Max-Min Problems in R^n and the Hessian Matrix

Prerequisite: Section 6.3, Orthogonal Diagonalization

In this section, we study the problem of finding local maxima and minima for real-valued functions on $\mathbb{R}^n$. The method we describe is the higher-dimensional analogue to finding critical points and applying the second derivative test to functions on $\mathbb{R}$ studied in first-semester calculus.

▶ Taylor's Theorem in $\mathbb{R}^n$

Let $f \in C^2(\mathbb{R}^n)$, where $C^2(\mathbb{R}^n)$ is the set of real-valued functions defined on $\mathbb{R}^n$ having continuous second partial derivatives. The method for solving for local extreme points of f relies upon Taylor's Theorem with second degree remainder terms, which we state here without proof. (In the following theorem, an **open hypersphere** centered at $\mathbf{x}_0$ is a set of the form $\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}_0\| < r\}$ for some positive real number r.)

THEOREM 1

(**Taylor's Theorem** in $\mathbb{R}^n$) Let A be an open hypersphere centered at $\mathbf{x}_0 \in \mathbb{R}^n$, let $\mathbf{u}$ be a unit vector in $\mathbb{R}^n$, and let $t \in \mathbb{R}$ such that $\mathbf{x}_0 + t\mathbf{u} \in A$. Suppose $f : A \to \mathbb{R}$ has continuous second partial derivatives throughout A; that is, $f \in C^2(A)$. Then there is a c with $0 \le c \le t$ such that

$$\begin{aligned} f(\mathbf{x}_0 + t\mathbf{u}) \;=\; & f(\mathbf{x}_0) + \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{x}_0} (tu_i) + \frac{1}{2}\sum_{i=1}^{n} \left.\frac{\partial^2 f}{\partial x_i^2}\right|_{\mathbf{x}_0 + c\mathbf{u}} (t^2 u_i^2) \\ & + \sum_{i=1}^{n}\sum_{j=i+1}^{n} \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{x}_0 + c\mathbf{u}} (t^2 u_i u_j). \end{aligned}$$

Taylor's Theorem in $\mathbb{R}^n$ is derived from the familiar Taylor's Theorem in $\mathbb{R}$ by applying it to the function $g(t) = f(\mathbf{x}_0 + t\mathbf{u})$. In $\mathbb{R}^2$, the formula in Taylor's Theorem is

$$\begin{aligned} f(\mathbf{x}_0 + t\mathbf{u}) \;=\; & f(\mathbf{x}_0) + \left.\frac{\partial f}{\partial x}\right|_{\mathbf{x}_0} (tu_1) + \left.\frac{\partial f}{\partial y}\right|_{\mathbf{x}_0} (tu_2) \\ & + \frac{1}{2}\left.\frac{\partial^2 f}{\partial x^2}\right|_{\mathbf{x}_0 + c\mathbf{u}} (t^2 u_1^2) + \frac{1}{2}\left.\frac{\partial^2 f}{\partial y^2}\right|_{\mathbf{x}_0 + c\mathbf{u}} (t^2 u_2^2) \\ & + \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{\mathbf{x}_0 + c\mathbf{u}} (t^2 u_1 u_2). \end{aligned}$$

Recall that the **gradient** of f is defined by $\nabla f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]$. If we let $\mathbf{v} = t\mathbf{u}$, then, in $\mathbb{R}^2$, $\mathbf{v} = [v_1, v_2] = [tu_1, tu_2]$, and so the sum $\left.\frac{\partial f}{\partial x}\right|_{\mathbf{x}_0} (tu_1) + \left.\frac{\partial f}{\partial y}\right|_{\mathbf{x}_0} (tu_2)$ simplifies to $\left(\nabla f\Big|_{\mathbf{x}_0}\right) \cdot \mathbf{v}$. Also, since f has continuous second partial derivatives, we have $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$. Therefore,

$$\begin{aligned} &\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t^2u_1^2) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(t^2u_2^2) + \frac{\partial^2 f}{\partial x \partial y}(t^2u_1u_2) \\ &= \frac{1}{2}v_1\left(\frac{\partial^2 f}{\partial x^2}v_1 + \frac{\partial^2 f}{\partial x \partial y}v_2\right) + \frac{1}{2}v_2\left(\frac{\partial^2 f}{\partial y \partial x}v_1 + \frac{\partial^2 f}{\partial y^2}v_2\right) \\ &= \frac{1}{2}\mathbf{v}^T \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \mathbf{v}, \end{aligned}$$

