DIRECTIONAL DERIVATIVES AND GRADIENT VECTORS

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.6.

What students should definitely get: Definition of directional derivative and gradient vector, gradient vector as direction with maximum magnitude of directional derivative, directional derivative as dot product of gradient vector and unit vector in the direction. Formulas for tangent plane and normal line at a point to a surface with a relational description.

Executive summary

Words ...

(1) The directional derivative of a scalar function f of two variables along a *unit* vector $\mathbf{u} = a\mathbf{i} + b\mathbf{j}$ at a point (x_0, y_0) is defined as the following limit of difference quotient, if the limit exists:

$$\lim_{h \to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h}$$

(2) The directional derivative of a differentiable scalar function f of two variables along a *unit* vector $\mathbf{u} = a\mathbf{i} + b\mathbf{j}$ at a point (x_0, y_0) is $D_{\mathbf{u}}(f) = af_x(x_0, y_0) + bf_y(x_0, y_0)$.

(3) The gradient vector for a differentiable scalar function f of two variables at a point (x_0, y_0) is $\nabla f(x_0, y_0) = f_x(x_0, y_0)\mathbf{i} + f_y(x_0, y_0)\mathbf{j}$.

(4) The directional derivative of f is the dot product of the gradient vector of ∇f and the unit vector $\mathbf{u}$.

(5) Suppose ∇f is nonzero. Then, if $\mathbf{u}$ makes an angle θ with ∇f, then $D_{\mathbf{u}}(f)$ is $|\nabla_f| \cos \theta$. The directional derivative is maximum in the direction of the gradient vector, zero in directions orthogonal to the gradient vector, and minimum in the direction opposite to the gradient vector.

(6) The level curves are orthogonal to the gradient vector.

(7) Similar formulas for gradient vector and directional derivative work in three dimensions.

(8) The level surfaces are orthogonal to the gradient vector for a function of three variables.

(9) For a surface given by $F(x, y, z) = 0$, if (x_0, y_0, z_0) is a point on the surface, and $F_x(x_0, y_0, z_0)$, $F_y(x_0, y_0, z_0)$, and $F_z(x_0, y_0, z_0)$ all exist and are nonzero, then the normal line is given by:

$$\frac{x - x_0}{F_x(x_0, y_0, z_0)} = \frac{y - y_0}{F_y(x_0, y_0, z_0)} = \frac{z - z_0}{F_z(x_0, y_0, z_0)}$$

The tangent plane is given by:

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0$$

1. Directional derivatives: definition and key facts

1.1. **Partial derivatives as derivatives along coordinate directions.** The partial derivative $f_x(x, y)$ is defined as the derivative of f with respect to x, keeping y constant. Thinking of the domain of f geometrically, this is the same as the derivative of f along a unit vector in the x-direction (which we archaically denote by $\mathbf{i}$). Similarly, the partial derivative $f_y(x, y)$ is defined as the derivative of f with respect to y, keeping x constant. This can be viewed as the derivative of f with respect to a unit vector in the y-direction (which we archaically denote by $\mathbf{j}$).

We may be interested in derivatives with respect to mixed directions, i.e., we may be interested in the question: if we move along the direction of the vector $\mathbf{i} + \mathbf{j}$, how does the function value change?

The correct notion is that of *directional derivative*. After defining this notion, we consider its implication both for functions that have physical significance and for the more abstract functions in economics (such as production and demand functions).

1.2. **Definition of directional derivative.** For a function f, the directional derivative of f at (x_0, y_0) in the direction of a unit vector $\mathbf{u} = \langle a, b\rangle$ (i.e., $a^2 + b^2 = 1$) is denoted $D_{\mathbf{u}}f(x_0, y_0)$ and is defined as:

$$D_{\mathbf{u}}f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

if the limit exists.

The special case of $\mathbf{u} = \mathbf{i} = \langle 1, 0\rangle$ gives the partial derivative $f_x(x, y)$ and the case $\mathbf{u} = \mathbf{j} = \langle 0, 1\rangle$ gives the partial derivative $f_y(x, y)$.

1.3. **Existence and computation of directional derivative.** It turns out that there is a notion of "differentiable" for a function of two variables (which are are avoiding discussion of) and if a function is differentiable at a point in that sense, then it has well-defined directional derivatives along all unit vectors.

