COMPOSITION, GRADIENT, AND EXTREMA

PAUL L. BAILEY

1. Composition of Functions

Let A, B, and C be sets and let $f : A \to B$ and $g : B \to C$. The *composition* of f and g is the function

$$g \circ f : A \to C$$

given by

$$g \circ f(a) = g(f(a)).$$

If f and g are injective, then $g \circ f$ is injective. If f and g are surjective, then $g \circ f$ is surjective.

The domain of $g \circ f$ is A and the codomain is C. The range of $g \circ f$ is the image under g of the image under f of the domain of f.

Example 1. Let A be the set of living things on earth, B the set of species, and C be the set of positive real numbers. Let $f : A \to B$ assign to each living thing its species, and let $g : B \to C$ assign to each species its average mass. Then $g \circ f$ guesses the mass of a living thing.

Example 2. Let $\nu : \mathcal{A}^n \to \mathcal{V}^n$ be the function that assigns an arrow to the vector it represents. Let $\phi : \mathcal{V}^n \to \mathbb{R}^3$ be a coordinatization of the set of n-dimensional vectors. Then $\phi \circ \nu : \mathcal{A}^n \to \mathbb{R}^n$ sends each arrow to a coordinatized point.

Example 3. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$ and let $g : \mathbb{R} \to \mathbb{R}$ be given by $g(x) = \sin x$. Then $g \circ f : \mathbb{R} \to \mathbb{R}$ is given by $g \circ f(x) = \sin x^2$ and $f \circ g : \mathbb{R} \to \mathbb{R}$ is given by $f \circ g(x) = \sin^2 x$.

Example 4. Let $\vec{r} : \mathbb{R} \to \mathbb{R}^2$ be given by $\vec{r}(t) : \langle 2\cos t, \sin t \rangle$. The image of $\vec{r}$ is an ellipse in the plane. Let $\vec{s} : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $\vec{s}(x, y) = \langle x, y, y^2 - x^2 \rangle$. The image of $\vec{s}$ is a saddle surface.

Then the image of $\vec{s} \circ \vec{r}$ is a curve in $\mathbb{R}^3$ whose shape is roughly the boundary of a potato chip.

We may think of the ellipse as a road on a plane. Then think of $\vec{s}$ as an earthquake which takes the plane and shifts it, warping its shape into a saddle. The road is carried along with the plane as it warps. The new position of the road is the image of the composition of the functions.

2. Chain Rule

Let $f : \mathbb{R}^2 \to \mathbb{R}$ and let $\vec{r} : \mathbb{R} \to \mathbb{R}^2$. Then $f \circ \vec{r} : \mathbb{R} \to \mathbb{R}$ is a real domained real valued function. Geometrically, we view the image of $\vec{r}$ as a curve on the plane and we view the graph of f as a surface in $\mathbb{R}^3$. If we vertically project the curve onto the surface, we obtain a curve in $\mathbb{R}^3$. The height of this curve is given by the

Date: July 19, 1995.

function $f \circ \vec{r}$. We may interpret the derivative of $f \circ \vec{r}$ to be the rate of change in the height of this curve.

The derivative of $\vec{r}$ is a tangent vector in $\mathbb{R}^2$. Let f vertically project this tangent vector onto the plane tangent to the surface over the point $\vec{r}(t_0)$ for some $t_0 \in \mathbb{R}$. This vertically projected vector is tangent to the vertically projected curve, and its z component is the rate of change in the height of the curve.

We know from our study of differentials that the z component is related to the x and y components. Suppose that the coordinate functions of $\vec{r}$ are $x(t)$ and $y(t)$, that is, $\vec{r}(t) = \langle x(t), y(t) \rangle$. Since the vertically projected vector's tail sits on the point $(x(t), y(t), f(x(t), y(t)))$, we think of the x component of the vertically projected vector as Δx, a change in x, and the y component of the vertically projected vector as Δy, a change in y. Then the z component is Δz, the change in z on the tangent plane, which is given by

$$dz = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y.$$

By our manner of taking the derivative of a path, we know that the components of a tangent vector are the derivatives of the component functions. Thus $\Delta x = \frac{dx}{dt}$ and $\Delta y = \frac{dy}{dt}$.

Thus we have

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

Example 5. Let $z = 9 - x^2 - y^2$ and let $\vec{r}(t) = \langle 2\cos t, \sin t \rangle$. Find $\frac{dz}{dt}$.

Solution #1. The partials for the paraboloid are $\frac{\partial z}{\partial x} = -2x$ and $\frac{\partial z}{\partial y} = -2y$. The derivatives of the component functions are $\frac{dx}{dt} = -2\sin t$ and $\frac{dy}{dt} = \cos t$. Thus

$$\begin{aligned}\frac{dz}{dt} &= (-2x)(-2\sin t) + (-2y)(\cos t)\\ &= 8\sin t\cos t - 2\sin t\cos t\\ &= 6\sin t\cos t = 3\sin 2t.\end{aligned}$$

□

Solution #2. We may explicitly write z as a function of t by substitution: $z = 9 - 4\cos^2 t - \sin^2 t$. Then $\frac{dz}{dt} = 8\cos t\sin t - 2\sin t\cos t$. □

Now consider functions $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}$. Then f is some transformation of the plane, and the graph of g is a surface. The lines $x = k$ and $y = k$, for some constant k, are sent to curves by f. Thus we may apply to above reasoning to partially visualize the chain rule in this case. Since $g \circ f : \mathbb{R}^2 \to \mathbb{R}$, we obtain the partial derivatives.

