4. The Gradient Interpretation: Examples 1 and 2

4.1 Setup

Recall that, for **r** such that rank[R(**r**)] < S, we can find Lagrange multipliers λ_i such that

$$\mathbf{f}(\mathbf{r}) + \lambda_1\mathbf{a}(\mathbf{r}) + \lambda_2\,\mathbf{b}(\mathbf{r}) + ... + \lambda_{S-1}\,\mathbf{q}(\mathbf{r}) = 0\ , \qquad (2.5)$$

where the bolded letters are the rows of the R matrix shown in (1.2). In component notation this equation reads

$$f_i(\mathbf{r}) + \lambda_1 a_i(\mathbf{r}) + \lambda_2\, b_i(\mathbf{r}) + ... + \lambda_{S-1}\, q_i(\mathbf{r}) = 0 \qquad i = 1,2...N \qquad (4.1.1)$$

where recall $f_i = \partial f/\partial x_i$. In the usual gradient operator ∇ notation this can be written

$$\nabla f(\mathbf{r}) + \lambda_1\nabla a(\mathbf{r}) + \lambda_2\,\nabla b(\mathbf{r}) + ... + \lambda_{S-1}\,\nabla q(\mathbf{r}) = 0 \qquad \nabla = (\partial_1,\partial_2...\partial_N)\ . \qquad (4.1.2)$$

Having obtained this equation by the matrix rank method, we shall now show that it makes perfect sense in terms of the gradients. We shall consider two Examples below in order to build up to the general claim made at the end in Section 4.4. The reader uninterested in examples may skip to that section.

Before continuing, here are a few Facts which perhaps are "old hat" to the reader but need stating:

Fact 1: The equation $F(x_1,x_2.....x_n) = K$ defines an n-1 dimensional surface in E^n . (4.1.3)

For example, $F(x,y,z) = x^2+y^2+z^2$ with $F = R^2$ describes a spherical 2D surface of radius R in E^3.
$F(x,y) = x^2+y^2$ with $F= R^2$ describes a 1D surface (curve) in E^2, a circle

Fact 2: $\nabla F(\mathbf{r})$ is always locally *normal* to the surface $F(\mathbf{r}) = K$, where $\mathbf{r} = (x_1,x_2.....x_n)$. (4.1.4)

Proof: Start at point **r** on the surface $F(\mathbf{r}) = K$ and move a small distance d**r** in an arbitrary direction along the surface. Since one stays on the surface, $F(\mathbf{r}+d\mathbf{r}) = K$. Then $dF = F(\mathbf{r}+d\mathbf{r})- F(\mathbf{r}) = 0$. But one knows that $dF = \nabla F \bullet d\mathbf{r}$ and, for dF = 0 for *all* d**r** displacements on the surface, $\nabla F(\mathbf{r})$ must be locally normal to the surface at point **r**.

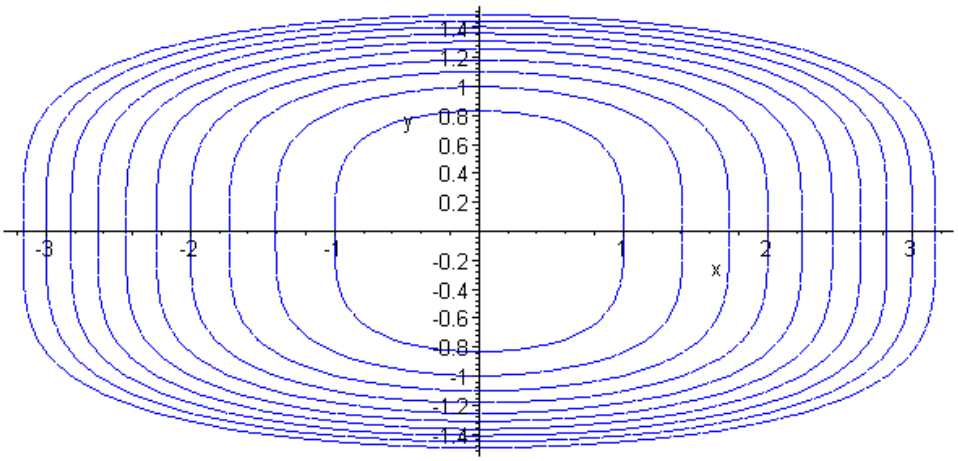
Fact 3: As K takes a set of different values K_1, K_2.....K_S, the equation $F(x_1,x_2.....x_n) = K$ describes a family of so-called **level surfaces**. If S is large and the range of the K_i is small, these surfaces will be closely spaced. Adjacent surfaces then have nearly the same shape, although in general *all* the surfaces in the set do not have the same shape. (4.1.5)

