Jim Lambers
MAT 419/519
Summer Session 2011-12
Lecture 3 Notes

These notes correspond to Section 1.3 in the text.

Positive and Negative Definite Matrices and Optimization

The following examples illustrate that in general, it cannot easily be determined whether a symmetric matrix is positive definite from inspection of the entries.

Example Consider the matrix

$$A = \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix}.$$

Then

$$Q_A(x, y) = x^2 + y^2 + 8xy$$

and we have

$$Q_A(1, -1) = 1^2 + (-1)^2 + 8(1)(-1) = 1 + 1 - 8 = -6 < 0.$$

Therefore, even though all of the entries of A are positive, A is not positive definite. □

Example Consider the matrix

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}.$$

Then

$$Q_A(x, y) = x^2 + 4y^2 - 2xy = x^2 - 2xy + y^2 + 3y^2 = (x - y)^2 + 3y^2$$

which can be seen to be always nonnegative. Furthermore, $Q_A(x, y) = 0$ if and only if $x = y$ and $y = 0$, so for all nonzero vectors (x, y), $Q_A(x, y) > 0$ and A is positive definite, even though A does not have all positive entries. □

Example Consider the *diagonal matrix*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Then

$$Q_A(x, y, z) = x^2 + 3y^2 + 2z^2,$$

which can plainly be seen to be positive except when $(x, y, z) = (0, 0, 0)$. Therefore, A is positive definite. □

The preceding example can be generalized as follows: if A is an $n \times n$ diagonal matrix

$$A = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{bmatrix},$$

then A is:

1. positive definite if and only if $d_i > 0$ for $i = 1, 2, \ldots, n$,
2. negative definite if and only if $d_i < 0$ for $i = 1, 2, \ldots, n$,
3. positive semidefinite if and only if $d_i \geq 0$ for $i = 1, 2, \ldots, n$,
4. negative semidefinite if and only if $d_i \leq 0$ for $i = 1, 2, \ldots, n$,
5. indefinite if and only if $d_i > 0$ for some indices i, $1 \leq i \leq n$, and negative for other indices.

We now consider a general 2×2 symmetric matrix

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

This matrix induces the quadratic form

$$Q_A(x, y) = ax^2 + 2bxy + cy^2.$$

If $y = 0$, then we have $Q_A(x, 0) = ax^2$, so we must certainly have $a > 0$ in order for A to be positive definite. If $y \neq 0$, then we use a change of variable $x = ty$ and obtain

$$Q_A(x, y) = Q_A(ty, y) = at^2y^2 + 2bty^2 + cy^2 = (at^2 + 2bt + c)y^2.$$

Thus $Q_A(x, y)$ is positive definite if and only if $\varphi(t) = at^2 + 2bt + c$ is positive.

From

$$\varphi'(t) = 2at + 2b, \quad \varphi''(t) = 2a,$$

we obtain the single critical point $t^* = -b/a$ and determine that it is a strict global minimizer of $\varphi(t)$. The minimum value is

$$\varphi(t^*) = a(-b/a)^2 + 2b(-b/a) + c = c - \frac{b^2}{a} = \frac{1}{a}\det(A).$$

We conclude that A is positive definite if and only if $a > 0$ and $\det(A) > 0$. This leads to the following theorem.

Theorem A 2×2 symmetric matrix A is

1. positive definite if and only if $a > 0$ and $\det(A) > 0$
2. negative definite if and only if $a < 0$ and $\det(A) > 0$
3. indefinite if and only if $\det(A) < 0$

A similar argument, combined with mathematical induction, leads to the following generalization.

Theorem Let A be an $n \times n$ symmetric matrix, and let A_k be the submatrix of A obtained by taking the upper left-hand corner $k \times k$ submatrix of A. Furthermore, let $\Delta_k = \det(A_k)$, the kth *principal minor* of A. Then

1. A is positive definite if and only if $\Delta_k > 0$ for $k = 1, 2, \ldots, n$;
2. A is negative definite if and only if $(-1)^k \Delta_k > 0$ for $k = 1, 2, \ldots, n$;
3. A is positive semidefinite if $\Delta_k > 0$ for $k = 1, 2, \ldots, n-1$ and $\Delta_n = 0$;
4. A is negative semidefinite if $(-1)^k \Delta_k > 0$ for $k = 1, 2, \ldots, n-1$ and $\Delta_n = 0$.

Note that the last two statements in the theorem are *not* "if and only if" statements. Furthermore, if A has nonnegative principal minors, then A is *not* necessarily positive semidefinite. For example, if

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & \frac{1}{2} \end{bmatrix}.$$

All of the principal minors are nonnegative, but $(1, 1, -2) \cdot A(1, 1, -2) < 0$, so A is not positive semidefinite; it is actually indefinite.

