# ARE Math Review Lecture 10: Quadratic Forms and Definite Matrices

Stephen M. Stohs

August 24, 2004

## 1 Quadratic Forms

Quadratic forms play a central role in optimization theory, by providing the key element in the multi-dimensional generalization of a second derivative test of whether a point is a maximum or a minimum. In this section, we review the definition of a quadratic form, then develop a classification system which leads to the higher dimension equivalent of a second derivative test for an interior maximum or minimum of a continuously differentiable function.

### 1.1 Review of the Definition of a Quadratic Form

We have already seen quadratic forms as an example of a type of function which generalizes the notion of a quadratic function $f : R^2 \to R^1$ to the case of functions from $R^n$ to $R^1$. We review this definition here, then see how certain properties of the matrix which defines a quadratic form can provide information about the extreme values and the curvature of the graph of the quadratic form.

A quadratic form may generally be written as

$$Q(x_1, x_2, \ldots, x_n) = \sum_{i \leq j} a_{ij} x_i x_j = \mathbf{x}^T A \mathbf{x}, \tag{1}$$

with

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

and

$$A = \begin{bmatrix} a_{11} & \frac{1}{2}a_{12} & \frac{1}{2}a_{13} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \frac{1}{2}a_{23} & \cdots & \frac{1}{2}a_{2n} \\ \frac{1}{2}a_{13} & \frac{1}{2}a_{23} & a_{33} & \cdots & a_{3n} \\ \frac{1}{2}a_{14} & \frac{1}{2}a_{24} & \frac{1}{2}a_{34} & \cdots & \frac{1}{2}a_{4n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{1n} & \frac{1}{2}a_{2n} & \frac{1}{2}a_{3n} & \cdots & a_{nn} \end{bmatrix}.$$

Exercises:

1. Write out enough terms in the summation expression for the quadratic form (as shown above) to clearly establish the pattern in the expansion.
2. What happens to the factors of $\frac{1}{2}$ when converting from the quadratic form in matrix notation to the expanded form?
3. Suppose you want to express the same quadratic form as shown above, except that you want to choose $A$ to be upper triangular. What would be the entries of $A$ in this case?
4. If the matrix $A$ in a quadratic form is chosen as a symmetric matrix, then it is uniquely determined for any particular quadratic form. Is the upper triangular version of $A$ unique[1]?

# 2 Definiteness of Quadratic Forms

## 2.1 Definiteness Definitions

We now consider a list of properties which classify quadratic forms in a manner which reveals whether they achieve global extreme values (global maxima or minima) on their domain[2]. To do this, we first need the following definitions:

A quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is said to be

(a) **positive definite** if $\mathbf{x}^T A\mathbf{x} > 0$ for all $\mathbf{x} \in R^n$ such that $\mathbf{x} \neq \mathbf{0}$;

[1] To prove uniqueness, assume there is some other upper-triangular matrix $B$ which is equal to $Q(x_1, \ldots, x_n)$. Use matrix algebra to answer the question of whether $A$ and $B$ must be equal.

[2] A global maximum is a value $\mathbf{x}_{max}$ such that $Q(\mathbf{x}) \leq Q(\mathbf{x}_{max})$ for all $\mathbf{x} \in R^n$. Similarly, a global mininum $\mathbf{x}_{min}$ satisfies $Q(\mathbf{x}) \geq Q(\mathbf{x}_{min})$ for all $\mathbf{x} \in R^n$.

(b) **positive semidefinite** if $\mathbf{x}^T A\mathbf{x} \geq 0$ for all $\mathbf{x} \in R^n$ such that $\mathbf{x} \neq \mathbf{0}$;

(c) **negative definite** if $\mathbf{x}^T A\mathbf{x} < 0$ for all $\mathbf{x} \in R^n$ such that $\mathbf{x} \neq \mathbf{0}$;

(d) **negative semidefinite** if $\mathbf{x}^T A\mathbf{x} \leq 0$ for all $\mathbf{x} \in R^n$ such that $\mathbf{x} \neq \mathbf{0}$;

(e) **indefinite** if $\mathbf{x}^T A\mathbf{x}$ takes on both positive and negative values as $\mathbf{x}$ ranges over all possible vectors in $R^n$.