Example: $F(x,y) = x^2+2y^4$ and F= K for K = 1 to 10. Here "level surfaces" means "level curves".

```
restart;with(plots):
F := x^2+2*y^4;
```

$$F := x^2 + 2y^4$$

```
implicitplot({seq(F = K,K=1..10)},x=-4..4,y=-2..2,scaling=constrained,numpoints=1000,
color = blue);
```



(4.1.6)

Fact 4: The gradient ∇F points in the direction in which F changes most rapidly. (4.1.7)

Proof: For tiny displacement d**r**, dF = ∇F • d**r** is largest positive when d**r** points in the direction of ∇F ("uphill"). And dF = ∇F • d**r** is largest negative when d**r** points opposite the direction of ∇F ("downhill").

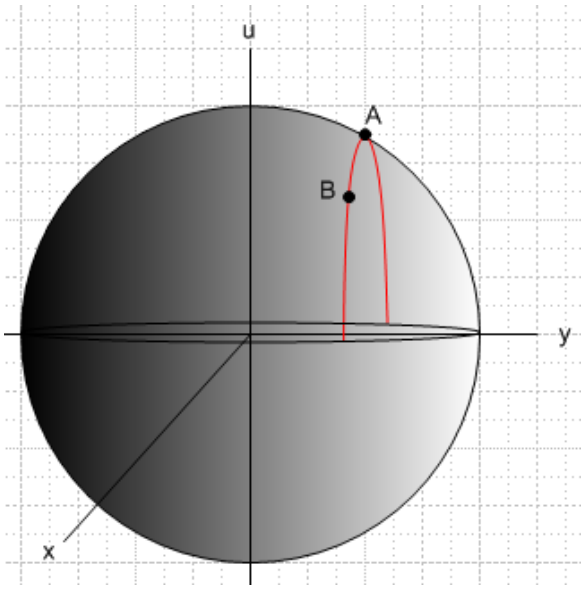
With these Facts established, we now consider a simple Lagrange Multiplier example.

4.2 Example 1: Half sphere with one simple constraint

Let u = f(x,y) represent the surface of a sphere (radius R = 2) in E^3, centered at the origin. We consider only the upper half of this surface, and we want to find (x,y) that maximizes f. If there are no constraints, then (4.1.2) above says ∇f = 0. The only place on our u = f surface having ∇f = 0 is the north pole of the sphere. This is just a regular "critical point" where $\partial_{\mathbf{x}}f = 0$ and $\partial_{\mathbf{y}}f = 0$, so we are happy with this interpretation of (4.1.2) with no constraints.

We now add a constraint a(x,y) = 0 where a(x,y) = y-1. This constraint is the line y=1 in the x-y plane E^2. We can extrude this line into a *plane* y = 1 in E^3. The hemispherical surface is a 2D surface in E^3, and the extruded constraint is also a 2D surface in E^3. These 2D surfaces intersect in a 1D surface which is a curve. There is hopefully some point on this curve that maximizes f.

Below is a picture of the sphere. The intersection of the upper spherical surface with the plane y = 1 is shown as a red curve (a half circle). Due to the constraint, we cannot get to the north pole so the maximum value of f is some value less than that u = 2 at the north pole. The extremum of this constrained problem will be at point **A**, and point **B** is not an extremum.

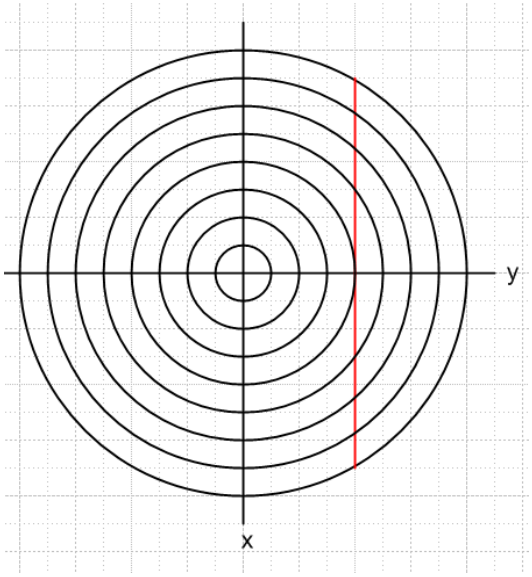


(4.2.1)