We now show how the preceding discussion can be applied to find minimizers and maximizers of functions.

Example Let

$$f(x, y, z) = x^2 + y^2 + z^2 - xy + yz - xz.$$

We have

$$\nabla f(x, y, z) = (2x - y - z, 2y - x + z, 2z + y - x).$$

Solving $\nabla f(x, y, z) = (0, 0, 0)$ yields only the trivial solution $(0, 0, 0)$. We also have

$$Hf(x, y, z) = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix},$$

which can be confirmed to be positive definite by examinination of the principal minors: $\Delta_1 = 2$, $\Delta_2 = 3$, $\Delta_3 = 4$. It follows that the critical point $(0, 0, 0)$ is a strict global minimizer of $f(x, y, z)$. $\square$

Example Let
$$f(x, y) = e^{x-y} + e^{y-x}.$$
We have
$$\nabla f(x, y) = (e^{x-y} - e^{y-x}, -e^{x-y} + e^{y-x}),$$
which yields the critical points (x, x) for all $x \in \mathbb{R}$. We also have
$$Hf(x, y) = \begin{bmatrix} e^{x-y} + e^{y-x} & -e^{x-y} - e^{y-x} \\ -e^{x-y} - e^{y-x} & e^{x-y} + e^{y-x} \end{bmatrix},$$
which yields
$$\Delta_1 = e^{x-y} + e^{y-x} > 0, \quad \Delta_2 = (e^{x-y} + e^{y-x})^2 - (-e^{x-y} - e^{y-x})^2 = 0.$$
That is, $Hf(x, x)$ is positive semidefinite, making (x, x) a global minimizer of $f(x, y)$. $\square$

Example Let
$$f(x, y, z) = e^{x-y} + e^{y-x} + e^{x^2} + z^2.$$
We have
$$\nabla f(x, y, z) = (e^{x-y} - e^{y-x} + 2xe^{x^2}, -e^{x-y} + e^{y-x}, 2z)$$
which yields the conditions $z = 0$, $x = y$ and $x = 0$ for a critical point. Therefore, $(0, 0, 0)$ is the only critical point. We then have
$$Hf(x, y, z) = \begin{bmatrix} e^{x-y} + e^{y-x} + 4x^2e^{x^2} + 2e^{x^2} & -e^{x-y} - e^{y-x} & 0 \\ -e^{x-y} - e^{y-x} & e^{x-y} + e^{y-x} & 0 \\ 0 & 0 & 2 \end{bmatrix},$$
and therefore
$$Hf(0, 0, 0) = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$
which is positive dfinite, in view of $\Delta_1 = 3$, $\Delta_2 = 2$, and $\Delta_3 = 4$. It can also be shown by direct computation of the minors that $Hf(x, y, z)$ is positive definite on all of $\mathbb{R}^3$, using the fact that $e^x > 0$ for all x. Therefore, the critical point $(0, 0, 0)$ is a strict global minimizer of $f(x, y, z)$. $\square$

Example Let
$$f(x, y) = e^{x-y} + e^{x+y}.$$

We have

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

which yields no critical points, as $\partial f/\partial x$ can never be zero. Therefore, there are no global minimizers. $\square$

We now consider how the Hessian can be used to establish the existence of a *local* minimizer or maximizer.

Theorem Suppose that $f(\mathbf{x})$ has continuous first and second partial derivatives on a set $D \subseteq \mathbb{R}^n$. Let $\mathbf{x}^*$ be an interior point of D that is a critical point of $f(\mathbf{x})$. Then $\mathbf{x}^*$ is a:

1. strict local minimizer of $f(\mathbf{x})$ if $Hf(\mathbf{x}^*)$ is positive definite;

2. strict local maximizer of $f(\mathbf{x})$ if $Hf(\mathbf{x}^*)$ is negative definite.

This theorem can be proved by using the continuity of the second partial derivatives to show that $Hf(\mathbf{x})$ is positive definite for $\mathbf{x}$ sufficiently close to $\mathbf{x}^*$, and then applying the multi-variable generalization of Taylor's Formula.