Let u and v be the coordinates in the domain of f and let x and y be the coordinates in the domain of g. Let $w = g \circ f$. Then x, y, and w are functions of u and v and a point in the image of w may be written $w(x(u,v), y(u,v))$. Since $w : \mathbb{R}^2 \to \mathbb{R}$, we obtain the partial derivatives

$$\frac{\partial w}{\partial u} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u}$$

and

$$\frac{\partial w}{\partial v} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial v}.$$

Example 6. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = z = y^2 - x^2$. Find the partials of f with respect to polar coordinates.

Solution. Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(r, \theta) = \langle r\cos\theta, r\sin\theta\rangle$. This function is the polar coordinate transformation. Then

$$\begin{aligned}\frac{\partial z}{\partial r} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} \\ &= (-2x)(\cos\theta) + (2y)(\sin\theta) \\ &= (-2r\cos\theta)(\cos\theta) + (2r\sin\theta)(\sin\theta) \\ &= -2r\cos^2\theta + 2r\sin^2\theta = -2r\cos 2\theta.\end{aligned}$$

Note that for $\theta = \pi/2$, this partial is zero. This corresponds to the radial lines at this angle on the saddle surface.

Also

$$\begin{aligned}\frac{\partial z}{\partial \theta} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \theta} \\ &= (-2x)(-r\sin\theta) + (2y)(r\cos\theta) \\ &= (-2r\cos\theta)(-r\sin\theta) + (2r\sin\theta)(r\cos\theta) \\ &= 4r\cos\theta\sin\theta = 2r\sin 2\theta.\end{aligned}$$

Note that this is zero whenever θ is a multiple of π; this corresponds to the extrema on concentric circles. □

Lastly, consider functions $f : \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$. Then $g \circ f : \mathbb{R}^m \to \mathbb{R}$. Label the coordinates of $\mathbb{R}^m$ by $t_1, \ldots, t_m$ and the coordinates of $\mathbb{R}^n$ by $x_1, \ldots, x_n$. When we consider $g \circ f$, each coordinate of $\mathbb{R}^n$ is a function of the coordinates of $\mathbb{R}^m$. Let $w = g \circ f$. Thus we may write a point in the image of w as

$$w(x_1(t_1, \ldots, t_m), \ldots, x_n(t_1, \ldots, t_m)).$$

The chain rule in this case is

$$\frac{\partial w}{\partial t_i} = \sum_{j=1}^{n} \frac{\partial w}{\partial x_j}\frac{\partial x_j}{\partial t_i}.$$

3. Tangent Vectors of Projected Curves

Let us recapitulate. Draw a corresponding picture using your favorite curves and surfaces as you read this.

Let $\vec{r} : \mathbb{R} \to \mathbb{R}^2$. The image of $\vec{r}$ is a curve on a plane. Let the component functions of $\vec{r}$ be $x(t)$ and $y(t)$ so that $\vec{r}(t) = \langle x(t), y(t)\rangle$. Let $f : \mathbb{R}^2 \to \mathbb{R}$. The graph of f is a surface. Then the composition $z = f \circ \vec{r} : \mathbb{R} \to \mathbb{R}$ is a real domained, real valued function which represents the height of the curve $\vec{c}(t) = \langle x(t), y(t), f(x(t), y(t))\rangle$.

Let $\vec{s} : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $\vec{s}(x, y) = \langle x, y, f(x, y)\rangle$, so that the image of $\vec{s}$ is the same surface as the graph of f. Note that $\vec{c} = \vec{s} \circ \vec{r}$. The image of $\vec{c}$ is the vertical projection of the curve which is the image of $\vec{r}$ onto the surface which is the image of $\vec{s}$.

Let $t_0 \in \mathbb{R}$ be a point in the domain of $\vec{r}$. Then $\vec{r}(t_0) = (x_0, y_0) \in \mathbb{R}^2$ is a point in the domain of f.

The velocity vector $\vec{c}\,'(t_0)$ is the vertical projection of the velocity vector $\vec{r}\,'(t_0)$ onto the tangent plane of the surface over the point (x_0, y_0). The z component of $\vec{c}\,'(t_0)$ is $z'(t_0)$, which is the rate of change of the height of the curve given by $\vec{c}$ at time t_0, and is given by the chain rule as

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dz}{dy}.$$

Note that the tangent plane is the graph of an injective function $\vec{p} : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$\vec{p}(x, y) = \langle x, y, z_0 + \frac{\partial z}{\partial x}(x - x_0) + \frac{\partial z}{\partial y}(y - y_0)\rangle,$$

where $z_0 = f(x_0, y_0)$. Thus there is a one to one correspondence between the vectors emanating from (x_0, y_0) and the tangent vectors on the tangent plane.