where $\mathbf{v}$ is considered to be a column vector. The matrix

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

in this expression is called the **Hessian matrix** for f. Thus, in the $\mathbb{R}^2$ case, with $\mathbf{v} = t\mathbf{u}$, the formula in Taylor's Theorem can be written as

$$f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + \left(\nabla f\Big|_{\mathbf{x}_0}\right) \cdot \mathbf{v} \;+\; \frac{1}{2}\mathbf{v}^T\left(\mathbf{H}\Big|_{\mathbf{x}_0 + k\mathbf{v}}\right)\mathbf{v},$$

for some k with $0 \leq k \leq 1$ (where $k = \frac{c}{t}$). While we have derived this result in $\mathbb{R}^2$, the same formula holds in $\mathbb{R}^n$, where the Hessian $\mathbf{H}$ is the matrix whose (i, j) entry is $\frac{\partial^2 f}{\partial x_i \partial x_j}$.

► Critical Points

If A is a subset of $\mathbb{R}^n$, then we say that $f : A \to \mathbb{R}$ has a **local maximum** at a point $\mathbf{x}_0 \in A$ if and only if there is an open neighborhood $\mathcal{U}$ of $\mathbf{x}_0$ such that $f(\mathbf{x}_0) \geq f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{U}$. A **local minimum** for a function f is defined analogously.

THEOREM 2
Let A be an open hypersphere centered at $\mathbf{x}_0 \in \mathbb{R}^n$, and let $f : A \to \mathbb{R}$ have continuous first partial derivatives on A. If f has a local maximum or a local minimum at $\mathbf{x}_0$, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$.

Proof If $\mathbf{x}_0$ is a local maximum, then $f(\mathbf{x}_0 + h\mathbf{e}_i) - f(\mathbf{x}_0) \leq 0$ for small h. Then, $\lim_{h \to 0^+} \frac{f(\mathbf{x}_0 + h\mathbf{e}_i) - f(\mathbf{x}_0)}{h} \leq 0$. Similarly, $\lim_{h \to 0^-} \frac{f(\mathbf{x}_0 + h\mathbf{e}_i) - f(\mathbf{x}_0)}{h} \geq 0$. Hence, for the limit to exist, we must have $\frac{\partial f}{\partial x_i}\Big|_{\mathbf{x}_0} = 0$. Since this is true for each i, $\nabla f\Big|_{\mathbf{x}_0} = \mathbf{0}$. A similar proof works for local minimums. **QED**

Points $\mathbf{x}_0$ at which $\nabla f(\mathbf{x}_0) = \mathbf{0}$ are called **critical points**.

EXAMPLE 1 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = 7x^2 + 6xy + 2x + 7y^2 - 22y + 23.$$

Then $\nabla f = [14x + 6y + 2,\ 6x + 14y - 22]$. We find critical points for f by solving $\nabla f = \mathbf{0}$. This is the linear system

$$\begin{cases} 14x + 6y + 2 = 0 \\ 6x + 14y - 22 = 0 \end{cases},$$

which has the unique solution $\mathbf{x}_0 = [-1, 2]$. Hence, by Theorem 2, $(-1, 2)$ is the only possible extreme point for f. (We will see later that $(-1, 2)$ is a local minimum.) ■

Sufficient Conditions for Local Extreme Points

If $\mathbf{x}_0$ is a critical point for a function f, how can we determine whether $\mathbf{x}_0$ is a local maximum or a local minimum? For functions on $\mathbb{R}$, we have the second derivative test from calculus, which says that if $f''(\mathbf{x}_0) < 0$, then $\mathbf{x}_0$ is a local maximum, but if $f''(\mathbf{x}_0) > 0$, then $\mathbf{x}_0$ is a local minimum. We now derive a similar test in $\mathbb{R}^n$.

