Save partially edited Section 4 on gradient stuff.
PhL 1.11.15

Note that page numbering is turned on in this template and view is 125%,
located in phil/roaming/microsoft/templates size about 219K.

4. The Gradient Interpretation: Example 1

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

$$f_i(\mathbf{r}) + \lambda_1 a_i(\mathbf{r}) + \lambda_2 b_i(\mathbf{r}) + ... + \lambda_{S-1} q_i(\mathbf{r}) = 0 \qquad i = 1,2...N \qquad (4.1)$$

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

$$\nabla f(\mathbf{r}) + \lambda_1 \nabla a(\mathbf{r}) + \lambda_2 \nabla b(\mathbf{r}) + ... + \lambda_{S-1} \nabla q(\mathbf{r}) = 0 \qquad \nabla = (\partial_1, \partial_2 ... \partial_N) \ . \qquad (4.2)$$

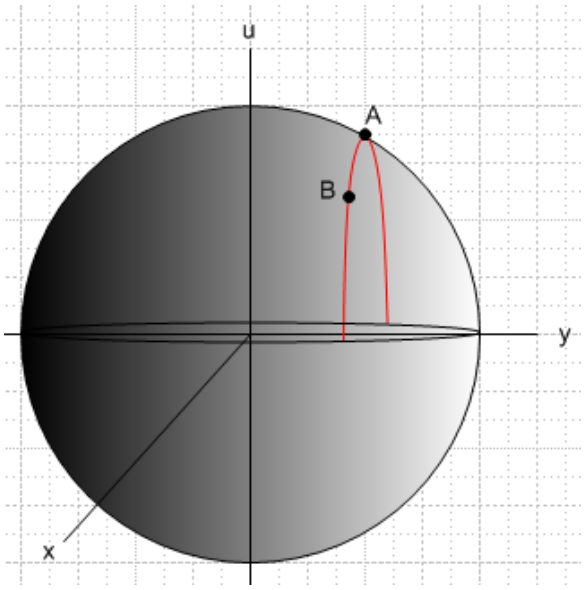
How might one interpret this gradient sum being 0? If we use N = 2 and S = 2, we can at least obtain some visualization, then we can generalize the conclusion in a logical fashion.

A Simple Example with One 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 (4.2) above says $\nabla f = 0$. The only place on our u = f surface having $\nabla f = 0$ is the north pole of the sphere. This is just a regular "critical point" where $\partial_x f = 0$ and $\partial_y f = 0$, so we are happy with this interpretation of (4.2) with no constraints.

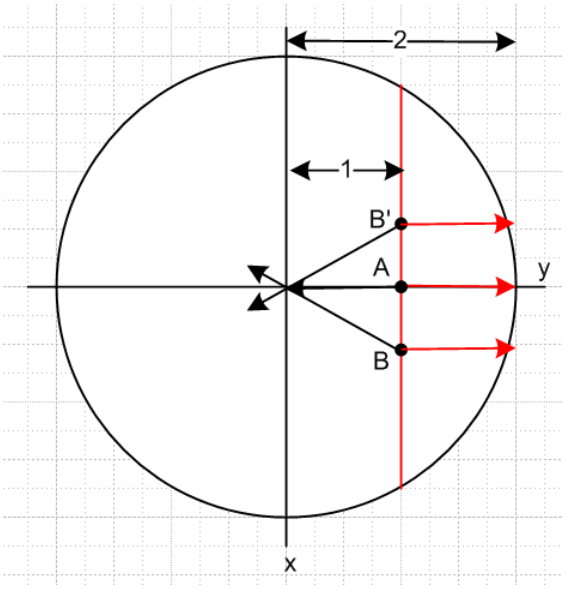
We now add a constraint a(x,y) = 0 where a(x,y) = y-1. This constraint is the line y=1 in the x-y plane E^2. We can extrude this line into a *plane* y = 1 in E^3. The hemispherical surface is a 2D surface in E^3, and the extruded constraint is also a 2D surface in E^3. These 2D surfaces intersect in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



(4.3)

Since N = 2, the gradients in (4.2) are 2D gradients and so are parallel to the x,y plane. To view these gradients, we draw a top view of the sphere, looking toward the sphere center from the positive u axis:



(4.4)

The three black arrows show ∇f at the points A, B and B'. It seems clear that on the spherical surface ∇f always points toward the north-south axis of the sphere. We can put the gradient arrow tails at the points of interest, but the gradient arrows are all parallel to the x,y plane since for example $\nabla f = (\partial_{\mathbf{x}} f, \partial_{\mathbf{y}} f)$.