Thus every vector on the tangent plane arises as the velocity vector of a vertically projected curve. We now find such a curve explicitly.

Let $\vec{w} = \langle w_1, w_2, w_3\rangle$ be an arbitrary vector on the tangent plane, which we think of as emanating from the point (x_0, y_0, z_0). Then $\vec{v} = \langle w_1, w_2\rangle$ is a vector in the domain of f emanating from the point (x_0, y_0). Let $\vec{r}(t) = \langle x_0 + w_1 t + y_0 + w_2 t\rangle$ be a line in the domain of f which passes through the point $(x_0, y_0$. Note that $\vec{r}(0) = (x_0, y_0)$ and that the velocity of this line in $\vec{v}$. Let $\vec{c}(t) = \vec{s} \circ \vec{r}(t)$ be the line vertically projected onto the surface. The velocity vector of $\vec{c}$ at $t = 0$ is $\vec{w}$, because w_3 is determined by w_1, w_2, and f by the differential formula

$$w_3 = \frac{\partial f}{\partial x} w_1 + \frac{\partial f}{\partial y} w_2.$$

Now suppose that $\vec{v}$ above is a unit vector. We wish to see that w_3 represents the rate of change of f in the direction $\vec{v}$.

4. Directional Derivatives

The partial derviatives of a function $f : \mathbb{R}^n \to \mathbb{R}$ represent the rate of change of the function in the direction of one of the coordinates axes. We may wish to find the rate of change in any direction, given by a unit vector $\vec{u}$.

Let $P \in \mathbb{R}^n$ be a point in the domain. A point near P in the direction of $\vec{u}$ is $P + h\vec{u}$, where h is a small positive real number. The average rate of change of the function f between P and $P + h\vec{u}$ is $\frac{f(P+h\vec{u})-f(P)}{h}$. Thus we define the *directional derivative* of f at the point P in the direction $\vec{u}$ to be

$$D_{\vec{u}} f(P) = \lim_{h\to 0} \frac{f(P + h\vec{u}) - f(P)}{h}.$$

The directional derivative is the instantaneous rate of change of the function f in the direction $\vec{u}$ at the point P.

To compute directional derivatives, we use the following proposition, which is readily generalized to functions with higher dimensional domains.

Proposition 1. *Let $D \subset \mathbb{R}^2$ and let $f : D \to \mathbb{R}$. Let $\vec{u} = \langle u_1, u_2 \rangle$. Then*

$$D_{\vec{u}} f = \frac{\partial f}{\partial x} u_1 + \frac{\partial f}{\partial y} u_2.$$

Proof. Let $P_0 = (x_0, y_0) \in D$ and let $\vec{r}(h) = P_0 + h\vec{u}$. Then $\vec{r}(h) = \langle x(h), y(h) \rangle$ where $x(h) = x_0 + hu_1$ and $y(h) = y_0 + hu_2$. Thus

$$\begin{aligned}
D_{\vec{u}} f(x_0, y_0) &= \lim_{h\to 0} \frac{f(P_0 + h\vec{u}) - f(P_0)}{h} && \text{by def of dir deriv} \\
&= \lim_{h\to 0} \frac{f(\vec{r}(h)) - f(\vec{r}(0))}{h} \\
&= \frac{d}{dh}(f \circ \vec{r})(0) && \text{by def of reg deriv} \\
&= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} && \text{by Chain Rule} \\
&= \frac{\partial f}{\partial x} u_1 + \frac{\partial f}{\partial y} u_2.
\end{aligned}$$

□

Viewing the above proof geometrically, we note that the graph of f is a surface in $\mathbb{R}^3$. The line $\vec{r}$ vertically projects onto this surface. The last component of the vertically projected curve is its height. The directional derivative is the last component of the tangent vector to this curve.

From this, we see that the directional derivative is linear in both positions.

Corollary 1. *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ and let $a \in \mathbb{R}$ be a constant. Let $\vec{u}, \vec{v} \in \mathcal{R}^n$ be unit vectors. Then*

1. $D_{\vec{u}}(f + g) = D_{\vec{u}} f + D_{\vec{u}} g$;
2. $D_{\vec{u}} af = aD_{\vec{u}} f$;
3. $D_{\vec{u}+\vec{v}} f = D_{\vec{u}} f + D_{\vec{v}} f$;
4. $D_{a\vec{u}} f = aD_{\vec{u}} f$.

Note that the directional derivative is a function. For a fixed $f : \mathbb{R}^n \to \mathbb{R}$ and $\vec{u} \in \mathcal{R}^n$,

$$D_{\vec{u}} f : \mathbb{R}^n \to \mathbb{R}.$$

Example 7. Let $f(x, y) = x^2 + y^2 + 4$ and let $\vec{v} = \langle 1, 2\rangle$. Find the directional derivative of f in the direction of $\vec{v}$ at the point $(1, 1)$.

