

Avoiding the Use of Lagrange Multipliers. II. Constrained Extrema of Functionals and the Evaluation of Constrained Derivatives

David S. Corti¹ , Ricardo Fariello² 

¹Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN, USA

²Department of Computer Science, State University of Montes Claros, Montes Claros, MG, Brazil

Email: dscorti@purdue.edu, ricardo.fariello@unimontes.br

How to cite this paper: Corti, D.S. and Fariello, R. (2024) Avoiding the Use of Lagrange Multipliers. II. Constrained Extrema of Functionals and the Evaluation of Constrained Derivatives. *Journal of Applied Mathematics and Physics*, 12, 2764-2788.

<https://doi.org/10.4236/jamp.2024.128165>

Received: July 5, 2024

Accepted: August 9, 2024

Published: August 12, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A method for determining the extrema of a real-valued and differentiable function for which its dependent variables are subject to constraints and that avoided the use of Lagrange multipliers was previously presented (Corti and Fariello, *Op. Res. Forum* 2 (2021) 59). The method made use of projection matrices, and a corresponding Gram-Schmidt orthogonalization process, to identify the constrained extrema. Furthermore, information about the second-derivatives of the given function with constraints was generated, from which the nature of the constrained extrema could be determined, again without knowledge of the Lagrange multipliers. Here, the method is extended to the case of functional derivatives with constraints. In addition, constrained first-order and second-order derivatives of the function are generated, in which the derivatives with respect to a given variable are obtained and, concomitantly, the effect of the variations of the remaining chosen set of dependent variables are strictly accounted for. These constrained derivatives are valid not only at the extrema points, and also provide another equivalent route for the determination of the constrained extrema and their nature.

Keywords

Constrained Extrema, Functional Derivatives, Projection Matrices

1. Introduction

Recently, the authors presented [1] a method for determining the extrema of real-valued and differentiable functions for which its dependent variables are subject to constraints and that avoids the use of Lagrange multipliers. The method made use of projection matrices, and a corresponding Gram-Schmidt orthogo-

nalization process, to identify the constrained extrema. Although Lagrange multipliers were not required, a comparison of this approach to the method of Lagrange multipliers nevertheless yielded expressions for the Lagrange multipliers in terms of the gradients of the given function and the additional functions that represent the constraints. Furthermore, information about the second-derivatives of the given function with constraints was generated, from which the nature (*i.e.*, maxima or minima) of the constrained extrema were determined, again without knowledge of the Lagrange multipliers.

Other methods for avoiding the use of Lagrange multipliers have been developed for constrained extrema problems. For example, one prior approach utilized wedge products [2], though only considered first-order variations of the given function. Another approach employed the chain rule and the implicit function theorem [3], and also yields information about second-order derivatives. The nature of the constrained extrema can also be found from the introduction of an extended Hessian matrix [4] [5]. Yet, the values of the Lagrange multipliers are still required when evaluating these matrices, and the general procedure for the case of multiple constraints was not provided. Our proposed method, with its use of well-known aspects of linear algebra and a relatively straightforward incorporation of multiple constraints, may prove computationally more convenient than these prior approaches for some optimization problems with a large number of constraints.

We again note that this method followed some of the ideas, as well as some of the convenient notation, provided in the work of Gál [6]-[11] for the evaluation of first- and second-order functional derivatives with constraints (and also without the use of Lagrange multipliers). To a certain extent, our approach for analyzing constrained extrema without Lagrange multipliers is the discrete version (*i.e.*, the number of variables is finite) of one of the continuum versions (*i.e.*, an “infinite” variable space requiring functional derivatives) presented by Gál. Some aspects of our current discrete approach were therefore hinted at in [11], though not fully developed (at least for multiple constraints).

Besides the direct benefit of solving constrained problems without the need for Lagrange multipliers, which also potentially simplifies the determination of the nature of the extrema, our approach may, upon further exploration, be of additional utility. In various scientific problems of interest that are formulated as constrained problems, the Lagrange multipliers typically have physical meanings. For example, in statistical physics, the maximization of the Gibbs entropy formula subject to certain constraints provides another route to generating the canonical and grand canonical ensembles. In both cases, one of the Lagrange multipliers is associated with the inverse temperature of the thermal reservoir, while in the latter case, another Lagrange multiplier is associated with the chemical potential of the particle reservoir [12]. Here, by direct comparison to the method of Lagrange multipliers, our approach leads to general expressions for the Lagrange multipliers themselves [1]. Given the often found connection between the

Lagrange multipliers and physical quantities, these new expressions may yield some different and useful insights into various constrained problems of physical interest [13]-[16].

We do not explore further these connections here, but instead consider other aspects of our approach that can be applied to other important constrained problems of interest. In particular, we present the extension of our approach to the case of functional derivatives with constraints. In addition, the explicitly constrained first-order and second-order derivatives of the given function (for the discrete variable set) are obtained, in which derivatives with respect to a given variable are generated and, concomitantly, the effect of the variations of the remaining chosen set of dependent variables are strictly accounted. These constrained derivatives are valid not only at the extrema points, thereby providing information that may be of use for various problems of interest, and may also be used for the determination of the constrained extrema and their nature.

In the remainder of this paper, we first present a summary of the method developed in [1], in which local extrema and their nature are determined with constraints, again without the use of Lagrange multipliers. We then present the extension of our approach to the continuum limit (*i.e.*, functional derivatives with constraints), some of the results of which are consistent with the analysis provided in [6]-[11]. Finally, we also consider the determination of constrained first- and second-order derivatives of the given function (for the discrete variable set), *i.e.*, derivatives for which the impact of the constraints are explicitly accounted. These derivatives are valid not just at the extrema points, several results of which have not been obtained previously.

2. Review of the Method Avoiding the Use of Lagrange Multipliers

Consider the real-valued function, f , that is dependent upon the n (initially) independent variables $\mathbf{x} = x_1, x_2, \dots, x_n$. In the multi-dimensional space \mathbf{x} , the (column) vector $\delta\mathbf{x} = \delta x_1, \dots, \delta x_n$ describes a general variation of all the n (initially) independent variables, such that the first-order variation of f or δf , is generated from $\delta f = \delta\mathbf{x}^T \nabla f$ (in which for notational convenience the corresponding transposed row vector is denoted by $(\delta\mathbf{x})^T \equiv \delta\mathbf{x}^T$ and $\nabla = \partial/\partial x_1, \dots, \partial/\partial x_n$). This first-order variation follows from the projection of ∇f onto the direction of $\delta\mathbf{x}$, which is just the directional derivative of f along $\delta\mathbf{x}$.

Let there now be $m < n$ constraints on \mathbf{x} , which are assumed to be represented by the following set of differentiable functions:

$$g_1(\mathbf{x}) = K_1, g_2(\mathbf{x}) = K_2, \dots, g_m(\mathbf{x}) = K_m,$$

and for which K_1, K_2, \dots, K_m are constants. These constraints describe m constraint surfaces in which the vectors $\nabla g_1, \nabla g_2, \dots, \nabla g_m$ are normal to their corresponding surface. To find a direction that is orthogonal to these resulting normal vectors, and as such resides in all of the corresponding m constraint sur-

faces, a Gram-Schmidt orthogonalization process was invoked [1], which generated the following projection matrix \mathbf{Q}_m that can be represented by the following “formal” determinant

$$\mathbf{Q}_m = \frac{1}{D_m} \begin{vmatrix} \nabla \mathbf{g}_1^T \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_1^T \nabla \mathbf{g}_m & \nabla \mathbf{g}_1^T \\ \vdots & \ddots & \vdots & \vdots \\ \nabla \mathbf{g}_m^T \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_m^T \nabla \mathbf{g}_m & \nabla \mathbf{g}_m^T \\ \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_m & \mathbf{I} \end{vmatrix} \quad (1)$$

which is evaluated along the bottom row. In Equation (1), \mathbf{I} is the identity matrix and

$$D_m \equiv \begin{vmatrix} \nabla \mathbf{g}_1^T \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_1^T \nabla \mathbf{g}_m \\ \vdots & \ddots & \vdots \\ \nabla \mathbf{g}_m^T \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_m^T \nabla \mathbf{g}_m \end{vmatrix} \quad (2)$$

which is also the $m \times m$ upper left minor of the matrix in Equation (1). (We are assuming that $D_m \neq 0$, which again also follows from the assumed linear independence of the various $\nabla \mathbf{g}_i$.) For a given vector, \mathbf{Q}_m projects that vector onto the K_1, K_2, \dots, K_m -constraint surfaces, or equivalently onto the K_1, K_2, \dots, K_m -conserving direction. As required, Equation (1) indicates that $\mathbf{Q}_m \nabla \mathbf{g}_k = 0$ for $k = 1, \dots, m$, since $\mathbf{Q}_m \nabla \mathbf{g}_k$ yields a determinant in which the k th and last columns are identical. (When \mathbf{Q}_m operates on a column vector only the last column of Equation (1) operates on that column vector.)

An alternative expression for \mathbf{Q}_m was also obtained [1]

$$\mathbf{Q}_m = \mathbf{I} - \frac{1}{D_m} \sum_{i=1}^m \sum_{j=1}^m \nabla \mathbf{g}_i \nabla \mathbf{g}_j^T A_{ji}^m \quad (3)$$

where A_{ji}^m is the co-factor of element ji in the $m \times m$ matrix used to determine D_m in Equation (2). With Equation (3), one can show, as required for a projection matrix, that \mathbf{Q}_m is idempotent, $\mathbf{Q}_m^T = \mathbf{Q}_m$, $\det(\mathbf{Q}_m) = 0$, and $\text{Tr}(\mathbf{Q}_m) = n - m$ (which also equals the number of non-zero eigenvalues of \mathbf{Q}_m , with each all being equal to one).

