Constrained Optimizations: Lagrange Multipliers vs Structured Differentiation

P. Reany

August 20, 2020

Abstract

Lagrange multipliers are examined from the dual perspective of the traditional way to present them vs the structured differentiation way. A little geometric algebra is used conveniently when the number of variants is greater than three. One of the more practical subjects dealt with here is maximum entropy (and the partition function).

1 Introduction

The traditional way to look at Lagrange multipliers is geometrical: it seeks to find a set of constraints points on which some given function is maximized or minimized on that base set of points (i.e., the critical points). If these conditions are consistent, then a solution(s) will be found when the function is varying the most (or the least) when its gradient is parallel to the gradient of the constraint function.

Note: This paper will avoid subtleties involved in finding maxima and/or minima of the problems listed. In particular, emphasis is placed on finding the critical point/s, not on the actual maxima or minima points.

Typically, the function to be optimized is given as $f(x, y, z)$ and is constrained to lie on or ‘over’ some constraint set usually defined implicitly as a locus of points, given by

$$g(x, y, z) = 0 \quad \text{or} \quad g(x, y, z) = c\,, \tag{1}$$

where c is a constant. Functions at a point vary the most in the direction of its gradient.

So, we have that

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z)\,, \tag{2}$$

where λ is just a number, which, except for rare occasions, we’re interested in only so far as it helps us determine the values of x, y, z that are the coordinates of the critical points.

However, more frequently, I shall use the equivalent forms

$$\nabla[f + \lambda g] = \mathbf{0}\,, \tag{3a}$$

$$\nabla f + \lambda \nabla g = \mathbf{0}\,. \tag{3b}$$

which are vector equations, and therefore each of the three components must be zero. Writing (3a) as components, we get

$$\partial_x[f + \lambda g] = 0 \quad \partial_y[f + \lambda g] = 0, \quad \partial_z[f + \lambda g]\,, \tag{4}$$

where the partial derivative ∂ is always an explicit derivative.

However, we shall see an alternative method to solve for these critical point by using what I call *Structured Differentiation* or SD. The method is simple. We begin by using the Implicit Function Theorem (IFT) to solve for one of variants of $g(x, y, z)$ in terms of the others. For example, if $\partial g/\partial z \neq 0$, then we can write

$$z = z(x, y)\,. \tag{5}$$

Since f is the function on which we will find critical points, we will take $\delta f/\delta x = 0$ and $\delta f/\delta y = 0$, where, by abuse of notation, we write,

$$\delta f(x, y, z(x, y))/\delta x = 0\,, \quad \delta f(x, y, z(x, y))/\delta y = 0\,. \tag{6}$$

Expanding these last equations, we get

$$\frac{\partial f(x, y, z(x, y))}{\partial x} + \frac{\partial f(x, y, z(x, y))}{\partial z}\frac{\delta z}{\delta x} = 0\,, \tag{7a}$$

$$\frac{\partial f(x, y, z(x, y))}{\partial y} + \frac{\partial f(x, y, z(x, y))}{\partial z}\frac{\delta z}{\delta y} = 0\,. \tag{7b}$$

where the deltal derivative δ is always a total derivative.

Now, since x and y are treated as independent variables, then

$$\frac{\delta z}{\delta x} = \frac{\partial z}{\partial x}\,, \tag{8a}$$

$$\frac{\delta z}{\delta y} = \frac{\partial z}{\partial y}\,. \tag{8b}$$

Therefore, with the above substitutions and some simplifications, we arrive at

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 0\,, \tag{9a}$$

$$\frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial y} = 0\,. \tag{9b}$$

Now, we don't want to deal with $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$, so, we can get rid of them by differentiating (1) through by, respectively, the total derivative by x and then the total derivative by y, to get

$$\frac{\partial g}{\partial x}+\frac{\partial g}{\partial z}\frac{\partial z}{\partial x}=0\,, \tag{10a}$$

$$\frac{\partial g}{\partial y}+\frac{\partial g}{\partial z}\frac{\partial z}{\partial y}=0\,. \tag{10b}$$

Therefore,

$$\frac{\partial z}{\partial x}=-\frac{\partial g/\partial x}{\partial g/\partial z}\,, \tag{11a}$$

$$\frac{\partial z}{\partial y}=-\frac{\partial g/\partial y}{\partial g/\partial z}\,, \tag{11b}$$

which explains why we can't have $\partial g/\partial z=0$.

On substituting the above into (9a, 9b), we get

$$\frac{\partial f}{\partial x}+\frac{\partial f}{\partial z}\left[-\frac{\partial g/\partial x}{\partial g/\partial z}\right]=0\,, \tag{12a}$$

$$\frac{\partial f}{\partial y}+\frac{\partial f}{\partial z}\left[-\frac{\partial g/\partial y}{\partial g/\partial z}\right]=0\,. \tag{12b}$$

Defining

$$\lambda\equiv-\frac{\partial f/\partial z}{\partial g/\partial z}\,, \tag{13}$$

then the previous pair of equations become

$$\frac{\partial f}{\partial x}+\lambda\frac{\partial g}{\partial x}=0\,, \tag{14a}$$

$$\frac{\partial f}{\partial y}+\lambda\frac{\partial g}{\partial y}=0\,. \tag{14b}$$

By rewriting (13), we can add to this

$$\frac{\partial f}{\partial z}+\lambda\frac{\partial g}{\partial z}=0\,. \tag{15}$$

And from these last three equations we get

$$\frac{\partial f}{\partial \mathbf{x}}+\lambda\frac{\partial g}{\partial \mathbf{x}}=0\,. \tag{16}$$

Or, put in usual Gibbs's vector form

$$\nabla F+\lambda\nabla G=\mathbf{0}\,, \tag{17}$$

which is Eq. (3b).

