Quadratic Optimization: Variant

Last time, we discussed finding the extremal (min & max) values of a quadratic form

\[ q(x) = \sum a_{ij} x_i x_j \]

subject to the constraint \( l = ||x||^2 = x_1^2 + \ldots + x_n^2 \).

Procedure: \( q(x) = x^T S x \) for \( S \) symmetric

orthogonally diagonalize: \( S = Q D Q^T \)

change variables: \( x = Q y \)

\[ \Rightarrow q(x) = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 \]

Answer:

maximum = \( \lambda_1 \), achieved at any unit \( \lambda_1 \)-eigenvector

maximum = \( \lambda_n \), achieved at any unit \( \lambda_n \)-eigenvector

Here's an (almost) equivalent variant of this problem that you can draw.

Quadratic Optimization Problem, Variant:

Given a quadratic form \( q(x) \), find the minimum & maximum values of \( ||x||^2 \) subject to \( q(x) = 1 \).

So we switched the function we're extremizing (\( ||x||^2 \)) and the constraint (\( q(x) = 1 \)).
How to draw this problem?

$q(x)=1$: this is a level set of the function $q(x)$

Extremizing $\|x\|^2$ just means finding the shortest & longest vectors on this level set.

Bad Eq: $q(x_1,x_2) = x_1^2 - x_2^2 = 1$ defines a hyperbola

→ Shortest vectors are $(1,0)$ and $(-1,0)$

So the minimum value of $\|x\|^2$ is $\|\pm(1,0)\|^2 = 1$.

→ There is no maximum $\|x\|^2$

subject to $q(x)=1$: there are arbitrarily long vectors on the hyperbola.

Good Eq: An equation of the form

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1 \quad (\lambda_1, \lambda_2 > 0)$$

defines an ellipse.

(This is a circle horizontally stretched by $\sqrt{\lambda_1}$, & vertically stretched by $\sqrt{\lambda_2}$)
If $\lambda \geq \lambda_2$ then $\frac{1}{\lambda_i} \leq \frac{1}{\lambda_2}$. The vectors
$$\pm \frac{1}{\sqrt{\lambda_1}} e_1 = (\pm \sqrt{\lambda_1}, 0) \quad \text{and} \quad \pm \frac{1}{\sqrt{\lambda_2}} e_2 = (0, \pm \sqrt{\lambda_2})$$
both lie on the ellipse $\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1$.

$\pm \frac{1}{\sqrt{\lambda_1}} e_1$ are the shortest vectors on the ellipse
$$\| \pm \frac{1}{\sqrt{\lambda_1}} e_1 \|^2 = \frac{1}{\lambda_1} = \text{minimum length}^2$$

$\pm \frac{1}{\sqrt{\lambda_2}} e_2$ are the longest vectors on the ellipse
$$\| \pm \frac{1}{\sqrt{\lambda_2}} e_2 \|^2 = \frac{1}{\lambda_2} = \text{maximum length}^2$$

In general, $q(x) = \lambda_1 x_1^2 + \ldots + \lambda_n x_n^2$ (all $\lambda_i > 0$) defines an ellipsoid ("egg"); extremizing $\|x\|^2$ subject to $q(x) > 1$ means finding the shortest & longest vectors.

$\pm \frac{1}{\sqrt{\lambda_1}} e_1$ are the shortest vectors on the ellipsoid
$$\| \pm \frac{1}{\sqrt{\lambda_1}} e_1 \|^2 = \frac{1}{\lambda_1} = \text{minimum length}^2$$

$\pm \frac{1}{\sqrt{\lambda_n}} e_n$ are the longest vectors on the ellipsoid
$$\| \pm \frac{1}{\sqrt{\lambda_n}} e_n \|^2 = \frac{1}{\lambda_n} = \text{maximum length}^2$$
What if $q(x)$ is not diagonal?

We still need the condition "All $\lambda_i > 0$"—otherwise a min or max may not exist.

**Def:** A quadratic form is **positive-definite** if $q(x) > 0$ for all $x \neq 0$.

**NB:** If $q(x) = x^TSx$ then

$q$ is positive-definite $\iff S$ is positive-definite

This is the **positive-energy criterion**.

Suppose that $q(x) = x^TSx$ is positive-definite. Let $\lambda_1 \geq \lambda_2 > 0$ be the eigenvalues of $S$ and $u_1, u_2$ orthonormal eigenvectors.

**Change variables:** $x = Qy$ $\Rightarrow$ $Q = (u_1, u_2)$

\[ \lambda_1 y_1^2 + \lambda_2 y_2^2 = 1 \quad \iff \quad q(x) = 1 \]
Upshot: If $g$ is positive-definite, then $q(x) = 1$ defines a (rotated) ellipse.

The minor axis is in the $u_1$-direction.

→ The shortest vectors are $\pm \frac{1}{\sqrt{\lambda_1}} u_1$.

The major axis is in the $u_2$-direction.

