Quadratic Optimization: Variant

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

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

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

Procedure:

\[ q(x) = x^T S x \quad \text{for } S \text{ symmetric} \]

orthogonally diagonalize: \( S = Q D Q^T \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ \vdots & \vdots \\ 0 & \lambda_n \end{pmatrix} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \)

change variables: \( x = Q y \)

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

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 $(11x^2)$ and the constraint $(gW=1)$.

In general the min/max may not exist:

- $q(x_1, x_2) = -x_1^2 - 2x_2^2$:
  
  there is no $x$ such that $q(x) = 1!$

- $q(x_1, x_2) = x_1^2 - x_2^2$:
  
  there is no maximum $\lVert x \rVert^2$ subject to $q(x) = 1$:

  \[ q \left( C, \sqrt{C^2 - 1} \right) = 1 \]

  for any (huge) $C$.

  (the min exists though)

Problem: $q(x)$ may be 0 or negative!

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

NB: If $q(x) = x^T S x$ then $q$ is positive-definite if $S$ is positive-definite; this is the positive-energy criterion.
In this case, the problem is equivalent to the previous one, as follows:

Recall: \( q(cx) = c^2 q(x) \)

**Fact:** If \( q \) is positive-definite then

\[
\begin{align*}
\text{u maximizes } q(u) \\
\text{subject to } \|u\| = 1 \\
\text{with maximum value } \lambda_1 \\
\end{align*}
\]

\[
\begin{align*}
\text{x} = \frac{1}{\sqrt{\lambda_1}} u \text{ minimizes} \\
\|x\|^2 \text{ subject to } q(x) = 1 \text{ with minimum value } \frac{1}{\lambda_1}.
\end{align*}
\]

and

\[
\begin{align*}
\text{u minimizes } q(u) \\
\text{subject to } \|u\| = 1 \\
\text{with minimum value } \lambda_n
\end{align*}
\]

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

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

\[
q(x) = q\left(\frac{1}{\sqrt{\lambda}} u\right) = \frac{1}{\lambda} q(u) = \frac{1}{\lambda} \cdot \lambda = 1.
\]

If \( \lambda \) is maximized then \( \|x\|^2 = \frac{1}{\lambda} \) is minimized and vice versa.
So we know exactly how to solve this QO problem variant: do the same procedure as in the original QO problem, and take reciprocals.

Eq: Extremize $\|x\|^2$ subject to

$$q(x_i x_2) = \frac{5}{3} x_1^2 + \frac{5}{2} x_2^2 - x_1 x_2 = 1$$

Diagonalize:

$$q(x) = x^T S x$$

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

$$Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$D = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

- $q$ is maximized (subject to $\|x\| = 1$) at $u_1 = \frac{1}{\sqrt{2}} (1, 1)^T$ with maximum value $3$.

$$q(u_1) = 3 \quad x_1 = \frac{u_1}{\|u_1\|} \quad \Rightarrow \quad q(x_1) = 1 \quad \|x_1\|^2 = \frac{1}{3}$$

The minimum value of $\|x\|^2$ subject to $q(x) = 1$ is $1/3$. It is achieved at $x_1 = \frac{1}{\sqrt{3}} u_1$.

- $q$ is minimized at $u_2 = \frac{1}{\sqrt{2}} (1, -1)^T$ with minimum value $2$.

$$q(u_2) = 2 \quad x_2 = \frac{1}{\sqrt{2}} u_2 \quad \Rightarrow \quad q(x_2) = 1 \quad \|x_2\|^2 = \frac{1}{2}$$

The maximum value of $\|x\|^2$ subject to $q(x) = 1$ is $1/2$. It is achieved at $x_2 = \frac{1}{\sqrt{2}} u_2$.

$\Rightarrow$ note $\frac{1}{3} > \frac{1}{3}$ ☑️
Geometric Interpretation

Recall: An equation of the form
\[ \lambda_1 x_1^2 + \lambda_2 x_2^2 = 1 \]
(\(\lambda_1, \lambda_2 > 0\)) defines an ellipse. (This is a circle horizontally stretched by \(\sqrt{\lambda_1}\) & vertically stretched by \(\sqrt{\lambda_2}\))

If \(q(x_1, x_2) = \lambda_1 x_1^2 + \lambda_2 x_2^2\) is diagonal & positive-definite, then \(q(x) = 1\) defines the ellipse above, and extremizing \(\|x\|^2 = 1\) subject to \(q(x) = 1\) amounts to finding the shortest (\(\pm x\)) & longest (\(\pm y\)) vectors on the ellipse.

