Gradient and directional derivatives

Math 251

October 8, 2015

1 The gradient vector

Definition 1.1. *Let $f(x, y)$ be a function that maps from $\mathbb{R}^2$ to $\mathbb{R}$. The gradient of f is defined as the vector*

$$\nabla f = \langle f_x, f_y \rangle.$$

If we have $f(x, y, z)$ as a function that maps from $\mathbb{R}^3$ to $\mathbb{R}$ then the gradient vector can similarly be defined as

$$\nabla f = \langle f_x, f_y, f_z \rangle.$$

Remark: We have seen that the equation of the tangent plane to a surface $z = f(x, y)$ at a point (a, b) is

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

That is the normal vector of the tangent plane at this point is

$$\langle f_x, f_y, -1 \rangle.$$

Thus the gradient vector can be seen geometrically as capturing the local behavior of the surface around the point.

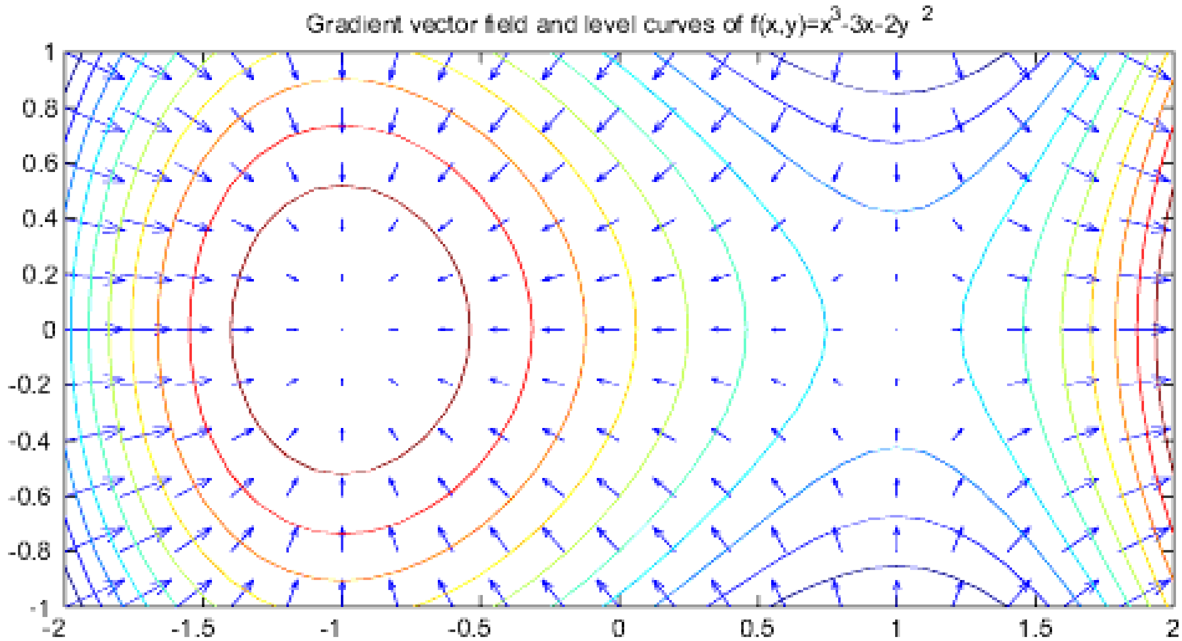

Figure 1.1: The gradient flow of a function

Before we develop further property of the gradient vector, we mention some basic calculus rules:

$$\begin{aligned} \nabla(f+g) &= \nabla f + \nabla g \\ \nabla(cf) &= c\nabla f, c \text{ a constant} \\ \nabla(fg) &= f\nabla g + g\nabla f. \end{aligned}$$

