Quadratic Forms

Wang Xiao

Shanghai University of Finance and Economics

May 10, 2026

The history of concerning quadratic forms is backed to ancient Greece and India, it is of greatest importance due to its connection to the shape of orbits of planets.

- Representation of integers as sums of two squares, motivated by Pythagoras theorem and geometry.
- Solution of Pell's equation. Pell's equation was considered by the Indian mathematician Brahmagupta in the 7th century CE. The Pell's equation, or called Pell-Fermat equation is

$$x^2 - ny^2 = 1$$

  where n is a given positive nonsquare integer, and integer solutions are sought for x and y.

- Specific Pell's equation was considered 400 BC in Greece. The positive integer solutions for
$$x^2 - 2y^2 = 1 \quad \text{or } x^2 - ny^2 = -1$$
can be used to approximate $\sqrt{2}$, i.e., using $\frac{x}{y}$. For example, $x = 17, y = 12$ and $x = 577, y = 408$ will give two approximations, $17/12$ and $577/408$, of $\sqrt{2}$.
- Lagrange proved that, as long as n is not a perfect square, Pell's equation has infinitely many distinct integer solutions. Actually, William Brouncker was the first European to solve this equation, but Euler mistakenly attributed Brouncker's solution to John Pell.
- In general, a binary quadratic form is a quadratic homegeneous polynomial in two variables
$$Q(x, y) = ax^2 + bxy + cy^2$$
the coefficients, of course, can go beyond integral numbers.

- Taylor's theorem provides a way to approximate a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ near a point $\mathbf{x}_0$ using polynomials. The expansion to second order is

$$f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top \mathbf{v} + \frac{1}{2}\mathbf{v}^\top \nabla^2 f(\mathbf{x}_0)\mathbf{v} + o(\|\mathbf{v}\|^2)$$

where $\nabla^2 f(\mathbf{x}_0)$ is the Hessian matrix with ij-entries

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0)$$

- The Hessian matrix defines the quadratic form $\mathbf{v}^\top H(\mathbf{x}_0)\mathbf{v}$ in the vector space with $\mathbf{x}_0$ as the origin. The positive and negative definite property encodes the local minimum or local maximum of f at $\mathbf{x}_0$.

Conics and Quadratic Surfaces

In this part, we will use the spectrual theorem to analyze the equations of conic sections and quadratic surfaces. Suppose we are given the quadratic equation

$$x_1^2 + 4x_1x_2 - 2x_2^2 = 6$$

Notice that we can write the quadratic expression

$$x_1^2 + 4x_1x_2 - 2x_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x}^\top A\mathbf{x}$$

where

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

is the symmetric matrix. To diagonalize A, we have to finish following three steps:

1. Find the eigenvalues of A. Set

$$\det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{bmatrix} == \lambda^2 + \lambda - 6 = 0$$

The eigenvalues are

$$\lambda_1 = 2, \quad \lambda_2 = -3.$$

2. Find the eigenvectors.

For $\lambda_1 = 2$,

$$\begin{bmatrix} 1-2 & 2 \\ 2 & -2-2 \end{bmatrix} \mathbf{v} = \mathbf{0}$$

Then

$$-v_1 + 2v_2 = 0,$$

and an eigenvector corresponding to $\lambda_1 = 2$ is $(2, 1)^\top$.

For $\lambda_2 = -3$,

$$\begin{bmatrix} 1-(-3) & 2 \\ 2 & -2-(-3) \end{bmatrix} \mathbf{v} = \mathbf{0}$$

gives

$$4v_1 + 2v_2 = 0$$

and then an eigenvector corresponding to $\lambda_2 = -3$ is $(-1, 2)^\top$
Thus, we have that

$$A = Q\Lambda Q^\top$$

where

$$Q = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \quad \text{and } \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}.$$

If we make the substitution $\mathbf{y} = Q^\top \mathbf{x}$, then we have

$$\mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top (Q\Lambda Q^\top)\mathbf{x} = (Q^\top \mathbf{x})^\top \Lambda (Q^\top \mathbf{x}) = \mathbf{y}^\top \Lambda \mathbf{y} = 2y_1^2 - 3y_2^2.$$

Note that the conic is much easier to understand in the $y_1 y_2$-coordinates.