Now, the projection of $\delta \mathbf{x}$ onto the K_1, K_2, \dots, K_m -conserving direction, or in the direction residing along the various constraint surfaces, is labeled as $\delta_{K_1 K_2 \dots K_m} \mathbf{x}$, or for convenience $\delta_K \mathbf{x}$, and is given by $\delta_K \mathbf{x} = \mathbf{Q}_m \delta \mathbf{x}$. As noted earlier, the first-order (unconstrained) variation of f or δf , is generated from $\delta f = \delta \mathbf{x}^T \nabla f$, which follows from the projection of ∇f onto the direction of $\delta \mathbf{x}$, which is just the directional derivative of f along $\delta \mathbf{x}$. Analogously, a first-order variation of f that maintains the constraints, denoted as $\delta_{K_1 \dots K_m} f$, or $\delta_K f$, can be generated by projecting ∇f onto the direction of $\delta_K \mathbf{x}$. Hence, the directional derivative of f residing along the constraint surface, or the K_1, K_2, \dots, K_m -conserving first-order variation of f is given by $\delta_K f = \delta_K \mathbf{x}^T \nabla f$, where $\delta_K \mathbf{x}^T \equiv (\delta_K \mathbf{x})^T$. This first-order variation can be, however, generated in a different way. To do so, we define the $K_1 K_2 \dots K_m$ -conserving gradient of f or what we label as $\nabla_K f$, to be the vector that yields, by construction, the

$K_1 K_2 \cdots K_m$ -conserving first-order variation of f when evaluated along the (unconstrained) variation $\delta \mathbf{x}$. In other words, $\nabla_{\mathbf{K}} f$ is defined via $\delta_{\mathbf{K}} f \equiv \delta \mathbf{x}^T \nabla_{\mathbf{K}} f$. Therefore,

$$\delta_{\mathbf{K}} f = \delta \mathbf{x}^T \nabla_{\mathbf{K}} f = \delta_{\mathbf{K}} \mathbf{x}^T \nabla f = (\mathbf{Q}_m \delta \mathbf{x})^T \nabla f \tag{4}$$

Given that $(\mathbf{Q}_m \delta \mathbf{x})^T = \delta \mathbf{x}^T \mathbf{Q}_m^T = \delta \mathbf{x}^T \mathbf{Q}_m$, then

$$\delta_{\mathbf{K}} f = \delta \mathbf{x}^T \nabla_{\mathbf{K}} f = \delta \mathbf{x}^T \mathbf{Q}_m \nabla f \tag{5}$$

or

$$\nabla_{\mathbf{K}} f = \mathbf{Q}_m \nabla f = \frac{1}{D_m} \begin{vmatrix} \nabla g_1^T \nabla g_1 & \cdots & \nabla g_1^T \nabla g_m & \nabla g_1^T \nabla f \\ \vdots & \ddots & \vdots & \vdots \\ \nabla g_m^T \nabla g_1 & \cdots & \nabla g_m^T \nabla g_m & \nabla g_m^T \nabla f \\ \nabla g_1 & \cdots & \nabla g_m & \nabla f \end{vmatrix} \tag{6}$$

(which is again evaluated along the bottom row). From the above, we conclude that $\nabla_{\mathbf{K}} f$ is the projection of ∇f onto the various constraint surfaces. Note that each component of $\nabla_{\mathbf{K}} f$ includes contributions from all the n variables, since $\nabla_{\mathbf{K}} f$ accounts implicitly for how all the variables must change due to the constraints. Furthermore, Equation (6) implies that the $K_1 K_2 \cdots K_m$ -conserving gradient operator, $\nabla_{\mathbf{K}}$, is equal to $\mathbf{Q}_m \nabla$.

At an extremum, $\nabla_{\mathbf{K}} f = \mathbf{Q}_m \nabla f = 0$, so we have that

$$\begin{vmatrix} \nabla g_1^T \nabla g_1 & \cdots & \nabla g_1^T \nabla g_m & \nabla g_1^T \nabla f \\ \vdots & \ddots & \vdots & \vdots \\ \nabla g_m^T \nabla g_1 & \cdots & \nabla g_m^T \nabla g_m & \nabla g_m^T \nabla f \\ \nabla g_1 & \cdots & \nabla g_m & \nabla f \end{vmatrix} = 0 \tag{7}$$

a condition for a constrained extremum that avoids the use of Lagrange multipliers. Equation (7) was shown to be equivalent to what follows from the method of Lagrange multipliers [1]. A comparison of Equation (7) to the method of Lagrange multipliers provides an immediate general expression for the Lagrange multipliers, showing that they are related to the co-factors of $\nabla g_1, \nabla g_2, \dots$ in the bottom row of the matrix in Equation (7) divided by D_m . Furthermore, due to various properties of the determinant (e.g., row and column interchanges), the choices of which constraints are labeled 1, 2, etc. are arbitrary (as to be expected), and will not affect the results presented above. Finally, $\nabla_{\mathbf{K}} f$ is, of course, by construction orthogonal to the vectors $\nabla g_1, \nabla g_2, \dots$. Given Equation (6), $\nabla g_k^T \nabla_{\mathbf{K}} f$ (for $k=1, \dots, m$) generates a determinant in which the k th and last rows are identical, and so $\nabla g_k^T \nabla_{\mathbf{K}} f = 0$. Finally, the introduction of the matrix $\tilde{\mathbf{Q}}_m \equiv D_m \mathbf{Q}_m$ proves convenient in various applications, where $\tilde{\mathbf{Q}}_m^2 = D_m \tilde{\mathbf{Q}}_m$, $\text{Tr}(\tilde{\mathbf{Q}}_m) = (n-m)D_m$, $\det(\tilde{\mathbf{Q}}_m) = 0$ and at an extremum $\tilde{\mathbf{Q}}_m \nabla f = 0$.

The nature of an extremum (*i.e.*, maximum or minimum or saddle point) can be determined from an appropriate matrix of the various corresponding 2nd-order derivatives. For the unconstrained case, the second-order variation of f or $\delta^2 f$, is generated from $\delta^2 f = \delta \mathbf{x}^T (\nabla \nabla^T f) \delta \mathbf{x} = \delta \mathbf{x}^T \mathbf{F} \delta \mathbf{x}$, where $\mathbf{F} \equiv \nabla \nabla^T f$ is the matrix of all the (unconstrained) second-order derivatives of f . This result can be

obtained in the following manner. As noted previously, the unconstrained first-order variation of f is given by $\delta f = \delta \mathbf{x}^T \nabla f$, which implies that one can introduce the (scalar) “variation operator” $\delta \equiv \delta \mathbf{x}^T \nabla$. Hence, the second-order variation operator (another scalar) is equal to

$$\delta^2 = \delta \delta = (\delta \mathbf{x}^T \nabla)(\delta \mathbf{x}^T \nabla) = (\delta \mathbf{x}^T \nabla)(\nabla^T \delta \mathbf{x}) = \delta \mathbf{x}^T (\nabla \nabla^T) \delta \mathbf{x},$$

and so $\delta^2 f = \delta \mathbf{x}^T \mathbf{F} \delta \mathbf{x}$. (Note when δ^2 operates on f , ∇^T does not act on $\delta \mathbf{x}$, but on f only.)

Now, similar to the previous derivation of the $K_1 \cdots K_m$ -conserving gradient, we likewise generate two different expressions for the second-order $K_1 \cdots K_m$ -conserving variation of f or $\delta_K^2 f$. First, a new matrix of constrained second-order derivatives, \mathbf{F}_K , is introduced, such that by construction when twice operated on by $\delta \mathbf{x}$, one obtains $\delta_K^2 f$. Hence, \mathbf{F}_K is defined in the following manner

$$\delta_K^2 f = \delta \mathbf{x}^T \mathbf{F}_K \delta \mathbf{x} \tag{8}$$

On the other hand, from Equation (4), $\delta_K f = \delta_K \mathbf{x}^T \nabla f = \delta \mathbf{x}^T \mathbf{Q}_m \nabla f$, which suggests that we introduce the $K_1 \cdots K_m$ -conserving variation operator $\delta_K \equiv \delta_K \mathbf{x}^T \nabla = \delta \mathbf{x}^T \mathbf{Q}_m \nabla$. Therefore, we also have that

$$\begin{aligned} \delta_K^2 f &= \delta_K \delta_K f = (\delta \mathbf{x}^T \mathbf{Q}_m \nabla)(\delta \mathbf{x}^T \mathbf{Q}_m \nabla f) \\ &= (\delta \mathbf{x}^T \mathbf{Q}_m \nabla)(\mathbf{Q}_m \nabla f)^T \delta \mathbf{x} = \delta \mathbf{x}^T \mathbf{Q}_m \nabla (\mathbf{Q}_m \nabla f)^T \delta \mathbf{x} \end{aligned} \tag{9}$$

where $(\mathbf{Q}_m \nabla f)^T$ is a row vector. Comparing Equations (8) and (9) implies that

$$\mathbf{F}_K = \mathbf{Q}_m \nabla (\mathbf{Q}_m \nabla f)^T = \mathbf{Q}_m \nabla (\nabla f^T \mathbf{Q}_m) \tag{10}$$

in which $(\nabla f^T \mathbf{Q}_m) = (\mathbf{Q}_m \nabla f)^T$. Since the elements of \mathbf{Q}_m may depend upon the variables \mathbf{x} , the order of operation of the gradient operator on \mathbf{Q}_m and ∇f needs to be maintained in Equation (10). Equation (10) can also be obtained in another way. Starting with the following definition of $\mathbf{F}_K \equiv \nabla_K \nabla_K^T f$ [7], and noting that $\nabla_K = \mathbf{Q}_m \nabla$, we have that

$$\mathbf{F}_K = (\mathbf{Q}_m \nabla)(\mathbf{Q}_m \nabla)^T f = \mathbf{Q}_m \nabla (\mathbf{Q}_m \nabla f)^T \tag{11}$$

When determining \mathbf{F}_K , first evaluating $(\mathbf{Q}_m \nabla f)^T$ and then operating on this row vector, initially with ∇ (to obtain a matrix) and subsequently with \mathbf{Q}_m , proves to be convenient. Nonetheless, \mathbf{F}_K can be expanded to yield alternative forms as shown in [1]. Since Equation (11) shows that \mathbf{F}_K is the product of \mathbf{Q}_m and another matrix, $\det(\mathbf{F}_K) = 0$, which is again consistent with the fact that while \mathbf{F}_K includes derivatives with respect to all n variables, ultimately not all of these $K_1 \cdots K_m$ -conserving variations are independent due to the constraint (where we now expect m eigenvalues to be equal to zero due to the m constraints). Finally, \mathbf{F}_K can also be expressed as [1]

$$\mathbf{F}_K = \frac{1}{D_m^2} \tilde{\mathbf{Q}}_m \nabla (\tilde{\mathbf{Q}}_m \nabla f)^T \tag{12}$$

an alternative form that proves convenient in some cases, and which is only ap-

appropriate for the constrained extrema, for which $\tilde{\mathbf{Q}}_m \nabla f = 0$. Ultimately, the nature of the constrained extrema can be obtained from the signs of the eigenvalues of \mathbf{F}_K .

