The Calculus of Functions of Several Variables

Section 3.4

Second-Order Approximations

In one-variable calculus, Taylor polynomials provide a natural way to extend best affine approximations to higher-order polynomial approximations. It is possible to generalize these ideas to scalar-valued functions of two or more variables, but the theory rapidly becomes involved and technical. In this section we will be content merely to point the way with a discussion of second-degree Taylor polynomials. Even at this level, it is best to leave full explanations for a course in advanced calculus.

Higher-order derivatives

The first step is to introduce higher order derivatives. If $f : \mathbb{R}^n \to \mathbb{R}$ has partial derivatives which exist on an open set U, then, for any $i = 1, 2, 3, \ldots, n$, $\frac{\partial f}{\partial x_i}$ is itself a function from $\mathbb{R}^n$ to $\mathbb{R}$. The partial derivatives of $\frac{\partial f}{\partial x_i}$, if they exist, are called *second-order partial derivatives* of f. We may denote the partial derivative of $\frac{\partial f}{\partial x_i}$ with respect to x_j, $j = 1, 2, 3, \ldots$, evaluated at a point $\mathbf{x}$, by either $\frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{x})$, or $f_{x_i x_j}(\mathbf{x})$, or $D_{x_i x_j} f(\mathbf{x})$. Note the order in which the variables are written; it is possible that differentiating first with respect to x_i and second with respect x_j will yield a different result than if the order were reversed. If $j = i$, we will write $\frac{\partial^2}{\partial x_i^2} f(\mathbf{x})$ for $\frac{\partial^2}{\partial x_i \partial x_i} f(\mathbf{x})$. It is, of course, possible to extend this notation to third, fourth, and higher-order derivatives.

Example Suppose $f(x, y) = x^2 y - 3x \sin(2y)$. Then

$$f_x(x, y) = 2xy - 3\sin(2y)$$

and

$$f_y(x, y) = x^2 - 6x\cos(2y),$$

so

$$f_{xx}(x, y) = 2y,$$
$$f_{xy}(x, y) = 2x - 6\cos(2y),$$
$$f_{yy}(x, y) = 12x\sin(2y),$$

and

$$f_{yx}(x, y) = 2x - 6\cos(2y).$$

Note that, in this example, $f_{xy}(x, y) = f_{yx}(x, y)$. For an example of a third-order derivative,

$$f_{yxy}(x, y) = 12\sin(2y).$$

Example Suppose $w = xy^2z^3 - 4xy\log(z)$. Then, for example,

$$\frac{\partial^2 w}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial w}{\partial x}\right) = \frac{\partial}{\partial y}(y^2z^3 - 4y\log(z)) = 2yz^3 - 4\log(z)$$

and

$$\frac{\partial^2 w}{\partial z^2} = \frac{\partial}{\partial z}\left(\frac{\partial w}{\partial z}\right) = \frac{\partial}{\partial z}\left(3xy^2z^2 - \frac{4xy}{z}\right) = 6xy^2z + \frac{4xy}{z^2}.$$

Also,

$$\frac{\partial^2 w}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial w}{\partial y}\right) = \frac{\partial}{\partial x}(2xyz^3 - 4x\log(z)) = 2yz^3 - 4\log(z),$$

and so

$$\frac{\partial^2 w}{\partial y \partial x} = \frac{\partial^2 w}{\partial x \partial y}.$$

In both of our examples we have seen instances where mixed second partial derivatives, that is, second-order partial derivatives with respect to two different variables, taken in different orders are equal. This is not always the case, but does follow if we assume that both of the mixed partial derivatives in question are continuous.

Definition We say a function $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open set U if $f_{x_j x_i}$ is continuous on U for each $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$.

Theorem If f is C^2 on an open ball containing a point $\mathbf{c}$, then

$$\frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{c}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{c})$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$.

Although we have the tools to verify this result, we will leave the justification for a more advanced course.

We shall see that it is convenient to use a matrix to arrange the second partial derivatives of a function f. If $f : \mathbb{R}^n \to \mathbb{R}$, there are n^2 second partial derivatives and this matrix will be $n \times n$.