Consider the following formula from Taylor's Theorem:

$$f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{v} \; + \; \frac{1}{2}\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0 + k\mathbf{v}} \right) \mathbf{v}.$$

At a critical point, $\nabla f(\mathbf{x}_0) = \mathbf{0}$, and so

$$f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + \frac{1}{2}\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0 + k\mathbf{v}} \right) \mathbf{v}.$$

Hence, if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0 + k\mathbf{v}} \right) \mathbf{v}$ is positive for all small nonzero vectors $\mathbf{v}$, then f will have a local minimum at $\mathbf{x}_0$. (Similarly, if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0 + k\mathbf{v}} \right) \mathbf{v}$ is negative, f will have a local maximum.) But since we assume that f has continuous second partial derivatives, $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0 + k\mathbf{v}} \right) \mathbf{v}$ is continuous in $\mathbf{v}$ and k, and will be positive for small $\mathbf{v}$ if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0} \right) \mathbf{v}$ is positive for all nonzero $\mathbf{v}$. Hence,

> THEOREM 3
> Given the conditions of Taylor's Theorem for a set A and a function $f : A \to \mathbb{R}$, f has a local minimum at a critical point $\mathbf{x}_0$ if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0} \right) \mathbf{v} > 0$ for all nonzero vectors $\mathbf{v}$. Similarly, f has a local maximum at a critical point $\mathbf{x}_0$ if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0} \right) \mathbf{v} < 0$ for all nonzero vectors $\mathbf{v}$.

Positive Definite Quadratic Forms

If $\mathbf{v}$ is a vector in $\mathbb{R}^n$, and $\mathbf{A}$ is an $n \times n$ matrix, the expression $\mathbf{v}^T\mathbf{A}\mathbf{v}$ is known as a **quadratic form**. (For more details on the general theory of quadratic forms, see Section 8.11.) A quadratic form such that $\mathbf{v}^T\mathbf{A}\mathbf{v} > 0$ for all nonzero vectors $\mathbf{v}$ is said to be **positive definite**. Similarly, a quadratic form such that $\mathbf{v}^T\mathbf{A}\mathbf{v} < 0$ for all nonzero vectors $\mathbf{v}$ is said to be **negative definite**.

Now, in particular, the expression $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0} \right) \mathbf{v}$ in Theorem 3 is a quadratic form. Theorem 3 then says that if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0} \right) \mathbf{v}$ is a positive definite quadratic form at a critical point $\mathbf{x}_0$, then f has a local minimum at $\mathbf{x}_0$. Theorem 3 also says that if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0} \right) \mathbf{v}$ is a negative definite quadratic form at a critical point $\mathbf{x}_0$, then f has a local maximum at $\mathbf{x}_0$. Therefore, we need a method to determine whether a quadratic form of this type is positive definite or negative definite.

Now, the Hessian matrix $\left(\mathbf{H}\Big|_{\mathbf{x}_0} \right)$, which we will abbreviate as $\mathbf{H}$, is symmetric because $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ (since $f \in C^2(A)$). Hence, by Theorem 6.20, $\mathbf{H}$ can be orthogonally diagonalized. That is, there is an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}\mathbf{H}\mathbf{P}^T = \mathbf{D}$, a diagonal matrix, and so, $\mathbf{H} = \mathbf{P}^T\mathbf{D}\mathbf{P}$. Hence, $\mathbf{v}^T\mathbf{H}\mathbf{v} = \mathbf{v}^T\mathbf{P}^T\mathbf{D}\mathbf{P}\mathbf{v} = (\mathbf{P}\mathbf{v})^T\mathbf{D}(\mathbf{P}\mathbf{v})$. Letting $\mathbf{w} = \mathbf{P}\mathbf{v}$, we get $\mathbf{v}^T\mathbf{H}\mathbf{v} = \mathbf{w}^T\mathbf{D}\mathbf{w}$. But $\mathbf{P}$ is nonsingular, so as $\mathbf{v}$ ranges over all of $\mathbb{R}^n$, so does $\mathbf{w}$, and vice-versa. Thus,