What is potentially misleading in this example (so far) is the notion of surface and gradient. For the above sphere, the surface is described by r = 2 in spherical coordinates, and then if f(**r**) = r, one finds that $\nabla f = \partial_{\mathbf{r}} f\ \hat{\mathbf{r}} = \hat{\mathbf{r}}$, and this is indeed normal to the surface at every point on it, as predicted by Fact 2 above. Similarly, the y=1 plane constraint function a(x,y) = y-1 = 0 has $\nabla a = -1\ \hat{\mathbf{y}}$ and this is everywhere normal to the constraint plane.

The potential confusion is that the Lagrange Multiplier Show does not play out on the stage shown in (4.2.1). It plays out in a space of one lower dimension. For Example 1, the Lagrange Multiplier Show plays out on the disk which lies at u = 0 in (4.2.1). The gradients involved with the Lagrange Multiplier method for Example 1 are 2D gradients which lie in this disk, not the 3D gradients to surfaces appearing in (4.2.1).

So let us move to the proper setting which is this disk in the x,y plane. Consider,



(4.2.2)

We wish to maximize the function $f(x,y) = \sqrt{4 - x^2 - y^2}$ subject to the constraint y = 1. In (4.2.1) the spherical surface is $u^2+x^2+y^2 = R^2 = 4$, so f(x,y) is then height u. So the function $\sqrt{4 - x^2 - y^2} = \sqrt{4 - r^2}$ is the height of the hemisphere in Fig (4.2.1) above the point (x,y). If we consider f(x,y) = K for a set of K values, we get a set of level curves in (4.2.2) which are circles. On each circle in (4.2.2), the height of the sphere lying over the surface in (4.2.1) is constant -- that is why they are called level curves.

The constraint is a(x,y) = 0 with constraint function a(x,y) = y-1, and the constraint y = 1 is shown as the red line in (4.4). We compute,

$$f = \sqrt{4 - x^2 - y^2} = \sqrt{4 - r^2}$$
$$a = y-1 \qquad (4.2.3)$$

so

$$\nabla f = \nabla\sqrt{4 - r^2} = [-r/\sqrt{4-r^2}\,]\,\hat{\mathbf{r}}$$
$$\nabla a = \nabla(y-1) = \hat{\mathbf{y}}\,. \qquad (4.2.4)$$

For any point **r** in the disk of (4.2.2), ∇f therefore points toward the disk center, and this is then the uphill direction for the hemisphere in (4.2.1). In Cartesian coordinates,

$$\nabla f = \nabla\sqrt{4 - x^2 - y^2} = -(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}} = -\frac{x}{\sqrt{4 - x^2 - y^2}}\,\hat{\mathbf{x}} - \frac{y}{\sqrt{4 - x^2 - y^2}}\,\hat{\mathbf{y}}\,. \qquad (4.2.5)$$

Now, equation (4.1.2) says

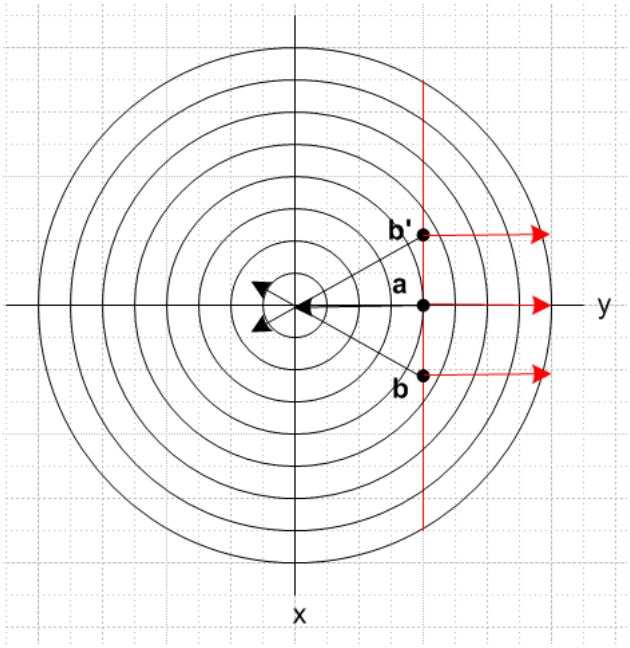
$$\nabla f(\mathbf{r}) = -\lambda_1 \nabla a(\mathbf{r}) = 0\,. \qquad (4.2.6)$$