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

\[ q(x) = x^T S x \quad \text{for } S \text{ positive-definite.} \]

Let \( \lambda_1 \geq \lambda_2 > 0 \) be the eigenvalues, \( u_1, u_2 \) orthonormal eigenvectors.

Change variables: \( x = Q y \quad Q = (u_1, u_2) \)

\[ \lambda_1 y_1^2 + \lambda_2 y_2^2 = 1 \quad \iff \quad q(x) = 1 \]

\[ \begin{aligned}
  & (0, \sqrt{\lambda_2}) \\
  & (\sqrt{\lambda_1}, 0)
\end{aligned} \]

\[ \text{ multiply by } Q \]

\[ \begin{aligned}
  & u_1 = Q e_1 \\
  & u_2 = Q e_2
\end{aligned} \]

\[ \text{(}x, y_2)\text{-plane} \quad \text{ and } \quad \text{(}x, y_1)\text{-plane} \]

**Upshot:** \( q(x) = 1 \) 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 \( \pm \frac{1}{\sqrt{\lambda_2}} u_2 \).

So we're drawn a picture of quadratic optimization problem (variant).

Everything works in higher dimensions; just get rotated ellipsoids.
E.g. \( q(x_1, x_2) = \frac{5}{2} x_1^2 + \frac{5}{2} x_2^2 - x_1 x_2 = x^T S x \)

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

\[ x = Q y \implies q = 3y_1^2 + 2y_2^2 \]

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

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.
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 \( u_i \). What is the maximum value of \( q(x) \) subject to \( \|x\|=1 \) and \( x \perp u_i \)?

This rules out the maximum value => 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 \( \lambda_i \)-eigenvec)
- Set \( x = Q y \)
  \[
  q = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n
  \]

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

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

- Otherwise, change variables $x = Qy$.

$Q$ is orthogonal so

\[ y \cdot e_i = 0 \iff 0 = (Qy) \cdot (Qe_i) = x \cdot u_i \]

\[ \|y\| = 1 \iff 1 = \|Qy\| = \|x\| \]

(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 $S = \begin{pmatrix} 2 & 1 & -4 \\ 1 & 4 & 2 \\ -4 & 2 & 5 \end{pmatrix}$

- $S = QDQ^T$ for

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

Second-largest value:

The maximum value of \( q(x) \) subject to \( \|x\| = 1 \) and \( x \perp u_1 \) is
\[
q(x) = 3 \quad \text{achieved at } x = \pm \frac{1}{\sqrt{6}} \frac{1}{2}
\]

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 \( \|x\| = 1 \) and \( x \perp u_n \) is \( \lambda_n. \) 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 $x \perp u_1$) at $u_2$.

What is the maximum value of $q(x)$ subject to $\|x\|=1$ and $x \perp u_1$ and $x \perp u_2$?

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

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

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 $q(x)=x^TSx$. Then

\[ q(x) = x^TSx = x^T(A^TA)x = (x^TATA)(Ax) = (Ax)^T(Ax) = (Ax)\cdot(Ax) = \|Ax\|^2 \]

\[
S=ATA \implies x^TSx = \|Ax\|^2
\]

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

**Procedure:** to extremize $\|Ax\|^2$ subject to $\|x\|=1$:

Orthogonally diagonalize $S=ATA$

using orthonormal eigenbasis $\{u_0, u_1, \ldots, u_n\}$,

- **Eigenvalues**: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$

  - 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 $u_2 \perp u_1 \perp \ldots$ 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

$$\|A\| = \text{maximum value of } |Ax| \text{ subject to } \|x\| = 1.$$ 

So $\|A\| = \|x\|_1$, $\lambda_1 =$ largest eigenvalue of $ATA$.

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

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

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

Eigenvector: $\begin{pmatrix} -b \\ a-\lambda \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

Unit eigenvector: $u_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = 2\sqrt{\frac{1}{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

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