Definition Suppose the second-order partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$ all exist at the point $\mathbf{c}$. We call the $n \times n$ matrix

$$Hf(\mathbf{c}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} f(\mathbf{c}) & \frac{\partial^2}{\partial x_2 \partial x_1} f(\mathbf{c}) & \frac{\partial^2}{\partial x_3 \partial x_1} f(\mathbf{c}) & \cdots & \frac{\partial^2}{\partial x_n \partial x_1} f(\mathbf{c}) \\ \frac{\partial^2}{\partial x_1 \partial x_2} f(\mathbf{c}) & \frac{\partial^2}{\partial x_2^2} f(\mathbf{c}) & \frac{\partial^2}{\partial x_3 \partial x_2} f(\mathbf{c}) & \cdots & \frac{\partial^2}{\partial x_n \partial x_2} f(\mathbf{c}) \\ \frac{\partial^2}{\partial x_1 \partial x_3} f(\mathbf{c}) & \frac{\partial^2}{\partial x_2 \partial x_3} f(\mathbf{c}) & \frac{\partial^2}{\partial x_3^2} f(\mathbf{c}) & \cdots & \frac{\partial^2}{\partial x_n \partial x_3} f(\mathbf{c}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_n} f(\mathbf{c}) & \frac{\partial^2}{\partial x_2 \partial x_n} f(\mathbf{c}) & \frac{\partial^2}{\partial x_3 \partial x_n} f(\mathbf{c}) & \cdots & \frac{\partial^2}{\partial x_n^2} f(\mathbf{c}) \end{bmatrix} \tag{3.4.1}$$

the *Hessian* of f at $\mathbf{c}$.

Put another way, the Hessian of f at $\mathbf{c}$ is the $n \times n$ matrix whose ith row is $\nabla f_{x_i}(\mathbf{c})$.

Example Suppose $f(x, y) = x^2 y - 3x\sin(2y)$. Then, using our results from above,

$$Hf(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix} = \begin{bmatrix} 2y & 2x - 6\cos(y) \\ 2x - 6\cos(2y) & 12x\sin(2y) \end{bmatrix}.$$

Thus, for example,

$$Hf(2, 0) = \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix}.$$

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open ball $B^2(\mathbf{c}, r)$ and let $\mathbf{h} = (h_1, h_2)$ be a point with $\|\mathbf{h}\| < r$. If we define $\varphi : \mathbb{R} \to \mathbb{R}$ by $\varphi(t) = f(\mathbf{c} + t\mathbf{h})$, then $\varphi(0) = f(\mathbf{c})$ and $\varphi(1) = f(\mathbf{c}+\mathbf{h})$. From the one-variable calculus version of Taylor's theorem, we know that

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(s), \tag{3.4.2}$$

where s is a real number between 0 and 1. Using the chain rule, we have

$$\varphi'(t) = \nabla f(\mathbf{c} + t\mathbf{h}) \cdot \frac{d}{dt}(\mathbf{c} + t\mathbf{h}) = \nabla f(\mathbf{c} + t\mathbf{h}) \cdot \mathbf{h} = f_x(\mathbf{c} + t\mathbf{h})h_1 + f_y(\mathbf{c} + t\mathbf{h})h_2 \tag{3.4.3}$$

and

$$\begin{aligned} \varphi''(t) &= h_1 \nabla f_x(\mathbf{c} + t\mathbf{h}) \cdot \mathbf{h} + h_2 \nabla f_y(\mathbf{c} + t\mathbf{h}) \cdot \mathbf{h} \\ &= (h_1 \nabla f_x(\mathbf{c} + t\mathbf{h}) + h_2 \nabla f_y(\mathbf{c} + t\mathbf{h}) \cdot \mathbf{h} \\ &= [\, h_1 \quad h_2 \,] \begin{bmatrix} f_{xx}(\mathbf{c} + t\mathbf{h}) & f_{xy}(\mathbf{c} + t\mathbf{h}) \\ f_{yx}(\mathbf{c} + t\mathbf{h}) & f_{yy}(\mathbf{c} + t\mathbf{h}) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ &= \mathbf{h}^T Hf(\mathbf{c} + t\mathbf{h})\mathbf{h}, \end{aligned} \tag{3.4.4}$$