A few comments are warranted:

1. Note that "semidefinite" is a weak version of "definite." So, for instance, all positive definite matrices are also positive semidefinite, but there are positive semidefinite matrices which are not positive definite.

2. The above categories are thus not mutually exclusive. However, they are inclusive – that is, they cover all possible cases. To classify a matrix by definiteness status, we agree to use one of the stronger (definite) criteria if applicable. Otherwise, we we ask whether one of the two weaker (semidefinite) criteria applies. A matrix which satisfies none of the strong or weak definite criteria is classified as indefinite.

Exercises:

1. Suppose that a quadratic form $Q(\mathbf{x})$ attains a global maximum at a point $\mathbf{x}_{max} \in R^n$. Let $X_{max} \equiv \{\mathbf{y} \in R^n : Q(\mathbf{y}) = Q(\mathbf{x}_{max})\}$. Prove that $\mathbf{0} \in X_{max}$.

2. Prove that $X_{max} \equiv \{\mathbf{0}\}$ if and only if the matrix $A$ in the quadratic form is negative definite.

3. Prove that if $\mathbf{0}$ does not maximize $Q(\mathbf{x})$, then $Q(\mathbf{x})$ does not have a maximum value.

4. Do the above assertions hold with "minimum" and "minimize" in place of "maximum" and "maximize"?

## 2.2 Simple Illustrations

Simon and Blume present simple cases which help to provide insight:

1. $Q_1(x_1, x_2) = x_1^2 + x_2^2$;

2. $Q_2(x_1, x_2) = -x_1^2 - x_2^2$;

3. $Q_3(x_1, x_2) = x_1^2 - x_2^2$;

4. $Q_4(x_1, x_2) = (x_1 + x_2)^2$;

5. $Q_5(x_1, x_2) = -(x_1 + x_2)^2$.

Rather than reproduce their graphs, I refer the reader to the illustrations of Figures 16.2 through 16.6. In order to gain a depth of understanding, perform the following simple exercises:

1. For each of the above quadratic forms, write the matrix $A_i$ needed to express the quadratic form as $Q_i(\mathbf{x}) = \mathbf{x}^T A_i \mathbf{x}$, for $i = 1, 2, \ldots, 5$.

2. Classify the matrices by definiteness type, noting the correspondence between definiteness category and the entries in the matrix for each case.

3. Figures 16.5 and 16.6 in Simon and Blume are not oriented correctly. Redraw them with the correct orientation. (Hint: Identify all values of $\mathbf{x} = (x_1, x_2)$ for which $Q_i(\mathbf{x}) = 0$, for the cases $i = 4, 5$).

## 2.3 Definiteness Tests

Next we address the question of whether there is a general method to determine the definiteness status of a matrix. The answer is yes, and by carrying out such a test, we can quickly determine in a given case whether the associated quadratic form attains a global extreme value and if so, whether the extremum is a maximum or minimum and whether it is unique.

The test is easiest when performed by a computer. Nonetheless, it is instructive to understand how it works. Before doing so, we must first clear some definitional underbrush.

### 2.3.1 Leading Principle Minors

The following definitions are needed to describe the definiteness test procedure. Let $A$ be an $n \times n$ square matrix.

1. A **principal submatrix** of order $k$ is formed by deleting the entries in $n - k$ rows of $A$ and the entries *in the same* $n - k$ columns.

2. A **principal minor** of order $k$ is the determinant of a principal submatrix of order $k$.

3. The $k^{th}$ order **leading principal submatrix**, denoted $A_k$, is formed by deleting the last $n-k$ rows and columns of $A$.

4. The $k^{th}$ order **leading principal minor** of $A$, denoted $|A_k|$, is the determinant of the $k^{th}$ order leading principal submatrix.