How do we make geometric sense of this equation? If **r** is an "extremum" location, we should have df = 0 for any legal small displacement d**r** starting at this point **r**. Eq. (4.2.6) dotted into d**r** then says,

$$0 = df = \nabla f(\mathbf{r}) \bullet d\mathbf{r} = -\lambda_1 \nabla a(\mathbf{r}) \bullet d\mathbf{r}\,. \qquad (4.2.7)$$

If we displace d**r** in any direction on the constraint surface (which from Fact 2 has normal $\nabla a(\mathbf{r})$), then $\nabla a(\mathbf{r}) \bullet d\mathbf{r} = \mathbf{0}$ and we find df = 0, so we are thus *at* an extremum of f. For other values of position **r** , we will find either that df > 0 or df < 0 if we move d**r** along the constraint surface. If we are looking for a maximum of f, for example, then it is to our advantage to move a little d**r** in a direction for which df > 0. Doing this repeatedly to find a maximum (minimum) of f is called the **method of gradient ascent** (**descent**)

In our example with only one constraint, (4.2.6) says that we have reached an extremum when ∇f and ∇a are collinear. Let's draw onto (4.2.2) the points **a** and **b** which lie under points **A** and **B** in (4.2.1), and we include a point **b'** which is a mirror image of **b**,



(4.2.8)

At each of the points **a**,**b**,**b'** we show ∇f in black and ∇a in red. At point **a** we have reached the closest distance to the center that is allowed for points on the red constraint line, so this will correspond to the maximum value of f, which is point **A** in (4.2.1). At point **a** we see that in fact the two gradients are collinear as required by (4.1.2) at an extremum.

Suppose we are at point **b**. We may compute df for a small displacement upwards (minus x direction)

$$df(\mathbf{b}) = \nabla f \bullet d\mathbf{r} \;= [\, -(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}}\,] \bullet |dx|(-\hat{\mathbf{x}}) = |dx|\,(x/f) \;> 0 \qquad \text{since } x>0 \text{ and } f>0\ . \qquad (4.2.9)$$

Since $df > 0$, it is to our advantage in finding max f to move upwards in (4.2.8). Conversely, suppose we are instead at point **b'**. Then moving toward **a** gives,

$$df(\mathbf{b}') = \nabla f \bullet d\mathbf{r} \;= [\, -(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}}\,] \bullet |dx|(\hat{\mathbf{x}}) = |dx|\,(-x/f) \;> 0 \qquad \text{since } x<0 \text{ and } f>0\ . \qquad (4.2.10)$$

and again we find that $df > 0$. From either starting position **b** or **b'**, moving toward point **a** is a win.

It is an easy matter to compute the solution value of the Lagrange Multiplier λ_1 and **r**. Inserting (4.2.4) into (4.2.6) gives

$$\nabla f(\mathbf{r}) = -\lambda_1 \nabla a(\mathbf{r}) \qquad \Rightarrow$$

$$-\frac{x}{\sqrt{4 - x^2 - y^2}}\,\hat{\mathbf{x}} \;-\frac{y}{\sqrt{4 - x^2 - y^2}}\,\hat{\mathbf{y}} \;= -\lambda_1\,\hat{\mathbf{y}}\ . \qquad (4.2.11)$$

This says $x = 0$, and then using the constraint $y = 1$,

$$-\frac{1}{\sqrt{4 - 1}}\,\hat{\mathbf{y}} \;= -\lambda_1\,\hat{\mathbf{y}} \qquad \Rightarrow \qquad \lambda_1 = 1/\sqrt{3}\ . \qquad (4.2.12)$$

The solution to the Example 1 extremum problem is then

$$\mathbf{r} = (0,1) \qquad \lambda_1 = 1/\sqrt{3}\ . \tag{4.2.13}$$

We have claimed in (2.1) that at the solution point, the Lagrangian function H(x,y) should have a maximum, since the whole theory is based on H being the function to maximize without constraints. One has,