For certain problems, this new method does yield a decrease in the number of steps needed to identify the constrained extrema, compared to the method of Lagrange multipliers and at least after the projection matrix \mathbf{Q}_m has been determined [1]. Moreover, the new method arguably provides a clear advantage over the method of Lagrange multipliers when determining the nature of the constrained extrema. As an example to illustrate the potential benefits of the new method, consider the following sample problem [1], in which we are interested in finding the extremum of the function $f = x_1^2 + x_2^2 + x_3^2$, subject to the two constraints $g_1 = 2x_1 + x_2 = 1$ and $g_2 = x_1 + x_3 = 2$. To begin, we first note that

$$\mathbf{Q}_2 = \frac{1}{6} \begin{vmatrix} 5 & 2 & \nabla g_1^T \\ 2 & 2 & \nabla g_2^T \\ \nabla g_1 & \nabla g_2 & \mathbf{I} \end{vmatrix} = \frac{1}{6} \begin{pmatrix} 1 & -2 & -1 \\ -2 & 4 & 2 \\ -1 & 2 & 1 \end{pmatrix} \quad (13)$$

Therefore,

$$\nabla_{K_1 K_2} f = \mathbf{Q}_2 \nabla f = \frac{1}{3} (x_1 - 2x_2 - x_3, -2x_1 + 4x_2 + 2x_3, -x_1 + 2x_2 + x_3) \quad (14)$$

At the constrained extremum, in which $\nabla_{K_1 K_2} f = 0$, we find with the constraints that $x_1 = 2/3$, $x_2 = -1/3$, and $x_3 = 4/3$. To determine the nature of the extremum, we note for the two given constraints that each element of \mathbf{Q}_2 is a constant, and so $\nabla \mathbf{Q}_2 = 0$. Therefore, from Equation (11),

$\mathbf{F}_{K_1 K_2} = \mathbf{Q}_2 \nabla \nabla f^T \mathbf{Q}_2 = \mathbf{Q}_2 \mathbf{F} \mathbf{Q}_2$. Now, for the given function f

$$\mathbf{F} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} = 2\mathbf{I} \quad (15)$$

Hence, $\mathbf{F}_{K_1 K_2} = 2\mathbf{Q}_2$. The rank of \mathbf{Q}_2 , and so $\mathbf{F}_{K_1 K_2}$, is 1, in which there are two zero eigenvalues (*i.e.*, two constraints). With the trace of $\mathbf{F}_{K_1 K_2}$ equaling 2, the only non-zero eigenvalue is then equal to 2. Thus, the extremum corresponds to a minimum. All of these results can be verified by direct substitution of the constraints into the function f . With 3 variables and 2 constraints, there is only a single degree of freedom for this problem (*i.e.*, only one of the variables can be varied independently of the other two). So, for example, we note that

$$f = x_1^2 + (1 - 2x_1)^2 + (2 - x_1)^2 = 6x_1^2 - 8x_1 + 5.$$

This function of a single variable has an extremum at $x_1 = 2/3$, and a second derivative equal to 12, confirming that this stationary point is a minimum.

3. Extrema for Functional Derivatives with Constraints

In the calculus of variations, which has important applications in mathematics and various physical problems, one attempts to find a particular function that yields an extremum of another function. The function that relates a particular value to each inputted function is called a *functional* [17]. The functional deriva-

tive indicates how the functional changes upon an infinitesimal change of the function upon which the functional depends. When there are additional constraints on the input functions, the extrema of the functionals can be obtained through the use of functional derivatives and the method of Lagrange multipliers. Yet, formally determining the nature of these extrema with this method is unclear, particularly for the case of functionals that depend upon several functions with multiple constraints. Hence, as noted for the extrema of functions with constraints, avoiding the use of Lagrange multipliers when considering the extrema of functionals with constraints, particularly when evaluating the nature of these extrema, may prove beneficial for various problems of interest.

The extension of our method to the case of finding extrema for functional derivatives with constraints can be obtained by analogy, or by considering the limit of $\delta \mathbf{x}$ when there are an “infinite number of (continuous) variables”. Thus, wherever a $\delta \mathbf{x}$ appears, it is replaced by $\delta y(\mathbf{x})$, where y is a function of the now continuous variable set \mathbf{x} . So, for f that is now a functional of y , or $f[y(\mathbf{x})]$, we have that

$$\begin{aligned}\nabla f &\rightarrow \frac{\delta f}{\delta y(\mathbf{x})} \\ \nabla g_i &\rightarrow \frac{\delta g_i}{\delta y(\mathbf{x})} \\ \nabla_{\kappa} f &\rightarrow \frac{\delta f}{\delta_{\kappa} y(\mathbf{x})}\end{aligned}\quad (16)$$

where each of the above can be thought of as a “column vector of infinite size”. Therefore, we also have, for example, that

$$\begin{aligned}\nabla g_i^{\top} \nabla g_j &\rightarrow \int d\mathbf{x} \frac{\delta g_i}{\delta y(\mathbf{x})} \frac{\delta g_j}{\delta y(\mathbf{x})} \\ \nabla g_i^{\top} \nabla f &\rightarrow \int d\mathbf{x} \frac{\delta g_i}{\delta y(\mathbf{x})} \frac{\delta f}{\delta y(\mathbf{x})}\end{aligned}\quad (17)$$

The last two expressions can be then used to evaluate the corresponding matrix elements in Equation (1). In addition, we have the following orthogonality condition

$$\nabla g_i^{\top} \nabla_{\kappa} f = 0 = \int d\mathbf{x} \frac{\delta g_i}{\delta y(\mathbf{x})} \frac{\delta f}{\delta_{\kappa} y(\mathbf{x})}\quad (18)$$

Hence, for the case of a single constraint, g , we have that

$$\frac{\delta f}{\delta_{\kappa} y(\mathbf{x})} = \frac{\delta f}{\delta y(\mathbf{x})} - \frac{\delta g}{\delta y(\mathbf{x})} \frac{\int d\mathbf{x}' \frac{\delta g}{\delta y(\mathbf{x}')} \frac{\delta f}{\delta y(\mathbf{x}')}}{\int d\mathbf{x}' \frac{\delta g}{\delta y(\mathbf{x}')} \frac{\delta g}{\delta y(\mathbf{x}')}}\quad (19)$$

and so at an extremum

$$\frac{\delta f}{\delta y(\mathbf{x})} = \frac{\delta g}{\delta y(\mathbf{x})} \frac{\int d\mathbf{x}' \frac{\delta g}{\delta y(\mathbf{x}')} \frac{\delta f}{\delta y(\mathbf{x}')}}{\int d\mathbf{x}' \frac{\delta g}{\delta y(\mathbf{x}')} \frac{\delta g}{\delta y(\mathbf{x}')}}\quad (20)$$

As the two integrals are the same for each \mathbf{x} , they are constants. Therefore, Equation (20) yields a result similar to what would be obtained with the use of Lagrange multipliers (though, in contrast, we again get an immediate expression for the multiplier with the new method). Note that Equations (19) and (20) are different from what follows from the analysis of Gál [6]-[11]. Gál hinted at their derivation [11], and does not explicitly include them, as it was noted that in some cases the integrals may be unbounded (which may occur when the integrals are not evaluated over a finite range of \mathbf{x}). Thus, this direct extension from the discrete case may not be valid in all situations, which was one of the motivations for Gál to obtain different, and more generally applicable, expressions for the functional derivatives with constraints. Nonetheless, at least for the case of a single constraint, the results following from Equation (20) (assuming the integrals are bounded) are equivalent to what follows from Gál's approach (as they should, as both the above and Gal's expression yield a constant that appears next to the functional derivative of the constraint, just like what appears from the method of Lagrange multipliers applied to functional derivatives).

We also consider the case of second-order functional derivatives. The matrix F of discrete (unconstrained) second derivatives now becomes the following in the continuum or functional limit

$$F \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} \rightarrow \frac{\delta^2 f}{\delta y(\mathbf{x}) \delta y(\mathbf{x}')} \quad (21)$$

(where i denotes the row and j the column of F for the discrete case). The determination of F_K in the same functional limit can be obtained, for example, from Equation (12), in which the discrete result is first generated followed by taking the continuum limit. The result is in general complicated, and is not provided here. But a significant simplification does arise for the case of a single constraint and when the second functional derivative of the constraint is always equal to zero. For this case (a single constraint in which its second functional derivative always vanishes), one finds that (as shown in the **Appendix**)

$$\begin{aligned} F_K &\equiv \frac{\delta^2 f}{\delta_K y(\mathbf{x}) \delta_K y(\mathbf{x}')} \\ &= \frac{\delta^2 f}{\delta y(\mathbf{x}) \delta y(\mathbf{x}')} - \frac{1}{D_1} \frac{\delta g}{\delta y(\mathbf{x})} \int d\mathbf{x}'' \frac{\delta g}{\delta y(\mathbf{x}'')} \frac{\delta^2 f}{\delta y(\mathbf{x}') \delta y(\mathbf{x}'')} \\ &\quad - \frac{1}{D_1} \frac{\delta g}{\delta y(\mathbf{x}')} \int d\mathbf{x}'' \frac{\delta g}{\delta y(\mathbf{x}'')} \frac{\delta^2 f}{\delta y(\mathbf{x}) \delta y(\mathbf{x}'')} \\ &\quad + \frac{1}{D_1^2} \frac{\delta g}{\delta y(\mathbf{x})} \frac{\delta g}{\delta y(\mathbf{x}')} \iint d\mathbf{x}'' d\mathbf{x}''' \frac{\delta g}{\delta y(\mathbf{x}'')} \frac{\delta g}{\delta y(\mathbf{x}''')} \frac{\delta^2 f}{\delta y(\mathbf{x}'') \delta y(\mathbf{x}''')} \end{aligned} \quad (22)$$

in which

$$D_1 = \int d\mathbf{x}'' \frac{\delta g}{\delta y(\mathbf{x}'')} \frac{\delta g}{\delta y(\mathbf{x}'')} \quad (23)$$

Through the use of dummy variables, each term in Equation (22) retains its dependence on the functional derivatives of $\delta y(\mathbf{x})$ and $\delta y(\mathbf{x}')$. The corresponding eigenvalue equation for the second functional derivatives is given by [17]

$$K\psi = \lambda\psi \quad (24)$$

where

$$K\psi(\mathbf{x}) = \int d\mathbf{x}' \psi(\mathbf{x}') \frac{\delta^2 f}{\delta_K y(\mathbf{x}) \delta_K y(\mathbf{x}')} \quad (25)$$

This is an integral equation, and gives rise to a continuum of eigenvalues (which can again be used to test the nature of the extrema).

In addition, there are cases for which f and the various constraints are functionals of multiple functions, where, for example, we have that $f[y_1(\mathbf{x}), \dots, y_n(\mathbf{x})]$ subject to the constraints of $g_i[y_1(\mathbf{x}), \dots, y_n(\mathbf{x})] = K_i$ for $i = 1, \dots, m$. Here, we can treat the various results as arising from the following “extended vectors”, in which

$$\begin{aligned} \delta y(\mathbf{x}) &\equiv \delta y_1(\mathbf{x}); \delta y_1(\mathbf{x}); \dots; \delta y_n(\mathbf{x}) \\ \nabla f &\equiv \frac{\delta f}{\delta y_1(\mathbf{x})}; \frac{\delta f}{\delta y_2(\mathbf{x})}; \dots; \frac{\delta f}{\delta y_n(\mathbf{x})} \\ \nabla g_i &\equiv \frac{\delta g_i}{\delta y_1(\mathbf{x})}; \frac{\delta g_i}{\delta y_2(\mathbf{x})}; \dots; \frac{\delta g_i}{\delta y_n(\mathbf{x})} \end{aligned} \quad (26)$$

The above are again “column vectors”, with each set of functional derivatives appearing one after the other in succession and each spanning the “same height” within the column. So then, we have, for example, that

$$\begin{aligned} \nabla g_i^T \nabla g_j &\rightarrow \sum_{k=1}^n \int d\mathbf{x} \frac{\delta g_i}{\delta y_k(\mathbf{x})} \frac{\delta g_j}{\delta y_k(\mathbf{x})} \\ \nabla g_i^T \nabla f &\rightarrow \sum_{k=1}^n \int d\mathbf{x} \frac{\delta g_i}{\delta y_k(\mathbf{x})} \frac{\delta f}{\delta y_k(\mathbf{x})} \end{aligned} \quad (27)$$

Second functional derivatives can also be obtained, but the resulting expressions become complicated very quickly.

