, we have that

$$\delta U / \delta U = 1, \quad \delta P / \delta P = 1, \quad (124)$$

and since U and P are independent of each other,

$$\delta U / \delta P = \delta P / \delta U = 0. \quad (125)$$

Therefore, upon substitution and simplification, we have

$$\begin{bmatrix} \delta U / \delta U & \delta U / \delta P \\ \delta P / \delta U & \delta P / \delta P \\ \delta V / \delta U & \delta V / \delta P \\ \delta T / \delta U & \delta T / \delta P \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \partial V / \partial U & \partial V / \partial P \\ \partial T / \partial U & \partial T / \partial P \end{bmatrix}, \quad (126)$$

where we used Rules 1 and 2 on page (11). Arguing similarly for the 4×2 matrix on the RHS, we have that

$$\delta V / \delta V = 1, \quad \delta T / \delta T = 1, \quad (127)$$

and since V and T are independent of each other,

$$\delta V/\delta T = \delta T/\delta V = 0. \quad (128)$$

Therefore, upon substitution and simplification, we have

$$\begin{bmatrix} \delta U/\delta V & \delta U/\delta T \\ \delta P/\delta V & \delta P/\delta T \\ \delta V/\delta V & \delta V/\delta T \\ \delta T/\delta V & \delta T/\delta T \end{bmatrix} \longrightarrow \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (129)$$

where, again, we used Rules 1 and 2 on page (11). Now we substitute these simpler expressions back into (123) to get

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \partial V/\partial U & \partial V/\partial P \\ \partial T/\partial U & \partial T/\partial P \end{bmatrix} = \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{J}, \quad (130)$$

where \mathbf{J} is equal to

$$\mathbf{J} = \frac{\delta \boldsymbol{\eta}}{\delta \boldsymbol{\eta}'} = \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\eta}'} = \begin{bmatrix} \partial V/\partial U & \partial V/\partial P \\ \partial T/\partial U & \partial T/\partial P \end{bmatrix}. \quad (131)$$

Now, the fast way to get the information we want out of these matrices is to take the determinants of 2×2 submatrices. Well, what is the exact information we need? Remember that the equation we started with is

$$\frac{\partial U}{\partial V} - T \frac{\partial P}{\partial T} + P = 0. \quad (132)$$

But the whole point of this exercise is to revamp this equation by replacing the partials $\frac{\partial U}{\partial V}$ and $\frac{\partial P}{\partial T}$ by their equivalent partials after we change the fundamental variables. (This is the fulfillment of the heads-up I made in the Introduction on page 1.)

Let's look first at $\frac{\partial U}{\partial V}$. If we could form the matrix equation made of submatrices formed by using only the first and fourth rows, we'd get

$$\begin{bmatrix} 1 & 0 \\ \partial T/\partial U & \partial T/\partial P \end{bmatrix} = \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ 0 & 1 \end{bmatrix} \mathbf{J}. \quad (133)$$

Before I show the reader how to justify this step, I want to show its value. By taking the determinant across this equation, we get

$$\frac{\partial T}{\partial P} = \frac{\partial U}{\partial V} J. \quad (134)$$

From this we get that

$$\frac{\partial U}{\partial V} = \frac{\partial T}{\partial P} J^{-1}. \quad (135)$$

That leaves the replacement for $\frac{\partial P}{\partial T}$. For this we take the second and third rows, to get

$$\begin{bmatrix} 0 & 1 \\ \partial V/\partial U & \partial V/\partial P \end{bmatrix} = \begin{bmatrix} \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \end{bmatrix} \mathbf{J}. \quad (136)$$

On taking the determinant across this equation, we get

$$-\frac{\partial V}{\partial U} = -\frac{\partial P}{\partial T} J. \quad (137)$$

From this we get that

$$\frac{\partial P}{\partial T} = \frac{\partial V}{\partial U} J^{-1}. \quad (138)$$

So, on substituting this result and the result from (135) into (132), we have

$$\frac{\partial T}{\partial P} J^{-1} - T \frac{\partial V}{\partial U} J^{-1} + P = 0, \quad (139)$$

which simplifies to

$$\frac{\partial T}{\partial P} - T \frac{\partial V}{\partial U} + PJ = 0. \quad (140)$$

This leaves only the justification for my extracting 2×2 submatrices from those 4×2 matrices in Eq. (130). Let's take the case where I extracted the second and third rows. Suppose we multiply on the left Eq. (130) by the 2×4 matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (141)$$

and use the key fact that matrix multiplication is associative to get

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \partial V/\partial U & \partial V/\partial P \\ \partial T/\partial U & \partial T/\partial P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \partial V/\partial U & \partial V/\partial P \end{bmatrix}. \quad (142)$$

Now, applying the matrix on the RHS, we get

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{J} = \begin{bmatrix} \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \end{bmatrix} \mathbf{J}, \quad (143)$$

It's obvious that by this technique that you can pick out any two rows you want. And this completes the justification.