On taking the cross product of this last equation by ∇G, we get

$$\nabla F\times\nabla G=\mathbf{0}\,, \tag{18}$$

with components,

$$f_y g_z - g_y f_z = 0\,, \tag{19a}$$
$$f_z g_x - g_z f_x = 0\,, \tag{19b}$$
$$f_x g_y - g_x f_y = 0\,. \tag{19c}$$

This following part is a recap of the previous development from a more typical SD presentation.

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

$$G(x, y, z) = 0\,, \tag{20}$$

where $\partial G/\partial z \neq 0$. (If you're interested, compare the SD approach to that of Taylor-Mann ([1]) (192–198).[1]

Since $\partial G/\partial z \neq 0$ and G is constrained according to (20), then by the *Implicit Function Theorem* we can write

$$z = z(x, y) = z(\boldsymbol{\eta})\,, \tag{21}$$

where z is in primitive form (i.e., the variables on which z is explicitly dependent are mutually independent of each other). Then

$$F(x, y, z) = F(\boldsymbol{\eta}, z(\boldsymbol{\eta})) = F(\mathbf{x})\,. \tag{22}$$

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

$$\delta F/\delta\boldsymbol{\eta} = \mathbf{0}\,, \tag{23}$$

which is a direct generalization from ordinary calculus. Therefore, applying the chain rule to expand (23), we get

$$\frac{\partial F}{\partial \boldsymbol{\eta}} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial \boldsymbol{\eta}} = \mathbf{0}\,, \tag{24}$$

where we used the fact that when we expanded $\delta F/\delta\boldsymbol{\eta}$ by the chain rule, we can relace $\delta z/\delta\boldsymbol{\eta}$ by $\partial z/\partial\boldsymbol{\eta}$ since z is in primitive form. Now, by differentiating (20) through by $\boldsymbol{\eta}$, we also have that

$$\frac{\delta G}{\delta \boldsymbol{\eta}} = \frac{\partial G}{\partial \boldsymbol{\eta}} + \frac{\partial G}{\partial z}\frac{\partial z}{\partial \boldsymbol{\eta}} = \mathbf{0}\,. \tag{25}$$

---

[1] It's important to understand that in this type of problem, the purpose of the relation $G(x, y, z) = 0$ 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.

Solving this system for $\partial z/\partial\boldsymbol{\eta}$ and substituting that result into (24) we obtain

$$\frac{\partial F}{\partial\boldsymbol{\eta}}+\lambda\frac{\partial G}{\partial\boldsymbol{\eta}}=\mathbf{0} \tag{26}$$

where

$$\lambda=-\frac{\partial F/\partial z}{\partial G/\partial z} \tag{27}$$

On rewriting this we have

$$\frac{\partial F}{\partial z}+\lambda\frac{\partial G}{\partial z}=0\,. \tag{28}$$

Combining this and (26) results in

$$\frac{\partial F}{\partial\mathbf{x}}+\lambda\frac{\partial G}{\partial\mathbf{x}}=\mathbf{0}\,. \tag{29}$$

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

$$\nabla F+\lambda\nabla G=\mathbf{0}\,. \tag{30}$$

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

So, if (20) and (30) 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 (20), there exists some nonzero scalar c such that the level surface defined by

$$F(x,y,z)=c \tag{31}$$

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

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

$$H(x,y,z)=F(x,y,z)+\lambda G(x,y,z) \tag{32}$$

such that the possible relative extrema of $F(x,y,z)$ subject to (20) can be found by solving the system of (20) and

$$\nabla H=\mathbf{0}\,. \tag{33}$$

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

$$J(x,y,z,\lambda)=F(x,y,z)+\lambda G(x,y,z) \tag{34}$$

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

$$\Box J=\mathbf{0}\,, \tag{35}$$

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

$$\Box \equiv (\partial_x, \partial_y, \partial_z, \partial_\lambda)\,. \tag{36}$$

Now, the first three components of (35) reproduce (30) and the fourth component reproduces the original constraint (20), 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 (35) 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?

My apologies to authors: I'm presenting these problems from my handwritten notes from a variety of sources from years ago, and I don't remember where they came from. Sorry. If anyone knows the sources of any of these individual problems, please let me know and I will gladly give the authors the credit for them. Thanks.

By the way, I have typically used three components for a generic representation of how these problems are handled, but any finite number of components from two or more will work. Now, as a general rule, the more components involved, the more algebraically involved the solutions will be, and we will usually resort to matrix solutions or geometric algebra.