### 2.3.2 The Test Procedure

The recipe is tedious to carry out by hand, but straightforward. We consider the signs on the list of leading principal minors $|A_1|, |A_2|, \ldots, |A_n| = |A|$.

1. $A$ is *positive definite* if $|A_i| > 0$ for $i = 1, 2, \ldots, n$.

2. $A$ is *positive semidefinite* if $|A_i| \geq 0$ for $i = 1, 2, \ldots, n$.

3. $A$ is *negative definite* if $(-1)^i |A_i| > 0$ for $i = 1, 2, \ldots, n$.

4. $A$ is *negative semidefinite* if $(-1)^i |A_i| \geq 0$ for $i = 1, 2, \ldots, n$.

5. $A$ is *indefinite* for any pattern of signs which does not fit one of the above cases.

Exercises:

1. Describe the pattern of signs for the negative definite case.

2. Directly describe the pattern of signs for the indefinite case.

# 3 Linear Constraints and Bordered Matrices

## 3.1 The Context of Constrained Optimization

Many problems encountered in economics involve maximizing or minimizing some objective function subject to a set of constraints. In the present context, we consider objective functions which are quadratic forms

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}, \tag{2}$$

and constraints which consist of a system of linear equations

$$Bx = 0. \tag{3}$$

We can easily visualize the situation in three dimensional Euclidean space. The quadratic form defines a surface with a shape such as those shown in Figures 16.2 through 16.6. A single linear constraint in three dimensions restricts the problem to a line through the origin

In general, the linear restrictions describe a subspace which necessarily contains the origin.. We then seek extreme values along the intersection of the vertical plane through this line with the graph of the quadratic form.

Exercises: Consider the quadratic form $Q_3(x_1, x_2) = x_1^2 - x_2^2$.

1. If we impose the linear restriction $x_1 = 0$, does this quadratic form attain an extreme value? If so, describe it.
2. Now consider the same quadratic form, but subject instead to the restriction $x_2 = 0$. Describe the extreme value, if there is one.
3. Next use the sole linear restriction $x_1 = x_2$. Is there an extreme value in this case? If so, describe it.
4. In light of the three examples considered above, is it possible to have extrema for a quadratic form subject to a set of linear constraints which are not extrema in the unrestricted case? Does imposing restrictions always guarantee the existence of extrema?

## 3.2 Definiteness Tests for Quadratic Forms Subject to Linear Constraints

Now we turn to the algorithm for definiteness testing of a quadratic form subject to a set of linear restrictions.

Let $A$ be an $n \times n$ matrix and $B$ be an $n \times m$ matrix. Consider the problem of seeking global extrema for

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} \tag{4}$$

on the **constraint set** described by

$$B\mathbf{x} = \mathbf{0}. \tag{5}$$

Form the bordered (partitioned) matrix

$$H = \begin{bmatrix} 0 & B \\ B^T & A \end{bmatrix}.$$

Compute the last $n-m$ leading principal minors in reverse (largest to smallest) order $|H| = |H_{n+m}|, |H_{n+m-1}|, \ldots, |H_{(n+m)-(n-m-1)}| = |H_{2m+1}|$. Consider the signs of these leading principal minors.

1. If the signs alternate and $|H|$ has the sign $(-1)^n$, then $Q$ is *negative definite* on the constraint set $B\mathbf{x} = \mathbf{0}$, and $Q$ attains a unique global maximum for $\mathbf{x} = \mathbf{0}$;

2. If the signs are all the same and match the sign of $(-1)^m$, then $Q$ is *positive definite* on the constraint set, and $Q$ attains a unique global minimum at $\mathbf{x} = \mathbf{0}$;

3. If either of the above patterns are violated by nonzero leading principal minors, then $Q$ is *indefinite* on the constraint set and $\mathbf{x} = \mathbf{0}$ is neither a max nor a min of $Q$ on the constraint set.

Exercise: The semidefinite cases are missing from the above classification. How would you guess we could include them?

# 4 The Big Picture

We are not covering a unit on optimization theory this year, but nonetheless, it seems like a brief discussion on how the material of this unit fits into that framework might help motivate its relevance.