Solution. First note that $|\vec{v}| = \sqrt{5}$, so we must divide by the $\sqrt{5}$. But since the directional derivative is linear, we may divide $\vec{v}$ first to obtain a unit vector or we may divide the result.

For clarity, let $\vec{u} = \frac{\vec{u}}{\sqrt{5}}$. Then $D_{\vec{u}}f = -2x\frac{1}{\sqrt{5}} - 2y\frac{2}{\sqrt{5}}$ so $D_{\vec{u}}f(1, 1) = \frac{-6}{\sqrt{5}}$. □

5. Gradient Vector

Let $f : \mathbb{R}^n \to \mathbb{R}$ and define the *gradient* of f to be

$$\nabla f = \langle \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\rangle.$$

We see that

$$D_{\vec{u}}f = \nabla f \cdot \vec{u}.$$

This is just a restatement of our above proposition in a more compact notation.

Note that the gradient is a function, just as a derivative is a function. Each point in the domain of f has its own gradient vector. Thus

$$\nabla f : \mathbb{R}^n \to \mathcal{R}^n.$$

Such a function is called a *vector field*.

Now we would like to find the direction in which f is increasing the most rapidly. We consider a sphere's worth of unit vectors emanating from a point in the domain of f, and find the one whose directional derivative is the greatest.

Thus let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\vec{u} \in \mathcal{R}^n$. Then at every point in the domain we have

$$D_{\vec{u}}f = \nabla f \cdot \vec{u} = |\nabla f||\vec{u}| \cos\theta = |\nabla f| \cos\theta,$$

where $\theta = \angle(\nabla f, \vec{u}) \in [0, \pi]$. Then the maximum value of $D_{\vec{u}}f$ is attained when $\theta = 0$; that is, when $\vec{u}$ is parallel to ∇f. Therefore the gradient points in the direction of maximum increase of the function f.

The maximum rate of change is the rate of change in the direction of the gradient vector. To see this, we unitize the gradient vector, letting

$$\vec{u} = \frac{\nabla f}{|\nabla f|},$$

and find that

$$D_{\vec{u}}f = \nabla f \cdot \vec{u} = \nabla f \cdot \frac{\nabla f}{|\nabla f|} = \frac{\nabla f \cdot \nabla f}{|\nabla f|} = \frac{|\nabla f|^2}{|\nabla f|} = |\nabla f|.$$

Now suppose that $f : \mathbb{R}^3 \to \mathbb{R}$. Now let $P_0 = (x_0, y_0, z_0) \in \mathbb{R}^3$ be a point in the domain and let $K = f(x_0, y_0, z_0)$. Consider the level surface of f at K. Let $\vec{r} : \mathbb{R} \to \mathbb{R}^3$ represent a curve which lies on the level surface and passes through the point P_0 at time t_0. The velocity vector of $\vec{r}(t_0)$ is tangent to the surface.

Let $x(t)$, $y(t)$, and $z(t)$ be the coordinate functions of $\vec{r}$ so that $\vec{r}(t) = \langle x(t), y(t), z(t)\rangle$. Now for every $t \in \mathbb{R}$, since the curve lies on the surface, we have that

$$f(x(t), y(t), z(t)) = K.$$

Differentiating both sides of this yields, by the chain rule,

$$\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = 0,$$

which may be rewritten as

$$\nabla f \cdot \vec{r}' = 0;$$

this says that the gradient is perpendicular to the velocity vector.

Every vector tangent to the surface may be represented by such a curve. Thus the gradient is perpendicular to the tangent plane of the surface. We say that the gradient is the normal vector of the surface.

It is equally true that the gradient vector is normal to a level curve of a function from $\mathbb{R}^2$ into $\mathbb{R}$.

We now have a way of computing the tangent line of a level curve and the tangent plane of a level surface.

Example 8. Find an equation for the line which is tangent to the ellipse $\frac{x^2}{4} + \frac{y^2}{9} = 1$ at the point $P_0(1, \frac{3\sqrt{3}}{2})$.

Solution. The line is perpendicular to the gradient vector of a function for which the ellipse is a level curve. Let $f(x, y) = \frac{x^2}{4} + \frac{y^2}{9}$ so that the curve is the level surface at 1 of f. Then the line is the set of point P such that $(P - P_0) \cdot \nabla f = 0$.

Note that $\nabla f = \langle \frac{x}{2}, \frac{2y}{9} \rangle$. We evaluated the gradient at our point and get $\nabla f(P_0) = \langle \frac{1}{2}, \frac{\sqrt{3}}{3} \rangle$. Then this equation becomes $\frac{1}{2}(x - 1) + \frac{\sqrt{3}}{3}(y - \frac{9}{2}) = 0$, which rearranges to

$$y = -\frac{\sqrt{3}}{2} + 3 + \frac{\sqrt{3}}{3}.$$

□

Example 9. The ray $\vec{r}(t) = \langle t, 2t, 2t \rangle$, where $t \geq 0$, intersects the hyperboloid $x^2 + y^2 - z^2 = 4$ in a point. Find the component of velocity of the line in the direction normal to the surface at the point of intersection.