The same trick can be used for any quadratic equation

$$\alpha x_1^2 + 2\beta x_1 x_2 + \gamma x_2^2 = \delta,$$

where $\alpha, \beta, \gamma, \delta$ are real numbers. Now we set

$$A = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix},$$

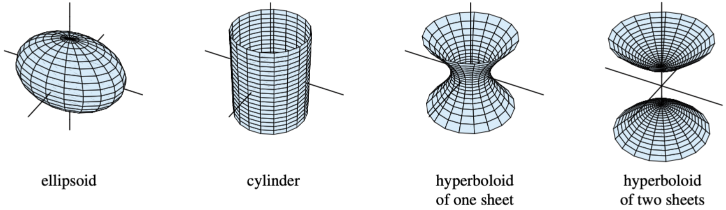
so that the equation can be written as

$$\mathbf{x}^\top A \mathbf{x} = \delta.$$

Since A is symmetric, we can find a diagonal matrix Λ and an orthogonal matrix Q so that $A = Q\Lambda Q^\top$. Thus, setting $\mathbf{y} = Q^\top \mathbf{x}$, we can rewrite equation as

$$\mathbf{y}^\top \Lambda \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 = \delta.$$

From this expression, we can infer that the "shape" of a quadratic form as a two-variable function is determined by the signs of eigenvalues of A. Quadratic surfaces include those shown in the following figure.



Consider the surface defined by the equation

$$2x_1x_2 + 2x_1x_3 + x_2^2 + x_3^2 = 2.$$

Observe that if

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

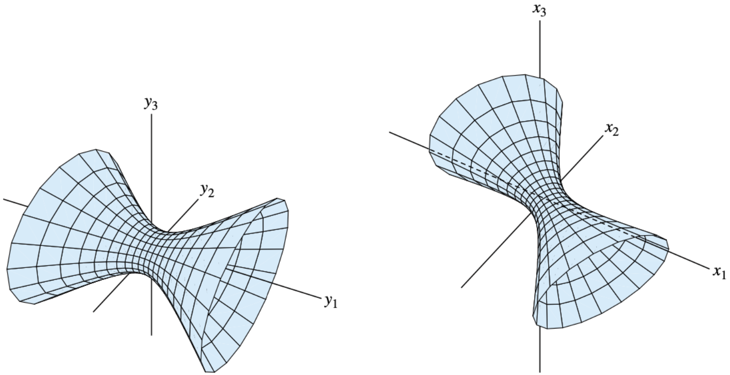
is the symmetric matrix, then

$$\mathbf{x}^\top A\mathbf{x} = 2x_1x_2 + 2x_1x_2 + x_2^2 + x_3^2,$$

and so we use the diagonalization and the substitution $\mathbf{y} = Q^\top \mathbf{x}$ as before to write

$$\mathbf{x}^\top A\mathbf{x} = \mathbf{y}^\top \Lambda \mathbf{y}, \quad \text{where } \Lambda = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

From the **y**-coordinates we can see that the graph is the hyperboloid of one sheet.



Quadratic Forms

Definition

A *quadratic form* in n variables $x_1, ..., x_n$ is a homogeneous second-degree polynomial in these variables. So the polynomial has the form

$$Q(x_1, ..., x_n) = \sum_{i,j}^{n} q_{ij} x_i x_j.$$

Every quadratic form can be viewed as a function of the vector $\mathbf{x} = x_1\mathbf{v}_1 + ... + x_n\mathbf{v}_n$, where $\mathbf{v}_1, ..., \mathbf{v}_n$ is some fixed basis of the vector space V of degree n.

Definition

A function $B(\mathbf{x}, \mathbf{y})$ that assigns to two vectors $\mathbf{x}, \mathbf{y} \in V$ a scalar value is called a *bilinear form* on V if it is linear in each of its arguments. In other words, the following conditions must be satisfied for all vectors of the space V and scalars c.

$$\begin{aligned} B(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) &= B(\mathbf{x}_1, \mathbf{y}) + B(\mathbf{x}_2, \mathbf{y}) \\ B(c\mathbf{x}, \mathbf{y}) &= cB(\mathbf{x}, \mathbf{y}) \\ B(\mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2) &= B(\mathbf{x}, \mathbf{y}_1) + B(\mathbf{x}, \mathbf{y}_2) \\ B(\mathbf{x}, c\mathbf{y}) &= cB(\mathbf{x}, \mathbf{y}). \end{aligned}$$

If $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is some basis of V, then we can express the bilinear form in terms of the coordinates of the vectors

$$B(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} = b_{ij} x_i y_i, \quad \text{where } b_{ij} = B(\mathbf{v}_i, \mathbf{v}_j).$$

In this case, the square matrix $B = (b_{ij})$ is called the matrix of the bilinear form B in the basis $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$.

Definition

A bilinear form $B(\mathbf{x}, \mathbf{y})$ is said to be *symmetric* if

$$B(\mathbf{x}, \mathbf{y}) = B(\mathbf{y}, \mathbf{x})$$

and *antisymmetric* if

$$B(\mathbf{x}, \mathbf{y}) = -B(\mathbf{y}, \mathbf{x}),$$

for all $\mathbf{x}, \mathbf{y} \in V$.