Finally, let's consider an example of the application of some of the above results. For a given curve in the y - x plane, the differential arc length ds is $(1 + (dy/dx)^2)^{1/2} dx$. Given two points a and b on the x -axis, the area enclosed by the curve with the x -axis is given by $\int_a^b y dx$. We can then ask the following question: what is the shape of the curve of fixed arc length L joining the given points, which, along with the x -axis, encloses the largest area? In other words, we want to minimize $f = \int_a^b y dx$ subject to the constraint of

$$g = \int_a^b [1 + (y')^2]^{1/2} dx = L,$$

where $y' = dy/dx$. The shape should be that of a circle that passes through the two points on the x -axis.

Now, with

$$\frac{\delta f}{\delta y(x)} = 1$$

$$\frac{\delta g}{\delta y(x)} = -\frac{y''}{[1+(y')^2]^{\frac{3}{2}}} = -\frac{d}{dx} \left(\frac{y'}{[1+(y')^2]^{\frac{1}{2}}} \right) \quad (28)$$

then, according to Equation (20), at an extremum,

$$\frac{\delta f}{\delta y(x)} = \frac{\delta g}{\delta y(x)} \frac{\int dx' \frac{\delta g}{\delta y(x')} \frac{\delta f}{\delta y(x')}}{\int dx' \frac{\delta g}{\delta y(x')} \frac{\delta g}{\delta y(x')}} \quad (29)$$

Instead of attempting to directly evaluate the above two integrals, we note they have the same values for each x (as they are evaluated over the same stationary curve y and limits of integration, $x' = a$ to $x' = b$). Their ratio is also equal to some constant σ (which is simply the Lagrange multiplier for this problem), and so Equations (28) and (29) indicate that

$$1 = -\frac{d}{dx} \left(\frac{\sigma y'}{[1+(y')^2]^{\frac{1}{2}}} \right) \quad (30)$$

Integrating both sides, after which appears the integration constant c_1 , leads upon rearrangement to

$$y' = \pm \frac{x + c_1}{[\sigma^2 - (x + c_1)^2]^{\frac{1}{2}}} \quad (31)$$

An additional integration, and further rearrangements with a new constant of integration c_2 , leads to the final result of

$$(y - c_2)^2 + (x + c_1)^2 = \sigma^2 \quad (32)$$

which is, as expected, the equation of a circle with radius σ . If so desired, we can determine σ using the constraint, which will be related to L , a , and b . Three points are needed to specify uniquely a given circle, or equivalently, two points and the circumference. If L is not much larger than $b - a$, the center of the circle will reside below the x -axis. If L is much larger than $b - a$, the center of the circle will reside above the x -axis. (By symmetry, the center of the circle should reside halfway between a and b , so that $c_1 = -(a + b)/2$.) Finally, this method provides some insight into why σ , or the Lagrange multiplier, is the radius of the circle. The second expression of Equation (28) indicates that the functional derivative of g is the negative of the curvature of the curve y , or the negative inverse of the radius of curvature [18]. Hence, with Equation (29), the Lagrange multiplier is related to a particular weighted average of the curvature of the sta-

tionary curve. Since the curvature of a circle is constant, the Lagrange multiplier (*ex post facto*) is seen to be equal to the radius of the circle.

4. Determination of the First- and Second-Order Explicitly Constrained Derivatives

Returning to the discrete case, or the finite variable set $\mathbf{x} = x_1, x_2, \dots, x_n$, we again note that the $K_1 K_2 \dots K_m$ -conserving gradient of f , or $\nabla_{\mathbf{K}} f$, is the projection of the unconstrained gradient, ∇f , onto the intersection of all the constraint surfaces. As a result, each one of the n components of $\nabla_{\mathbf{K}} f$ is, in general, not equal to zero, even though the n variables cannot be independently varied. Since the effect of the constraint is implicitly, and not explicitly, accounted for in $\nabla_{\mathbf{K}} f$, the specific meaning of its components is unclear (besides simply being the projection of $\nabla_{\mathbf{K}} f$ onto each of the separate n initially independent directions). While it is tempting to represent the i th component of $\nabla_{\mathbf{K}} f$ as $(\partial f / \partial x_i)_{\mathbf{K}}$, this partial derivative cannot be meaningful as written. Unlike what was implied for $\partial g / \partial x_i$, $(\partial f / \partial x_i)_{\mathbf{K}}$ cannot also imply that all variables besides x_i are held fixed, otherwise the \mathbf{K} -conserving constraints cannot be satisfied. Such a \mathbf{K} -conserving partial derivative, if it is to be meaningfully written, must indicate explicitly which variables are held fixed and which variables are still allowed to vary. For the case of n variables with a single constraint, only $n - 2$ variables can be held constant upon partial differentiation, with the variation of the remaining 2 variables being coupled via the constraint. For example, a meaningful constrained first derivative in this case can be represented by $(\partial f / \partial x_i)_{\mathbf{K}, x[i, n]}$, in which we have arbitrarily chosen x_n to be the additional parameter that is allowed to vary and $x[i, n]$ denotes that all \mathbf{x} are to be held fixed except x_i and x_n (with $i \neq n$). There are of course several choices for the dependent or “floating” variable in these partial derivatives. But as presently derived, $\nabla_{\mathbf{K}} f$ does not immediately provide information about any one of these possible derivatives $(\partial f / \partial x_i)_{\mathbf{K}, x[i, n]}$.

4.1. Case of a Single Constraint

When there is a single constraint, the derivation of $(\partial f / \partial x_i)_{\mathbf{K}, x[i, n]}$ is straightforward. Using the chain rule, we have that

$$\left(\frac{\partial f}{\partial x_i} \right)_{\mathbf{K}, x[i, n]} = \left(\frac{\partial f}{\partial x_i} \right)_{x[i]} + \left(\frac{\partial f}{\partial x_n} \right)_{x[n]} \left(\frac{\partial x_n}{\partial x_i} \right)_{\mathbf{K}, x[i, n]} \quad (33)$$

where, for example, $(\partial f / \partial x_i)_{x[i]} = \partial f / \partial x_i$ is the partial derivative in which all variables are held fixed except x_i (*i.e.*, the standard unconstrained partial derivatives). Noting for the single constraint that $dg = \nabla g^T \delta \mathbf{x} = 0$, then

$$\left(\frac{\partial x_n}{\partial x_i} \right)_{\mathbf{K}, x[i, n]} = - \frac{\partial g / \partial x_i}{\partial g / \partial x_n} \quad (34)$$

a result that is only meaningful if x_n appears explicitly within the constraint,

i.e., $\partial g/\partial x_n \neq 0$. Equation (34) also follows from the implicit function theorem, or implicit differentiation [19]. Equations (33) and (34) can be combined to evaluate the constrained first-order derivative for any appropriate choice of i and n .

Since \mathbf{Q}_m is the projection matrix onto the K -conserving direction, and thus contains information on variations that reside within the intersection of all those surfaces, Equations (33) and (34) can likewise be obtained using various properties of \mathbf{Q}_m . We therefore repeat the derivation of these equations, along with generating some additional new relations, making use of \mathbf{Q}_m . While for a single constraint this approach is admittedly less straightforward than how Equation (33) was already derived, generating these relations in another way nevertheless proves to be useful. For the case of multiple constraints, the use of \mathbf{Q}_m and its various properties greatly simplifies the derivation of the explicitly constrained first- and second-order derivatives.

To begin, and to illustrate the approach, we again consider the case in which there is a single constraint, g . Now, a set of vectors that reside within, or are tangent to, the constraint surface can be generated from the eigenvectors of \mathbf{Q}_1 . If $\mathbf{u}^T = [u_1, u_2, \dots, u_n]$ denotes an eigenvector of \mathbf{Q}_1 corresponding to one of its $n-1$ eigenvalues with value 1, then by definition $\mathbf{Q}_1 \mathbf{u} = \mathbf{u}$, *i.e.*, \mathbf{u} already resides within the constraint surface. With Equation (1), $\mathbf{Q}_1 \mathbf{u} = \mathbf{u}$ reduces to $\nabla g(\nabla g^T \mathbf{u}) = 0$, which yields n identical equations, all of which are given by

$$\left(\frac{\partial g}{\partial x_1}\right)u_1 + \left(\frac{\partial g}{\partial x_2}\right)u_2 + \dots + \left(\frac{\partial g}{\partial x_n}\right)u_n = 0 \tag{35}$$

Since \mathbf{u} , as an eigenvector of \mathbf{Q}_1 , is orthogonal to ∇g , the vector normal to the constraint surface, we have that $\nabla g^T \mathbf{u} = 0$. To generate a set of $n-1$ eigenvectors that all correspond to Equation (35), we need to set the values of $n-2$ components of \mathbf{u} in the above equation.

For example, if we are interested in the relative variations of, say, x_1 and x_n , such that x_2 through x_{n-1} are constant and the constraint is maintained, then an eigenvector that corresponds to such a variation is obtained from Equation (35) upon setting u_2 through u_{n-1} equal to zero, or $\left(\frac{\partial g}{\partial x_1}\right)u_1 = -\left(\frac{\partial g}{\partial x_n}\right)u_n$. If

we are further interested in the resulting variation in x_n relative to a given variation in x_1 , then we set $u_1 = 1$ and solve for $u_n = -(\partial g/\partial x_1)/(\partial g/\partial x_n)$, which corresponds to Equation (34). The resulting eigenvector that is in the direction of this variation is $\mathbf{u}^T = [1, 0, 0, \dots, -(\partial g/\partial x_1)/(\partial g/\partial x_n)]$. More formally, the variations of x_1 and x_n (based on a change in x_1), such that x_2 through x_{n-1} are constant and the constraint is maintained, is given by $\delta_{K,[1,n]} \mathbf{x} = \delta x_1 \mathbf{u}$. Now, the K -conserving first-order variation of f such that x_2 through x_{n-1} are also constant, or $\delta_{K,[1,n]} f$, is generated by projecting ∇f onto the direction of this eigenvector. Hence,

$$\delta_{K,[1,n]} f = \delta x_1 \left(\frac{\partial f}{\partial x_1}\right)_{K,x[1,n]} = \delta_{K,[1,n]} \mathbf{x}^T \nabla f = \delta x_1 \mathbf{u}^T \nabla f \tag{36}$$

or $(\partial f / \partial x_1)_{K,x[1,n]} = \mathbf{u}^T \nabla f$, a result which upon evaluation matches Equations (33) and (34).