where we have used the notation

$$\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

and

$$\mathbf{h}^T = [\, h_1 \quad h_2 \,],$$

the latter being called the *transpose* of $\mathbf{h}$ (see Problem 12 of Section 1.6). Hence

$$\varphi'(0) = \nabla f(\mathbf{c}) \cdot \mathbf{h} \tag{3.4.5}$$

and

$$\varphi''(s) = \frac{1}{2}\mathbf{h}^T Hf(c + s\mathbf{h})\mathbf{h}, \tag{3.4.6}$$

so, substituting into (3.4.2), we have

$$f(\mathbf{c} + \mathbf{h}) = \varphi(1) = f(\mathbf{c}) + \nabla f(\mathbf{c}) \cdot \mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{c} + s\mathbf{h})\mathbf{h}. \tag{3.4.7}$$

This result, a version of Taylor's theorem, is easily generalized to higher dimensions.

Theorem Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open ball $B^n(\mathbf{c}, r)$ and let $\mathbf{h}$ be a point with $\|\mathbf{h}\| < r$. Then there exists a real number s between 0 and 1 such that

$$f(\mathbf{c}+\mathbf{h}) = f(\mathbf{c}) + \nabla f(\mathbf{c}) \cdot \mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf(\mathbf{c}+s\mathbf{h})\mathbf{h}. \tag{3.4.8}$$

If we let $\mathbf{x} = \mathbf{c} + \mathbf{h}$ and evaluate the Hessian at $\mathbf{c}$, (3.4.8) becomes a polynomial approximation for f.

Definition If $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open ball about the point $\mathbf{c}$, then we call

$$P_2(\mathbf{x}) = f(\mathbf{c}) + \nabla f(\mathbf{c}) \cdot (\mathbf{x}-\mathbf{c}) + \frac{1}{2}(\mathbf{x}-\mathbf{c})^T Hf(\mathbf{c})(\mathbf{x}-\mathbf{c}) \tag{3.4.9}$$

the *second-order Taylor polynomial* for f at $\mathbf{c}$.

Example To find the second-order Taylor polynomial for $f(x,y) = e^{-2x+y}$ at $(0,0)$, we compute

$$\nabla f(x,y) = (-2e^{-2x+y}, e^{-2x+y})$$

and

$$Hf(x,y) = \begin{bmatrix} 4e^{-2x+y} & -2e^{-2x+y} \\ -2e^{-2x+y} & e^{-2x+y} \end{bmatrix},$$

from which it follows that

$$\nabla f(0,0) = (-2, 1)$$

and

$$Hf(0,0) = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}.$$

Then

$$\begin{aligned} P_2(x,y) &= f(0,0) + \nabla f(0,0) \cdot (x,y) + \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix} Hf(0,0) \begin{bmatrix} x \\ y \end{bmatrix} \\ &= 1 + (-2,1)\cdot(x,y) + \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= 1 - 2x + y = \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 4x-2y \\ -2x+y \end{bmatrix} \\ &= 1 - 2x + y + \frac{1}{2}(4x^2 - 2xy - 2xy + y^2) \\ &= 1 - 2x + y + 2x^2 - 2xy + \frac{1}{2}y^2. \end{aligned}$$

Symmetric matrices

Note that if $f : \mathbb{R}^2 \to \mathbb{R}$ is C^2 on an open ball about the point $\mathbf{c}$, then the entry in the ith row and jth column of $Hf(\mathbf{c})$ is equal to the entry in the jth row and ith column of $Hf(\mathbf{c})$ since

$$\frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{c}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{c}).$$

Definition We call a matrix $M = [a_{ij}]$ with the property that $a_{ij} = a_{ji}$ for all $i \neq j$ a *symmetric matrix.*

Example The matrices

$$\begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & -7 \end{bmatrix}$$

are both symmetric, while the matrices

$$\begin{bmatrix} 2 & -1 \\ 3 & 4 \end{bmatrix}$$

and

$$\begin{bmatrix} 2 & 1 & 3 \\ 2 & 3 & 4 \\ -2 & 4 & -6 \end{bmatrix}$$

are not symmetric.