$\mathbf{v}^T\mathbf{H}\mathbf{v} > 0$ for all nonzero $\mathbf{v}$ if and only if $\mathbf{w}^T\mathbf{D}\mathbf{w} > 0$ for all nonzero $\mathbf{w}$. Now, $\mathbf{D}$ is diagonal, and so $\mathbf{w}^T\mathbf{D}\mathbf{w} = d_{11}w_1^2 + d_{22}w_2^2 + \cdots + d_{nn}w_n^2$. But the d_{ii}'s are the eigenvalues of $\mathbf{H}$. Thus, it follows that $\mathbf{w}^T\mathbf{D}\mathbf{w} > 0$ for *all* nonzero $\mathbf{w}$ if and only if all of these eigenvalues are positive. (Set $\mathbf{w} = \mathbf{e}_i$ for each i to prove the "only if" part of this statement.) Similarly, $\mathbf{w}^T\mathbf{D}\mathbf{w} < 0$ for *all* nonzero $\mathbf{w}$ if and only if all of these eigenvalues are negative. Hence,

THEOREM 4
A symmetric matrix $\mathbf{A}$ defines a positive definite quadratic form $\mathbf{v}^T\mathbf{A}\mathbf{v}$ if and only if all of the eigenvalues of $\mathbf{A}$ are positive. A symmetric matrix $\mathbf{A}$ defines a negative definite quadratic form $\mathbf{v}^T\mathbf{A}\mathbf{v}$ if and only if all of the eigenvalues of $\mathbf{A}$ are negative.

Hence, Theorem 3 can be restated as follows:

Given the conditions of Taylor's Theorem for a set A and a function $f : A \to \mathbb{R}$:
(1) if all of the eigenvalues of $\mathbf{H}$ are positive at a critical point $\mathbf{x}_0$, then f has a local minimum at $\mathbf{x}_0$, and
(2) if all of the eigenvalues of $\mathbf{H}$ are positive at a critical point $\mathbf{x}_0$, then f has a local minimum at $\mathbf{x}_0$.

EXAMPLE 2 Consider the function

$$f(x, y) = 7x^2 + 6xy + 2x + 7y^2 - 22y + 23.$$

In Example 1, we found that f has a critical point at $\mathbf{x}_0 = [-1, 2]$. Now, the Hessian matrix for f at $\mathbf{x}_0$ is

$$\mathbf{H} = \left[\begin{array}{cc} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{array}\right]\Bigg|_{\mathbf{x}_0} = \left[\begin{array}{cc} 14 & 6 \\ 6 & 14 \end{array}\right].$$

But $p_{\mathbf{H}}(x) = x^2 - 28x + 160$, which has roots $x = 8$ and $x = 20$. Thus, $\mathbf{H}$ has all eigenvalues positive, and hence, $\mathbf{v}^T\mathbf{H}\mathbf{v}$ is positive definite. Theorem 4 then tells us that $\mathbf{x}_0 = [-1, 2]$ is a local minimum for f. ■

Local Maxima and Minima in $\mathbb{R}^2$

It can be shown (see Exercise 3) that a 2×2 symmetric matrix $\mathbf{A}$ defines a positive definite quadratic form ($\mathbf{v}^T\mathbf{A}\mathbf{v} > 0$ for all nonzero $\mathbf{v}$) if and only if $a_{11} > 0$ and $|\mathbf{A}| > 0$. Similarly, a 2×2 symmetric matrix defines a negative definite quadratic form ($\mathbf{v}^T\mathbf{A}\mathbf{v} < 0$ for all nonzero $\mathbf{v}$) if and only if $a_{11} < 0$ and $|\mathbf{A}| > 0$.

EXAMPLE 3 Suppose $f(x, y) = 2x^2 - 2x^2y^2 + 2y^2 + 24y - x^4 - y^4$. First, we look for critical points by solving the system

$$\begin{cases} \frac{\partial f}{\partial x} = 4x - 4xy^2 - 4x^3 = 4x(1 - (y^2 + x^2)) = 0 \\ \frac{\partial f}{\partial y} = -4x^2y + 4y + 24 - 4y^3 = -4y(x^2 + y^2) + 4y + 24 = 0 \end{cases}.$$

Now $\frac{\partial f}{\partial x} = 0$ yields $x = 0$ or $y^2 + x^2 = 1$. If $x = 0$, then $\frac{\partial f}{\partial y} = 0$ gives $4y + 24 - 4y^3 = 0$. The unique real solution to this equation is $y = 2$. Thus, $[0, 2]$ is a critical point.