We now consider the implications of an indefinite Hessian at a critical point. Suppose that $f(\mathbf{x})$ has continuous second partial derivatives on a set $D \subseteq \mathbb{R}^n$. Furthermore, let $\mathbf{x}^*$ be an interior point of D that is a critical point of $f(\mathbf{x})$. If $Hf(\mathbf{x}^*)$ is indefinite, then there exist vectors $\mathbf{y}$, $\mathbf{w}$ such that

$$\mathbf{y} \cdot Hf(\mathbf{x}^*)\mathbf{y} > 0, \quad \mathbf{w} \cdot Hf(\mathbf{x}^*)\mathbf{w} < 0.$$

By continuity of the second partial derivatives, there exists an $\epsilon > 0$ such that

$$\mathbf{y} \cdot Hf(\mathbf{x}^* + t\mathbf{y})\mathbf{y} > 0, \quad \mathbf{w} \cdot Hf(\mathbf{x}^* + t\mathbf{w})\mathbf{w} < 0$$

for $|t| < \epsilon$. If we define

$$Y(t) = f(\mathbf{x}^* + t\mathbf{y}), \quad W(t) = f(\mathbf{x}^* + t\mathbf{w}),$$

then $Y'(0) = W'(0) = 0$, while $Y''(0) > 0$ and $W''(0) < 0$. Therefore, $t = 0$ is a strict local minimizer of $Y(t)$ and a strict local maximizer of $W(t)$.

It follows that $\mathbf{x}^*$ is a strict local minimizer along a curve contained in the graph of f that passes through $\mathbf{x}^*$ in the direction of $\mathbf{y}$, and a strict local maximizer along a curve contained in the graph of f that passes through $\mathbf{x}^*$ in the direction of $\mathbf{w}$. This gives the graph of f the appearance of a saddle near $\mathbf{x}^*$, which leads to the following definition.

Definition A *saddle point* for $f(\mathbf{x})$ is a critical point $\mathbf{x}^*$ for $f(\mathbf{x})$ such that there are vectors $\mathbf{y}$, $\mathbf{w}$ for which $t = 0$ is a strict local minimizer for $Y(t) = f(\mathbf{x}^* + t\mathbf{y})$ and a strict local maximizer for $W(t) = f(\mathbf{x}^* + t\mathbf{w})$.

The result of the preceding discussion is summarized in the following theorem.

Theorem If $f(\mathbf{x})$ is a function with continuous second partial derivatives on a set $D \subseteq \mathbb{R}^n$, if $\mathbf{x}^*$ is an interior point of D that is also a critical point of $f(\mathbf{x})$, and if $Hf(\mathbf{x}^*)$ is indefinite, then $\mathbf{x}^*$ is a saddle point of $\mathbf{x}^*$.

Example Let
$$f(x, y) = x^3 - 12xy + 8y^3.$$
We have
$$\nabla f(x, y) = (3x^2 - 12y, -12x + 24y^2),$$
which yields the critical points $(0, 0)$ and $(2, 1)$. We also have
$$Hf(x, y) = \begin{bmatrix} 6x & -12 \\ -12 & 48x \end{bmatrix},$$
and therefore
$$Hf(0, 0) = \begin{bmatrix} 0 & -12 \\ -12 & 0 \end{bmatrix}, \quad Hf(2, 1) = \begin{bmatrix} 12 & -12 \\ -12 & 48 \end{bmatrix}.$$
We see that $Hf(2, 1)$ is positive definite, because its principal minors are positive, but $Hf(0, 0)$ is not, as $\Delta_1 = 0$ and $\Delta_2 = -144$. That is, $Hf(0, 0)$ is indefinite, so $(0, 0)$ is a saddle point.

Furthermore,
$$\lim_{x\to\infty} f(x, y) = +\infty, \quad \lim_{x\to-\infty} f(x, y) = -\infty,$$
so $f(x, y)$ has no global minimizer on $\mathbb{R}^2$. We can conclude, however, that $(2, 1)$ is a strict local minimizer. □

It should be emphasized that if the Hessian is positive semidefinite or negative semidefinite at a critical point, then it cannot be concluded that the critical point is necessarily a minimizer, maximizer or saddle point of the function.

Example Let $f(x, y) = x^4 - y^4$. We have
$$\nabla f(x, y) = (4x^3, -4y^3),$$
which yields the critical point $(0, 0)$. We then have
$$Hf(x, y) = \begin{bmatrix} 12x^2 & 0 \\ 0 & -12y^2 \end{bmatrix}.$$
Therefore $Hf(0, 0)$ is the zero matrix, which is positive semidefinite. However, $f(x, y)$ increases from $(0, 0)$ along the x-direction, and decreases along the y-direction, so $(0, 0)$ is neither a local minimizer nor maximizer. □

Example Let $f(x, y) = x^4 + y^4$. As in the previous example, $Hf(0, 0)$ is the zero matrix, but it can be seen from a graph that $(0, 0)$ is a strict global minimizer. □

Exercises

1. Chapter 1, Exercise 2
2. Chapter 1, Exercise 6
3. Chapter 1, Exercise 7
4. Chapter 1, Exercise 9
5. Chapter 1, Exercise 11
6. Chapter 1, Exercise 19