Theorem

Every quadratic form $Q(\mathbf{x})$ on the space V can be represented in the form $Q(\mathbf{x}) = B(\mathbf{x}, \mathbf{x})$, where B is a symmetric bilinear form, and moreover, for the given quadratic form Q, the bilinear form B is unique.

Proof.

To construct such a bilinear form B from a given quadratic form, we can do the following,

$$B(\mathbf{x}, \mathbf{y}) = \frac{1}{4}\left(Q(\mathbf{x}+\mathbf{y}) - Q(\mathbf{x}-\mathbf{y})\right)$$

□

The following result is for antisymmetric forms.

Theorem

For every antisymmetric bilinear form $Q(\mathbf{x}, \mathbf{y})$ on the space V, we have

$$Q(\mathbf{x}, \mathbf{x}) = 0.$$

Conversely, if above equality is satisfied for every vector in V, then the bilinear form $Q(\mathbf{x}, \mathbf{y})$ is antisymmetric.

Reduction to Canonical Form

Let $Q(\mathbf{x}, \mathbf{y})$ be a symmetric bilinear form defined on a finite dimensional vector space V.

Definition

Vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be orthogonal if $Q(\mathbf{x}, \mathbf{y}) = 0$.

The dimensionality and space decomposition property of V with respect to Q is the same as that for Euclidean inner product.

Theorem

For every quadratic form $\tilde{Q}(\mathbf{x})$, there exists a basis in which the form can be written as

$$\tilde{Q}(\mathbf{x}) = \lambda_1 x_1^2 + ... + \lambda_n x_n^2,$$

where $x_1, ..., x_n$ are the coordinates of the vector $\mathbf{x}$ in this basis.

Proof.

Let $Q(\mathbf{x}, \mathbf{y})$ be a symmetric bilinear form associated with the quadratic form $\tilde{Q}(\mathbf{x})$. If $\tilde{Q}(\mathbf{x})$ is identically zero, then the theorem is true. If the quadratic form $\tilde{Q}(\mathbf{x})$ is not identically zero, then there exists a vector $\mathbf{v}_1$ such that $\tilde{Q}(\mathbf{v}_1) = Q(\mathbf{v}_1, \mathbf{v}_1) \neq 0$. This implies that the restriction of the bilinear form Q to the subspace $V' = \langle \mathbf{v}_1 \rangle$ is nonsingular, and therefore, for the subspace $V' = \langle \mathbf{v}_1 \rangle$ we have the decomposition

$$V = \langle \mathbf{v}_1 \rangle \oplus \langle \mathbf{v}_1 \rangle^{\perp}.$$

Since $\dim \langle \mathbf{v}_1 \rangle = 1$, then we have that $\dim \langle \mathbf{v}_1 \rangle^{\perp} = n - 1$.
By induction, we may assume the theorem to have been proved for the space $\langle \mathbf{v}_1 \rangle^{\perp}$. Thus in this space there exists a basis $\mathbf{v}_2, \ldots, \mathbf{v}_n$ such that $Q(\mathbf{v}_i, \mathbf{v}_j) = 0$ for all $i \neq j \geq 2$. Then in the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$, the quadratic form $\tilde{Q}(\mathbf{x})$ can be written as the form in the theorem. □

Applications: Quadratic Forms and Max/Min Problems

Definition

Let $X \subset \mathbb{R}^n$, and let $\mathbf{a} \in X$. The function $f : X \to \mathbb{R}$ has a *global maximum* at $\mathbf{a}$ if $f(\mathbf{x}) \leq f(\mathbf{a})$; the function f has a *local maximum* at $\mathbf{a}$ if, for some $\delta > 0$, we have $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in B(\mathbf{a}, \delta) \cap X$. We say $\mathbf{a}$ is a local or global maximum point of f. The minimum can be defined analogously. If $\mathbf{a}$ is either a local maximum or local minimum point, we say it is an *extremum*.

Lemma

Suppose f is defined on some neighborhood of the extremum $\mathbf{a}$ and f is differentiable at $\mathbf{a}$. Then $Df(\mathbf{a}) = \mathbf{0}$, or, equivalently, $\nabla f(\mathbf{a}) = \mathbf{0}$.

Definition

Suppose f is differentiable at $\mathbf{a}$. We say $\mathbf{a}$ is a *critical point* if $Df(\mathbf{a}) = \mathbf{0}$. A critical point $\mathbf{a}$ with the property that $f(\mathbf{x}) < f(\mathbf{a})$ for some $\mathbf{x}$ near $\mathbf{a}$ and $f(\mathbf{x}) > f(\mathbf{a})$ for other $\mathbf{x}$ near $\mathbf{a}$ is called a *saddle point*.