→ The longest vectors are $\pm \frac{1}{\sqrt{\lambda_2}} u_2$.

Orthogonally diagonalizing $S = Q D Q^T$ found the major & minor axes & radii!

Eq: $q(x_1, x_2) = \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 - x_1 x_2 = x^T S x$

$S = \frac{1}{2} \begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix} = Q D Q^T \quad Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad D = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}$

$x = Q y \quad \Rightarrow \quad q = 3 y_1^2 + 2 y_2^2$

$3 y_1^2 + 2 y_2^2 = 1$

$q(x) = 1$

---

Diagram:

- $q(x) = 1$
- $S = Q D Q^T$
- $Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
- $D = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}$
- $u_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
- $u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
shortest vectors: \( \pm \frac{1}{\sqrt{3}} \mathbf{u}_1 = \pm \frac{1}{\sqrt{3}} (-1) \mathbf{v}_1 \)

longest vectors: \( \pm \frac{1}{\sqrt{5}} \mathbf{u}_2 = \pm \frac{1}{\sqrt{5}} (1) \mathbf{v}_2 \)

(subject to \( q(\mathbf{v}) = 1 \))

The orthogonal diagonalization procedure took the ellipse

\[ q(x_1, x_2) = \frac{5}{2} x_1^2 + \frac{5}{2} x_2^2 - x_1 x_2 \]

and found its major & minor axes & radii: the change of variables

\[ x = Q y = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \text{us} \quad x_1 = \frac{1}{\sqrt{2}} (-y_1 + y_2) \]
\[ x_2 = \frac{1}{\sqrt{2}} (y_1 + y_2) \]

made \( q(\mathbf{v}) = 1 \) into the standard (non-rotated) ellipse

\[ 3y_1^2 + 2y_2^2 = 1. \]

**Relationship to the original QO problem:**

How is this "almost equivalent" to extremizing \( q(\mathbf{v}) \) subject to \( \| \mathbf{v} \| = 1 \)?

**Recall:** \( q(c \mathbf{x}) = c^2 q(\mathbf{x}) \)
**Fact:** If \( q \) is positive-definite then

\[
\begin{align*}
&\text{If } q \text{ is positive-definite then} \\
&\ u \text{ maximizes } q(u) \\
&\text{subject to } \|u\| = 1 \\
&\text{with maximum value } \lambda. \\
\end{align*}
\]

\[
\begin{align*}
&x = \frac{1}{\sqrt{\lambda}} u \text{ minimizes} \\
&\|x\|^2 \text{ subject to} \\
&\ q(x) = 1 \text{ with minimum value } \frac{1}{\lambda}. \\
\end{align*}
\]

\[
\begin{align*}
&\text{and} \\
&x = \frac{1}{\sqrt{\lambda_n}} u \text{ maximizes} \\
&\|x\|^2 \text{ subject to} \\
&\ q(x) = 1 \text{ with maximum value } \frac{1}{\lambda_n}. \\
\end{align*}
\]

**Why?** if \( q(u) = \lambda > 0 \) and \( x = \frac{1}{\sqrt{\lambda}} u \) then

\[
\begin{align*}
&\text{If } \lambda = \lambda_n, \text{ maximized then } \|x\|^2 = \frac{1}{\lambda} \text{ is minimized} \\
&\text{and vice-versa.}
\end{align*}
\]

So the QO variant gives us a picture of the original QO problem, at least when \( q \) is positive-definite—we're just finding axes & radii of ellipsoids.
Additional Constraints

These come up naturally in practice (see the spectral graph theory problem on the HW) and in the PCA.

“Second-largest” value:
Suppose \( q(x) \) is maximized (subject to \( ||x||=1 \) ) at \( u_1 \).
What is the maximum value of \( q(x) \) subject to \( ||x||=1 \) and \( x \perp u_1 \)?

This rules out the maximum value to get “second-largest” value.

How to solve this?
- Write \( q(x) = x^T S x \)
- Orthogonally diagonalize \( S = Q \Sigma Q^T \)
- Suppose \( u_1 \) is the first column of \( Q \) (1st \( \lambda_1 \)-eigenvector)
- Set \( x = Q y \)

\[ q = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \]

Answer: The maximum value of \( q(x) \) subject to \( \|x\|=1 \) and \( x \perp u_1 \) is \( \lambda_2 \). It is achieved at any unit \( \lambda_2 \)-eigenvector \( u_2 \) that is \( \perp u_1 \).
NB: If $x_i > \lambda_2$ then $u_2 \perp u_1$ automatically. Why?

- If $q = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2$ is diagonal then $u_i = e_i = (1,0,\ldots)$ so $x^T u_i$ means $y_1 = 0$
- As extremizing $\lambda_2 y_2^2 + \lambda_3 y_3^2 + \ldots + \lambda_n y_n^2$.
- Otherwise, change variables $x = Q y$.

