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 $1 = \|x\|^2 = x_1^2 + \cdots + x_n^2$.

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

orthogonally diagonalize: $S = QDQ^T$ $\quad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

change variables: $x = Qy$

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

Answer:

maximum $= \lambda_1$, achieved at any unit λ_1-eigenvector

maximum $= \lambda_n$, achieved at any unit λ_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.

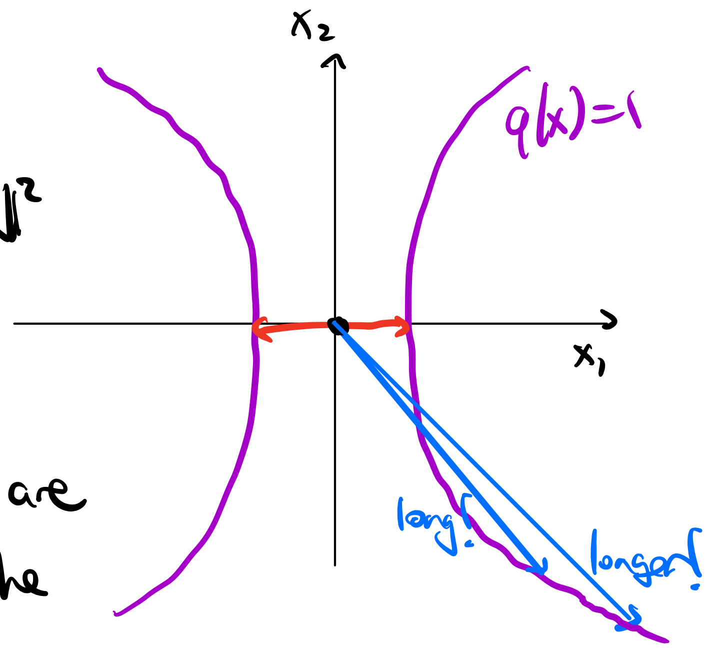
Bad Eg: $q(x_1,x_2) = x_1^2 - x_2^2 = 1$ defines a hyperbola

→ 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.

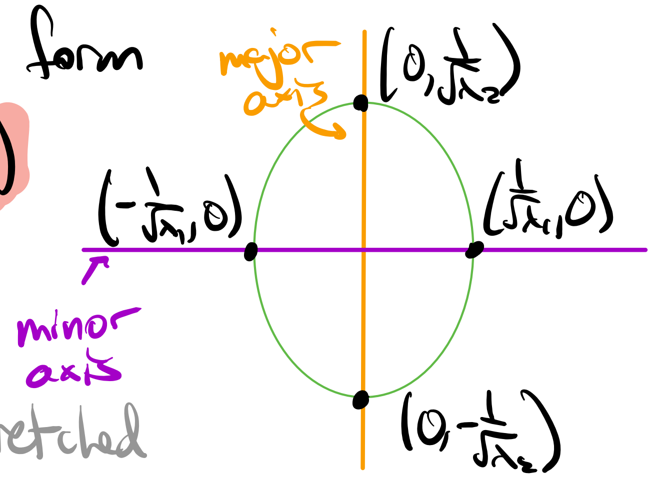


Good Eg: An equation of the form

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1 \qquad \boxed{(\lambda_1 \geq \lambda_2 > 0)}$$

defines an ellipse.

(This is a circle horizontally stretched by $1/\sqrt{\lambda_1}$ & vertically stretched by $1/\sqrt{\lambda_2}$)



If $\lambda_1 \geq \lambda_2$ then $\frac{1}{\lambda_1} \leq \frac{1}{\lambda_2}$. The vectors

$$\pm\frac{1}{\sqrt{\lambda_1}}e_1 = \left(\pm\frac{1}{\sqrt{\lambda_1}}, 0\right) \quad \text{and} \quad \pm\frac{1}{\sqrt{\lambda_2}}e_2 = \left(0, \pm\frac{1}{\sqrt{\lambda_2}}\right)$$

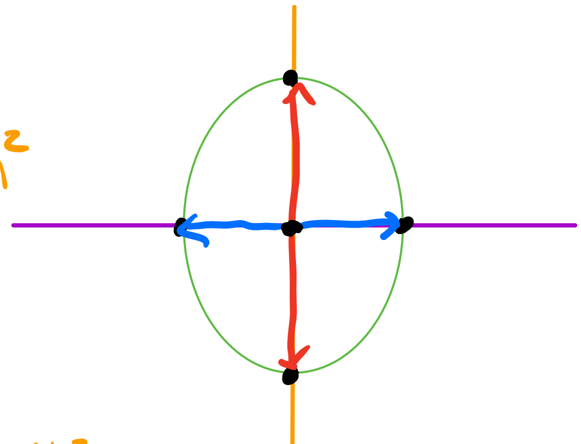
both lie on the ellipse $\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1$.