By again selecting x_n to be that variable which depends upon the variation of one of the other variables, we can generate the remaining $n-2$ eigenvectors corresponding to Equation (35) in exactly the same manner. The full set of $n-1$ eigenvectors, for the arbitrary choice of x_n as the dependent variable, is $\mathbf{u}_1^T = [1, 0, 0, \dots, -(\partial g / \partial x_1) / (\partial g / \partial x_n)]$, $\mathbf{u}_2^T = [0, 1, 0, \dots, -(\partial g / \partial x_2) / (\partial g / \partial x_n)]$, \dots , $\mathbf{u}_{n-1}^T = [0, 0, 0, \dots, 1, -(\partial g / \partial x_{n-1}) / (\partial g / \partial x_n)]$. Note that these $n-1$ eigenvectors are linearly independent since $c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \dots + c_{n-1} \mathbf{u}_{n-1} = 0$ requires each of the constants to equal zero or $c_1 = c_2 = \dots = c_{n-1} = 0$. Hence, these eigenvectors form a basis set to describe all other variations that reside in the constraint surface. (There are of course other linearly independent basis sets that can be generated, based on the choice of which variable is selected as the common dependent variable.) The projection of ∇f onto each of these eigenvectors generates in turn the following $n-1$ constrained first-order derivatives of f : $(\partial f / \partial x_1)_{K,x[1,n]}$, $(\partial f / \partial x_2)_{K,x[2,n]}$, \dots , $(\partial f / \partial x_{n-1})_{K,x[n-1,n]}$.

With n choices for the first-order variation of interest, followed by $n-1$ additional choices for the dependent variable, there are $n(n-1)$ K -conserving first-order derivatives of f that can be generated. Nevertheless, Equations (33) and (34) imply, for example, that $(\partial g / \partial x_n)(\partial f / \partial x_i)_{K,x[i,n]} + (\partial g / \partial x_i)(\partial f / \partial x_n)_{K,x[i,n]} = 0$. Hence, not all of these derivatives are independent. The generation of a linearly independent set of $n-1$ eigenvectors indicates that only $n-1$ of these derivatives are independent (as the eigenvector corresponding to one of the $n(n-1)$ derivatives can therefore be obtained via a linear combination of the chosen $n-1$ basis set).

Thus, at an extremum, each one of these linearly independent K -conserving first-order derivatives should vanish. Unlike the extremal condition based on $\nabla_K f = 0$, which corresponds to n equations, the use of, for example, $(\partial f / \partial x_i)_{K,x[i,n]} = 0$ ($i=1, \dots, n-1$) at an extrema, corresponds to only $n-1$ equations (which is consistent with the fact that there are only $n-1$ degrees of freedom for this system with a single constraint). Although there is a reduction in the number of conditions to solve at an extrema, additional work is required for this approach. The application of $\nabla_K f = 0$ is rather straightforward, and still seems to be the preferred approach, as it avoids the additional required step of generating the eigenvectors of \mathcal{Q}_1 (although the eigenvectors for the case of a single constraint are rather straightforward to obtain).

All of these $n-1$ independent variations can be represented in the following manner. Let $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_{n-1}]$ be the $n \times (n-1)$ matrix whose columns are the linearly independent eigenvectors obtained earlier. Hence, $\delta_{K,[i,n]} \mathbf{x}_{n-1} = \mathbf{U} \delta \mathbf{x}_{n-1}$, where $\delta \mathbf{x}_{n-1}$ indicates that we are now only dealing with the first $n-1$ variations, and does not include the variation of the n th variable. The constrained first-order variation of f based on these linearly independent $n-1$ variations is therefore given by

$$\delta_K f = \delta_{K,[i,n]} \mathbf{x}_{n-1}^T \nabla f = \delta \mathbf{x}_{n-1}^T \mathbf{U}^T \nabla f \tag{37}$$

Hence, $\nabla_{K,[i,n]} f = \mathbf{U}^T \nabla f$, where the i th component of $\nabla_{K,[i,n]}$ is $(\partial/\partial x_i)_{K,x[i,n]}$. Note that, just by considering their dimensions, $\nabla_K f = \mathbf{Q}_1 \nabla f \neq \mathbf{U}^T \nabla f = \nabla_{K,[i,n]} f$. But a connection between these two gradients can be obtained. Since the columns of \mathbf{U} are the eigenvectors of \mathbf{Q}_1 , then $\mathbf{Q}_1 \mathbf{U} = \mathbf{U}$, and so (after transposing both sides) $\mathbf{U}^T \mathbf{Q}_1 = \mathbf{U}^T$. Therefore,

$$\nabla_{K,[i,n]} f = \mathbf{U}^T \nabla f = \mathbf{U}^T \mathbf{Q}_1 \nabla f = \mathbf{U}^T \nabla_K f \tag{38}$$

which indicates that the individual terms of $\nabla_{K,[i,n]} f$ are the further projections of $\nabla_K f$ onto each of the linearly independent eigenvectors of \mathbf{Q}_1 , e.g., $\nabla_{K,[1,n]} f = (\partial f/\partial x_1)_{K,x[1,n]} = \mathbf{u}_1^T \nabla_K f = \mathbf{u}_1^T \nabla f$.

Now, by the construction of the eigenvectors, we have that

$$\begin{aligned} \left(\frac{\partial g}{\partial x_i}\right) \left(\frac{\partial g}{\partial x_n}\right) \mathbf{u}_i^T \nabla f &= \left(\frac{\partial g}{\partial x_i}\right) \left(\frac{\partial g}{\partial x_n}\right) \left(\frac{\partial f}{\partial x_i}\right) - \left(\frac{\partial g}{\partial x_i}\right)^2 \left(\frac{\partial f}{\partial x_n}\right) \\ &= \left(\frac{\partial g}{\partial x_i}\right) \left(\frac{\partial g}{\partial x_n}\right) \left(\frac{\partial f}{\partial x_i}\right)_{K,x[i,n]} \end{aligned} \tag{39}$$

Summing the above from $i = 1, \dots, n-1$, and then comparing with $\nabla_K f = \mathbf{Q}_1 \nabla f$, we find after rearrangement that the n th component of $\nabla_K f$ is given by

$$\left(\frac{\partial f}{\partial x_n}\right)_K \equiv \left(\frac{\partial f}{\partial x_n}\right) - \frac{1}{D_1} \left(\frac{\partial g}{\partial x_n}\right) (\nabla g^T \nabla f) = -\frac{1}{D_1} \left(\frac{\partial g}{\partial x_n}\right) \sum_{i=1}^{n-1} \left(\frac{\partial g}{\partial x_i}\right) \left(\frac{\partial f}{\partial x_i}\right)_{K,x[i,n]} \tag{40}$$

which provides a connection between the components of $\nabla_K f$ and the various constrained first-order derivatives. Equation (40) applies to any component of $\nabla_K f$ or for which variable is allowed to “float”, simply by replacing n with j .

Explicitly constrained second-order derivatives of f , such as $(\partial^2 f/\partial x_1^2)_{K,x[1,n]}$, can also be obtained, and are best done so by using the operator formalism and the repeated application of the chain rule. For example, Equations (33) and (36) imply that

$$\left(\frac{\partial}{\partial x_1}\right)_{K,x[1,n]} = \frac{\partial}{\partial x_1} + \left(\frac{\partial}{\partial x_n}\right) \left(\frac{\partial x_n}{\partial x_1}\right)_{K,x[1,n]} = \mathbf{u}_1^T \nabla \tag{41}$$

Thus, by making sure to maintain the repeated and correct ordering of the gradient operator, one can show that

$$\left(\frac{\partial^2}{\partial x_1^2}\right)_{K,x[1,n]} = \mathbf{u}_1^T \nabla (\mathbf{u}_1^T \nabla) = \mathbf{u}_1^T (\nabla \mathbf{u}_1^T) \nabla + \mathbf{u}_1 \mathbf{u}_1^T : \nabla \nabla^T \tag{42}$$

where the second term on the far-right side corresponds to the Frobenius inner product of the two matrices. Consequently,

$$\left(\frac{\partial^2 f}{\partial x_1^2}\right)_{K,x[1,n]} = \mathbf{u}_1^T (\nabla \mathbf{u}_1^T) \nabla f + \mathbf{u}_1 \mathbf{u}_1^T : \mathbf{F} = \mathbf{u}_1^T (\nabla \mathbf{u}_1^T) \nabla f + \mathbf{u}_1^T \mathbf{F} \mathbf{u}_1 \tag{43}$$

Similar operations can be performed to obtain the mixed second-order deriva-

tives, e.g., first performing $(\partial/\partial x_1)_{K,x[1,n]}$ followed by $(\partial/\partial x_2)_{K,x[2,n]}$. In general, the result becomes

$$\left(\frac{\partial}{\partial x_i}\right)_{K,x[i,n]}\left(\frac{\partial f}{\partial x_j}\right)_{K,x[j,n]} = \mathbf{u}_i^T (\nabla \mathbf{u}_j^T) \nabla f + \mathbf{u}_i \mathbf{u}_j^T : F = \mathbf{u}_i^T (\nabla \mathbf{u}_j^T) \nabla f + \mathbf{u}_i^T F \mathbf{u}_j \quad (44)$$

As these are derivatives that explicitly maintain the constraint (for an arbitrary subset of variables), Equation (44) indicates that these mixed second derivatives ($i \neq j$) are not necessarily equal, as in general $\nabla \mathbf{u}_i^T \neq \nabla \mathbf{u}_j^T$. There are various types of constraints, or forms of the constraint function g , in which $\nabla \mathbf{u}_i^T = \nabla \mathbf{u}_j^T$, and so the symmetric nature of the mixed second derivatives is recovered in these cases.

A matrix of these explicit K -conserving second-order derivatives of f can be obtained from Equation (44). For example, the K -conserving second-order variation of f based on the chosen linearly independent $n-1$ variations is, as before, given by or again noting that

$$\begin{aligned} \delta_K^2 f &= \delta_K \delta_K f = \left(\delta_{K,[i,n]} \mathbf{x}_{n-1}^T \nabla\right) \left(\delta_{K,[i,n]} \mathbf{x}_{n-1}^T \nabla f\right) \\ &= \left(\delta \mathbf{x}_{n-1}^T \mathbf{U}^T \nabla\right) \left(\delta \mathbf{x}_{n-1}^T \mathbf{U}^T \nabla f\right) = \delta \mathbf{x}_{n-1}^T \mathbf{U}^T \nabla \left(\mathbf{U}^T \nabla f\right)^T \delta \mathbf{x}_{n-1} \end{aligned} \quad (45)$$

which allows us to introduce the following new matrix

$$\mathbf{F}_{K,n-1} = \mathbf{U}^T \nabla \left(\mathbf{U}^T \nabla f\right)^T = \mathbf{U}^T \nabla f^T (\nabla \mathbf{U}) + \mathbf{U}^T \mathbf{F} \mathbf{U} = \mathbf{U}^T \nabla f^T (\nabla \mathbf{U}) + \mathbf{U} \mathbf{U}^T : \mathbf{F} \quad (46)$$

The above also could have been obtained from defining $\mathbf{F}_{K,n-1} \equiv \nabla_{K,[i,n]} \nabla_{K,[i,n]}^T f$ and again noting that $\nabla_{K,[i,n]} f = \mathbf{U}^T \nabla f$. This is an $(n-1) \times (n-1)$ matrix that is not necessarily symmetric. The eigenvalues of $\mathbf{F}_{K,n-1}$ can likewise be used to determine the nature of the extrema, which given its smaller size may be more convenient to evaluate than \mathbf{F}_K .