Example The Hessian of any C^2 scalar valued function is a symmetric matrix. For example, the Hessian of $f(x, y) = e^{-2x+y}$, namely,

$$Hf(x, y) = \begin{bmatrix} 4e^{-2x+y} & -2e^{-2x+y} \\ -2e^{-2x+y} & e^{-2x+y} \end{bmatrix},$$

is symmetric for any value of (x, y).

Given an $n \times n$ symmetric matrix M, the function $q : \mathbb{R}^n \to \mathbb{R}$ defined by

$$q(\mathbf{x}) = \mathbf{x}^T M \mathbf{x}$$

is a quadratic polynomial. When M is the Hessian of some function f, this is the form of the quadratic term in the second-order Taylor polynomial for f. In the next section it will be important to be able to determine when this term is positive for all $\mathbf{x} \neq \mathbf{0}$ or negative for all $\mathbf{x} \neq \mathbf{0}$.

Definition Let M be an $n \times n$ symmetric matrix and define $q : \mathbb{R}^n \to \mathbb{R}$ by

$$q(\mathbf{x}) = \mathbf{x}^T M \mathbf{x}.$$

We say M is *positive definite* if $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$, *negative definite* if $q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$ in $\mathbb{R}^n$, and *indefinite* if there exists an $\mathbf{x} \neq 0$ for which $q(\mathbf{x}) > 0$ and an $\mathbf{x} \neq \mathbf{0}$ for which $q(\mathbf{x}) < 0$. Otherwise, we say M is *nondefinite.*

In general it is not easy to determine to which of these categories a given symmetric matrix belongs. However, the important special case of 2×2 matrices is straightforward. Consider

$$M = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

and let

$$q(x,y) = [\,x \quad y\,]\, M \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + cy^2. \tag{3.4.10}$$

If $a \neq 0$, then we may complete the square in (3.4.10) to obtain

$$\begin{aligned} q(x,y) &= a\left(x^2 + \frac{2b}{a}xy\right) + cy^2 \\ &= a\left(\left(x + \frac{b}{a}y\right)^2 - \frac{b^2}{a^2}y^2\right) + cy^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \frac{ac - b^2}{a}y^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \frac{\det(M)}{a}y^2 \qquad (3.4.11). \end{aligned}$$

Now suppose $\det(M) > 0$. Then from (3.4.11) we see that $q(x,y) > 0$ for all $(x,y) \neq (0,0)$ if $a > 0$ and $q(x,y) < 0$ for all $(x,y) \neq (0,0)$ if $a < 0$. That is, M is positive definite if $a > 0$ and negative definite if $a < 0$. If $\det(M) < 0$, then $q(1,0)$ and $q\left(-\frac{b}{a}, 1\right)$ will have opposite signs, and so M is indefinite. Finally, suppose $\det(M) = 0$. Then

$$q(x,y) = a\left(x + \frac{b}{a}y\right)^2,$$

so $q(x,y) = 0$ when $x = -\frac{b}{a}y$. Moreover, $q(x,y)$ has the same sign as a for all other values of (x,y). Hence in this case M is nondefinite.

Similar analyses for the case $a = 0$ give us the following result.

Theorem Suppose

$$M = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

If $\det(M) > 0$, then M is positive definite if $a > 0$ and negative definite if $a < 0$. If $\det(M) < 0$, then M is indefinite. If $\det(M) = 0$, then M is nondefinite.

Example The matrix

$$M = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

is positive definite since $\det(M) = 5 > 0$ and $2 > 0$.

Example The matrix

$$M = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}$$

is negative definite since $\det(M) = 7 > 0$ and $-2 < 0$.

Example The matrix

$$M = \begin{bmatrix} -3 & 1 \\ 1 & 2 \end{bmatrix}$$

is indefinite since $\det(M) = -7 < 0$.