2 Problem 1

Find the critical point of $f(x, y, z) = x^3 + y^3 + z^3$, subject to the constraint

$$g(x, y, z) = x^2 + y^2 + z^2 - 1 = 0\,. \tag{37}$$

Solution 1: Employing (4), we get

$$3x^2 + \lambda(2x) = 0, \quad 3y^2 + \lambda(2y) = 0, \quad 3z^2 + \lambda(2z) = 0\,. \tag{38}$$

Solving this last equation for x, y, z, we get

$$x = y = z = -2\lambda/3\,. \tag{39}$$

On substituting these values into (37), we get

$$\lambda = \pm\sqrt{3}/2\,. \tag{40}$$

Therefore,

$$x = y = z = \pm 1/\sqrt{3}\,, \tag{41}$$

from which we get the eight critical points:

$$(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}),\ (-1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}),\ (1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3}),\ \text{etc.}$$

Solution 2:

Using the Implicit Function Theorem, we write

$$z = z(x, y)\,. \tag{42}$$

Then we differentiate $f(x, y, z)$ by x and y to get the critical points:

$$\frac{\delta f}{\delta x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 3x^2 + 3z^2\frac{\partial z}{\partial x} = 0\,, \tag{43a}$$

$$\frac{\delta f}{\delta y} = \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial y} = 3y^2 + 3z^2\frac{\partial z}{\partial y} = 0\,. \tag{43b}$$

Since we neither want nor need $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$, we can eliminate these by simply taking the total derivatives of (37) by x and y, respectively, to get

$$\frac{\delta g}{\delta x} = \frac{\partial g}{\partial x} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial x} = 2x + 2z\frac{\partial z}{\partial x} = 0\,, \tag{44a}$$

$$\frac{\delta g}{\delta y} = \frac{\partial g}{\partial y} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial y} = 2y + 2z\frac{\partial z}{\partial y} = 0\,. \tag{44b}$$

On eliminating $\frac{\partial z}{\partial x}$ between (43a) and (44a), to get

$$x = z\,. \tag{45}$$

And, on eliminating $\frac{\partial z}{\partial y}$ between (43b) and (44b), to get

$$y = z\,. \tag{46}$$

Putting these last two equations together, we have that $x = y = z$. And, using this result in (37), we have that

$$x = y = z = \pm 1/\sqrt{3}\,, \tag{47}$$

from which we get the same critical points as before.

3 Problem 2

Find the critical point of $f(x, y, z) = x^3 + y^3 + z^3$, subject to the constraints

$$g(x, y, z) = x^2 + y^2 + z^2 - 1 = 0\,, \tag{48}$$

and

$$h(x, y, z) = x + y + z = 0\,. \tag{49}$$

Solution 1: By the method of Lagrange multipliers, we get[2]

$$\nabla(f + \lambda g + \mu h) = \mathbf{0}\,, \tag{50}$$

which, when broken down into components, gives us

$$\partial_x(f + \lambda g + \mu h) = 0\,, \tag{51a}$$
$$\partial_y(f + \lambda g + \mu h) = 0\,, \tag{51b}$$
$$\partial_z(f + \lambda g + \mu h) = 0\,. \tag{51c}$$

Expanding these, we get

$$3x^2 + \lambda(2x) + \mu(1) = 0\,, \tag{52a}$$
$$3y^2 + \lambda(2y) + \mu(1) = 0\,, \tag{52b}$$
$$3z^2 + \lambda(2z) + \mu(1) = 0\,. \tag{52c}$$

This homogeneous system can be put into matrix form, to get

$$\begin{bmatrix} 3x^2 & 2x & 1 \\ 3y^2 & 2y & 1 \\ 3z^2 & 2z & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\,. \tag{53}$$

From matrix theory, we know that the determinant of the coefficient matrix must be zero. So,

$$\begin{vmatrix} 3x^2 & 2x & 1 \\ 3y^2 & 2y & 1 \\ 3z^2 & 2z & 1 \end{vmatrix} = 0\,. \tag{54}$$

Therefore, to solve for the critical points, we need to use this last equation together with equations (48) and (49) to derived a simultaneous solution set.

Solution 2: For the SD solution, we begin wiht the fact that with two contraint eqautions, we end up with only one independent variable when using the Implicit Function Theorem (IFT). We choose that variable to be x (why not?). Therefore, $y = y(x)$ and $z = z(x)$ to find the critical points of $f(x, y, z)$, we differentiate by x and set that result to zero.

$$\frac{\delta f}{\delta x} = \frac{d}{dx}(x^3 + y^3 + z^3) = 3x^2 + 3y^2 y' + 3z^2 z' = 0\,, \tag{55a}$$

where the primes mean differentiation by x. But we can also differentiate across (48) and (49) by x, to get

$$2x + 2yy' + 2zz' = 0\,, \tag{55b}$$
$$1 + y' + z' = 0\,. \tag{55c}$$

---

[2] Both λ and μ are Lagrange multipliers.

This homogeneous system can be put into matrix form, to get

$$\begin{bmatrix} 3x^2 & 2y^2 & 3z^2 \\ 2x & 2y & 2z \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} . \tag{56}$$

As before, the determinant of the coefficient matrix of zero, therefore,

$$y^2 z - yz^2 + (y - z)x^2 + (z^2 - y^2) = 0 . \tag{57}$$

On eliminating x between (49) and (57), we get

$$2y^2 + 5yz + 2z^2 = 0 . \tag{58}$$

On eliminating x between (48) and (49), we get

$$2y^2 + 2yz + 2z^2 = 1 . \tag{59}$$

From these last two equations, we get that $y = -1/3z$. Substituting this result into (58), we get

$$\frac{2}{9z^2} - \frac{5}{3} + 2z^2 = 0 . \tag{60}$$

The solutions for z^2 are 2/3 and 1/6. It should be easy now to solve for x and y values of the critical points.