Now, for an exercise, what information could we get by extracting on rows one and three from Eq. (130)? (Try to do this by inspection only.) Answer:

$$\frac{\partial V}{\partial P} = -\frac{\partial U}{\partial T} J. \quad (144)$$

Now, a final comment on the SD approach. I trust the reader can see why in Eq. (129) I had to use total derivatives to get 1's and 0's as entries, and thus

$$\begin{bmatrix} \delta U/\delta V & \delta U/\delta T \\ \delta P/\delta V & \delta P/\delta T \\ \delta V/\delta V & \delta V/\delta T \\ \delta T/\delta V & \delta T/\delta T \end{bmatrix} \longrightarrow \begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (145)$$

And this justifies why I had to use the equation

$$\frac{\delta\psi'}{\delta\eta'} = \frac{\delta\psi}{\delta\eta} \frac{\delta\eta}{\delta\eta'}, \quad (146)$$

instead of the equation

$$\frac{\delta\psi'}{\delta\eta'} = \frac{\partial\psi}{\partial\eta} \frac{\delta\eta}{\delta\eta'}. \quad (147)$$

So, I can just hear a defender of the conventional way of writing these derivatives argue that convention is just fine in how it does this. So, instead of the SD version of (145), one could use the conventional 'partial' derivative the way SD used the delta derivative and write

$$\begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ \partial V/\partial V & \partial V/\partial T \\ \partial T/\partial V & \partial T/\partial T \end{bmatrix} \longrightarrow \begin{bmatrix} ? & ? \\ ? & ? \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (148)$$

In SD the partial derivative is an explicit derivative, but by convention, it is a total derivative. Therefore, to make an explicit derivative out of a partial derivative comes with a price. One gets

$$\begin{bmatrix} \partial U/\partial V & \partial U/\partial T \\ \partial P/\partial V & \partial P/\partial T \\ \partial V/\partial V & \partial V/\partial T \\ \partial T/\partial V & \partial T/\partial T \end{bmatrix} \longrightarrow \begin{bmatrix} (\partial U/\partial V)_T & (\partial U/\partial T)_V \\ (\partial P/\partial V)_T & (\partial P/\partial T)_V \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (149)$$

Okay, this system works, but I consider it cumbersome and ugly. Anyway, what else must change by use of the standard convention? Ans: First off, we'd probably never see an equation like (146) or like (147), but if we did, it would probably appear as

$$\frac{\partial\psi'}{\partial\eta'} = \frac{\partial\psi}{\partial\eta} \frac{\partial\eta}{\partial\eta'}. \quad (150)$$

But then what about Rules 1 and 2? They'll have to be modified to apply to the partial derivative instead of the delta derivative. Additional changes would involve the two constraint equations. Eq. (132) becomes

$$\left(\frac{\partial U}{\partial V}\right)_T - T\left(\frac{\partial P}{\partial T}\right)_V + P = 0, \quad (151)$$

and (140) becomes

$$\left(\frac{\partial T}{\partial P}\right)_U - T\left(\frac{\partial V}{\partial U}\right)_P + PJ = 0. \quad (152)$$

Okay, I now make a confession to the notation of conventional thermodynamics: When thermodynamic formulas are derived which require one change in fundamental to follow another in rapid succession, then these ugly subscripts are necessary to keep track of what is being held constant (which reflects the current fundamental). And as I said before, subscripts that indicate a process of constant volume, pressure, entropy, or the like, are indeed useful.

References

- [1] K. Huang. *Statistical Mechanics*. John Wiley & Sons. New York. 126 (1963).
- [2] S.N. Rasband. *Dynamics*. John Wiley & Sons. New York. 152 (1983).
- [3] C. Page. *Physical Mathematics*. D. Van Nostand Co. Princeton, New Jersey. 44 (1955).
- [4] M. Abbott and H. Van Ness. *Theory and Problems of Thermodynamics*. McGraw Hill Book Co. 66-97 (1972).
- [5] A. Taylor and W.R. Mann. *Advanced Calculus*, 2ed. John Wiley & Sons. New York (1972).
- [6] R.C. Buck. *Advanced Calculus*, 3ed. McGraw-Hill Book Co. (1984).