Quadratic Forms & Constrained Optimization

Introduction

- Quadratic forms generalize taking the inner (dot, scalar) product of a vector with itself.
- They come up in engineering applications involving optimization and signal processing, utility functions in economics, confidence ellipsoids in statistics, etc.
- There is a natural progression from diagonalization of symmetric matrices through this topic to the SVD.
- Key result: By an (orthogonal) change of variable, any quadratic form is equivalent to a form without cross terms;
- Key result: The "sign" of a quadratic form (positive definite, negative definite, indefinite) is determined by the sign of its *spectrum* (i.e., eigenvalues);
- Key result: The max and min value attained by a QF on the set of *unit vectors* is just the largest and smallest eigenvalue of the corresponding matrix.

Quadratic forms from symmetric matrices

Definition 1. A **quadratic form** $\mathcal{Q}$ on $\mathbb{R}^n$ is a function $\mathcal{Q} : \mathbb{R}^n \to \mathbb{R}$ of the form $\mathcal{Q}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$, where A is a *symmetric matrix*, called the **matrix of the quadratic form**.

Example 2. What are the quadratic forms corresponding to the following matrices?

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & -3 \\ -3 & 4 \end{bmatrix}, \quad \text{and } C = \begin{bmatrix} 4 & -3 \\ -5 & 4 \end{bmatrix}.$$

Example 3. What is the matrix A of the quadratic form $\mathcal{Q} : \mathbb{R}^3 \to \mathbb{R}$ given by $\mathcal{Q}(\mathbf{x}) = -4x_1^2 + 7x_2^2 - 5x_3^2 - 6x_1x_2 + 3x_2x_3$?

Change of variables & Principal Axes Theorem

Recall that any $n \times n$ *invertible* matrix P (whose columns $\mathcal{B}$ are a basis for $\mathbb{R}^n$), represents a change of basis from standard $\mathcal{E}$-coordinates to $\mathcal{B}$-coordinates:

$$\mathbf{x} = P\mathbf{y} \iff \mathbf{y} = P^{-1}\mathbf{x} \iff [\mathbf{y}]_{\mathcal{B}} = [\mathbf{x}]_{\mathcal{E}}$$

Question 4. *What happens if we change variable in a quadratic form? Why is this so great? How can we use that A is symmetric?*

Example 5. Orthogonally diagonalize the QF on $\mathbb{R}^2$ given by $\mathcal{Q}(\mathbf{x}) = 2x_1^2 - 4x_1x_2 + 5x_2^2$.

Theorem 6. *Let $\mathcal{Q}$ be a quadratic form on $\mathbb{R}^n$ corresponding to the (symmetric) matrix A. Then we can find an orthogonal (change of basis) matrix P such that $\mathbf{x} = P\mathbf{y}$, transforming $\mathcal{Q}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ to $\mathcal{Q}(\mathbf{y}) = \mathbf{y}^\top D\mathbf{y}$, with no cross term. ($D$ is diagonal.)*

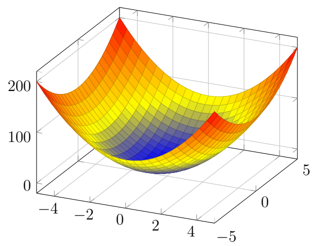
The columns of P are called the **principal axes** of the quadratic form $\mathcal{Q}$.

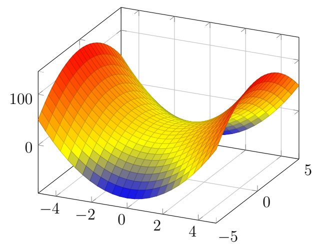
See the text for pictures explaining how quadratic forms correspond to conic sections. Those with nonzero cross terms are rotated relative to the standard position. Eliminating the cross terms is equivalent to the rotation-of-axes change of variables technique that used to be taught in some HS "college algebra" classes.

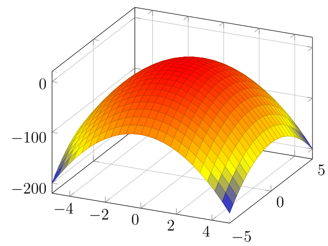
Classifying quadratic forms

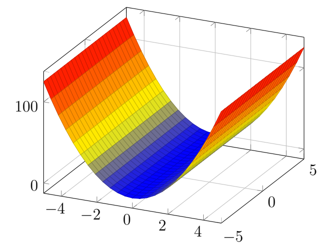
Pictured below are graphs of the quadratic forms

(a) $5x_1^2 + 3x_2^2$, (b) $5x_1^2 - 3x_2^2$, (c) $-5x_1^2 - 3x_2^2$ and (d) $5x_1^2$









Definition 7. Call a quadratic form $\mathcal{Q}$:

a. **positive definite** if $\mathcal{Q}(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$;

b. **positive semidefinite** if $\mathcal{Q}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$;

c. **negative definite** if $\mathcal{Q}(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$;

d. **negative semidefinite** if $\mathcal{Q}(\mathbf{x}) \leq 0$ for all $\mathbf{x} \neq \mathbf{0}$; and

e. **indefinite** if $\mathcal{Q}$ assume both positive and negative values.

Example 8. Classify each of the quadratic forms above.

Theorem 9. *For an $n \times n$ symmetric matrix A, the quadratic form $Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ is:*

a. *positive definite if and only if all the eigenvalues of A are positive.*

b. *positive semidefinite if and only if all the eigenvalues of A are ≥ 0.*

c. *negative definite if and only if all the eigenvalues of A are negative.*

d. *negative semidefinite if and only if all the eigenvalues of A are ≤ 0.*

e. *indefinite if and only if A has both positive and negative eigenvalues.*

Proof. Use the Principal Axes Theorem.