Solution. Lets find the point of intersection by substitution. The time t of intersection satisfies the equation $t^2 + 4t^2 - 4t^2 = 4$; thus $t = 2$ and $\vec{r}(1) = (2, 4, 4)$. The velocity of the line is always $\vec{v} = \langle 1, 2, 2 \rangle$. The speed of the line is always 9.

Let $f(x, y, z) = x^2 + y^2 - z^2$. Then the hyperboloid is a the level surface of f at $f = 1$. The vector normal to the the level surfaces of f is the gradient vector. We have that $\nabla f = \langle 2x, 2y, -2z \rangle$. Evaluating this at the point of intersection gives us the vector which is normal to the surface at this point; it is $\vec{n} = \langle 4, 8, -8 \rangle$.

The component of $\vec{v}$ in the direction of $\vec{n}$ is

$$\text{proj}_{\vec{n}}(\vec{v}) = \frac{\vec{n} \cdot \vec{v}}{|\vec{n}|} = \frac{8 + 32 - 32}{\sqrt{4}\sqrt{1 + 4 + 4}} = \frac{4}{3}.$$

□

There are some common misunderstandings about the gradient vector. We recap what we know, and expel some of these misunderstandings.

The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ IS:

- a vector which lies in the DOMAIN of a function;
- a vector which we may dot with a unit vector to obtain a directional derivative;
- a vector which points in the direction of maximum increase of the function;
- a vector whose length is the maximum rate of change of the function;
- a vector which "points into the hill" on the GRAPH of f;
- normal to the LEVEL SETS of the function.

The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ IS NOT:

- a vector which lies in the RANGE or in the DOMAIN $\times$ RANGE of the function f;
- a vector which is normal to the GRAPH of the function f;
- a vector which "points uphill" on the GRAPH of f.

6. Differentiability

Consider this section to be optional. You should read it, but we will not study functions which are not differentiable.

We are now in a position to define what it means for a function $f : \mathbb{R}^n \to \mathbb{R}$ to be differentiable at a point $P \in \mathbb{R}^n$. Recall that we intuitively wish this to mean that the graph of the function is approximated by the graph of a linear function (such as a line if $n = 1$ or a plane if $n = 2$) near the point P, and that this approximation should become better and better the closer we get to P.

First, we generalize our terminology by saying that a subset $D \subset \mathbb{R}^n$ is a *hyperplane* if D is a "flat" set in $\mathbb{R}^n$ of dimension $n - 1$. Such a set is the span of $n - 1$ vectors whose tails are at $P \in D$.

Thus let $f : \mathbb{R}^n \to \mathbb{R}$ and let $P \in \mathbb{R}^n$. Suppose that all of the partial derivatives of f exist at P, so that ∇f exists at P. We define the *linear approximation* of f at P to be the function $T_P : \mathbb{R}^n \to \mathbb{R}$ which is given by

$$T_P(Q) = f(P) + \nabla f(P) \cdot (Q - P);$$

the graph of this function is called the *tangent hyperplane* of f at P.

Consider the case $n = 2$, $z = f$, $P = (x_0, y_0)$, $Q = (x, y)$, and $f(x_0, y_0) = z_0$. Let $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ be the partial derivatives of f evaluated at (x_0, y_0), so that $\nabla f(P) = \langle \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y} \rangle$. Thus in this case

$$\begin{aligned} T_P(Q) &= f(P) + \nabla f(P) \cdot (Q - P) \\ &= z_0 + \langle \partial z \partial x, \frac{\partial z}{\partial y} \rangle \cdot (x - x_0, y - y_0) \\ &= z_0 + \frac{\partial z}{\partial x}(x - x_0) + \frac{\partial z}{\partial y}(y - y_0), \end{aligned}$$

which is the equation of the tangent plane which we have previously seen.

We say that $f : \mathbb{R}^n \to \mathbb{R}$ is *differentiable* at $P \in \mathbb{R}^n$ if all the partials of f exist at P (so that ∇f exists at P and T_P exists) and also

$$\lim_{Q \to P} \frac{|f(Q) - T_P(Q)|}{|Q - P|} = 0;$$

that is, not only does $T_P(Q)$ approach $f(P)$ as P approaches Q, but it does so faster than the approach.

We may restate this in terms of differentials by saying that $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at (x_0, y_0) if

$$\lim_{(\Delta x, \Delta y) \to (0,0)} \frac{\Delta z - dz}{|\langle \Delta x, \Delta y \rangle|} = 0.$$

One may ask why we don't simply require all the partials to exist, or that $T_P(Q)$ approach $f(P)$ as Q approaches P. The answer is that these definitions are insufficient to guarantee that the locus of T_P is a unique hyperplane, or to give us the following proposition, which is true under our definition:

Proposition 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be differentiable. Then $g \circ f$ is differentiable.*

The idea of differentiability extends to functions $f : \mathbb{R}^m \to \mathbb{R}^n$; however, in this case, the map T_P becomes a *linear transformation* of multiple dimensions which is described by a matrix of partial derivatives. Understanding this in any detail requires linear algebra.