As an illustration of the new approach, consider $f = x_1 x_2$ subject to the constraint of $g = x_1^2 + x_2^2 = 1$. Upon direct substitution using the constraint, $f = \pm x_1 (1 - x_1^2)^{1/2}$ (where the + sign corresponds to $x_2 > 0$ and the - sign to $x_2 < 0$). Therefore, $(df/dx_1)_{K,x[1,2]} = \pm (1 - 2x_1^2) / (1 - x_1^2)^{1/2} = (x_2^2 - x_1^2) / x_2$ and $(d^2 f/dx_1^2)_{K,x[1,2]} = \pm x_1 (2x_1^2 - 3) / (1 - x_1^2)^{3/2} = x_1 (x_1^2 - x_2^2 - 2) / x_2^3$. This function has several constrained extrema [1], at which $(\partial f/\partial x_1)_{K,x[1,2]} = 0$ (and taking $x_2 \neq 0$), which implies that $x_2^2 = x_1^2$. With the constraint, there are then four extrema occurring at $x_1 = \pm \frac{1}{\sqrt{2}}$ and $x_2 = \pm \frac{1}{\sqrt{2}}$.

Now, for this problem, and using Equation (1),

$$\mathbf{Q}_1 = \begin{pmatrix} 1 - x_1^2 & -x_1 x_2 \\ -x_1 x_2 & 1 - x_2^2 \end{pmatrix} = \begin{pmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_2 & x_1^2 \end{pmatrix} \quad (47)$$

Thus, $\mathbf{U}^T = \mathbf{u}_1^T = (1, -x_1/x_2)$ and so

$$\left(\frac{\partial f}{\partial x_1}\right)_{K,x[1,2]} = \mathbf{u}_1^T \nabla f = \frac{\partial f}{\partial x_1} - \frac{x_1}{x_2} \frac{\partial f}{\partial x_2} = x_2 - \frac{x_1^2}{x_2} = \frac{x_2^2 - x_1^2}{x_2} \tag{48}$$

matching the result obtained by direct substitution. In addition, from Equation (43)

$$\begin{aligned} \left(\frac{\partial^2 f}{\partial x_1^2}\right)_{K,x[1,2]} &= \begin{pmatrix} 1 & -\frac{x_1}{x_2} \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{x_2} \\ 0 & \frac{x_1}{x_2^2} \end{pmatrix} \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} + \begin{pmatrix} 1 & -\frac{x_1}{x_2} \\ -\frac{x_1}{x_2} & \frac{x_1^2}{x_2^2} \end{pmatrix} : \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= -\frac{x_1}{x_2} \left(1 + \frac{x_1^2}{x_2^2}\right) - \frac{2x_1}{x_2} = -\frac{x_1(1+2x_2^2)}{x_2^3} = \frac{x_1(x_1^2 - x_2^2 - 2)}{x_2^3} \end{aligned} \tag{49}$$

again matching the previous result. In addition, at the extrema, Equation (49) indicates that $(\partial^2 f / \partial x_1^2)_{K,x[1,2]} = -2x_1/x_2^3 = -4x_1/x_2$. Hence, the two extrema in which x_1 and x_2 have the same sign correspond to maxima, while those with opposite signs correspond to minima.

4.2. Case of Multiple Constraints

Meaningful constrained first-order derivatives of f for the case of multiple constraints can be obtained in a manner similar to what was done for a single constraint. With n total variables and m constraints ($m < n$), there are now only $n - m$ degrees of freedom. Hence, we can determine an explicitly constrained first-order derivative of f with respect to a variation in x_i , with m additional variables also allowed to vary because of the constraints, and with the remaining $n - (m + 1)$ variables held fixed. If, for example, $x_{n-(m-1)}$ through x_n are chosen to be the m dependent or “floating parameters”, then via the chain rule we have that

$$\begin{aligned} \left(\frac{\partial f}{\partial x_i}\right)_{K,x[i,n-(m-1),\dots,n]} &= \frac{\partial f}{\partial x_i} + \left(\frac{\partial f}{\partial x_{n-(m-1)}}\right) \left(\frac{\partial x_{n-(m-1)}}{\partial x_i}\right)_{K,x[i,n-(m-1),\dots,n]} + \dots \\ &\quad + \left(\frac{\partial f}{\partial x_n}\right) \left(\frac{\partial x_n}{\partial x_i}\right)_{K,x[i,n-(m-1),\dots,n]} \end{aligned} \tag{50}$$

in which $x[i, n - (m - 1), \dots, n]$ again denotes the variables that are not held constant in the derivatives, $K \equiv K_1, \dots, K_m$ and, for example, $\partial f / \partial x_i = (\partial f / \partial x_i)_{x[i]}$. Unlike the case of a single constraint, due to the complicated and possible coupled nature of the multiple constraints, a direct determination of, for example, $(\partial x_n / \partial x_i)_{K,x[i,n-(m-1),\dots,n]}$ may be cumbersome. But as before, information about such derivatives are contained within \mathbf{Q}_m , or more precisely by the eigenvectors of \mathbf{Q}_m .

Again, a set of vectors that reside within, or are tangent to, the constraint surfaces can be generated from the eigenvectors of \mathbf{Q}_m that have an eigenvalue equal to one, for which $\mathbf{Q}_m \mathbf{u} = \mathbf{u}$. Since these $n - m$ eigenvectors are already

in the \mathbf{K} -conserving direction, they are of course also orthogonal to all the gradient vectors of the various constraint surfaces. In other words, this eigenvalue equation is equivalent to the following set of orthogonality conditions, $\nabla \mathbf{g}_k^T \mathbf{u} = 0$ for $k = 1, \dots, m$, which proves more convenient for generating the various eigenvectors. (With $\mathbf{u}^T \mathbf{Q}_m = \mathbf{u}^T$, then $\mathbf{u}^T \mathbf{Q}_m \nabla \mathbf{g}_k = \mathbf{u}^T \nabla \mathbf{g}_k = 0$ as $\mathbf{Q}_m \nabla \mathbf{g}_k = 0$. Hence, $\mathbf{u}^T \nabla \mathbf{g}_k = \nabla \mathbf{g}_k^T \mathbf{u} = 0$. Or, $\nabla \mathbf{g}_k^T \mathbf{Q}_m \mathbf{u} = \nabla \mathbf{g}_k^T \mathbf{u}$. But $\nabla \mathbf{g}_k^T \mathbf{Q}_m = 0$, so $\nabla \mathbf{g}_k^T \mathbf{u} = 0$. Again, there is a connection between these eigenvectors and the implicit function theorem for these multiple constraints.) We therefore have m separate equations with n variables, for a total of $n - m$ degrees of freedom. For each of the possible $n - m$ eigenvectors, we therefore set the value of one component equal to 1 (the variable x_i whose derivative we are evaluating in Equation (50)), while setting equal to zero the values of $n - (m + 1)$ other components of the eigenvector (those held strictly fixed in Equation (50)). The orthogonality conditions are then solved to obtain the relative variations of the remaining m dependent variables (again all relative to changes in x_i).

If any one of these eigenvectors is denoted as \mathbf{u}_i , then as before

$$\left(\frac{\partial f}{\partial x_i} \right)_{\mathbf{K}, x[i, n-(m-1), \dots, n]} = \mathbf{u}_i^T \nabla f \quad (51)$$

Generating the $n - m$ eigenvectors in this way again yields a set of linearly independent vectors, and will therefore form a basis set to describe other variations that reside in all of the constraint surfaces. And as before, all of these $n - m$ relevant variations can be represented through the use of the matrix \mathbf{U} , which is now the $n \times (n - m)$ matrix whose columns are the linearly independent eigenvectors just obtained. Hence, as before,

$$\nabla_{\mathbf{K}, [i, n-(m-1), \dots, n]} f = \mathbf{U}^T \nabla f = \mathbf{U}^T \mathbf{Q}_m \nabla f = \mathbf{U}^T \nabla_{\mathbf{K}} f \quad (52)$$

Furthermore, explicitly constrained second-order derivatives of f can again be generated. If, for example, $x_{n-(m-1)}$ through x_n are the m “floating parameters”, then

$$\begin{aligned} & \left(\frac{\partial}{\partial x_i} \right)_{\mathbf{K}, x[i, n-(m-1), \dots, n]} \left(\frac{\partial f}{\partial x_j} \right)_{\mathbf{K}, x[j, n-(m-1), \dots, n]} \\ & = \mathbf{u}_i^T \left(\nabla \mathbf{u}_j^T \right) \nabla f + \mathbf{u}_i \mathbf{u}_j^T : \mathbf{F} = \mathbf{u}_i^T \left(\nabla \mathbf{u}_j^T \right) \nabla f + \mathbf{u}_i^T \mathbf{F} \mathbf{u}_j \end{aligned} \quad (53)$$

Again, these mixed second derivatives ($i \neq j$) are not necessarily equal, as in general $\nabla \mathbf{u}_i^T \neq \nabla \mathbf{u}_j^T$.

Finally, we introduce the following matrix

$$\mathbf{F}_{\mathbf{K}, n-m} = \mathbf{U}^T \nabla \left(\mathbf{U}^T \nabla f \right)^T = \mathbf{U}^T \nabla f^T (\nabla \mathbf{U}) + \mathbf{U}^T \mathbf{F} \mathbf{U} = \mathbf{U}^T \nabla f^T (\nabla \mathbf{U}) + \mathbf{U} \mathbf{U}^T : \mathbf{F} \quad (54)$$

which is formally identical to Equation (46). The eigenvalues of $\mathbf{F}_{\mathbf{K}, n-m}$ can again be used to determine the nature of the extrema, which given its smaller size may again be more convenient to evaluate.

As an example of the application of some of the above results, consider the

function $f = x_1^2 + x_2^2 + x_3^2$ subject to the two constraints of $g_1 = 2x_1 + x_2 = 1$ and $g_2 = x_1 + x_3 = 2$. From direct substitution of the two constraints into the function, $f = x_1^2 + (1 - 2x_1)^2 + (2 - x_1)^2 = 6x_1^2 - 8x_1 + 5$, and so $df/dx_1 = 12x_1 - 8$ and $d^2f/dx_1^2 = 12$. For this problem [1], a constrained extremum exists at $x_1 = 2/3$, where $df/dx_1 = 12x_1 - 8 = 0$, and which resides at a minimum.