Solution 3: Unless you actually want to know the value/s of the Lagrange multiplier/s, it's probably better not to solve for them at all. First, because to solve for them will likely make the solution longer, and, second, because solving for them will likely add into the solution additional round-off error.

Let's return to Eq. (50), and rewrite it as

$$\nabla f + \lambda \nabla g + \mu \nabla h = \mathbf{0} . \tag{61}$$

Next, we multiply through by the cross product ∇g, to get

$$\nabla f \times \nabla g + \mu \nabla h \times \nabla g = \mathbf{0} , \tag{62}$$

where we used that ∇g is just a vector and crossing any vector with itself is identically zero. Now, we use the vector identity that $\mathbf{a} \times \mathbf{b} \cdot \mathbf{a} = 0$ for arbitrary vectors $\mathbf{a}$ and $\mathbf{b}$, and then dot the last equation through by ∇h, to get

$$\nabla f \times \nabla g \cdot \nabla h = 0 , \tag{63}$$

When you work this out as a determinant, you get

$$\begin{vmatrix} f_x & f_y & f_z \\ g_x & g_y & g_z \\ h_x & h_y & h_z \end{vmatrix} = 0 . \tag{64}$$

which gives you the same equation as before in (56) by taking the determinant of the coefficient matrix.

But what is Eq. (64) in terms of our goal of finding solutions? It is the third algebraic equation that couples to the two constraint equations, providing for us a system of three simultaneous equations that can be solved for, in principle, to gives of values of x, y, z for critical points.

4 Problem 3

This problem is a simple extension of where we started from.

Given the function $f(x, y, z, w)$ to be optimized, with constraint function

$$g(x, y, z, w) = 0\,. \tag{65}$$

Show that, by use of SD, we can arrive at the Lagrange multiplier form

$$\Box f(x, y, z, w) = \lambda \Box\, g(x, y, z, w)\,, \tag{66}$$

where λ is just a number, and $\Box = (\partial_x, \partial_y, \partial_z, \partial_w)$

Proof: Using the IFT on (65) and that $\partial g/\partial w \neq 0$, we can write

$$w = w(x, y, z)\,. \tag{67}$$

For critical points, we need to set

$$\frac{\delta f}{\delta x} = \frac{\delta f}{\delta y} = \frac{\delta f}{\delta z} = 0\,. \tag{68}$$

From this we have that

$$f_x + f_w \frac{\delta w}{\delta x} = 0\,, \tag{69a}$$

$$f_y + f_w \frac{\delta w}{\delta y} = 0\,, \tag{69b}$$

$$f_z + f_w \frac{\delta w}{\delta z} = 0\,. \tag{69c}$$

Taking like derivatives of (65), we get

$$g_x + g_w \frac{\delta w}{\delta x} = 0\,, \tag{70a}$$

$$g_y + g_w \frac{\delta w}{\delta y} = 0\,, \tag{70b}$$

$$g_z + g_w \frac{\delta w}{\delta z} = 0\,. \tag{70c}$$

On multiplying (69a) by g_w and (70a) by f_w, we get

$$f_x g_w + f_w g_w \frac{\delta w}{\delta x} = 0\,, \quad f_x g_w + f_w g_w \frac{\delta w}{\delta x} = 0\,. \tag{71}$$

Now, eliminating the total derivatives between these equations, we are left with

$$f_x g_w = g_x f_w\,. \tag{72a}$$

Pairing up the other two equations, we end up with

$$f_y g_w = g_y f_w \,, \tag{72b}$$

$$f_z g_w = g_z f_w \,. \tag{72c}$$

Putting these last three equations together, we end up with

$$\frac{f_x}{g_x} = \frac{f_y}{g_y} = \frac{f_z}{g_z} = \frac{f_w}{g_w} \equiv \lambda \,. \tag{73}$$

We can put these relations into one vector equation:

$$(f_x, f_y, f_z, f_w) = \lambda(g_x, g_y, g_z, g_w) \,, \tag{74}$$

which is equivalent to (66).

Using geometric algebra, we can recoup the cross product relations from (66) by first taking the wedge product of (66) by $\Box\, g(x, y, z, w)$ to get

$$\Box f(x, y, z, w) \wedge \Box\, g(x, y, z, w) = \mathbf{0} \,. \tag{75}$$

Next, we dot this last equation through by $\sigma_{ij} = \sigma_i \wedge \sigma_j$ $(i \neq j = 1, 2, 3, 4)$

$$\cdot \sigma_{ji} \cdot \Box f \wedge \Box\, g = 0 \,, \tag{76}$$

where $\Box f = (f_1, f_2, f_3, f_4) = (f_x, f_y, f_z, f_w)$, and likewise for $\Box f$. Up to a sign, there are only six nonzero independent bivectors σ_{ij}

5 Problem 4

Given the curve

$$g(x, y) = x^2 + xy - 1 = 0 \,, \tag{77}$$

what is the point of this curve in the first quadrant that is closest to the origin?

Note: I found this problem on a PDF handout posted on the Internet a few years ago, but at this time I cannot find its source url.

Solution 1: We will want to minimize the distance function on planar coordinates from the origin to some arbitrary point on curve defined by (77).

$$D(x, y) \equiv \sqrt{(x - 0)^2 + (y - 0)^2} = \sqrt{x^2 + y^2} \,. \tag{78}$$

But wait! Do we really want to differentiate a square root? Only if we have to. But in this case, we don't have to. Define

$$h(x, y) \equiv x^2 + y^2 \,. \tag{79}$$

This works because the two functions share the same minimums.