Meanwhile, the three red arrows show the direction of ∇a at points A, B and B'. Since a = y-1, this direction is of course $\nabla a = (\partial_{\mathbf{x}} a \; \partial_{\mathbf{y}} a) = (0,1)$ which points to the right no matter where on the extruded constraint surface the arrow tail is placed.

As point B or B' moves toward point A, the black and red arrows become more aligned and finally at point A (the extremum) they are exactly collinear.

Equation (4.2) says that at the solution point $\mathbf{r}$ = A for the extremum, we have

$$\nabla f(\mathbf{r}=A) + \lambda_1 \nabla a(\mathbf{r}=A) = 0 \qquad (4.5)$$

and indeed, this equation says that at the solution point the two gradients must be collinear.

More details of the Example 1.

$f(x,y) = \sqrt{4 - x^2 - y^2}$ since the sphere has $u^2+x^2+y^2 = R^2 = 4$

$\nabla f(x,y) = -(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}}$

$f(r,\theta) = \sqrt{4 - r^2}$ polar coordinates

$\nabla f(r,\theta) = [-r/\sqrt{4-r^2}\,]\,\hat{\mathbf{r}}$ points toward the sphere's vertical axis (and $\nabla f = 0$ at $\mathbf{r} = 0$, north pole)

$a(x,y) = y-1 = 0$ the constraint

$\nabla a = [1]\,\hat{\mathbf{y}}$ points toward the right

At point A = (x,y) = (0,1) one has r = 1, $\hat{\mathbf{r}} = \hat{\mathbf{y}}$ and $f = \sqrt{3}$. Then (4.5) reads

$$[-1/\sqrt{3}\,]\hat{\mathbf{y}} + \lambda_1[\,1/\sqrt{3}\,]\,\hat{\mathbf{y}} = 0 \quad \Rightarrow \quad \lambda_1 = 1/\sqrt{3}\,. \qquad (4.6)$$

At the solution point A the Lagrange multiplier λ_1 can thus be interpreted as the negative of the ratio of the two gradient vectors ∇f and ∇a. Note that at the solution point, the two gradients really are collinear.

At point B, we can consider a differential movement $d\mathbf{r} = |dx|(-\hat{\mathbf{x}})$ along the constraint surface toward point A (in Fig (4.4) the x axis points down). We find that

$$df(B) = \nabla f \bullet d\mathbf{r} = [\,-(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}}\,] \bullet |dx|(-\hat{\mathbf{x}}) = |dx|\,(x/f) > 0 \quad \text{since } x>0 \text{ and } f>0\,. \qquad (4.7)$$

Thus, as we move toward point A, f increases since df>0, suggesting that A is indeed a maximum.

Consider the mirror point B' located above point A in Fig (4.4). The differential vector $d\mathbf{r} = |dx|\,\hat{\mathbf{x}}$ lies along the constraint surface and points from B' toward A. Then

$$df(B') = \nabla f \bullet d\mathbf{r} = [\,-(x/f)\hat{\mathbf{x}} - (y/f)\hat{\mathbf{y}}\,] \bullet |dx|\,\hat{\mathbf{x}} = |dx|\,(-x/f) > 0 \quad \text{since } -x>0 \text{ and } f>0\,. \qquad (4.8)$$

Once again, df > 0 showing indeed that point A is a maximum.

Finally, we have claimed 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,

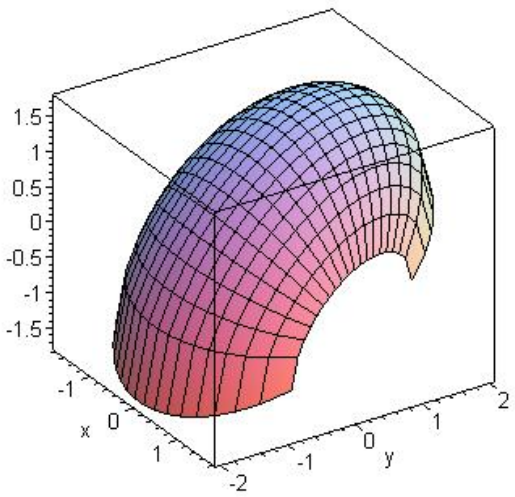
$$H(x,y) = f(\mathbf{r}) + \lambda_1 a(\mathbf{r}) = \sqrt{4 - x^2 - y^2} + (1/\sqrt{3}\,)\,(y-1). \qquad (4.9)$$