Now, for this problem,

$$\begin{aligned} \mathbf{Q}_2 &= \frac{1}{6} \begin{vmatrix} 5 & 2 & \nabla g_1^T \\ 2 & 2 & \nabla g_2^T \\ \nabla g_1 & \nabla g_2 & \mathbf{I} \end{vmatrix} \\ &= \frac{1}{3} \nabla g_1 \nabla g_2^T - \frac{1}{3} \nabla g_1 \nabla g_1^T - \frac{5}{6} \nabla g_1 \nabla g_1^T + \frac{1}{3} \nabla g_1 \nabla g_1^T + \mathbf{I} \quad (55) \\ &= \frac{1}{6} \begin{pmatrix} 1 & -2 & -1 \\ -2 & 4 & 2 \\ -1 & 2 & 1 \end{pmatrix} \end{aligned}$$

The single eigenvector of \mathbf{Q}_2 with eigenvalue of 1 can also be obtained from the two conditions, $\nabla g_1 \mathbf{u} = 0$ and $\nabla g_2 \mathbf{u} = 0$, or $2u_1 + u_2 = 0$ and $u_1 + u_3 = 0$. If we are interested in obtaining the explicitly constrained derivative with respect to x_1 (i.e., $u_1 = 1$), the eigenvector satisfying these two conditions is $\mathbf{u}^T = (1, -2, -1)$. (Note that g_1 represents a plane parallel to the x_3 -direction with a normal vector $(2, 1, 0)$, while g_2 represents a plane parallel to the x_2 -direction with a normal vector $(1, 0, 1)$. The vector residing along the line resulting from the intersection of these two constraint planes must be orthogonal to both normal vectors, which in this case can also be obtained from their cross product, or $\nabla g_1 \times \nabla g_2 = (1, -2, -1)$, and is identical to \mathbf{u} .) Thus,

$$\left(\frac{\partial f}{\partial x_1}\right)_{\mathbf{K}, x[1,2,3]} = \mathbf{u}^T \nabla f = 2x_1 - 4x_2 - 2x_3 = 12x_1 - 8$$

(upon the subsequent use of the constraints). In addition,

$$\left(\frac{\partial^2 f}{\partial x_1^2}\right)_{\mathbf{K}, x[1,2,3]} = \mathbf{u}^T (\nabla \mathbf{u}^T) \nabla f + \mathbf{u} \mathbf{u}^T : \mathbf{F} = 12$$

(also noting that $\nabla \mathbf{u}^T = 0$), as found before. We can equivalently generate, for this problem, the now 1×1 matrix from Equation (54), $\mathbf{F}_{\mathbf{K}, n-2} = \mathbf{u}^T \mathbf{F} \mathbf{u} = 12$, again confirming that the extremum is a minimum.

Finally, again consider finding the extrema of $f = x_1^2 + x_2^2 + x_3^2$, but now subject to the two constraints of $g_1 = x_1 - x_2 + x_3 = 2$ and $g_2 = x_1^2 + x_2^2 = 4$. (For this problem, we are identifying those points along the intersection of a plane and a cylinder, or an ellipse oriented in the three-dimensional space, that are closest to and furthest from the origin.) As shown in Ref. [1], using Equations (7) and (12), there are four extrema located at $(0, -2, 0)$, $(2, 0, 0)$, $(\sqrt{2}, -\sqrt{2}, 2 - 2\sqrt{2})$, and $(-\sqrt{2}, \sqrt{2}, 2 + 2\sqrt{2})$, for which $(\pm\sqrt{2}, \mp\sqrt{2}, 2 \mp 2\sqrt{2})$ correspond to the two furthest points from the origin (maxima), while the other two extremal points are the two closest points to the origin (minima).

The single eigenvector of \mathbf{Q}_2 with eigenvalue of 1 can be obtained from the two conditions, $\nabla g_1 \mathbf{u} = 0$ and $\nabla g_2 \mathbf{u} = 0$, or $u_1 - u_2 + u_3 = 0$ and $x_1 u_1 + x_2 u_2 = 0$.

If we are interested in obtaining the explicitly constrained derivative with respect to x_1 (*i.e.*, $u_1 = 1$), the eigenvector satisfying these two conditions is

$$\mathbf{u}^T = \left(1, -\frac{x_1}{x_2}, -\frac{x_1 + x_2}{x_2} \right).$$

(Again note that the vector residing in the direction along the intersection of these two constraint surfaces is proportional to $\nabla g_1 \times \nabla g_2 = (-2x_2, 2x_1, 2x_1 + 2x_2)$, and is identical to $-2x_2 \mathbf{u}$.) Thus,

$$\left(\frac{\partial f}{\partial x_1} \right)_{\mathbf{K}, x[1,2,3]} = \mathbf{u}^T \nabla f = -2x_3 \frac{(x_1 + x_2)}{x_2} \quad (56)$$

which also follows from the direct substitution of the constraints into f . At an extremum, this derivative is equal to zero, which can be satisfied for either $x_3 = 0$ or $x_1 + x_2 = 0$, and when $x_2 \neq 0$. For $x_1 + x_2 = 0$, or $x_1 = -x_2$, the second constraint indicates that $x_1 = \pm\sqrt{2}$, and so $x_2 = \mp\sqrt{2}$, which then requires that $x_3 = 2\mp 2\sqrt{2}$ upon using the first constraint. For $x_3 = 0$, the first constraint reduces to $x_1 - x_2 = 2$, which when substituted into the second constraint leads to the following two possible cases of $x_1 = 0$, $x_2 = -2$ and $x_1 = 2$, $x_2 = 0$. While the latter case is one of the extrema identified in Ref. [1], this result must nevertheless be ignored. Hence, for this problem, Equation (56) can be used to identify three of the previously four determined extremal points. In addition,

$$\left(\frac{\partial^2 f}{\partial x_1^2} \right)_{\mathbf{K}, x[1,2,3]} = \mathbf{u}^T (\nabla \mathbf{u}^T) \nabla f + \mathbf{u} \mathbf{u}^T : \mathbf{F} = \frac{2[x_2(x_1 + x_2)^2 - x_3(x_1^2 + x_2^2)]}{x_2^3} \quad (57)$$

At $(0, -2, 0)$, $(\partial^2 f / \partial x_1^2)_{\mathbf{K}, x[1,2,3]} > 0$, which corresponds to a local minimum. At $(\sqrt{2}, -\sqrt{2}, 2 - 2\sqrt{2})$ and $(-\sqrt{2}, \sqrt{2}, 2 + 2\sqrt{2})$, both result in $(\partial^2 f / \partial x_1^2)_{\mathbf{K}, x[1,2,3]} < 0$, indicating these are local maxima (all results that were determined in Ref. [1]).

(Note that the method in Ref. [1] makes use of a vector in the direction along the intersection of the two constraints that always has a finite magnitude. Hence, such a vector remains well-behaved all along the ellipse, even at $x_1 = 0$ or $x_2 = 0$, so that Equations (7) and (12) are themselves able to identify all four extremal points. While the eigenvector \mathbf{u} also points in the same direction along the intersection of these two constraints, it is not necessarily required to have a finite magnitude. Since these explicitly constrained derivatives consider how the dependent, or “floating”, variables change due to a “unit variation” of the chosen independent variable, \mathbf{u} is instead required to always have a finite value (equal to 1) for one of its components. In order to keep \mathbf{u} pointing in the correct direction around the ellipse, some of the other components of \mathbf{u} may therefore diverge, as occurs for example when $x_2 = 0$. Thus, these explicitly constrained derivatives may not, in general, be able to identify all relevant extremal points.

To identify the other extremal point, we evaluate the explicitly constrained derivatives with respect to x_2 . With $u_2 = 1$, the appropriate eigenvector is now

given by $\mathbf{u}^T = \left(-\frac{x_2}{x_1}, 1, \frac{x_1 + x_2}{x_1} \right)$, which leads to

$$\left(\frac{\partial f}{\partial x_2} \right)_{\mathbf{K}, x[1,2,3]} = \mathbf{u}^T \nabla f = 2x_3 \frac{(x_1 + x_2)}{x_1} \quad (58)$$

Equation (58) again identifies the two extremal points corresponding to $x_1 + x_2 = 0$. But now with $x_1 \neq 0$, the case of $x_3 = 0$ can be used to identify the fourth extremal point at $x_1 = 2$, $x_2 = 0$. Moreover,

$$\left(\frac{\partial^2 f}{\partial x_2^2} \right)_{\mathbf{K}, x[1,2,3]} = \frac{2 \left[x_1 (x_1 + x_2)^2 + x_3 (x_1^2 + x_2^2) \right]}{x_1^3} \quad (59)$$

which indicates that $(2, 0, 0)$ corresponds to a local minimum. These constrained derivatives allow for the straightforward determination of both the extrema and their nature, all of which were obtained without any knowledge of the Lagrange multipliers.

Finally, for completeness, we also consider the explicitly constrained derivatives with respect to x_3 . Now with $\mathbf{u}^T = \left(-\frac{x_2}{x_1 + x_2}, \frac{x_1}{x_1 + x_2}, 1 \right)$, then $(\partial f / \partial x_3)_{\mathbf{K}, x[1,2,3]} = 2x_3$ and $(\partial^2 f / \partial x_3^2)_{\mathbf{K}, x[1,2,3]} = 2$. (Both of these results directly follow from the substitution of the second constraint into the function, or $f = x_1^2 + x_2^2 + x_3^2 = 4 + x_3^2$.) These constrained derivatives only provide information about two of the four extremal points, in which $x_3 = 0$.

5. Conclusions

Through the use of linear algebra, and in particular the Gram-Schmidt orthogonalization process, the local extrema of functions for which its independent variables are also subject to constraints were determined without the use of Lagrange multipliers. This approach was extended to the continuum limit, thereby obtaining the extrema of functionals subject to one or more constraints. The direct determination of the explicitly constrained first- and second-order derivatives was also considered, and various additional expressions were obtained for generating information about the nature of the extrema (again without the need to evaluate any Lagrange multipliers).