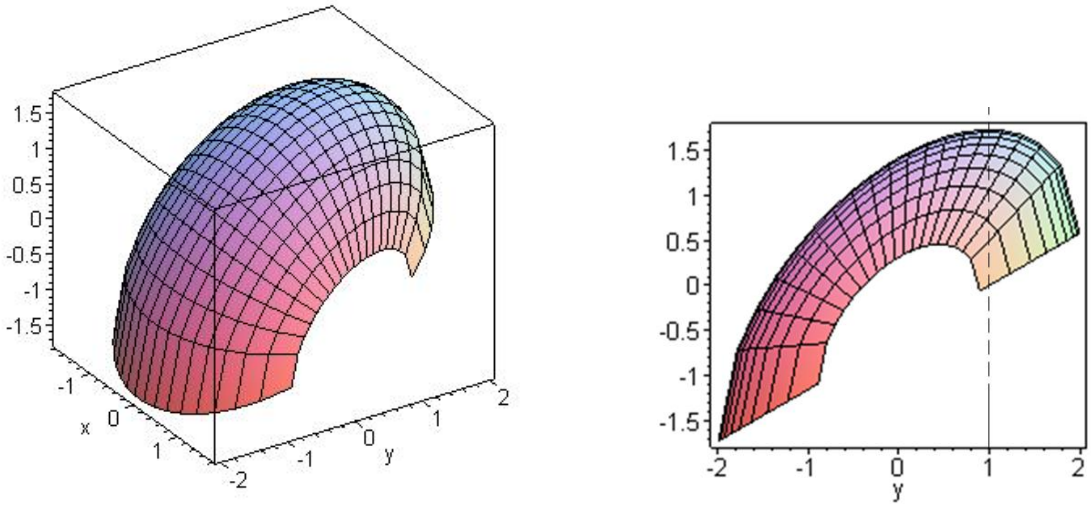
$$H(x,y) = f(\mathbf{r}) + \lambda_1 a(\mathbf{r}) = \sqrt{4 - x^2 - y^2} + (1/\sqrt{3}\ )\,(y-1). \tag{4.2.14}$$

Now $u = (1/\sqrt{3}\ )\,(y-1)$ is a plane sloping up to the right in Fig (4.2.1). This is different from the plane y = 1 which is the vertical constraint plane in that figure. In (4.2.14) we are adding a spherical surface to a plane sloping up to the right, and the result is an ellipsoidal-like surface (really a quartic surface) which has a maximum at the point **A** = (0,1). Here is a plot of that surface:

```
H := sqrt(4-x^2-y^2) + (1/sqrt(3))*(y-1);
```

$$H := \sqrt{4 - x^2 - y^2} + \frac{1}{3}\sqrt{3}\,(y-1)$$

```
plot3d(H,x=-2..2,y = -sqrt(4-x^2)+1e-8..sqrt(4-x^2)-1e-8,axes=boxed,grid = [20,20]);
```



(4.2.15)

One can see from these plots that the maximum of H(x,y) occurs at x = 0 and y = 1.

A direct method of confirming the maximum is provided by examining H_{ii} :

$$H_i = f_i + \lambda_1 a_i \qquad H_{ii} = f_{ii} + \lambda_1 a_{ii} \qquad a_{ii} = \partial^2(y-1)/\partial x_i^2 = 0$$

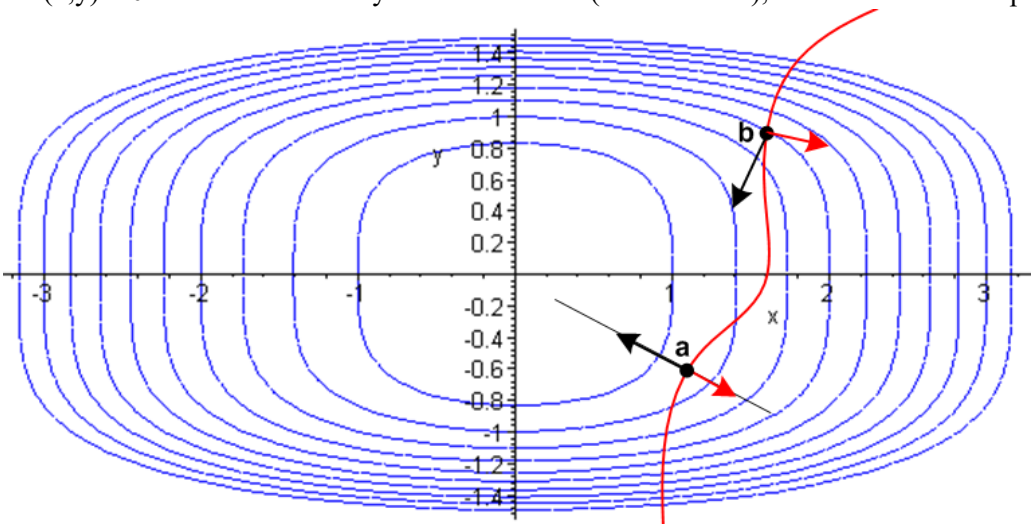
$$f = \sqrt{4 - x^2 - y^2} = \sqrt{4 - x_1^2 - x_2^2} \qquad \Rightarrow \qquad f_i = -x_i/f \qquad f > 0$$