$\pm\frac{1}{\sqrt{\lambda_1}}e_1$ are the shortest vectors on the ellipse

$$\left\|\frac{\pm 1}{\sqrt{\lambda_1}}e_1\right\|^2 = \frac{1}{\lambda_1} = \text{minimum length}^2$$

$\pm\frac{1}{\sqrt{\lambda_2}}e_2$ are the longest vectors on the ellipse

$$\left\|\frac{\pm 1}{\sqrt{\lambda_2}}e_2\right\|^2 = \frac{1}{\lambda_2} = \text{maximum length}^2$$



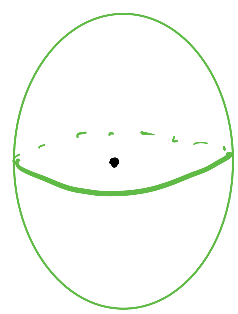
In general, $q(x) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$ (all $\lambda_i > 0$) defines an ellipsoid ("egg"); extremizing $\|x\|^2$ subject to $q(x) = 1$ means finding the shortest & longest vectors.

$\pm\frac{1}{\sqrt{\lambda_1}}e_1$ are the shortest vectors on the ellipsoid

$$\left\|\frac{\pm 1}{\sqrt{\lambda_1}}e_1\right\|^2 = \frac{1}{\lambda_1} = \text{minimum length}^2$$

$\pm\frac{1}{\sqrt{\lambda_n}}e_n$ are the longest vectors on the ellipsoid

$$\left\|\frac{\pm 1}{\sqrt{\lambda_n}}e_n\right\|^2 = \frac{1}{\lambda_n} = \text{maximum length}^2$$



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

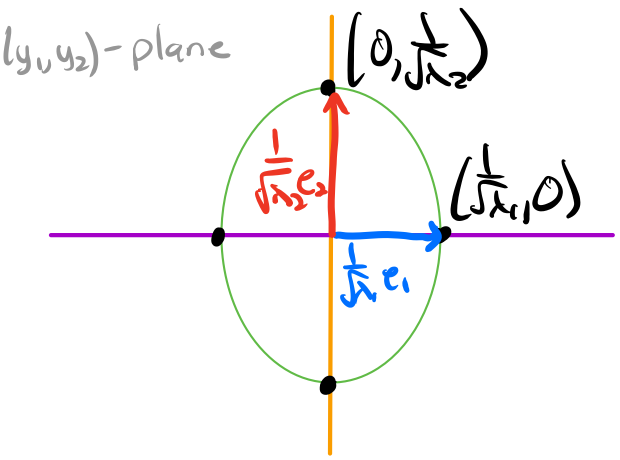
$$\boxed{q \text{ is positive-definite} \iff S \text{ is positive-definite}}$$

This is the *positive-energy* criterion.

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

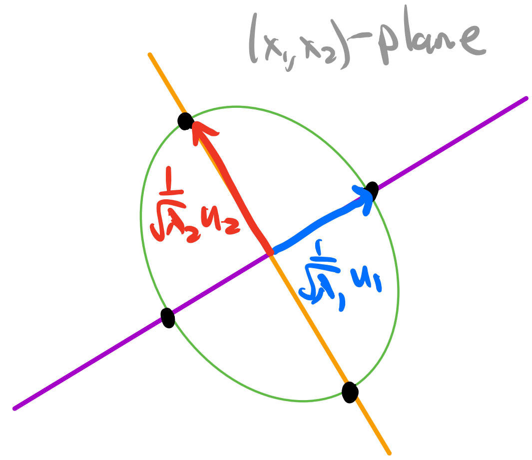
Change variables: $x = Qy$ $\quad Q = \begin{pmatrix} | & | \\ u_1 & u_2 \\ | & | \end{pmatrix}$

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1 \iff q(x) = 1$$



$\xrightarrow[\text{by } Q]{\text{multiply}}$

$$\boxed{\begin{aligned} u_1 &= Q e_1 \\ u_2 &= Q e_2 \end{aligned}}$$



Upshot: If q 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 $S=QDQ^T$ found the major & minor axes & radii!

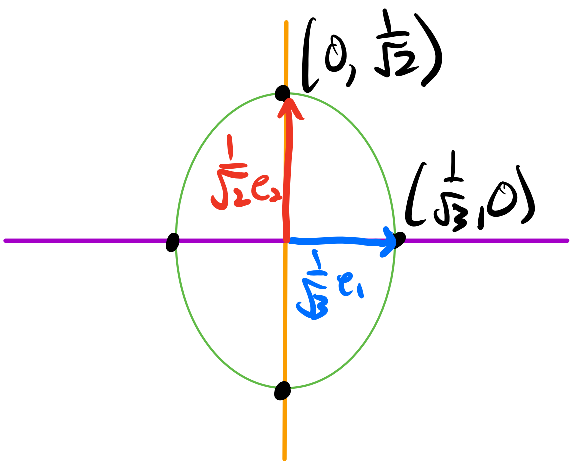
Eg: $q(x_1,x_2)=\frac{5}{2}x_1^2+\frac{5}{2}x_2^2-x_1x_2=x^TSx$