As noted previously, avoiding the use of Lagrange multipliers may simplify the analysis of various problems of interest. As our method can also be used to obtain immediate expressions for the Lagrange multipliers (in terms of the gradients of the function and constraints), the new relations provided here may yield some new and insightful relations into several physical problems of longstanding interest, which also require the imposition of various constraints. Such claims remain to be tested, and future work should be directed towards the application of the results provided here to standard problems of interest (e.g., the use of the Gibbs entropy formula along with appropriate constraints to generate the probability distributions for the various ensembles of statistical me-

chanics), as well as some new ones.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Corti, D.S. and Fariello, R. (2021) Avoiding the Use of Lagrange Multipliers: I—Evaluating the Constrained Extrema of Functions with Projection Matrices. *Operations Research Forum*, **2**, Article No. 59. <https://doi.org/10.1007/s43069-021-00100-0>
- [2] Zizza, F. (1998) Differential Forms for Constrained Max-Min Problems: Eliminating Lagrange Multipliers. *The College Mathematics Journal*, **29**, 387-396. <https://doi.org/10.1080/07468342.1998.11973974>
- [3] Gigena, S. (2013) Constrained Local Extrema without Lagrange Multipliers and the Higher Derivative Test.
- [4] Spring, D. (1985) On the Second Derivative Test for Constrained Local Extrema. *The American Mathematical Monthly*, **92**, 631-643. <https://doi.org/10.1080/00029890.1985.11971702>
- [5] Nerenberg, M.A.H. (1991) The Second Derivative Test for Constrained Extremum Problems. *International Journal of Mathematical Education in Science and Technology*, **22**, 303-308. <https://doi.org/10.1080/0020739910220215>
- [6] Gál, T. (2001) Differentiation of Density Functionals That Conserves the Normalization of the Density. *Physical Review A*, **63**, Article 022506. <https://doi.org/10.1103/physreva.63.022506>
- [7] Gál, T. (2002) Functional Differentiation under Conservation Constraints. *Journal of Physics A: Mathematical and General*, **35**, 5899-5905. <https://doi.org/10.1088/0305-4470/35/28/309>
- [8] Gál, T. (2007) Functional Differentiation under Simultaneous Conservation Constraints. *Journal of Physics A: Mathematical and Theoretical*, **40**, 2045-2052. <https://doi.org/10.1088/1751-8113/40/9/010>
- [9] Gál, T. (2007) Differentiation of Functionals with Variables Coupled by Conservation Constraints: Analysis through a Fluid-Dynamical Model. *Journal of Mathematical Physics*, **48**, Article 053520. <https://doi.org/10.1063/1.2737265>
- [10] Gál, T. (2010) Stability of Equilibrium under Constraints: Role of Second-Order Constrained Derivatives. *Journal of Physics A: Mathematical and Theoretical*, **43**, Article 425208. <https://doi.org/10.1088/1751-8113/43/42/425208>
- [11] Gál, T. (2012) On Constrained Second Derivatives.
- [12] Chandler, D. (1987) Introduction to Modern Statistical Mechanics. Oxford University Press.
- [13] Baxley, J.V. and Moorhouse, J.C. (1984) Lagrange Multiplier Problems in Economics. *The American Mathematical Monthly*, **91**, 404-412. <https://doi.org/10.1080/00029890.1984.11971446>
- [14] Zheng, Y. and Meng, Z. (2017) A New Augmented Lagrangian Objective Penalty Function for Constrained Optimization Problems. *Open Journal of Optimization*, **6**, 39-46. <https://doi.org/10.4236/ojop.2017.62004>
- [15] Cacuci, D.G. (2023) Fourth-Order Predictive Modelling: I. General-Purpose

Closed-Form Fourth-Order Moments-Constrained Maxent Distribution. *American Journal of Computational Mathematics*, **13**, 413-438.

<https://doi.org/10.4236/ajcm.2023.134024>

- [16] Jarab'ah, O.A. (2023) Lagrangian Formulation of Fractional Nonholonomic Constrained Damping Systems. *Advances in Pure Mathematics*, **13**, 552-558.
<https://doi.org/10.4236/apm.2023.139037>
- [17] Davis, H.T. (1996) *Statistical Mechanics of Phases, Interfaces, and Thin Films*. Wiley-VCH.
- [18] Kreyszig, E. (1991) *Differential Geometry*. Dover Publications.
- [19] Boas, M.L. (1983) *Mathematical Methods in the Physical Sciences*. 2nd Edition, Wiley.

Appendix. Derivation of the Functional Derivatives by Analogy

Starting with

$$\tilde{\mathbf{Q}}_m = \begin{pmatrix} \nabla \mathbf{g}_1^T \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_1^T \nabla \mathbf{g}_m & \nabla \mathbf{g}_1^T \\ \vdots & \ddots & \vdots & \vdots \\ \nabla \mathbf{g}_m^T \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_m^T \nabla \mathbf{g}_m & \nabla \mathbf{g}_m^T \\ \nabla \mathbf{g}_1 & \cdots & \nabla \mathbf{g}_m & \mathbf{I} \end{pmatrix} \quad (60)$$

and

$$\mathbf{F}_K = \frac{1}{D_m^2} \tilde{\mathbf{Q}}_m \nabla (\tilde{\mathbf{Q}}_m \nabla f)^T \quad (61)$$

which is only appropriate at the constrained extrema, we note for the case of a single constraint that

$$\tilde{\mathbf{Q}}_1 = D_1 \mathbf{I} - \nabla \mathbf{g} \nabla \mathbf{g}^T \quad (62)$$

where $D_1 = \nabla \mathbf{g}^T \nabla \mathbf{g}$. Thus, $(\tilde{\mathbf{Q}}_1 \nabla f)^T = D_1 \nabla f^T - (\nabla \mathbf{g}^T \nabla f) \nabla \mathbf{g}^T$. Hence,

$$\nabla (\tilde{\mathbf{Q}}_1 \nabla f)^T = D_1 \nabla \nabla f^T + \nabla D_1 \nabla f^T - (\nabla \mathbf{g}^T \nabla f) \nabla \nabla \mathbf{g}^T - \nabla (\nabla \mathbf{g}^T \nabla f) \nabla \mathbf{g}^T \quad (63)$$

With $\mathbf{F} = \nabla \nabla f^T$ and now defining $\mathbf{G} = \nabla \nabla \mathbf{g}^T$, and also noting that $\nabla D_1 = 2\mathbf{G} \nabla \mathbf{g}$ and $\nabla (\nabla \mathbf{g}^T \nabla f) = \mathbf{G} \nabla f + \mathbf{F} \nabla \mathbf{g}$, Equation (63) can also be expressed as

$$\nabla (\tilde{\mathbf{Q}}_1 \nabla f)^T = D_1 \mathbf{F} - \mathbf{F} \nabla \mathbf{g} \nabla \mathbf{g}^T + 2\mathbf{G} \nabla \mathbf{g} \nabla f^T - (\nabla \mathbf{g}^T \nabla f) \mathbf{G} - \mathbf{G} \nabla f \nabla \mathbf{g}^T \quad (64)$$

So with $D_1^2 \mathbf{F}_K = \tilde{\mathbf{Q}}_1 \nabla (\tilde{\mathbf{Q}}_1 \nabla f)^T$, then

$$\begin{aligned} D_1^2 \mathbf{F}_K &= D_1^2 \mathbf{F} - D_1 \mathbf{F} \nabla \mathbf{g} \nabla \mathbf{g}^T - D_1 \nabla \mathbf{g} \nabla \mathbf{g}^T \mathbf{F} + \nabla \mathbf{g} \nabla \mathbf{g}^T \mathbf{F} \nabla \mathbf{g} \nabla \mathbf{g}^T \\ &\quad + 2D_1 \mathbf{G} \nabla \mathbf{g} \nabla f^T - D_1 (\nabla \mathbf{g}^T \nabla f) \mathbf{G} - D_1 \mathbf{G} \nabla f \nabla \mathbf{g}^T \\ &\quad - 2\nabla \mathbf{g} \nabla \mathbf{g}^T \mathbf{G} \nabla \mathbf{g} \nabla f^T + \nabla \mathbf{g} \nabla \mathbf{g}^T \mathbf{G} \nabla f \nabla \mathbf{g}^T \end{aligned} \quad (65)$$

Since, for example, $\nabla \mathbf{g}^T \mathbf{F} \nabla \mathbf{g}$ is a scalar, the above can also be rewritten as

$$\begin{aligned} D_1^2 \mathbf{F}_K &= D_1^2 \mathbf{F} - D_1 \mathbf{F} \nabla \mathbf{g} \nabla \mathbf{g}^T - D_1 \nabla \mathbf{g} \nabla \mathbf{g}^T \mathbf{F} + (\nabla \mathbf{g}^T \mathbf{F} \nabla \mathbf{g}) \nabla \mathbf{g} \nabla \mathbf{g}^T \\ &\quad + 2D_1 \mathbf{G} \nabla \mathbf{g} \nabla f^T - D_1 (\nabla \mathbf{g}^T \nabla f) \mathbf{G} - D_1 \mathbf{G} \nabla f \nabla \mathbf{g}^T \\ &\quad - 2(\nabla \mathbf{g}^T \mathbf{G} \nabla \mathbf{g}) \nabla \mathbf{g} \nabla f^T + (\nabla \mathbf{g}^T \mathbf{G} \nabla f) \nabla \mathbf{g} \nabla \mathbf{g}^T \end{aligned} \quad (66)$$

Finally, for the case in which $\mathbf{G} = 0$, Equation (66) reduces to

$$D_1^2 \mathbf{F}_K = D_1^2 \mathbf{F} - D_1 \mathbf{F} \nabla \mathbf{g} \nabla \mathbf{g}^T - D_1 \nabla \mathbf{g} \nabla \mathbf{g}^T \mathbf{F} + (\nabla \mathbf{g}^T \mathbf{F} \nabla \mathbf{g}) \nabla \mathbf{g} \nabla \mathbf{g}^T \quad (67)$$

Now, the matrix \mathbf{F} of discrete (unconstrained) second derivatives becomes the following in the continuum or functional limit

$$\mathbf{F} \equiv \frac{\partial^2 f}{\partial x_i \partial x_j} \rightarrow \frac{\delta^2 f}{\delta y(\mathbf{x}) \delta y(\mathbf{x}')} \quad (68)$$

(where i denotes the row and j the column of \mathbf{F} for the discrete case). Thus, within Equation (67),

$$\mathbf{F}\nabla\mathbf{g} = \sum_j F_{ij} \frac{\partial\mathbf{g}}{\partial x_j} \rightarrow \int d\mathbf{x}' \frac{\delta^2 f}{\delta y(\mathbf{x})\delta y(\mathbf{x}')\delta y(\mathbf{x}')} \frac{\delta\mathbf{g}}{\delta y(\mathbf{x}')} \quad (69)$$

$$\nabla\mathbf{g}^T \mathbf{F} = \sum_i \frac{\partial\mathbf{g}}{\partial x_i} F_{ij} \rightarrow \int d\mathbf{x} \frac{\delta\mathbf{g}}{\delta y(\mathbf{x})\delta y(\mathbf{x}')\delta y(\mathbf{x}')} \frac{\delta^2 f}{\delta y(\mathbf{x})\delta y(\mathbf{x}')\delta y(\mathbf{x}')} \quad (70)$$

$$\nabla\mathbf{g}^T \mathbf{F}\nabla\mathbf{g} = \sum_i \sum_j \frac{\partial\mathbf{g}}{\partial x_i} F_{ij} \frac{\partial\mathbf{g}}{\partial x_j} \rightarrow \iint d\mathbf{x}d\mathbf{x}' \frac{\delta\mathbf{g}}{\delta y(\mathbf{x})\delta y(\mathbf{x}')\delta y(\mathbf{x}')} \frac{\delta^2 f}{\delta y(\mathbf{x})\delta y(\mathbf{x}')\delta y(\mathbf{x}')} \frac{\delta\mathbf{g}}{\delta y(\mathbf{x}')} \quad (71)$$

The other terms appearing in Equation (66) can be evaluated similarly as in the above.