7. Extrema

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Let $P \in \mathbb{R}^n$ be a point in the domain of f. Since the range $\mathbb{R}$ is an ordered set (unlike $\mathbb{R}^n$ itself), we may compare $f(P)$ to $f(Q)$ for points Q near P. If $f(P)$ is the largest value attained by f in some neighborhood of P, we say that P is a *local maximum* of f, and that $f(P)$ is a *local maximum value* of f. In symbols, P is a local maximum if there exists some real number $\delta > 0$ such that for every Q with $|Q - P| < \delta$, we have $f(Q) \le f(P)$.

Similarly, P is a *local minimum* if $f(Q)$ is the smallest value attained by f in some neighborhood of P. If P is either a local maximum or a local minimum, we say that P is a *local extremum.*

Example 10. The function of $f(x, y) = (4 - x^2 - y^2)^2$ has a local maximum at the origin; the local maximum value there is 16.

If $f(P) \ge f(Q)$ for every $Q \in D$, we say that P is a *global maximum* of f; if $f(P) \le f(Q)$ for every $Q \in D$, we say that P is a *global minimum.*

It is possible that global extrema do not exist. For example, a plane has no local nor global extrema.

A set $D \subset \mathbb{R}^n$ is called *compact* if it is closed and bounded. Here, closed means that the set contains all of its boundary points and bounded means that it may be contained in a sphere of finite radius.

Proposition 3. *Let $D \subset \mathbb{R}^n$ be a compact set and let $f : D \to \mathbb{R}$ be continuous. Then f attains a global maximum value and a global minimum value somewhere on D.*

Proof. We suppose that D is “connected”, that is, it is all in one piece. The gist of the proof is that a continuous function maps compact sets to compact sets and connected sets to connected sets. Thus the image of f is some closed interval of real numbers. Then the minimum of the function is the lower endpoint of the this interval. □

8. Critical Points

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable and suppose that $P \in D$. We say that P is a *critical point* of f if the gradient vector at P is the zero vector, that is, if the partial derivatives are all equal to zero at P.

First suppose that $f : \mathbb{R} \to \mathbb{R}$ and that $f'(x_0) = 0$ for some $x_0 \in \mathbb{R}$. For simplicity, we assume that f' is continuous near x_0. Thus on either side of x_0, there is a small interval where f' is either positive, negative, or zero.

If the sign of the derivative is the same on either side, then we have a local extremum; the second derivative detects whether it is a minimum or a maximum. If the second derivative is positive, then the derivative is increasing, so function is sloping less and less downward as we approach x_0, and is sloping more and more upward as we leav x_0. This is a local minimum. A negative second derivative detects a local maximum.

It is possible that the signs differ on either side, such as is the case for the function $f(x) = x^3$ at $x_0 = 0$. In this case, the second derivative is also zero.

It is also possible that the derivative is zero on one side of x_0 but nonzero on the other. For example, this occurs at $x_0 = 0$ with the function

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } x \geq 0 \end{cases}$$

In this case, the second derivative is discontinuous and undefined at $x_0 = 0$.

Lastly, the derivative may be zero everywhere in a neighborhood of x_0, in which case we say that the function is *locally constant.*

Now let us examine what the partial derivatives being zero implies for function $f : \mathbb{R}^2 \to \mathbb{R}$; the ideas are generalizable to higher dimensions.

Thus let $f : \mathbb{R}^2 \to \mathbb{R}$ and let $P_0 = (x_0, y_0)$ be a critical point. Define $g(x) = f(x, y_0)$ and $h(y) = f(x_0, y)$. Then we have $g' = f_x$ and $h' = f_y$. Now if (x_0, y_0) is a local extremum of f, it is clear that x is a local extremum of g and y is a local extremum of h.

On the other hand, suppose that x_0 is a local extremum of g and y_0 is a local extremum of h. It turns out that, if these extrema are of the same type (either both maxima or both minima), this implies that (x_0, y_0) is a local extremum of f.

To see this, we consider the second order directional derivative. We assume here that the second partials of f are continuous. Let $\vec{u} = \langle u_1, u_2 \rangle$ be a unit vector. Then $D_{\vec{u}} f = f_x u_1 + f_y u_2$. Note that $D_{\vec{u}} f$ is a real valued function of two variables. Thus we may apply the operator $D_{\vec{u}}$ a second time:

$$\begin{aligned} D^2_{\vec{u}} f &= D_{\vec{u}}(f_x u_1 + f_y u_1) \\ &= D_{\vec{u}}(f_x)u_1 + D_{\vec{u}}(f_y)u_1 \qquad \text{because } D_{\vec{u}} \text{ is linear} \\ &= (f_{xx}u_1 + f_{xy}u_2)u_1 + (f_{yx}u_1 + f_{yy}u_2)u_2 \\ &= f_{xx}u_1^2 + 2f_{xy}u_1u_2 + f_{yy}u_2^2 \qquad \text{because } f_{xy} = f_{yx} \end{aligned}$$

We are interested in whether or not $D^2_{\vec{u}}$ has the same sign for all unit vectors $\vec{u}$, because this operator $D^2_{\vec{u}}$ tells us if the slice of the surface in the direction of $\vec{u}$ is concave up or concave down. If the sign of the second order directional derivative is the same in all directions, this will tell us that the critical point is a local extremum.