So, if we are to solve this problem by the method of Lagrange Multipliers, we would write down

$$\nabla(h(x,y) + \lambda g(x,y)) = \mathbf{0}\,. \tag{80}$$

Since there are no surprises following this course, I will, instead, continue with the SD approach (at least in the sense that it doesn't use λ's).

Solution 2: We go back to Equation (19c), to get

$$f_x g_y - g_x f_y = 0\,, \tag{81}$$

where, in this case $f = h$, thus

$$h_x g_y - g_x h_y = 0\,. \tag{82}$$

From this we have

$$x^2 - 2xy - y^2 = 0\,. \tag{83}$$

On elilminating xy between (77) and (83), we get

$$y^2 = 3x^2 - 2\,. \tag{84}$$

Going back to (77), solving it for xy and squaring that result, gives

$$x^2 y^2 = (1 - x^2)^2\,. \tag{85}$$

Combining this last equation with (84), we have

$$x^2(3x^2 - 2) = 1 - 2x^2 + x^4\,. \tag{86}$$

From this we get

$$x^4 = \frac{1}{2} \quad \text{or} \quad x^2 = \pm\frac{1}{\sqrt{2}}\,. \tag{87}$$

We will ignore the negative root to keep x^2 real. Then, with (84), we get that

$$y^2 = \frac{3}{\sqrt{2}} - 2\,. \tag{88}$$

So, our roots in the first quadrant are:

$$x = \frac{1}{\sqrt[4]{2}}\,, \quad y = \sqrt{\frac{3}{\sqrt{2}} - 2}\,. \tag{89}$$

Then, expressing these values in their decimal approximations, yields

$$x = 0.840896\,, \quad y = 0.348311\,. \tag{90}$$

6 Problem 5

Given the point $(2, 1)$, find the point of closest approach on the curve

$$g(x, y) = x^2 + 4xy - 5x + 2y^2 - 3y = 0\,. \tag{91}$$

Note: I found this problem on the same handout as the last problem.)

Solution 1: This time, we will skip doing the solution by the method of Lagrange multipliers.

Similar to the last propblem, we want to minimize the square-distance function on planar coordinates from the point $(2, 1)$ to some arbitrary point on curve defined by (91).

$$f(x, y) \equiv x^2 + y^2\,. \tag{92}$$

This works because the two functions share the same minimums.

Now, we need two coordinates top solve for x and y. We already h ave one equation, given by (91). So, all we need is one more algebraic eqaution to solve simultaneously with (91).

We can easily get this second equation by using

$$\begin{vmatrix} f_x & f_y \\ g_x & g_y \end{vmatrix} = \begin{vmatrix} 2(x-2) & 2(y-1) \\ 2x+4y-5 & 4x+4y-3 \end{vmatrix} = 0\,. \tag{93}$$

or

$$4x^2 + 2xy - 9x + y - 4y^2 + 1 = 0\,. \tag{94}$$

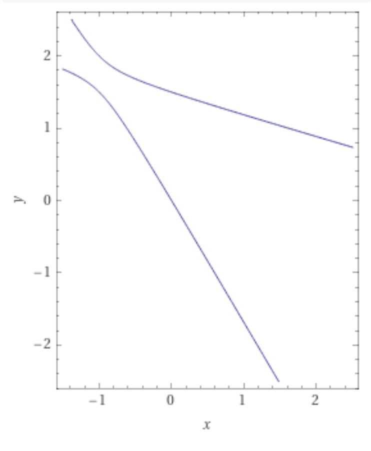

Figure 1. This graph is provided to us courtesy of WolframAlph.com. Clearly, the point on the graph closest to the point (2,1) is given by (96).

I used WolframAlpha.com to simultaneously solve for x and y between this last equation and (91), to get

$$x \approx 0.0863868\,, \quad y \approx -0.144235\,, \tag{95}$$

and

$$x \approx 1.96881\,, \quad y \approx 0.895289\,. \tag{96}$$

7 Problem 6

Suppose we have a function $f(x, y, z, w)$ of four variants we need to optimize, constrained by two constraint functions

$$g(x, y, z, w) = 0\,, \tag{97}$$

and

$$h(x, y, z, w) = 0\,, \tag{98}$$

We can write our Lagrange multiplier form as

$$\Box f(x, y, z, w) + \lambda \Box g(x, y, z, w) + \mu \Box h(x, y, z, w) = 0\,, \tag{99}$$

where λ and μ are just numbers, and $\Box = (\partial_x, \partial_y, \partial_z, \partial_w)$.

Find a way to augment Equations (97) and (98) by two determinant equations similar to (64).