$$f_{ii} = -[f * 1 - x_i f_i] / f^2 = -[f - x_i(-x_i/f)] / f^2 = -(1/f) - (x_i/f)^2$$

$$H_{ii} = f_{ii} + \lambda_1 a_{ii} = f_{ii} = -(1/f) - (x_i/f)^2 < 0\ . \tag{4.2.16}$$

This shows that the function H(x,y) is "cupping down" at all locations, as the graph suggests. The point A where $H_i = 0$ (see (4.6)) thus also has $H_{ii} < 0$ and is thus a maximum.

If we had a surface in (4.2.1) more complicated than a sphere, and a constraint more complicated than y = 1, the nature of the interpretation of (4.1.2) does not change. For example, if (4.2.1) contained a surface u = f(x,y) whose level curves were those of (4.1.6) with f = K values increasing toward the center, and if the constraint a(x,y) = 0 were some arbitrary constraint curve (shown in red), we would have this picture:



(4.2.17)

In this case the red arrow ∇a changes direction as one moves along the red curve, but the solution point is shown as the black dot **a** where the red arrow and the black arrow representing ∇f are collinear. At point **b** the gradient arrows do not line up, and there is advantage for increasing f by moving toward point **a**.

Notice that saying the two arrows are collinear is to say they are linearly dependent. Equation (4.1.2) says that the gradients are linearly dependent in the general case as well as this simple case.

4.3 Example 2: Half 4D sphere with two simple constraints

In order to better demonstrate the interpretation of (4.1.2) concerning the gradients. we upgrade Example 1 and then allow two constraints. We take the surface of (4.2.1) to be the "upper half" of a 4D sphere, whatever that means. Since we cannot draw such a thing in E^4 even in a projection drawing like (4.2.1), we don't even try. However, the Lagrange Multiplier scenario (the analog of (4.2.2)) exists in E^3, not E^4, and we can at least draw that. In Example 1 the level curves were a set of concentric circles in E^2. For Example 2 the level surfaces are a set of concentric spheres in E^3 each labeled by a value of K, where f(x,y,z) = K and where $f(x,y,z) = \sqrt{4 - x^2 - y^2 - z^2}$. We are not proving this claim, it is just based on analogy: if the slice of a 3D sphere is a 2D sphere (a disk), then a slice of a 4D sphere ought to be a 3D sphere.

We take as our two constraints the equations y = 1 and x = 1, just to keep it simple. Thus