Now suppose that $f_{xx}(x_0, y_0) \neq 0$ and $f_{yy}(x_0, y_0) \neq 0$, as happens when g and h have local extrema at (x_0, y_0). We may think of our last expression of $D^2_{\vec{u}}$ as a polynomial in u_1. We wish to see if this polynomial has any roots. If it has no roots, then the $D^2_{\vec{u}}$ never changes sign as $\vec{u}$ rotates around the critical point. If it is always positive, the critical point is a *local minimum* of f. If it is always negative, the critical point is a *local maximum* of f.

If it has two roots, then the $D^2_{\vec{u}}$ detects a local minima in some directional slices and a local maximum in others. In this case, the critical point is called a *saddle point.* If it has exactly one root, it is called a *degenerate* critical point.

The roots of the polynomial are given by the quadratic formula as

$$\frac{-2f_{xy}u_2 \pm \sqrt{4f_{xy}^2 - 4f_{xx}f_{yy}u_2^2}}{2f_{xx}} = u_2\Big(\frac{-f_{xy} \pm \sqrt{f_{xy}^2 - f_{xx}f_{yy}}}{f_{xx}}\Big).$$

Now u_2 depends on u_1 and is never zero when u_1 is. Thus to find how many roots this polynomial has, it suffices to look at the discriminant; by tradition we let

$$\Delta = f_{xx}f_{yy} - f_{xy}^2.$$

Note that this is the negative of the value under the radical above.

If $\Delta > 0$, the polynomial has no real roots, because then the value under the radical is negative. In this case we have a local extremum. Whether such a quadratic polynomial is always positive or always negative depends on the sign of the coefficient in front of u_1^2, in this case, f_{xx}. If $f_{xx} > 0$, then the polynomial is always positive, which means that all slices are concave down, so we have a local minimum. If $f_{xx} < 0$, then we have a local maximum.

If $\Delta < 0$, the polynomial has two real roots. Some slices are concave up and some are concave down, so we have a saddle point.

If $\Delta = 0$, more complex things are happening, and we have a degenerate critical point.

If $f_{xx} = 0$ at our critical point, we may use the same test with f_{yy}. If this is also zero, the second derivative test fails. We cannot conclude that we have a degenerate critical point.

Example 11. Find and classify the critical points of $f(x, y) = x^2 + y^2$.

Solution. The partials are $f_x(x, y) = 2x$ and $f_y = 2y$. These are zero only at the origin. The second order partials are $f_{xx} = 2$, $f_{yy} = 2$, and $f_{xy} = 0$. Thus $\Delta = 4 > 0$, and $f_{xx} = 2 > 0$, so we have a local minimum. □

Example 12. Find and classify the critical points of $f(x, y) = x^2 - y^2$.

Solution. The partials are $f_x(x, y) = 2x$ and $f_y = -2y$. These are zero only at the origin. The second order partials are $f_{xx} = 2$, $f_{yy} = -2$, and $f_{xy} = 0$. Thus $\Delta = -4 > 0$, and we have a saddle point. □

Example 13. Find and classify the critical points of $f(x, y) = x^2 + y^3$.

Solution. The partials are $f_x(x, y) = 2x$ and $f_y = 3y^2$. These are zero only at the origin. The second order partials are $f_{xx} = 2$, $f_{yy} = 6y$, and $f_{xy} = 0$. Thus $\Delta(0, 0) = 12y \mid_{(0,0)} = 0$, and we have a degenerate critical point. □

Example 14. Find and classify the critical points of $f(x, y) = x^2 + y^3 + 3xy$.

Solution. The partials are $f_x(x, y) = 2x + 3y$ and $f_y = 3y^2 + 3x$. First find all solutions to $f_x = f_y = 0$. The first equations gives that $y = -\frac{2}{3}x$. Plugging this into the second gives that $\frac{4}{3}x^2 + 3x = 0$. Thus $x = 0$, in which case $y = 0$, or $x = -\frac{9}{4}$, in which case $y = \frac{3}{2}$. Thus our critical points are $(0, 0)$ and $(-\frac{9}{4}, \frac{3}{2})$.

The second order partials are $f_{xx} = 2$, $f_{yy} = 6y$, and $f_{xy} = 3$. Thus $\Delta(x, y) = 12y - 3$. So $\Delta(0, 0) = -3 < 0$, implying that the origin is a saddle point. Also $\Delta(-\frac{9}{4}, \frac{3}{2}) = 15 > 0$ and $f_{xx}(-\frac{9}{4}, \frac{3}{2}) = 2 > 0$, so $(-\frac{9}{4}, \frac{3}{2})$ is a local minimum with minimum value $f(-\frac{9}{4}, \frac{3}{2}) = -\frac{27}{16}$. □

9. Lagrange Multipliers

We now simultaneously consider two different problems, which initially seem unrelated. The first is geometric: we ask, "when are the level sets of to functions tangent?". The second is analytic: we ask, "how may a maximize or minimize a function when constrained by an extra condition?"