Just as the second derivative test in single-variable calculus often allows us to differentiate between local minima and local maxima, there si something quite analogous in the multivariable case.

Lemma

Suppose $g : [0, 1] \to \mathbb{R}$ is twice differentiable. Then

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(\xi) \quad \textit{for some } 0 < \xi < 1.$$

The second derivative in the multivariable setting becomes a quadratic form.

Definition

Assume $f \in C^2$ in a neighborhood of $\mathbf{a}$. Define the symmetric matrix

$$\mathrm{Hess}(f)(\mathbf{a}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})\right].$$

Define the associated quadratic form $H_{f,\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}$ by

$$H_{f,\mathbf{a}}(\mathbf{h}) = \mathbf{h}^\top \left(\mathrm{Hess}(f)(\mathbf{a})\right) \mathbf{h} = \sum_{i,j}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) h_i h_j.$$

Proposition

Suppose $f : B(\mathbf{a}, r) \to \mathbb{R}$ *is* C^2. *Then for all* $\mathbf{h}$ *with* $\|h\| < r$ *we have*

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}H_{f,\mathbf{a}+\xi\mathbf{h}}(\mathbf{h}) \quad \textit{for some } 0 < \xi < 1.$$

Consequently,

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \frac{1}{2}H_{f,\mathbf{a}}(\mathbf{h}) + \boldsymbol{\epsilon}(\mathbf{h})$$

where

$$\boldsymbol{\epsilon}(\mathbf{h}) / \|h\|^2 \to 0 \quad \textit{as } \mathbf{h} \to \mathbf{0}.$$

Proof

Using the chain rule,

$$g'(t) = Df(\mathbf{a} + t\mathbf{h})\mathbf{h} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a} + t\mathbf{h})h_i$$

$$g''(t) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a} + t\mathbf{h})h_j \right) h_i = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a} + t\mathbf{h})h_i h_j = H_{f,\mathbf{a}+t\mathbf{h}}.$$

Substitution yields the first result.

Proof. cont.

Since f is C^2, given any $\epsilon > 0$, there is $\delta > 0$ such that whenever $\mathbf{v} < \delta$ we have

$$\left\|\text{Hess}(f)(\mathbf{a}+\mathbf{v}) - \text{Hess}(f)(\mathbf{a})\right\| < \epsilon.$$

Using the Cauchy-Schwarz inequality, we have that

$$\left|\mathbf{h}^\top A\mathbf{h}\right| \leq \|A\| \, \|h\|^2.$$

So whenever $\|h\| < \delta$, we have, for any $0 < \xi < 1$,

$$|H_{f,\mathbf{a}+\xi\mathbf{h}}(\mathbf{h}) - H_{f,\mathbf{a}}\mathbf{h}| < \epsilon \|h\|^2.$$

By definition, $\boldsymbol{\epsilon}(\mathbf{h}) = \frac{1}{2}\left(H_{f,\mathbf{a}+\xi\mathbf{h}} - H_{f,\mathbf{a}}(\mathbf{h})\right)$, so

$$\frac{|\boldsymbol{\epsilon}(\mathbf{h})|}{\|\mathbf{h}\|^2} = \frac{|H_{f,\mathbf{a}+\xi\mathbf{h}}(\mathbf{h}) - H_{f,\mathbf{a}}(\mathbf{h})|}{2\|h\|^2} < \frac{\epsilon}{2}$$

whenever $\|\mathbf{h}\| < \delta$.

Definition

Given a symmetric $n \times n$ matrix A, we say the associated quadratic form $Q : \mathbb{R}^n \to \mathbb{R}$, $Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$, is

- *positive definite* if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,
- *negative definite* if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,
- *positive semidefinite* if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ and $= 0$ for some $\mathbf{x} \neq \mathbf{0}$,
- *negative semidefinite* if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$ and $= 0$ for some $\mathbf{x} \neq \mathbf{0}$,
- *indefinite* if $Q(\mathbf{x}) > 0$ for some $\mathbf{x}$ and $Q(\mathbf{x}) < 0$ for other $\mathbf{x}$.

Theorem

Suppose $f : B(\mathbf{a}, r) \to \mathbb{R}$ is C^2 and $\mathbf{a}$ is a critical point. If $H_{f,\mathbf{a}}$ is positive (resp., negative) definite, then $\mathbf{a}$ is a local minimum (resp., maximum) point; if $H_{f,\mathbf{a}}$ is indefinite, then $\mathbf{a}$ is a saddle point. If $H_{f,\mathbf{a}}$ is semidefinite, we can draw no conclusions.