Example The matrix

$$M = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

is nondefinite since $\det(M) = 0$.

In the next section we will see how these ideas help us identify local extreme values for scalar valued functions of two variables.

Problems

1. Let $f(x, y) = x^3y^2 - 4x^2e^{-3y}$. Find the following.
   (a) $\dfrac{\partial^2}{\partial x \partial y} f(x, y)$ (b) $\dfrac{\partial^2}{\partial y \partial x} f(x, y)$
   (c) $\dfrac{\partial^2}{\partial x^2} f(x, y)$ (d) $\dfrac{\partial^3}{\partial x \partial y \partial x} f(x, y)$
   (e) $\dfrac{\partial^3}{\partial x \partial y^2} f(x, y)$ (f) $\dfrac{\partial^3}{\partial y^3} f(x, y)$
   (g) $f_{yy}(x, y)$ (h) $f_{yxy}(x, y)$
2. Let $f(x, y, z) = \dfrac{xy}{x^2 + y^2 + z^2}$. Find the following.
   (a) $\dfrac{\partial^2}{\partial z \partial x} f(x, y, z)$ (b) $\dfrac{\partial^2}{\partial y \partial z} f(x, y, z)$
   (c) $\dfrac{\partial^2}{\partial z^2} f(x, y, z)$ (d) $\dfrac{\partial^3}{\partial x \partial y \partial z} f(x, y, z)$
   (e) $f_{zyx}(x, y, z)$ (f) $f_{yyy}(x, y, z)$
3. Find the Hessian of each of the following functions.
   (a) $f(x, y) = 3x^2y - 4xy^3$ (b) $g(x, y) = 4e^{-x}\cos(3y)$
   (c) $g(x, y, z) = 4xy^2z^3$ (d) $f(x, y, z) = -\log(x^2 + y^2 + z^2)$
4. Find the second-order Taylor polynomial for each of the following at the point $\mathbf{c}$.
   (a) $f(x, y) = xe^{-y}$, $\mathbf{c} = (0, 0)$ (b) $g(x, y) = x\sin(x + y)$, $\mathbf{c} = (0, 0)$

(c) $f(x,y) = \dfrac{1}{x+y}$, $\mathbf{c} = (1,1)$

(d) $g(x,y,z) = e^{x-2y+3z}$, $\mathbf{c} = (0,0,0)$

5. Classify each of the following symmetric 2×2 matrices as either positive definite, negative definite, indefinite, or nondefinite.

(a) $\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} -2 & 3 \\ 3 & -5 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(f) $\begin{bmatrix} 8 & 4 \\ 4 & 2 \end{bmatrix}$

6. Let M be an $n \times n$ symmetric nondefinite matrix and define $q : \mathbb{R}^n \to \mathbb{R}$ by

$$q(\mathbf{x}) = \mathbf{x}^T M \mathbf{x}.$$

Explain why (1) there exists a vector $\mathbf{a} \neq \mathbf{0}$ such that $q(\mathbf{a}) = 0$ and (2) either $q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$ or $q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

7. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open ball $B^n(\mathbf{c}, r)$, $\nabla f(\mathbf{c}) = \mathbf{0}$, and $Hf(\mathbf{x})$ is positive definite for all $\mathbf{x}$ in $B^n(\mathbf{c}, r)$. Show that $f(\mathbf{c}) < f(\mathbf{x})$ for all $\mathbf{x}$ in $B^n(\mathbf{c}, r)$. What would happen if $Hf(\mathbf{x})$ were negative definite for all $\mathbf{x}$ in $B^n(\mathbf{c}, r)$? What does this say in the case $n = 1$?

8. Let

$$f(x,y) = \begin{cases} \dfrac{xy(x^2 - y^2)}{x^2 + y^2}, & \text{if } (x,y) \neq (0,0), \\ 0, & \text{if } (x,y) = (0,0). \end{cases}$$

(a) Show that $f_x(0,y) = -y$ for all y.

(b) Show that $f_y(x,0) = x$ for all x.

(c) Show that $f_{yx}(0,0) \neq f_{xy}(0,0)$.

(d) Is f C^2?