We have a twice continuously differentiable function (i.e., of order $C^2$) $f : R^m \to R^1$ which is differentiable in all of its arguments up to order three. We want to find candidates for locally extreme values of $f$, that is, points $\mathbf{x}_0$ such that $f(\mathbf{x}_0)$ is either larger or smaller than the value of $f$ at all points in a surrounding ball $B_\varepsilon(\mathbf{x}_0)$. To do this, we first identify candidates for extrema by the first order conditions:

$$Df(\mathbf{x}) = 0. \tag{6}$$

Any point $\mathbf{x}$ which satisfies these necessary first order conditions[3] is a candidate for an interior extremum. Intuitively, to be at the top of a hill or the bottom of a valley, we require that it is impossible to change the value of the function by moving a small amount away from the top of the hill or bottom of the value. Because a function with a nonzero gradient will always increase (decrease) in value if we move a sufficiently small distance in the direction of (opposite) the gradient, we must have a zero gradient at an interior extremum.

Now recall that we can locally approximate a change in the value of a function by the differential:

$$f(\mathbf{x} + \mathbf{dx}) - f(\mathbf{x}) \approx Df(\mathbf{x}) \cdot \mathbf{dx}, \tag{7}$$

which in principle shows how the function changes as we move away from $\mathbf{x}$. However, if $\mathbf{x}$ is a candidate for an extremum, then the gradient condition implies the right hand side of the above is equal to 0! We thus learn nothing about the local behavior of a function from the differential approximation about a candidate for an extremum.

What is needed is a local approximation to the value of the function which takes into consideration the curvature. This is accomplished using *one of the most powerful tools in the Newtonian arsenal*: A second-order Taylor expansion. We write

$$f(\mathbf{x} + \mathbf{dx}) = f(\mathbf{x}) + Df(\mathbf{x}) \cdot \mathbf{dx} + \mathbf{dx}^T D^2 f(\mathbf{x}) \mathbf{dx} + o(\|\mathbf{dx}\|^2), \tag{8}$$

where $o(x)$ is a function which goes to zero faster than $x$:

$$\lim_{x \to 0} \frac{o(x)}{x} = 0. \tag{9}$$

What this means is that for sufficiently small $\|\mathbf{dx}\|$, the remainder term in the second-order Taylor expansion is negligibly small compared to the quadratic form.

Now suppose we have identified a point $\mathbf{x}_0$ which satisfies the first order conditions for an extremum. Using the fact that $Df(\mathbf{x}_0) = \mathbf{0}$, we may rearrange the second-order Taylor expansion to write

$$f(\mathbf{x_0} + \mathbf{dx}) - f(\mathbf{x_0}) \approx \mathbf{dx}^T D^2 f(\mathbf{x_0}) \mathbf{dx}. \tag{10}$$

[3]This discussion only pertains to the case of interior extrema. The Kuhn-Tucker conditions consider the possibility of extreme values on the boundary of the domain, but this is beyond the scope of the discussion.

This says that the change in the value of $f(\cdot)$ in a local neighborhood of a candidate for an extremum is closely approximated by a quadratic form in the Hessian matrix. This leads us directly to a test for whether the function attains a local extremum based on the definite type of the Hessian.

Exercises:

1. Use the second-order Taylor expansion with remainder to prove that a $C^2$ function $f : R^n \to R^1$ attains a unique local maximum at an interior point $\mathbf{x}_0$ if and only if $Df(\mathbf{x}_0) = \mathbf{0}$ and $D^2 f(\mathbf{x}_0)$ is negative definite.
2. Revise the above assertion and proof thereof for the case of a local minimum at $\mathbf{x}_0$.
3. What can we say if $D^2 f(\mathbf{x}_0)$ is merely positive (negative) semidefinite?
4. Prove that if $D^2 f(\mathbf{x}_0)$ is indefinite, then $\mathbf{x}_0$ is a saddle point[4].

[4] A **saddle point** is a point which satisfies the necessary first order conditions but is neither a local maximum nor a local minimum of the function in question.