If $x \neq 0$, then $y^2 + x^2 = 1$. From $\frac{\partial f}{\partial y} = 0$, we have $0 = -4y(1) + 4y + 24 = 24$, a contradiction, so there is no critical point when $x \neq 0$.

Next, we compute the Hessian matrix at the critical point $[0, 2]$.

$$\mathbf{H} = \left.\begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}\right|_{[\mathbf{0},\mathbf{2}]}$$

$$= \left.\begin{bmatrix} 4 - 4y^2 - 12x^2 & -8xy \\ -8xy & -4x^2 + 4 - 12y^2 \end{bmatrix}\right|_{[\mathbf{0},\mathbf{2}]} = \begin{bmatrix} -12 & 0 \\ 0 & -44 \end{bmatrix}.$$

Since the $(1,1)$ entry is negative and $|\mathbf{H}| > 0$, $\mathbf{H}$ defines a negative definite quadratic form and so f has a local maximum at $[0, 2]$. ■

▶ An Example in $\mathbb{R}^3$

EXAMPLE 4 Consider the function

$$g(x, y, z) = 5x^2 + 2xz + 4xy + 10x + 3z^2 - 6yz - 6z + 5y^2 + 12y + 21.$$

We find the critical points by solving the system

$$\begin{cases} \frac{\partial g}{\partial x} = 10x + 2z + 4y + 10 = 0 \\ \frac{\partial g}{\partial y} = 4x - 6z + 10y + 12 = 0 \\ \frac{\partial g}{\partial z} = 2x + 6z - 6y - 6 = 0 \end{cases}.$$

Using row reduction to solve this linear system yields the unique critical point $[-9, 12, 16]$. The Hessian matrix at $[-9, 12, 16]$ is

$$\mathbf{H} = \left.\begin{bmatrix} \frac{\partial^2 g}{\partial x^2} & \frac{\partial^2 g}{\partial x \partial y} & \frac{\partial^2 g}{\partial x \partial z} \\ \frac{\partial^2 g}{\partial y \partial x} & \frac{\partial^2 g}{\partial y^2} & \frac{\partial^2 g}{\partial y \partial z} \\ \frac{\partial^2 g}{\partial z \partial x} & \frac{\partial^2 g}{\partial z \partial y} & \frac{\partial^2 g}{\partial z^2} \end{bmatrix}\right|_{[-9,12,16]} = \begin{bmatrix} 10 & 4 & 2 \\ 4 & 10 & -6 \\ 2 & -6 & 6 \end{bmatrix}.$$

A lengthy computation produces $p_{\mathbf{H}}(x) = x^3 - 26x^2 + 164x - 8$. The roots of $p_{\mathbf{H}}(x)$ are approximately 0.04916, 10.6011, and 15.3497. Since all of these eigenvalues for $\mathbf{H}$ are positive, $[-9, 12, 16]$ is a local minimum for g. ■

▶ Failure of the Hessian Matrix Test

In calculus, we discovered that the second derivative test fails when the second derivative is zero at a critical point. A similar situation is true in $\mathbb{R}^n$. If the Hessian matrix at a critical point has 0 as an eigenvalue, and all other eigenvalues have the same sign, then the function f could have a local maximum, a local minimum, or neither at this critical point. Of course, if the Hessian matrix at a critical point has two eigenvalues with opposite signs, the critical point is not a local extreme point (why?). Exercise 2 illustrates these concepts.

▶ New Vocabulary

$C^2(\mathbb{R}^n)$ (functions from $\mathbb{R}^n$ to $\mathbb{R}$ having continuous second partial derivatives)
critical point (of a function)
gradient (of a function on $\mathbb{R}^n$)
Hessian matrix
local maximum (of a function on $\mathbb{R}^n$)
local minimum (of a function on $\mathbb{R}^n$)
negative definite quadratic form

open hypersphere (in $\mathbb{R}^n$)
positive definite quadratic form
Taylor's Theorem (in $\mathbb{R}^n$)

