Lag Mult Gradient Section

4. The Gradient Interpretation

Recall that when 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 derivative 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

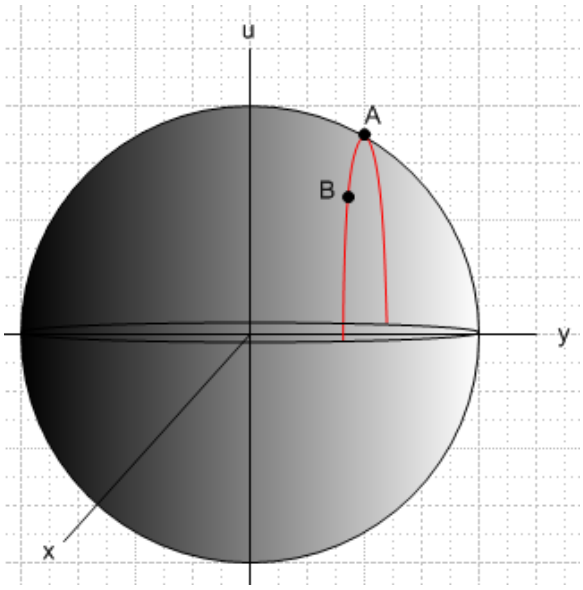
$$\nabla f(\mathbf{r}) + \lambda_1\nabla a(\mathbf{r}) + \lambda_2\,\nabla b(\mathbf{r}) + ... + \lambda_{S-1}\,\nabla q(\mathbf{r}) = 0 \qquad \nabla = (\partial_1,\partial_2...\partial_N) \qquad (4.2)$$

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: Let u = f(x,y) represent the surface of a sphere (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 normal "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.

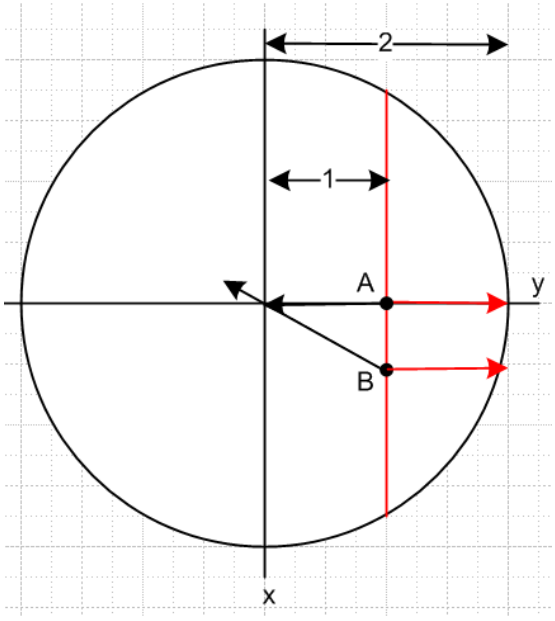
We now add a constraint a(x,y) = 0 where a(x,y) = y-1. This constraint is the line y=1 in the x-y plane E^2. We can extrude this line into a *plane* y = 1 in E^3. The hemispherical surface is a 2D surface in E^3, and the extruded constraint is also a 2D surface in E^3. These 2D surfaces intersect in 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. So view these gradients, we draw a top view of the sphere, looking toward the sphere center from the positive u axis:



(4.4)

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

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

Equation (4.2) says that at the solution point **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.

In this example the equation of the sphere surface is $u^2+x^2+y^2 = R^2$ so $u = f(x,y) = \sqrt{2 - x^2 - y^2}$. which in polar coordinates is $f(r,\theta) = \sqrt{2-r^2}$. The 2D gradient is then $\nabla f = [-r/\sqrt{2-r^2}\,]\,\hat{\mathbf{r}}$ and thus points toward the sphere's vertical axis as claimed. Also as claimed, at the north pole $\nabla f = 0$. At point A one finds $\nabla f = [-1/\sqrt{2-1^2}\,]\,\hat{\mathbf{r}} = [-1]\hat{\mathbf{r}}$. Since $\nabla a = [1]\,\hat{\mathbf{r}}$ at point A, we write ** as

$$[-1]\hat{\mathbf{r}} + \lambda_1[1]\,\hat{\mathbf{r}} = 0 \qquad \Rightarrow \qquad \lambda_1 = 1 \qquad (4.6)$$

At the solution point the Lagrange multiplier λ_1 can thus be interpreted as the negative of the ratio of the two gradient vectors ∇f and ∇a.

Now look back at Fig (4.4). At point B, since ∇f has some component *along* the constraint path, we know that f increases if we move along that path toward point A. But at point A, ∇f is normal to the constraint path and thus has no component along the path, so there is then no advantage to going away from point A in terms of increasing f. So at point A, we don't have the $\nabla f = 0$ (a normal critical point without constraints), but we *do* have $\nabla f \bullet d\mathbf{r} = 0$ where dr is a differential motion along the red intersection curve.

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 a $d\mathbf{r}$ along this curve, that $d\mathbf{r}$ will be perpendicular to both ∇a and ∇b, which says

$$(\lambda_1 \nabla a + \lambda_2 \nabla b)\bullet d\mathbf{r} = 0 \qquad (4.7)$$

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

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

Since this $d\mathbf{r}$ 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,

$$\nabla f \bullet d\mathbf{r} = -(\lambda_1 \nabla a + \lambda_2 \nabla b + \ldots + \lambda_{s-1} \nabla q) \bullet d\mathbf{r} \qquad (4.8)$$

In general the intersection surface of f and all the constraint surfaces will be a surface of dimension N-S+1 in E^n. A tiny displacement $d\mathbf{r}$ 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 $d\mathbf{r}$ in any constraint-legal direction results in $df = 0$, hence we are at an extremum.