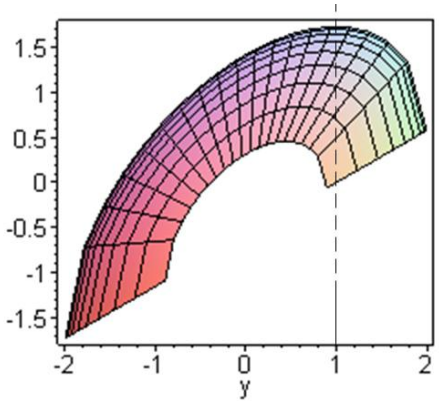
Now $u = (1/\sqrt{3}\,)\,(y-1)$ is a plane sloping up to the right in Fig (4.3). This is different from the plane $y = 1$ which is a vertical plane in that figure. In (4.9) we are adding a spherical surface to a plane sloping up to the right, and the result is an ellipsoidal-like surface (really a quartic 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]);
```





(4.10)

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

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

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

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

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

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.

<u>The general case</u>

Consider first a 2D surface in 3D space defined by F(x,y,z) = C. If one is on this surface and moves around on the surface, one always has F = C.

Fact: At any point **r**, $\nabla F(\mathbf{r})$ is locally *normal* to the surface.

The reason is very simple. Suppose an ant starts at point **r** on the surface and move d**r** in some arbitrary direction *on the surface*. Saying the ant stays on the surface says that dF = 0 for this motion. But dF = d**r** • ∇F. Since dF= 0, and since this is true for any on-surface direction d**r**, ∇F must be normal to the surface at that point **r**.

Now consider $F(\mathbf{x_1},\mathbf{x_2}.....\mathbf{x_n}) = C$. This is an n-1 dimensional surface in E^n. For any motion d**r** along this surface, one finds the same thing as above, dF = d**r** • ∇F = 0 so ∇F is normal to the surface in E^n.

In the extremum problem with constraints in n dimensions, for any constant C the equation $f(\mathbf{x_1},\mathbf{x_2}.....\mathbf{x_n}) = C$ defines an n-1 dimensional surface. This surface is called a **level surface** of f. If one takes a set values for C, one gets a set of level surfaces. They might look like a set of concentric spherical shells in E^3, but in general the shape changes slightly as C makes a small change in value. In Example 1 above, the half sphere shown in Fig (4.3) is a level surface for x^2+y^2+

For a surface of dimension n-1 in E^n that same thing is true.

Given a surface F(x,y,z) = C in 3D space, at any point on that surface∇F is a *normal vector*, so if an ant crawls some small distance d**r** away from the point that surface, d**r** • ∇F = 0. The reason this is true is very simple. A motion of d**r** *along* the surface results in a change in F of dF = 0 and since dF = d**r** • ∇F, and since dr could be any direction along the surface. it must be that ∇F must be a normal vector (in general it is not a unit normal vector).

If we were to upgrade this example to u = f(x,y,z) and have two constraints a=0 and b=0, the final intersection path of the hypersphere surface with the two constraint surfaces is again a 1D curve. If we consider d**r** along this curve, that d**r** will be perpendicular to both ∇a and ∇b, which says

$$(\lambda_1\nabla a + \lambda_2\nabla b) \bullet d\mathbf{r} = 0 \quad . \tag{4.11}$$

At an extremum point, if (4.2) is valid, we find that

$$\nabla f \bullet d\mathbf{r} = -(\lambda_1\nabla a + \lambda_2\nabla b) \bullet d\mathbf{r} = 0 \quad . \tag{4.12}$$

Since this d**r** movement is perpendicular to ∇f, there can be no change in f by moving along the intersecting constraint surface, and that is why we are at an extremum.

So this then provides an interpretation for the general solution case where this equation is true:

$$\nabla f + \lambda_1 \nabla a + \ldots \lambda_{s-1} \nabla q = 0 \qquad (4.2)$$

and thus,

$$\nabla f \bullet d\mathbf{r} = -(\lambda_1 \nabla a + \lambda_2 \nabla b + \ldots + \lambda_{s-1} \nabla q) \bullet d\mathbf{r} \qquad (4.13)$$

In general the intersection surface of f with all the constraint surfaces will be of dimension N-S+1 in E^n. A tiny displacement d**r** along this surface is perpendicular to *all* the local constraint surface gradients, so the right side of (4.8) is 0. At a solution point where equation (4.2) is true, we then have $\nabla f \bullet d\mathbf{r} = 0$ so a small displacement d**r** in any constraint-legal direction results in df = 0, hence we are at an extremum.