▶ Highlights

- The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined by $\nabla f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]$.
- Let A be an open hypersphere about $\mathbf{x}_0$, and let f be a function on A with continuous partial derivatives. If f has a local maximum or minimum at $\mathbf{x}_0$, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$.
- For a function $f : \mathbb{R}^n \to \mathbb{R}$, its corresponding Hessian matrix $\mathbf{H}$ is the $n \times n$ matrix whose (i, j) entry is $\frac{\partial^2 f}{\partial x_i \partial x_j}$. In particular, for a function $f : \mathbb{R}^2 \to \mathbb{R}$, the Hessian matrix $\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$.
- Taylor's Theorem in $\mathbb{R}^n$: Let A be an open hypersphere centered at $\mathbf{x}_0 \in \mathbb{R}^n$, let $\mathbf{u}$ be a unit vector in $\mathbb{R}^n$, and let $t \in \mathbb{R}$ such that $\mathbf{x}_0 + t\mathbf{u} \in A$. Suppose $f : A \to \mathbb{R}$ has continuous second partial derivatives throughout A; that is, $f \in C^2(A)$. Then there is a c with $0 \le c \le t$ such that $f(\mathbf{x}_0 + t\mathbf{u}) = f(\mathbf{x}_0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}\Big|_{\mathbf{x}_0} (tu_i) + \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}\Big|_{\mathbf{x}_0 + c\mathbf{u}} (t^2 u_i^2) + \sum_{i=1}^n \sum_{j=i+1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}\Big|_{\mathbf{x}_0 + c\mathbf{u}} (t^2 u_i u_j)$. In particular, we have $f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + \left(\nabla f\Big|_{\mathbf{x}_0}\right) \cdot \mathbf{v} \; + \; \frac{1}{2}\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0 + k\mathbf{v}}\right) \mathbf{v}$, for some k with $0 \le k \le 1$.
- Let A be an open hypersphere centered at $\mathbf{x}_0 \in \mathbb{R}^n$. If $f : A \to \mathbb{R}$ has continuous second partial derivatives throughout A, then $f : A \to \mathbb{R}$, f has a local minimum at a critical point $\mathbf{x}_0$ if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0}\right) \mathbf{v} > 0$ for all nonzero vectors $\mathbf{v}$. Similarly, f has a local maximum at a critical point $\mathbf{x}_0$ if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0}\right) \mathbf{v} < 0$ for all nonzero vectors $\mathbf{v}$.
- A quadratic form is an expression of the form $\mathbf{v}^T\mathbf{A}\mathbf{v}$, where $\mathbf{v}$ is a vector in $\mathbb{R}^n$, and $\mathbf{A}$ is an $n \times n$ matrix. A positive definite quadratic form is one such that $\mathbf{v}^T\mathbf{A}\mathbf{v} > 0$ for all nonzero vectors $\mathbf{v}$. Similarly, a negative definite quadratic form is one such that $\mathbf{v}^T\mathbf{A}\mathbf{v} < 0$ for all nonzero vectors $\mathbf{v}$.
- For a function $f : \mathbb{R}^n \to \mathbb{R}$ having Hessian matrix $\mathbf{H}$, if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0}\right) \mathbf{v}$ is a positive definite quadratic form at a critical point $\mathbf{x}_0$, then f has a local minimum at $\mathbf{x}_0$. Similarly, if $\mathbf{v}^T \left(\mathbf{H}\Big|_{\mathbf{x}_0}\right) \mathbf{v}$ is a negative definite quadratic form at a critical point $\mathbf{x}_0$, then f has a local maximum at $\mathbf{x}_0$.
- A symmetric matrix $\mathbf{A}$ defines a positive definite quadratic form $\mathbf{v}^T\mathbf{A}\mathbf{v}$ if and only if all of the eigenvalues of $\mathbf{A}$ are positive.
- A symmetric matrix $\mathbf{A}$ defines a negative definite quadratic form $\mathbf{v}^T\mathbf{A}\mathbf{v}$ if and only if all of the eigenvalues of $\mathbf{A}$ are negative.
- If $f : \mathbb{R}^n \to \mathbb{R}$ has Hessian matrix $\mathbf{H}$, and all eigenvalues of $\mathbf{H}$ are positive at a critical point $\mathbf{x}_0$, then f has a local minimum at $\mathbf{x}_0$.
- If $f : \mathbb{R}^n \to \mathbb{R}$ has Hessian matrix $\mathbf{H}$, and all eigenvalues of $\mathbf{H}$ are negative at a critical point $\mathbf{x}_0$, then f has a local maximum at $\mathbf{x}_0$.