Further, then the directional derivative in the direction of $\mathbf{u} = \langle a, b\rangle$ is given by:

$$D_{\mathbf{u}}f(x_0, y_0) = af_x(x_0, y_0) + bf_y(x_0, y_0)$$

A *sufficient* condition for a function to be differentiable is that both the partial derivative f_x and f_y exist and are continuous *around* the point.

1.4. **Geometric sense of directional derivatives.** Directional derivatives make direct geometric sense in cases where the functions actually have physical significance. For instance, consider a function whose input domain is a flat surface and whose output value at any point on the surface is the temperature at that point. The *directional* derivative at a point with respect to a direction at that point can be thought of as the rate at which the temperature changes if you move along that direction (physically) starting at that point.

More generally, suppose you move along a curve in the surface that's the domain of the function. You want to find the rate at which the temperature is changing along the curve that you are moving along. This rate of temperature change is the product of the (directional derivative along the unit vector tangent direction to the curve at the point) times the (length of the tangent vector, i.e., the speed of motion).

1.5. **Sense of directional derivatives in non-physical contexts.** Consider the case of a production function $f(L, K)$ with inputs L (the expenditure on labor) and K (the expenditure on capital). We already made sense of the partial derivatives $f_L(L, K)$ and $f_K(L, K)$. The partial derivative $f_L(L, K)$ is the marginal change in output for a marginal change in the labor input (i.e., it is the marginal product of labor). Similarly, the partial derivative $f_K(L, K)$ is the marginal chane in output for a marginal change in the capital input (i.e., it is the marginal product of capital).

Let's think of what it means to take the directional derivative along the vector $\langle a, b\rangle$ with $a^2 + b^2 = 1$. What this basically means is the following: we want to measure the marginal change in output if the inputs L and K are changed marginally in the ratio $a : b$. In other words, it measures the impact on output of a particular *combined trajectory of change* in the values of L and K.

2. The gradient vector

2.1. **The direction of change, the direction of no change.** The *gradient vector* for a differentiable function f of two variables is defined as follows:

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y)\rangle = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$$

Note that the gradient vector, thus viewed, is a *vector-valued function of two variables*, i.e., it has type $2 \to 2$. The gradient vector at a particular point is, however, an actual vector, i.e., an actual tuple of two real numbers.

If the gradient vector is nonzero, the *unit vector* in this direction can be computed by dividing this vector by its length.

Here are some key observations that hold if the gradient vector is nonzero:

- Of all the unit vector directions, the directional derivative is maximum (and positive) along the unit vector in the same direction as the gradient vector, and is minimum (and negative) along the unit vector in the opposite direction to the gradient vector.
- The directional derivative is zero along the directions *perpendicular* to the gradient vector.

Intuitively the gradient vector is telling us the direction along which the change/action is happening, and also telling us that there's no action happening orthogonal to it.

2.2. **Special case of function depending on only one variable.** If the function f depends only on the variable x and has no dependence on the variable y, then the gradient vector, where nonzero, will always point parallel to the x-axis (in a positive or negative direction, depending on whether the function is increasing or decreasing).

2.3. **Writing directional derivative in terms of gradient vector.** The directional derivative along a unit vector $\mathbf{u}$ can be defined as the dot product $(\nabla f) \cdot \mathbf{u}$. Since $\mathbf{u}$ is a unit vector, this can be interpreted as the scalar projection of ∇f along $\mathbf{u}$.

Intuitively, what this means is that the extent to which the function is changing in a particular direction depends on the component of the gradient vector that falls in that direction.

Another way of thinking of the directional derivative $D_{\mathbf{u}}(f)$ for a unit vector $\mathbf{u}$ is as $|\nabla f| \cos \theta$ where θ is the angle between ∇f and $\mathbf{u}$. Note that this is equal to $|\nabla f|$ when $\mathbf{u}$ is in the direction of ∇f, and it is $-|\nabla f|$ when $\mathbf{u}$ is opposite to ∇f. It is 0 when $\mathbf{u}$ is orthogonal to ∇f.

Note that if the gradient vector is zero, then there is no direction to it, and the directional derivative along *every* direction is zero.

2.4. **Relationship of gradient vector and level curves.** Recall that the gradient vector represents the direction along which all the change in the function value is happening. It should thus come as no surprise that at any point, the level curve through that point is orthogonal to the gradient vector at that point (note that the statement is trivially true if the gradient vector is zero, but gives no geometric information in that case). Further, if we consider the picture of level curves along with the function values for each curve, then the gradient vector points in the direction of increasing function values.