1.1 The chain rules

There are many different ways one can compose functions of the type we have discussed up to now. First if we have a function $F : \mathbb{R} \to \mathbb{R}$ and $f : \mathbb{R}^n \to \mathbb{R}, n = 2, 3$ then $F \circ f$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$. Thus the gradient of $F \circ f$ makes sense and we have

$$\nabla(F \circ f) = F'\nabla f.$$

Example: Let $F(z) = \sqrt{z}$ and $f(x, y) = x^2 + y^2$ then

$$F(f(x, y)) = \sqrt{x^2 + y^2}$$

and

$$\nabla F(f(x,y)) = \Big\langle \frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}} \Big\rangle.$$

Second, we can have $f(x,y,z) : \mathbb{R}^3 \to \mathbb{R}$ and we can imagine evaluating f along a curve. That is (x,y,z) are functions of t : $\big\langle x(t), y(t), z(t) \big\rangle$. In this case $f(x(t), y(t), z(t))$ is a function from $\mathbb{R}$ to $\mathbb{R}$ and we have

$$\frac{d}{dt} f(x(t), y(t), z(t)) = \nabla f \cdot \big\langle x'(t), y'(t), z'(t) \big\rangle.$$

The above equation can be written compactly as

$$\frac{d}{dt} f(\mathbf{r}(t)) = \nabla f \cdot \mathbf{r}'(t).$$

<u>Example:</u> Let $f(x,y) = \frac{y}{x}$ and $\big\langle x(t), y(t) \big\rangle = \big\langle \cos(t), \sin(t) \big\rangle$. Then

$$\begin{aligned} f'(\mathbf{r}(t)) &= \Big\langle \frac{-y}{x^2}, \frac{1}{x} \Big\rangle \Big|_{(x,y)=(\cos t, \sin t)} \cdot \big\langle -\sin t, \cos t \big\rangle \\ &= \frac{\sin^2 t}{\cos^2 t} + 1 = \sec^2 t. \end{aligned}$$

The last chain rule we discuss is the generalization of the second case mentioned above. We also have a function $f(x,y,z) : \mathbb{R}^3 \to \mathbb{R}$. But this time we do not simply assume x, y, z are uni-variate function (function of 1 variable t). They themselves can be surfaces, that is x, y, z can be functions of $t_1, t_2, \cdots, t_n$. The composition

$$f(x(t_1, \cdots, t_n), y(t_1, \cdots, t_n), z(t_1, \cdots, t_n))$$

becomes a function from $\mathbb{R}^n$ to $\mathbb{R}$. Thus it also makes sense to dicuss the gradient of such a function. The result is

$$\nabla f(x(t_1, \cdots, t_n), y(t_1, \cdots, t_n), z(t_1, \cdots, t_n)) = f_x \nabla x + f_y \nabla y + f_z \nabla z.$$

It may be clearer to state the above formula in terms of the partial derivative of f with respect to each variable $t_1, \cdots t_n$:

$$f_{t_k} = f_x x_{t_k} + f_y y_{t_k} + f_z z_{t_k}.$$

1.2 Implicit differentiation

Suppose we have a surface that is described by an equation $F(x, y, z) = 0$. The understanding is that there exists $z = f(x, y)$ a surface so that $F(x, y, f(x, y)) = 0$. The point is we can NOT find such a function f explicitly. Actually writing $z = f(x, y)$ is not precise because it can be the case that for a point (x, y) there are many values of z that satisfies $F(x, y, z) = 0$. The notation $z = f(x, y)$ is just to motivate the idea that if we limit ourselves to a particular part of the surface, then indeed we can view $z = f(x, y)$ and discuss the partials of z with respect to (x, y). The simplest example is a sphere: $x^2 + y^2 + z^2 = 1$. Here z is NOT a function of x, y but if we limit ourselves to the upper or lower part of the sphere then it is.

So our ultimate goal is to find the partials of z with respec to (x, y) at a point. The technique is to differentiate the equation $F(x, y, z) = 0$ with respect to x and keeping in mind:

$$\begin{aligned}
\frac{\partial}{\partial x}F(x, y, z) &= \frac{\partial}{\partial y}F(x, y, z) = 0 \\
\frac{\partial}{\partial x}y &= \frac{\partial}{\partial y}x = 0 \\
\frac{\partial}{\partial x}x &= \frac{\partial}{\partial y}y = 1.
\end{aligned}$$

Thus it follows that

$$\frac{\partial}{\partial x}F(x, y, z) = F_x + F_z z_x.$$

This gives you an equation to solve for z_x. Similarly

$$\frac{\partial}{\partial x}F(x, y, z) = F_y + F_z z_y$$

gives you an equation to solve for z_y.