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. Suppose that we wish to minimize f under the condition that our minimum $P \in \mathbb{R}^n$ also has to satisfy an equation of the form $g(P) = K$ for some constant K. The condition $g(P) = K$ is called a *constraint.* This is the same as asking for the minimum value attained by f on the level set of g at K.

Note that no two distinct level surfaces of f intersect. We think of the level surfaces of f as moving through space as time passes. The value of f on the surface at time t is t; for smaller t, f has smaller values. Some of these level surfaces of f will intersect the set where $g = K$ and some will not. The first one that does will be the one with the smallest values of f, that is, the minimum value of f on the set where $g = K$. The surfaces will be tangent when they first touch. Thus there normal vectors will be parallel there. Since these are level surfaces, the normal vectors are given by the gradient vector. Thus $\nabla f \| \nabla g$ at the minimum point. Similar comments apply to maxima.

The condition $\nabla f \| \nabla g$ may be stated as $\nabla f = \lambda \nabla g$ for some real number λ. This equation gives us n equations in $n+1$ variables (the n coordinate variables plus λ). The constraint gives us another equation, so we search for solutions to this system of equations. This is called the *method of Lagrange multipliers.*

Example 15. Find the minimum and maximum values of $f(x, y) = x^2 + y^2$ subject to the constraint $x^2 - y^2 = 1$.

Solution. First view this geometrically. The level curves of f are circles, which we view as expanding with time. The minimum value of f on the hyperbola $x^2 - y^2 = 1$ is attained on a circle centered at the origin which is tangent to the hyperbola. This occurs for a circle of radius one, which is the level surface of f at 1. There is no maximum value.

Now we find this analytically via the method of Lagrange multipliers. Let $g(x, y) = x^2 - y^2$. Thus our hyperbola becomes the level surface of g at 1, and the equation $g(x, y) = 1$ is our constraint. Then $\nabla f = \langle 2x, 2y \rangle$ and $\nabla g = \langle 2x, -2y \rangle$. We solve the following system of equations:

(1) $2x = \lambda 2x$;
(2) $2y = -\lambda 2y$;
(3) $x^2 - y^2 = 1$.

Immediately from equation (1), either $x = 0$ or $\lambda = 1$. But $x = 0$ does not satisfy equation (3), so $\lambda = 1$. Then equation (2) gives that $y = 0$, so $x = 1$ and $y = 0$. Thus the minimum occurs at $(1, 0)$, as we suspected, and the minimum value of f on the hyperbola is $f(1, 0) = 1$. □

Example 16. Find all level surfaces of $f(x, y, z) = x^2 + (y - 1)^2 - z^2$ which are tangent to the sphere of radius 2 centered at the origin. In the process, find the minimum and maximum values attained by f on the sphere and the points where they occur.

Solution. We search for level surfaces of f which intersect the unit sphere in points where their normal vectors are parallel via the method of Lagrange Multipliers.

Let $g(x, y, z) = x^2 + y^2 + z^2$. Then our sphere is the level surface of g at 4. We have $\nabla f = \langle 2x, 2(y-1), -2z\rangle$ and $\nabla g = \langle 2x, 2y, 2z\rangle$. We solve the system of equations

(1) $2x = \lambda 2x$;
(2) $2(y-1) = \lambda 2y$;
(3) $-2z = \lambda 2z$;
(4) $x^2 + y^2 + z^2 = 4$.

Equation (1) gives that either $x = 0$ or $\lambda = 1$. If $\lambda = 1$, then equation (2) gives that $-2 = 0$, a contradiction. Thus $\lambda \neq 1$, and $x = 0$.

Equation (3) gives that either $z = 0$ or $\lambda = -1$. We test these cases separately.

Suppose that $z = 0$. Since x is also zero, we have that $y^2 = 4$ so $y = \pm 2$. Thus $(0, -2, 0)$ and $(0, 2, 0)$ at points where the sphere is tangent to some level surface of f. These surfaces are at $f(0, -2, 0) = 9$ and $f(0, 2, 0) = 1$ respectively.

Suppose that $\lambda = -1$. Then from equation (2), $2y - 2 = -2y$ so $y = \frac{1}{2}$. Plugging this and $x = 0$ into equation (4) gives that $z = \pm\frac{\sqrt{15}}{2}$. Thus the points $(0, \frac{1}{2}, \pm\frac{\sqrt{15}}{2})$ at points of tangency between the sphere and the level surface of f at $f(0, \frac{1}{2}, \pm\frac{\sqrt{15}}{2}) = -\frac{14}{4}$. This is the minimum value of f on the sphere, and 9 is the maximum value.

Geometrically, we know that the level surfaces of f at t are one sheeted hyperboloids for $t > 0$ and two sheeted hyperboloids for $t < 0$. The axis of symmetry of the hyperboloids intersects the sphere off center; thus there are two one sheeted hyperboloids which are tangent the the sphere and there is one such two sheeted hyperboloid. □

Department of Mathematics, University of California, Irvine
E-mail address: `pbailey@@math.uci.edu`