Second derivative

Let $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ be a differentiable scalar field. As we have discussed, at each point $p$ in $D$, the derivative of $f$ is a linear functional on $\mathbb{R}^n$. This linear functional can be expressed in two ways, either as multiplication by a $1 \times n$ matrix, $J_f(p)$, or as the dot product with an $n$-vector called the gradient of $f$ at $p$, $\nabla f(p)$. One consequence of the second representation is that it allows us to view the derivative of a scalar-valued function as a vector field. That is, the derivative of a scalar-valued function is a vector-valued function with the same domain.

A natural question to ask is: when is a vector field the gradient of something? For example, $h(x, y) = (y + 1, x + 1)$ is the gradient of $f(x, y) = xy + x + y + C$ for any constant $C$. On the other hand, $h(x, y) = (xy, x + y)$ is not the gradient of any scalar-valued function. How can we tell? If $\nabla f = h$, then $f_x = xy$ and $f_y = x + y$. So, $f_{xy} = x$ and $f_{yx} = 1$. But, since $h$ is a nice function, we should have $f_{xy} = f_{yx}$. This contradiction means that $h$ is not a gradient.

Thus, we have found a necessary condition for a vector field to be a gradient. If $h(x, y) = (a(x, y), b(x, y))$ is a gradient, then $a_y = b_x$. Such a vector field is called closed. (It is easy to write down the analagous condition in higher dimensions.) The really interesting question is: is this condition also sufficient? In other words, if $a_y = b_x$, is there a function $f$ so that $f_x = a$ and $f_y = b$? (In this case we say that $h$ is exact.) It turns out that, if the domain is all of $\mathbb{R}^2$ then the answer is ‘yes.’ However, if the domain has one or more “holes” (e.g. all of $\mathbb{R}^2$ except $(0, 0)$), then there are vector fields that satisfy the condition but are not gradients. (In other words, there are closed vector fields that are not exact.) In fact, there is such a strong connection between closed forms that are not exact and the shape of the domain that they have become one of the main tools in topology, the mathematical theory of shape. This approach to topology is called cohomology.

* * *

As we have discussed, the derivative of a scalar-valued function, $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, can be viewed as the vector-valued function $\nabla f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$. Since we know how to differentiate vector fields, this allows us to compute the second derivative of $f$. Clearly, the second derivative of $f$ at a point $p$ in $D$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. It is given by the Jacobian matrix of $\nabla f$ at the point $p$. This matrix is
called the Hessian matrix of $f$ at $p$, $Hf_p$. Explicitly,

$$
Hf_p = 
\begin{bmatrix}
\frac{\partial^2 f}{\partial x_1 \partial x_1} (p) & \frac{\partial^2 f}{\partial x_1 \partial x_2} (p) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} (p) \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} (p) & \frac{\partial^2 f}{\partial x_2 \partial x_2} (p) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} (p) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} (p) & \frac{\partial^2 f}{\partial x_n \partial x_2} (p) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} (p)
\end{bmatrix}.
$$

REMARK. Since $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$, for all $i$ and $j$, the Hessian matrix is symmetric (i.e. equal to its own transpose).

For example, let’s compute the Hessian of $f(x, y) = \ln(2x^2 + y^2)$ at $p = (1, 2)$. We have

$$
\frac{\partial f}{\partial x} = \frac{4x}{2x^2 + y^2}, \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{2y}{2x^2 + y^2}.
$$

So,

$$
\frac{\partial^2 f}{\partial x \partial x} = \frac{4y^2 - 8x^2}{(2x^2 + y^2)^2}, \quad \frac{\partial^2 f}{\partial y \partial y} = \frac{4x^2 - 2y^2}{(2x^2 + y^2)^2},
$$

and

$$
\frac{\partial^2 f}{\partial x \partial y} (x, y) = \frac{-8xy}{(2x^2 + y^2)^2}.
$$

Thus,

$$
Hf_{(1, 2)} = \begin{bmatrix}
8/36 & -16/36 \\
-16/36 & -4/36
\end{bmatrix} = \frac{1}{9} \begin{bmatrix}
2 & -4 \\
-4 & -1
\end{bmatrix}.
$$

We will be particularly interested in the Hessian matrix at points where the gradient vanishes. Ultimately, we will state a theorem (the second derivative test) that classifies critical points as local maximum, local minimum, or saddle point, in terms of the Hessian matrix.

Recall the second derivative test in one-variable. If $f'(a) = 0$ then $f''(a) > 0$ implies local minimum at $a$, and $f''(a) < 0$ implies local maximum at $a$. ($f''(a) = 0$ gives no conclusion.) The most simple-minded generalization to higher dimensions would be that $f$ has a local minimum (maximum) if all partial derivatives are positive (negative). If some are positive and others are negative, we have a saddle point. This simple idea is correct if it happens that all mixed partials are zero, that is, if the Hessian matrix is diagonal, but in general things are more complicated. However, if we change basis so that the Hessian matrix becomes diagonal, then we can apply this idea. To say it differently, the naive idea works if, instead of looking at the standard directions, we look at certain special directions that depend on the function itself.
Examples. $f(x, y) = x^2 + y^2$. $f$ has a critical point at $(0, 0)$. Here, $f_{xx}(0, 0) > 0$ and $f_{yy}(0, 0) > 0$ and we have a local minimum as we expect. Note that $Hf_{(0,0)} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

On the other hand, if $f(x, y) = x^2 - y^2$, $f_{xx}(0, 0) > 0$ but $f_{yy}(0, 0) < 0$ so we expect a saddle point, which is indeed the case. Here $Hf_{(0,0)} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$.

Now suppose $f(x, y) = x^2 + 3xy + y^2$. Once again, $(0, 0)$ is a critical point. Also, $f_{xx}(0, 0)$ and $f_{yy}(0, 0)$ are positive. However, in this case we have a saddle point as we can see from the graph below.

The Hessian matrix is $Hf_{(0,0)} = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$. Suppose that we change basis from $e_1, e_2$ to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Let $M = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Then,

$M^{-1}Hf_{(0,0)}M = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}$.

After changing basis, the Hessian matrix has become diagonal. Since one diagonal entry is positive and the other is negative, we have a saddle point.