$f(x,y,z) = \sqrt{4 - x^2 - y^2 - z^2} = \sqrt{4 - r^2}$

a(x,y,z) = y-1

$$b(x,y,z) = x-1 \ . \qquad (4.3.1)$$

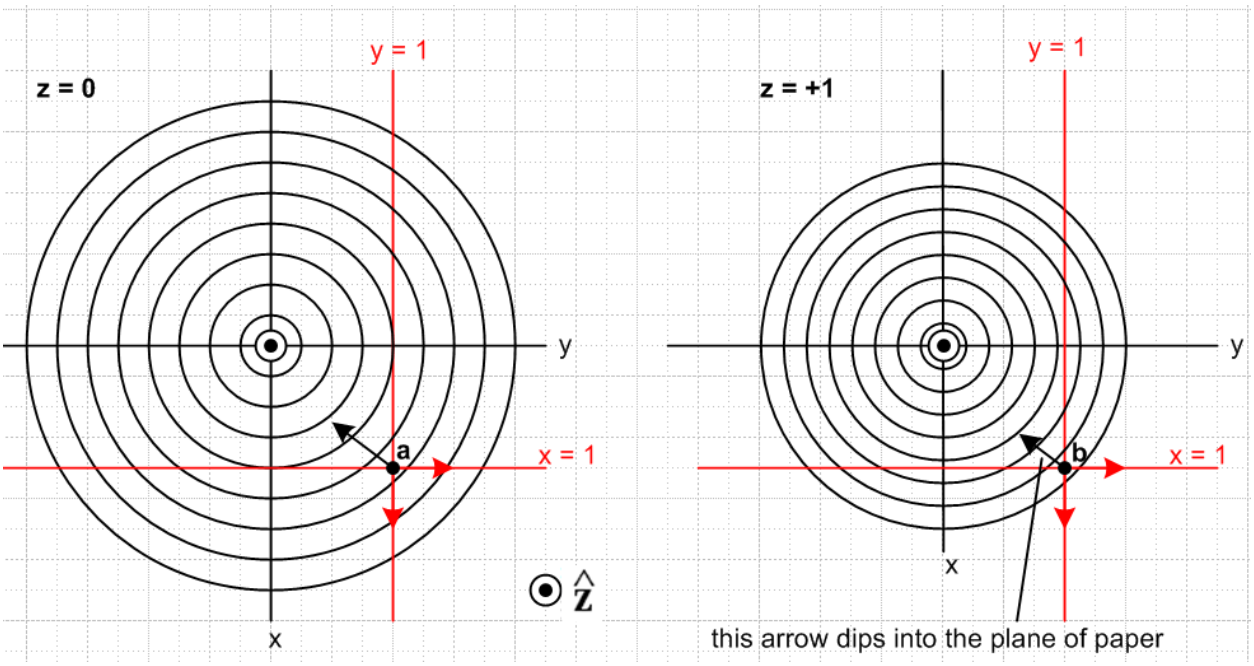
The gradients of interest are now 3D gradients in E^3 space, so

$$\nabla f = \nabla(\sqrt{4 - r^2}\,) = [-r/\sqrt{4-r^2}\,]\,\hat{\mathbf{r}} = -(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}} - (z/f)\hat{\mathbf{z}} \qquad \text{points to sphere center}$$
$$\nabla a = \nabla(\,y-1) = \hat{\mathbf{y}}$$
$$\nabla b = \nabla(\,x-1) = \hat{\mathbf{x}}\ . \qquad (4.3.2)$$

Equation (4.1.2) reads

$$\nabla f(\mathbf{r}) + \lambda_1 \nabla a(\mathbf{r}) + \lambda_2\, \nabla b(\mathbf{r}) = 0 \ . \qquad (4.3.3)$$

This says that the three gradients must be linearly dependent, which means they must be coplanar! Before solving the problem, we would like to draw a picture corresponding to (4.2.2) for Example 1. Although one could in fact draw such a picture with some effort, we shall decline this task and instead draw two z = constant slices of the desired picture. On the left below is a slice at z = 0, while on the right is a slice at z = 1



(4.3.4)

In these pictures, we are viewing the two red constraint planes y = 1 and x = 1 edge on. The surface which satisfies both constraints is a line normal to the plane of paper which is the intersection of the two constraint planes. The circles on the left are slices of the family of concentric spheres taken at the equator z = 0. At z = +1 since our slicing plane is closer to the north pole, the sphere slices have smaller diameter as shown.

Consider now the point **r** = **b** shown as a black dot on the right. The two red arrows are $\nabla a = \hat{\mathbf{y}}$ and $\nabla b = \hat{\mathbf{x}}$, each normal to its constraint plane. The black arrow dips into the plane of paper because ∇f =

$[-r/\sqrt{4-r^2}\,]\ \hat{\mathbf{r}}$ which points toward the sphere center. This black point is located outside the r = 7 units sphere (count the rings). The three arrows ∇f, ∇a,∇b are therefore not coplanar, so they are linearly independent. That means that equation (4.3.3) cannot exist [see (1.6)] for this point **r**.

Consider next the point **r** = **a** shown as a black dot on the left. Since this is the equatorial slice, the black arrow ∇f lies in the plane of paper so the three arrows are coplanar. This means the three vectors ∇f, ∇a,∇b are coplanar so they are linearly *dependent*. For this point **a**, the equation (4.3.3) can and does exist, and therefore point **a** is the problem solution. This point is located inside the r = 6 unit sphere.