■

Example 10. Classify the QF $\mathcal{Q} = x_1^2 + 3x_2^2 + x_3^2 + 2x_1x_2 + 6x_1x_3 + 2x_2x_3$.

Constrained Optimization (Sec. 7.3)

- Maximizing and minimizing is a major application of mathematical modeling.
- Here try to maximize a quadratic form $\mathcal{Q}(\mathbf{x})$ on $\mathbb{R}^n$ subject to the constraint that $\mathbf{x}$ is a unit vector, i.e., $\mathbf{x}^\top\mathbf{x} = 1 \iff x_1^2 + \cdots + x_n^2 = 1$.
- *Key idea:* It's easy to maximize when no cross terms, just get max/min is largest/smallest eigenvalue of the matrix of $\mathcal{Q}$. By diagonalization, this holds for *every* QF.
- Also can understand other eigenvalues in terms of optimization with additional constraints.

Example 11. Find the max and min values of $\mathcal{Q}(\mathbf{x}) = 7x_1^2 - 5x_2^2 + 4x_3^2$, subject to the constraint $\mathbf{x}^\top\mathbf{x} = 1$.

Geometrically a QF on $\mathbb{R}^2$ can be viewed as a paraboloid, which we are interesecting with the cylinder $x_1^2 + x_2^2$, see Fig. 7.3.2, p. 409.

Theorem 12. *Let A be a (real) symmetric $n \times n$ matrix, and define the constrained extrema of the QF $\mathcal{Q}(\mathbf{x}) := \mathbf{x}^\top A\mathbf{x}$ by*

$$m = \min\{\mathbf{x}^\top A\mathbf{x} : \|\mathbf{x}\| = 1\}, \qquad M = \max\{\mathbf{x}^\top A\mathbf{x} : \|\mathbf{x}\| = 1\},$$

Then M is the greatest eigenvalue λ_1 and m is the least eigenvalue λ_n of A. These values are obtained by setting $\mathbf{x}$ to be the corresponding unit eigenvector.

Proof. Orthogonally diagonalize $A = PDP^{-1}$, and use P as the change-of-variable matrix that eliminates the cross-terms in $\mathcal{Q}$,

∎

Example 13. Find the max and min of the QF $\mathcal{Q} = x_1^2 + 3x_2^2 + x_3^2 + 2x_1x_2 + 6x_1x_3 + 2x_2x_3$, subject to $\|\mathbf{x}\| = 1$.

There is an interpretation of the other eigenvalues as well, in terms of adding additional constraints.

Theorem 14. *Let A be a (real) symmetric $n \times n$ matrix with orthogonal diagonalization $A = PDP^{-1}$, where WLOG diagonal entries of D are in decreasing order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. (If $P = [\mathbf{u}_1 \cdots \mathbf{u}_n]$, this means that $\mathbf{u}_i$ is the eigenvector corresponding to λ_i.) Then for $k = 2, 3, \ldots n$, the maximum value of $\mathcal{Q}(\mathbf{x}) := \mathbf{x}^\top A\mathbf{x}$ subject to*

$$\mathbf{x}^\top\mathbf{x} = 1, \quad \mathbf{x}^\top\mathbf{u_1} = 0, \quad \ldots, \quad \mathbf{x}^\top\mathbf{u_{k-1}} = 0$$

is the eigenvalue λ_k, attained when $\mathbf{x} = \mathbf{u}_k$.

Example 15. Find max of $\mathcal{Q}(\mathbf{x}) = 7x_1^2 - 5x_2^2 + 4x_3^2$, subject to $\mathbf{x}^\top\mathbf{x} = 1$ **and** $\mathbf{x}^\top\mathbf{u_1} = 0$

Singular Value Decomposition

These are partial notes for Section 7.4 of Lay's text.

Logistics & notes for the instructor

1. No Quiz! Keep your Homework! Thanks for doing the SET!
2. Remind them of the review session on Thursday 12/15 (Reading Day) at 3:30, final Saturday at 10:30.
3. Very sad day for linear algebra.

Introduction

- We made it! The SVD is a modern topic, hugely important in applications, e.g., *Principal Component Analysis* (§ 7.5), signal processing, linear least squares, polar decomposition, robust computations of rank, and many more.
- We can only diagonalize **some square** matrices A to get $A = PDP^{-1}$, and only **symmetric** A with *orthogonal* P. But the SVD factors **any rectangular** $m \times n$ matrix as $A = QDP^{-1}$, with P, Q orthogonal (when A is real).
- For matrices over $\mathbb{C}$, replace $M \mapsto M^T$ with $M \mapsto M^*$ (conjugate-transpose), and "orthogonal" with "Hermetian".
- The positive (diagonal) entries in the SVD, $\sigma_1, \ldots, \sigma_r$ are the *singular values*; they are the square roots of the eigenvalues of $A^\top A$, **not** A itself.

Maximizing effect of arbitrary matrix

Eigenvalues measure how much a matrix A stretches/shrinks vectors pointing in a certain direction. If $\mathbf{x}$ is a unit vector.

$$A\mathbf{x} = \lambda\mathbf{x} \implies \|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\| = |\lambda|$$

What happens with diagonal matrices?

Example 1. Let $A = [\ ]$.