Solution: From (99), we can write

$$\underset{(4\times3)}{\begin{bmatrix} f_x & g_x & h_x \\ f_y & g_y & h_y \\ f_z & g_z & h_z \\ f_w & g_w & h_w \end{bmatrix}} \underset{(3\times1)}{\begin{bmatrix} 1 \\ \lambda \\ \mu \end{bmatrix}} = \underset{(4\times1)}{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}} \tag{100}$$

Now, if we can reduce the 4×3 coefficient matrix to a 3×3 then we can take a determinant and set it equal to zero. To that end, let's multiply on the left of both sides of (100) by

$$\underset{(3\times4)}{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}, \tag{101}$$

which gives us

$$\underset{(3\times4)}{\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}}\underset{(4\times3)}{\begin{bmatrix}f_x&g_x&h_x\\f_y&g_y&h_y\\f_z&g_z&h_z\\f_w&g_w&h_w\end{bmatrix}}\underset{(3\times1)}{\begin{bmatrix}1\\\lambda\\\mu\end{bmatrix}}=\underset{(3\times1)}{\begin{bmatrix}0\\0\\0\end{bmatrix}} \tag{102}$$

This simplifies to

$$\underset{(3\times3)}{\begin{bmatrix}f_x&g_x&h_x\\f_y&g_y&h_y\\f_z&g_z&h_z\end{bmatrix}}\underset{(3\times1)}{\begin{bmatrix}1\\\lambda\\\mu\end{bmatrix}}=\underset{(3\times1)}{\begin{bmatrix}0\\0\\0\end{bmatrix}} \tag{103}$$

And the determinant of this coefficient matrix gives us (64).

Now, suppose we multiply through on the left of (100) by

$$\underset{(3\times4)}{\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}} \tag{104}$$

$$\underset{(3\times4)}{\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}}\underset{(4\times3)}{\begin{bmatrix}f_x&g_x&h_x\\f_y&g_y&h_y\\f_z&g_z&h_z\\f_w&g_w&h_w\end{bmatrix}}\underset{(3\times1)}{\begin{bmatrix}1\\\lambda\\\mu\end{bmatrix}}=\underset{(3\times1)}{\begin{bmatrix}0\\0\\0\end{bmatrix}} \tag{105}$$

which, by similar reasoning, gives us

$$\underset{(3\times3)}{\begin{bmatrix}f_x&g_x&h_x\\f_z&g_z&h_z\\f_w&g_w&h_w\end{bmatrix}}\underset{(3\times1)}{\begin{bmatrix}1\\\lambda\\\mu\end{bmatrix}}=\underset{(3\times1)}{\begin{bmatrix}0\\0\\0\end{bmatrix}} \tag{106}$$

Taking the determinant of the coefficient matrix gives

$$\begin{vmatrix} f_x & g_x & h_x \\ f_z & g_z & h_z \\ f_w & g_w & h_w \end{vmatrix} = 0\,. \tag{107}$$

Thus, (64) and (107) give us the additional equations we need to solve for the four variables x, y, z, w.

8 Problem 7

Optimize

$$f(x, y, z, w) = 8x + 6y + z - w \tag{108}$$

with constraints

$$g(x, y, z, w) = xy + z + w - 10 = 0\,, \tag{109}$$

and

$$h(x, y, z, w) = x + y - z - 2 = 0\,. \tag{110}$$

Solution:

Using the methods of the last problem, we get that

$$\begin{vmatrix} f_x & f_y & f_z \\ g_x & g_y & g_z \\ h_x & h_y & h_z \end{vmatrix} = \begin{vmatrix} 8 & 6 & 1 \\ y & x & 1 \\ 1 & 1 & -1 \end{vmatrix} = 0\,, \tag{111}$$

which simplifies to

$$9x - 7y + 2 = 0\,. \tag{112}$$

Similarly,

$$\begin{vmatrix} f_x & f_z & f_w \\ g_x & g_z & g_w \\ h_x & h_z & h_w \end{vmatrix} = \begin{vmatrix} 8 & 1 & -1 \\ y & 1 & 1 \\ 1 & -1 & 0 \end{vmatrix} = 0\,, \tag{113}$$

which simplifies to

$$y + 10 = 0\,. \tag{114}$$

Solving (109), (110), (112) and (114) simultaneously, we get

$$x = -8\,, \quad y = -10\,, \quad z = -20\,, \quad w = -50\,. \tag{115}$$

9 Problem 8

Suppose we let $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)^t$ and $\Box = (\partial_1, \partial_2, \partial_3, \partial_4, \partial_5)$. Given the function $f(\mathbf{x})$ to be optimized, with constraint functions

$$g(\mathbf{x}) = 0\,, \quad h(\mathbf{x}) = 0\,, \quad k(\mathbf{x}) = 0\,. \quad . \tag{116}$$

Starting off with the Lagrange multiplier form, we have

$$\Box f(x,y,z,w)+\lambda\,\Box\, g(x,y,z,w)+\mu\,\Box\, g(x,y,z,w)+\nu\,\Box\, g(x,y,z,w)=0\,. \quad (117)$$

Derive two additional equations to go with the three in (116) so that all together they can be used to solve for the critical points x_1, x_2, x_3, x_4, x_5.

Solution: This problem is a simple extension of of previous generalizations,. however, rather than use Gibbs's vector algebra, we'll instead use the wedge of geometric algebra. Therefore, we can place Equation (117) by the alternative form

$$\Box f \wedge \Box\, g \wedge \Box\, h \wedge \Box\, p = 0\,. \quad (118)$$

To extract from this the additional equations we need, we can dot this wedge product by $\sigma_{ijkm} = \sigma_i \wedge \sigma_j \wedge \sigma_k \wedge \sigma_m$, where i, j, k, m are integers from 1 to 5. For σ_{ijkm} to be nonzero, the subscripts must be all different.