Example: Consider the equation $x^2 + y^2 + z^2 = 2$. Differentiate both sides with respect to x gives

$$2x + 2zz_x = 0.$$

Thus $z_x = -x/z$, for $z \neq 0$. Thus the derivative of z with respect to x at $(1, 0, 1)$ is -1.

2 Directional derivative

Consider $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$. We can imagine f as described the "height" at a point (x, y) relative to ground zero (if f is negative we're just going below the ground). Now imagine we start at a point (x_0, y_0) at time $t = t_0$ and walk in a direction described by a vector $\mathbf{u} = \langle u_1, u_2 \rangle$. The interest is of course how fast the terrain changes as we walk. More specifically as the direction we follow is described as:

$$\mathbf{r}(t) = \langle x_0, y_0 \rangle + \mathbf{u}(t - t_0),$$

we are interested in measuring

$$\frac{d}{dt} f(\mathbf{r}(t)), t = t_0.$$

It is of pratical interest, for example to find the direction $\mathbf{u}$ that has the steepest rise (or the steepest descend); or for a smooth climb a direction that has relatively low rate of change. All of these information can be captured via the directional derivative vector. We discussed here the example for $f : \mathbb{R}^2 \to \mathbb{R}$ but you can think of a similar interpretation for $f : \mathbb{R}^3 \to \mathbb{R}$. We interpret in that case $f(x, y, z)$ as measure, for example the temperature at a point (x, y, z) and we are interested in finding out how fast the temperature is changing as we move along a direction $\mathbf{u}$ in space.

Just one remark before we give the definition: we would want $\mathbf{u}$ to be the **unit** vector. This is because clearly how fast or slow we go would affect how fast we see the terrain change. But this is NOT an intrinsic property of the terrain. Making $\mathbf{u}$ be the unit vector will guarantee that we only measure the rate of change of the terrain itself and not being influenced by how fast we walk.

Applying the chain rule we have

$$\frac{d}{dt} f(\mathbf{r}(t)) = \nabla f \cdot \mathbf{r}'(t) = \nabla f \cdot \mathbf{u}.$$

We call this the *directional derivative of f* along the direction of $\mathbf{u}$ and denote it as $\nabla_{\mathbf{u}} f$.

2.1 The gradient as the direction of steepest ascent

We now discuss the problem of finding the vector $\mathbf{u}$ that achieves the maximum rate of change (steepest ascent) and minimum rate of change (steepest descent). Recall that

$$\frac{d}{dt} f(\mathbf{r}(t)) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{u}.$$

We want to choose $\mathbf{u}$ to maximize $\nabla f(\mathbf{r}(t)) \cdot \mathbf{u}$. But we know that

$$\nabla f(\mathbf{r}(t)) \cdot \mathbf{u} = \|\nabla f(\mathbf{r}(t))\| \cos(\theta),$$

where θ is the angle between $\mathbf{u}$ and $\nabla f(\mathbf{r}(t))$. We have used the property that $\mathbf{u}$ is a unit vector here. But since $-1 \leq \cos(\theta) \leq 1$ clearly $\nabla f(\mathbf{r}(t)) \cdot \mathbf{u}$ is maximized when $\theta = 0$, that is when we move along $\nabla f(\mathbf{r}(t))$ and minimized when $\theta = \pi$, that is when we move in opposite direction of $\nabla f(\mathbf{r}(t))$. We say that the gradient is the direction of steepest ascent. Note that this result also tells us that the direction of smallest change (in terms of magnitude) is when $\theta = \pi/2$ that is in the orthogonal direction with $\nabla f(\mathbf{r}(t))$.