Once again, at an extremum point we should have df = 0. Moving an amount d**r** which is consistent with both constraints (meaning d**r** is in the z direction2) we then have from (4.3.3),

$$df = \nabla f \bullet d\mathbf{r} \;=\; -\lambda_1[\nabla a(\mathbf{r})\bullet d\mathbf{r}] \;-\; \lambda_2\,[\nabla b(\mathbf{r})\bullet d\mathbf{r}] \;= -\lambda_1\,[0] \;-\lambda_2[0] \;= 0 \qquad (4.3.5)$$

and sure enough, df = 0. At point **b** one finds for a d**r** pointed toward the equatorial plane,

$$df = \nabla f \bullet d\mathbf{r} \;=\; [\,-(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}} \;-(z/f)\hat{\mathbf{z}}\,] \bullet (-|dr|\,\hat{\mathbf{z}}\,) = (z/f)dr \;> 0 \qquad (4.3.6)$$

and so it is advantageous to move $d\mathbf{r} = |dr|\hat{\mathbf{z}}$ toward **a** and thereby increase f, so **b** is not an extremum. Evaluating (4.3.6) at z = 0 for point **a** again shows df = 0.

Finally, we solve the problem. Inserting (4.3.2) into (4.3.3) gives

$$\nabla f(\mathbf{r}) = -\lambda_1\nabla a(\mathbf{r}) \;-\lambda_2\,\nabla b(\mathbf{r}) \qquad \Rightarrow$$

$$[-\frac{x}{\sqrt{4 - x^2 - y^2 - z^2}}\,\hat{\mathbf{x}} - \frac{y}{\sqrt{4 - x^2 - y^2 - z^2}}\,\hat{\mathbf{y}} \;-\frac{z}{\sqrt{4 - x^2 - y^2 - z^2}}\,\hat{\mathbf{z}}\,] = -\;\lambda_1\hat{\mathbf{y}} \;-\; \lambda_2\,\hat{\mathbf{x}}\,. \qquad (4.3.7)$$

We see at once that z = 0 and then the above becomes

$$\frac{x}{\sqrt{4 - x^2 - y^2}} = \lambda_2 \qquad \frac{y}{\sqrt{4 - x^2 - y^2}} = \lambda_1 \quad . \qquad (4.3.8)$$

But the constraints say x = 1 and y = 1 so,

$$\frac{1}{\sqrt{4 - 1^2 - 1^2}} = \lambda_2 \qquad \frac{1}{\sqrt{4 - 1^2 - 1^2}} = \lambda_1 \qquad \Rightarrow \qquad \lambda_1 = \lambda_2 = 1/\sqrt{2} \qquad (4.3.9)$$

Therefore the solution to Example 2 is this:

$$\mathbf{r} = (1,1,0) \qquad \lambda_1 \;=\; 1/\sqrt{2} \qquad \lambda_2 \;=\; 1/\sqrt{2} \qquad (4.3.10)$$

4.4 The General Case

Here we summarize what has been demonstrated in the above examples. We have the general gradient equation (4.1.2) emerging from the Lagrange Multiplier analysis,

$$\nabla f(\mathbf{r}) + \lambda_1\nabla a(\mathbf{r}) \;+\lambda_2\,\nabla b(\mathbf{r}) + \ldots\ldots + \lambda_{s-1}\,\nabla q(\mathbf{r}) \;= 0 \qquad \nabla \;= (\partial_1, \partial_2 \ldots \partial_N) \quad . \qquad (4.1.2)$$

This equation only exists if **r** is an extremum!! Solving for $\nabla f(\mathbf{r})$ and dotting with d**r** gives

$$df(\mathbf{r}) = \nabla f(\mathbf{r}) \bullet d\mathbf{r} = -\lambda_1[\nabla a(\mathbf{r}) \bullet d\mathbf{r}] - \lambda_2 [\nabla b(\mathbf{r}) \bullet d\mathbf{r}] - ..... - \lambda_{s-1} [\nabla q(\mathbf{r}) \bullet d\mathbf{r}] . \quad (4.4.1)$$

If d**r** at point **r** is chosen so *all* the constraints are respected (that is, we stay on all the constraint surfaces as we displace d**r** from **r**) , then $\nabla a(\mathbf{r}) \bullet d\mathbf{r} = 0$, $\nabla b(\mathbf{r}) \bullet d\mathbf{r} = 0$ ..... and $\nabla q(\mathbf{r}) \bullet d\mathbf{r} = 0$. Then

$$df(\mathbf{r}) = \nabla f(\mathbf{r}) \bullet d\mathbf{r} = -\lambda_1[0] - \lambda_2 [0] - ..... - \lambda_{s-1} [0] = 0 . \quad (4.4.2)$$

Since $df(\mathbf{r}) = 0$, the point **r** is an extremum of $f(\mathbf{r})$ subject to these constraints.