3. Case of functions of three or more variables

3.1. **Description.** The same expressions for gradient vector and directional derivative apply to functions of three or more variables.

Basically:

- The directional derivative along a unit vector can be defined as a limit of a difference quotient. In three variables: For a function $f(x, y, z)$ at a point (x_0, y_0, z_0) with unit vector $\mathbf{u} = \langle a, b, c \rangle$, this becomes:

$$\lim_{h \to 0} \frac{f(x_0 + ah, y_0 + bh, z_0 + ch)}{h}$$

- For a differentiable function, the gradient vector is defined as the vector obtained by adding, for each coordinate direction, the partial derivative in that direction times the unit vector in that direction. In three variables: For a function $f(x, y, z)$, this becomes:

$$\nabla f(x, y, z) = f_x(x, y, z)\mathbf{i} + f_y(x, y, z)\mathbf{j} + f_z(x, y, z)\mathbf{k}$$

  This is a vector-valued function, i.e., it has type $3 \to 3$ (and more generally $n \to n$). At a particular point in the domain, it gives an actual vector, i.e., a tuple of 3 (respectively n) real numbers.
- For a differentiable function, and in particular for a function where all the first partials exist and are continuous around the point, the directional derivative along a unit vector is the dot product of the gradient vector and that unit vector.
- The gradient vector at a point (if nonzero) is orthogonal to the level surface for the function at the point, or equivalently, it is orthogonal to the tangent plane for the level surface.

3.2. **Application to finding normal direction and tangent plane.** The fact that the gradient vector points in the normal direction and is hence orthogonal to the tangent plane provides a strategy to compute the tangent plane for any point on a surface in $\mathbb{R}^3$ given by a top-down (relational) description of the form $F(x, y, z) = 0$. Namely, to find the normal vector at a point (x_0, y_0, z_0), we compute $\nabla F(x_0, y_0, z_0)$. If this is nonzero, it is a normal vector. We can convert it to a unit vector if desired. Next, we use the technique for finding the scalar equation of a plane to obtain the scalar equation of the tangent plane.

Instead of working things out in each case using vectors, we can also directly determine the scalar version and apply these directly. For a relational description $F(x, y, z) = 0$ and a point (x_0, y_0, z_0), the symmetric equations of the normal line are:

$$\frac{x - x_0}{F_x(x_0, y_0, z_0)} = \frac{y - y_0}{F_y(x_0, y_0, z_0)} = \frac{z - z_0}{F_z(x_0, y_0, z_0)}$$

The equation of the tangent plane is:

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0$$

Note that this method does not work if ∇F *is* 0. In other words, it does not work if all the three first partials $F_x(x_0, y_0, z_0)$, $F_y(x_0, y_0, z_0)$, and $F_z(x_0, y_0, z_0)$ equal 0. In this case, it may happen either that there is no tangent plane, or it may happen that the tangent plane exists but cannot be found through this procedure.

Recall that we had earlier determined the equation to the tangent plane for $z = f(x, y)$, which is a special case of the above with $F(x, y, z) = f(x, y) - z$. It can be verified (See the book) that the earlier formula is consistent with the formula obtained above.

Note: application to temperature. We consider the example of temperature.

Consider the temperature function defined on the surface of the earth that sends a point to the surface of the earth to the surface temperature at that point.

The level curves for this are the *isothermal lines* and represent curves of constant temperature. The gradient vector at any point is a vector tangential to the sphere and represents the direction in which temperature is changing fastest. The positive direction along the gradient vector is the direction of fastest temperature increase. The negative direction is the direction of fastest temperature decrease.

The directional derivative along a direction tangential to the surface of the earth describes the rate at which temperature changes if we move along that direction.

The same example of temperature can be adapted to a three-dimensional setting instead of the surface of the earth. For instance, in a three-dimensional container that is not in thermal equilibrium (so different parts have different temperatures) we can consider the temperature as a function of the location in space. Then, the level *surfaces* are the isothermal surfaces, and the gradient vector at a point is orthogonal to the level surface and describes the direction along which temperature is changing fastest. The directional derivative along a particular direction is the rate at which temperature changes if we move along that direction.