Thus, one equation we could try is

$$\sigma_{4321} \cdot \Box f \wedge \Box\, g \wedge \Box\, h \wedge \Box\, p = 0\,, \quad (119)$$

which, in conventional form, is

$$\begin{vmatrix} f_1 & f_2 & f_3 & f_4 \\ g_1 & g_2 & g_3 & g_4 \\ h_1 & h_2 & h_3 & h_4 \\ p_1 & p_2 & p_3 & p_4 \end{vmatrix} = 0\,. \quad (120)$$

We can get our second equation, say, by dropping the 3 subscript and replacing it with 5, to get

$$\begin{vmatrix} f_1 & f_2 & f_4 & f_5 \\ g_1 & g_2 & g_4 & g_5 \\ h_1 & h_2 & h_4 & h_5 \\ p_1 & p_2 & p_4 & p_5 \end{vmatrix} = 0\,. \quad (121)$$

10 Problem 9

A discrete probability distribution with maximal information entropy is given by

$$f(p_1,\ldots,p_n) = -\sum_{k=1}^{n} p_k \log_2 p_k = -\sum_{k=1}^{n} p_k \frac{\ln p_k}{\ln 2}\,, \quad (122)$$

with constraint equation

$$g(p_1,\ldots,p_n) = \sum_{k=1}^{n} p_k - 1 = 0\,, \quad (123)$$

where this contraint enforces the rule that the sum of probabilities is equal to unity.

Now, show that for f to be at maximum, all the $p's$ must be the same.

Solution: Similar to the last problem, we have that

$$\sigma_{ji} \cdot \Box f \wedge \Box g = 0\,, \tag{124}$$

from which we get that

$$\begin{vmatrix} f_i & f_j \\ g_i & g_j \end{vmatrix} = 0\,. \tag{125}$$

Now, to compute these derivatives. First,

$$f_i = \frac{\partial f}{\partial p_i} = \frac{\partial}{\partial p_i}\Big[-\sum_{k=1}^{n} p_k \frac{\ln p_k}{\ln 2}\Big] = -\sum_{k=1}^{n}\Big[\delta_{ik}\frac{\ln p_k}{\ln 2} + p_k \frac{\delta_{ik}}{\ln 2 p_k}\Big] \tag{126}$$

$$= -\Big[\frac{\ln p_i}{\ln 2} + \frac{1}{\ln 2}\Big] = -[\log_2 p_i + \frac{1}{\ln 2}]\,, \tag{127}$$

Second,

$$g_m = \frac{\partial}{\partial p_m}\big[\sum_{k=1}^{n} p_k - 1\big] = 1\,. \tag{128}$$

Therefore, Eq. (125) becomes

$$\begin{vmatrix} -[\log_2 p_i + \frac{1}{\ln 2}] & -[\log_2 p_j + \frac{1}{\ln 2}] \\ 1 & 1 \end{vmatrix} = 0\,, \tag{129}$$

from which we have that

$$p_i = p_j\,, \tag{130}$$

for all i and j.

11 Problem 10

Suppose[3]

$$\nabla f(\mathbf{x}) = \sum_{\beta} \lambda_\beta \nabla g^\beta(\mathbf{x})\,, \tag{131}$$

with constraint equations

$$g^\beta(\mathbf{x}) = \kappa^\beta\,, \tag{132}$$

where we assume that the κ^β are continuous mutually independent parameters. Show that

$$\lambda_\alpha = \frac{df}{d\kappa^\alpha}\,. \tag{133}$$

Note:

$$\frac{d\mathbf{x}}{d\kappa^\alpha} \cdot \nabla g^\beta(\mathbf{x}) = \frac{dg^\beta}{d\kappa^\alpha} = \delta^\beta_\alpha\,. \tag{134}$$

---

[3] For what it's worth, I wrote on the page where I got my notes for this theorem the descriptor 'Objective Function Sensitivity Theorem'.

Solution: Dot through Eq. (131) by $d\mathbf{x}/d\kappa^\alpha$.

$$\frac{d\mathbf{x}}{d\kappa^\alpha} \cdot \nabla f(\mathbf{x}) = \frac{d\mathbf{x}}{d\kappa^\alpha} \cdot \sum_\beta \lambda_\beta \nabla g^\beta(\mathbf{x}) = \sum_\beta \lambda_\beta \frac{d\mathbf{x}}{d\kappa^\alpha} \cdot \nabla g^\beta(\mathbf{x}) = \sum_\beta \lambda_\beta \delta^\beta_\alpha = \lambda_\alpha \,. \tag{135}$$

Hence,

$$\frac{df}{d\kappa^\alpha} = \lambda_\alpha \,. \tag{136}$$

12 Problem 11

Problem: Minimize

$$F(\mathbf{x}) = \sum_\beta A_i^2 x_i \,, \quad \text{with} \quad A_i > 0 \quad \forall\; i \,, \tag{137}$$

with constraint equation

$$G(\mathbf{x}) = \frac{1}{c} - \sum_{i=1}^{n} \frac{1}{x_i - k} = 0 \,. \tag{138}$$

Solution: Using the method of Lagrange multipliers, we can write

$$\nabla F = \lambda \nabla G \,, \tag{139}$$

which produces

$$A_i^2 = \lambda \left(\frac{1}{(x_i - k)^2} \right) \,. \tag{140}$$

From this last equation we can write down a result we'll soon need:

$$\sqrt{\lambda} = |\, A_i \,| \, (x_i - k) \,. \tag{141}$$

Solving (138)

$$\frac{1}{c} = \sum_{i=1}^{n} \frac{1}{x_i - k} = \sum_{i=1}^{n} \frac{|\, A_i \,|}{\sqrt{\lambda}} = \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{n} |\, A_i \,| \,. \tag{142}$$