- A 2×2 symmetric matrix $\mathbf{A}$ has a positive definite quadratic form $\mathbf{v}^T\mathbf{A}\mathbf{v}$ if and only if $a_{11} > 0$ and $|\mathbf{A}| > 0$. Similarly, a 2×2 symmetric matrix has a negative definite quadratic form $\mathbf{v}^T\mathbf{A}\mathbf{v}$ if and only if $a_{11} < 0$ and $|\mathbf{A}| > 0$.

▶ EXERCISES

1. In each part, solve for all critical points for the given function. Then, for each critical point, use the Hessian matrix to determine whether the critical point is a local maximum, a local minimum, or neither.

★ **a)** $f(x, y) = x^3 + x^2 + 2xy - 3x + y^2$

b) $f(x, y) = 6x^2 + 4xy + 3y^2 + 8x - 9y$

★ **c)** $f(x, y) = 2x^2 + 2xy + 2x + y^2 - 2y + 5$

d) $f(x, y) = x^3 + 3x^2y - x^2 + 3xy^2 + 2xy - 3x + y^3 - y^2 - 3y$ (Hint: To solve for critical points, first set $\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y} = 0$.)

★ **e)** $f(x, y, z) = 2x^2 + 2xy + 2xz + y^4 + 4y^3z + 6y^2z^2 - y^2 + 4yz^3 - 4yz + z^4 - z^2$

a) Show that $f(x, y) = (x - 2)^4 + (y - 3)^2$ has a local minimum at $[2, 3]$, but its Hessian matrix at $[2, 3]$ has 0 as an eigenvalue.

b) Show that $f(x, y) = -(x - 2)^4 + (y - 3)^2$ has a critical point at $[2, 3]$, its Hessian matrix at $[2, 3]$ has all nonnegative eigenvalues, but $[2, 3]$ is not a local extreme point for f.

c) Show that $f(x, y) = -(x+1)^4 - (y+2)^4$ has a local maximum at $[-1, -2]$, but its Hessian matrix at $[-1, -2]$ is $\mathbf{O}$ and thus has all of its eigenvalues equal to zero.

d) Show that $f(x, y, z) = (x-1)^2 - (y-2)^2 + (z-3)^4$ does not have any local extreme points. Then verify that its Hessian matrix has eigenvalues of opposite sign at the function's only critical point.

a) Prove that a symmetric 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ defines a positive definite quadratic form if and only if $a > 0$ and $|\mathbf{A}| > 0$. (Hint: Compute $p_{\mathbf{A}}(x)$ and show that both roots are positive if and only if $a > 0$ and $|\mathbf{A}| > 0$.)

b) Prove that a symmetric 2×2 matrix $\mathbf{A}$ defines a negative definite quadratic form if and only if $a_{11} < 0$ and $|\mathbf{A}| > 0$.

★ **2.** True or False:

a) If $f : \mathbb{R}^n \to \mathbb{R}$ has continuous second partial derivatives, then the Hessian matrix is symmetric.

b) Every symmetric matrix $\mathbf{A}$ defines either a positive definite or a negative definite quadratic form.

c) A Hessian matrix for a function with continuous second partial derivatives evaluated at any point is diagonalizable.

d) $\mathbf{v}^T \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \mathbf{v}$ is a positive definite quadratic form.

e) $\mathbf{v}^T \begin{bmatrix} 3 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & 4 \end{bmatrix} \mathbf{v}$ is a positive definite quadratic form.

▶ Answers to Selected Exercises

(1) (a) Critical points: $(1,-1)$, $(-1,1)$; local minimum at $(1,-1)$

(c) Critical point: $(-2,3)$; local minimum at $(-2,3)$

(e) Critical points: $(0,0,0)$, $(\frac{1}{2},-\frac{1}{2},-\frac{1}{2})$, $(-\frac{1}{2},\frac{1}{2},\frac{1}{2})$; local minimums at $(\frac{1}{2},-\frac{1}{2},-\frac{1}{2})$, $(-\frac{1}{2},\frac{1}{2},\frac{1}{2})$

(4) (a) T

(b) F

(c) T

(d) T

(e) F