$Q$ is orthogonal, so

$$y \cdot e_i = 0 \implies 0 = (Q y) \cdot (Q e_i) = x \cdot a_i, \quad \|y\| = 1 \implies 1 = \|Q y\| = \|x\|$$

(relate constraints on $x \& y$)

Eg: Find the largest and second-largest values of $q(x) = 2x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 8x_1x_3 + 8x_2x_3$
subject to $x_1^2 + x_2^2 + x_3^2 = 1$.

- $q = x^T S x$ for $S = \begin{pmatrix} 2 & 1 & -4 \\ 1 & 2 & 4 \\ -4 & 4 & 5 \end{pmatrix}$

- $S = Q D Q^T$ for

$$Q = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{pmatrix}, \quad D = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$
Largest value is \( q(x) = 9 \) at \( x = \pm \frac{1}{\sqrt{6}} \left( \frac{1}{2} \right) = \pm u_1 \).

Second-largest value:

The maximum value of \( q(x) \) subject to \( ||x|| = 1 \) & \( x \perp u_1 \) is

\[ q(x) = 3 \quad \text{achieved at} \quad x = \pm \frac{1}{\sqrt{6}} \left( \frac{1}{2} \right) \]

This also works for minimizing.

Second-smallest value:

Suppose \( q(x) \) is minimized (subject to \( ||x|| = 1 \)) at \( u_n \).

What is the minimum value of \( q(x) \) subject to \( ||x|| = 1 \) and \( x \perp u_n \)?

Answer: The minimum value of \( q(x) \) subject to \( ||x|| = 1 \) & \( x \perp u_n \) is \( \lambda_{n-1} \). It is achieved at any unit \( \lambda_{n-1} \)-eigenvector \( u_{n-1} \) that is \( \perp u_n \). (Automatic if \( \lambda_{n-1} > \lambda_n \))
You can keep going:

Third-largest value:

Suppose $q(x)$ is maximized (subject to $\|x\|=1$) at $u_1$ and $q(x)$ is maximized (subject to $\|x\|=1$ and $x^Tu_1$) at $u_2$.

What is the maximum value of $q(x)$ subject to $\|x\|=1$ and $x^Tu_1$, and $x^Tu_2$?

**NB:** This "rules out" the largest & second-largest values.

**Answer:** The maximum value of $q(x)$ subject to $\|x\|=1$ & $x^Tu_1$, & $x^Tu_2$ is $\lambda_3$. It is achieved at any unit $\lambda_3$-eigenvector $u_3$ that is $\perp u_1$ and $u_2$. (automatic if $\lambda_2 > \lambda_3$)

This also works for the variant problem, except you have to take reciprocals.

*Et cetera...*
Quadratic Optimization for $S=ATA$

This is what we'll use for PCA.

Let $S=ATA$ and $q(x) = x^T S x$. Then

$$q(x) = x^T S x = x^T (A^T A) x = (x^T A^T A) x$$
$$= (A x)^T (A x) = A x \cdot A x = \|A x\|^2$$

$q(x) = \|A x\|^2$ is a quadratic form with $S=ATA$

In this case, extremizing $q(x)$ subject to $\|x\|=1$ means extremizing $\|A x\|^2$ subject to $\|x\|=1$.

Procedure: to extremize $\|A x\|^2$ subject to $\|x\|=1$:
- Orthogonally diagonalize $S=ATA$
  - as orthonormal eigenbasis $\{u_1, \ldots, u_n\}$,
  - eigenvalues $\lambda_1, \lambda_2, \ldots \geq \lambda_n \geq 0$ (ATA is positive semidefinite)
- The largest value is $\lambda_1$, achieved at any unit $\lambda_1$-eigenvector $u_1$.
- The smallest value is $\lambda_n$, achieved at any unit $\lambda_n$-eigenvector $u_n$.
- The second-largest value is $\lambda_2$, achieved at any unit $\lambda_2$-eigenvector $u_2 \perp u_1$, $\ldots$ etc.
NB: these are eigenvectors/eigenvalues of $S=ATA$, not of $A$ (which need not be square).

Def: The matrix norm of a matrix $A$ is $\|A\|_2 = \text{the maximum value of } \|Ax\|_2$ subject to $\|x\|_2 = 1$.

So $\|A\|_2 = \delta$, $\lambda_1 = \text{largest eigenvalue of } ATA$.

Eg: Compute $\|A\|_2$ for $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

$ATA = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}$ $p(\lambda) = \lambda^2 - 6\lambda + 5 = (\lambda - 5)(\lambda - 1)$

The largest eigenvalue is $\lambda = 5$, so $\|A\|_2 = \sqrt{5}$.

Eigenvector: $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$ $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Unit eigenvector: $u_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

Check: $Au_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

has length $\frac{1}{\sqrt{5}} \sqrt{(-1)^2 + (-1)^2 + (-1)^2} = \frac{1}{\sqrt{5}} \cdot \sqrt{3} = \sqrt{5}$.