So,

$$\sqrt{\lambda} = c \sum_{i=1}^{n} |\, A_i \,| \,. \tag{143}$$

From (144), we write

$$x_i = \frac{\sqrt{\lambda}}{|\, A_i \,|} + k = \frac{\sum_{j=1}^{n} |\, A_j \,|}{|\, A_i \,|} + k \,. \tag{144}$$

13 Problem 12

Problem: Maximize the entropy function of a discrete distribution

$$H(\mathbf{p}) = \sum_{j=1}^{n} p_j \, \ln \frac{1}{p_j} \,, \tag{145}$$

where p_j is the jth probability. We also have constraint equations

$$G_1(\mathbf{x}) = \sum_{j=1}^{n} p_j - 1 = 0 \,, \tag{146}$$

and

$$G_2(\mathbf{x}) = \sum_{j=1}^{n} p_j g_j - G = 0 \,, \tag{147}$$

where G is a constant, being interpreted as the total energy.

Solution: First, we set up the main Lagrange equation

$$\nabla H(\mathbf{p}) = \alpha \nabla G_1 + \beta \nabla G_2 \,. \tag{148}$$

Now, we use geometric algebra to solve for α and β, starting with α. Thus, we wedge both sides by ∇G_2, to get

$$\nabla H \wedge \nabla G_2 = \alpha \nabla G_1 \wedge \nabla G_2 \,. \tag{149}$$

Now, dotting both sides by σ_{ji} and then solving for α, yields

$$\alpha = \frac{\sigma_{ji} \cdot \nabla H \wedge \nabla G_2}{\sigma_{ji} \cdot \nabla G_1 \wedge \nabla G_2} = \frac{\begin{vmatrix} H_i & H_j \\ G_{2i} & G_{2j} \end{vmatrix}}{\begin{vmatrix} G_{1i} & G_{1j} \\ G_{2i} & G_{2j} \end{vmatrix}} = \frac{g_j \ln \frac{1}{p_i} - g_i \ln \frac{1}{p_j}}{g_j - g_i} + 1 \,. \tag{150}$$

From this we have that

$$e^{(g_j - g_i)(\alpha - 1)} = \frac{p_j^{g_i}}{p_i^{g_j}} \,. \tag{151}$$

Next, we solve for β. We wedge both sides by ∇G_1, to get

$$\nabla H \wedge \nabla G_1 = -\beta \nabla G_1 \wedge \nabla G_2 \,. \tag{152}$$

Now, dotting both sides by σ_{ji} and then solving for β, yields

$$\beta = -\frac{\sigma_{ji} \cdot \nabla H \wedge \nabla G_1}{\sigma_{ji} \cdot \nabla G_1 \wedge \nabla G_2} = \frac{(\ln \frac{1}{p_i} + 1) - (g_i \ln \frac{1}{p_j} + 1)}{g_j - g_i} = \frac{\ln(p_i/p_j)}{g_j - g_i} \,. \tag{153}$$

From this we have that

$$e^{(g_j-g_i)\beta} = \frac{p_j}{p_i}\,. \tag{154}$$

From this we get that

$$p_j e^{g_j\beta} = p_i e^{g_i\beta} \quad \text{for all } i,j\,. \tag{155}$$

This last equality necessitates that we set the quantity $p_i e^{g_i\beta}$ equal to a constant, say, c:

$$p_i e^{g_i\beta} = c \quad \text{for all } i\,. \tag{156}$$

Solving this for p_i, we get

$$p_i = ce^{-g_i\beta} \quad \text{for all } i\,. \tag{157}$$

From (146), we have that

$$\sum_{j=1}^{n} p_j = c\sum_{j=1}^{n} e^{-\beta g_j} = 1\,. \tag{158}$$

We next introduce the *partition function* z by

$$z \equiv \sum_{j=1}^{n} e^{-\beta g_j}\,, \tag{159}$$

and, finally, we have that

$$z = \frac{1}{c}\,. \tag{160}$$

Corollary: (For the reader to solve.)

Show that with[4]

$$\nabla G_1 \wedge \nabla G_2 \wedge \nabla H = \mathbf{0}\,, \tag{161}$$

and, using that the g's are being treated as independent of the p's under 'partial differentiation', we get the (maybe useful?) result that

$$p_\nu^{g_\tau} p_\tau^{g_\mu} p_\mu^{g_\nu} = p_\tau^{g_\nu} p_\mu^{g_\tau} p_\nu^{g_\mu}\,. \tag{162}$$

Hint: The proof is a little messy, but straightforeward. Just expand the determinant

$$\sigma_{\tau\mu\nu} \cdot \nabla G_1 \wedge \nabla G_2 \wedge \nabla H = 0\,. \tag{163}$$

---

[4] We get this by wedging through (152) on the right by ∇H.

14 Conclusion

The point of this paper has been to produce some theory and applications of constrained optimization. I have not been averse to using (along with the usual Lagrange Multipliers) Structured Differentiation, matrix theory, Gibbs's vector theory, the Implicit Function Theorem, and Geometric Algebra to accomplish this goal. I could have dispensed with Gibb's's vector algebra altogether, but I used it because it is readily accessible to a larger audience than is Geometric Algebra.

Structured Differentiation isn't squirmish about using whatever mathematical tool is best fit to accomplish the job.

References

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