$S=\frac{1}{2}\begin{pmatrix}5&-1\\-1&5\end{pmatrix}=QDQ^T \quad Q=\frac{1}{\sqrt{2}}\begin{pmatrix}-1&1\\1&1\end{pmatrix} \quad D=\begin{pmatrix}3&0\\0&2\end{pmatrix}$

$x=Qy \rightsquigarrow q=3y_1^2+2y_2^2$

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

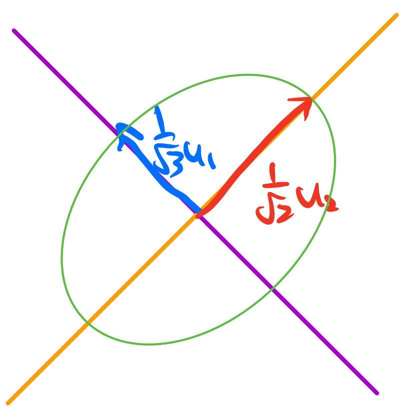
$q(x)=1$



(y_1,y_2)-plane

$\xrightarrow[\text{by } Q]{\text{multiply}}$

$\boxed{\begin{array}{l}u_1=Qe_1\\u_2=Qe_2\end{array}}$



$u_1=\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix}$

$u_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$

(x_1,x_2)-plane

shortest vectors: $\pm\frac{1}{\sqrt{3}}u_1 = \pm\frac{1}{\sqrt{6}}\begin{pmatrix}-1\\1\end{pmatrix}$ length$^2 = 1/3$

longest vectors: $\pm\frac{1}{\sqrt{3}}u_2 = \pm\frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}$ length$^2 = 1/2 > \frac{1}{3}$

(subject to $q(x)=1$)

The orthogonal diagonalization procedure took the ellipse

$$q(x_1, x_2) = \frac{5}{2}x_1^2 + \frac{5}{2}x_2^2 - x_1x_2$$

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

$$x = Qy = \frac{1}{\sqrt{2}}\begin{pmatrix}-1 & 1\\1 & 1\end{pmatrix}\begin{pmatrix}y_1\\y_2\end{pmatrix} \quad \text{vs} \quad \begin{aligned} x_1 &= \tfrac{1}{\sqrt{2}}(-y_1+y_2)\\ x_2 &= \tfrac{1}{\sqrt{2}}(y_1+y_2)\end{aligned}$$

made $q(x)=1$ 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 $q(x)$ subject to $\|x\|=1$?

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

Fact: If q is positive-definite then

u maximizes $q(u)$ subject to $\|u\|=1$ with maximum value λ_1	$\Longleftrightarrow$	$x = \frac{1}{\sqrt{\lambda_1}} u$ minimizes $\|x\|^2$ subject to $q(x)=1$ with minimum value $1/\lambda_1$.

and

u minimizes $q(u)$ subject to $\|u\|=1$ with minimum value λ_n	$\Longleftrightarrow$	$x = \frac{1}{\sqrt{\lambda_n}} u$ maximizes $\|x\|^2$ subject to $q(x)=1$ with maximum value $1/\lambda_n$

Why? if $\underset{\|u\|=1}{q(u)=\lambda} > 0$ and $\underset{\|x\|^2=\frac{1}{\lambda}}{x = \frac{1}{\sqrt{\lambda}} u}$ then

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

If λ is maximized then $\|x\|^2 = \frac{1}{\lambda}$ is minimized and vice-versa.

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 $\rightsquigarrow$ get "second-largest" value.

How to solve this?

- Write $q(x) = x^T S x$
- Orthogonally diagonalize $S = QDQ^T$
  Suppose u_1 is the first column of Q (1st λ_1-eigenvec)
- Set $x = Qy$
  $\rightsquigarrow q = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

Answer: The maximum value of $q(x)$ subject to $\|x\|=1$ & $x \perp u_1$ is λ_2. It is achieved at any unit λ_2-eigenvector u_2 that is $\perp u_1$.

NB: If $\lambda_1 > \lambda_2$ then $u_2 \perp u_1$ automatically.

Why?

- If $q = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$ is diagonal then $u_1 = e_1 = (1, 0, \ldots)$ so $x \perp u_1$ means $y_1 = 0$ 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

$$y \cdot e_1 = 0 \iff 0 = (Qy) \cdot (Qe_1) = x \cdot u_1$$

$$\|y\| = 1 \iff 1 = \|Qy\| = \|x\|$$

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

$$Q = \begin{pmatrix} -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{pmatrix} \quad D = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{pmatrix}$$

Largest value is $q(x)=9$ at $x=\pm\frac{1}{\sqrt{6}}\begin{pmatrix}-1\\1\\2\end{pmatrix}=\pm u_1$

Second-largest value:

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

$q(x)=3$ achieved at $x=\pm\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}$

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 λ_{n-1}. It is achieved at any unit λ_{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 λ_3. It is achieved at any unit λ_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 $S = A^TA$

This is what we'll use for PCA.

Let $S = A^TA$ and $q(x) = x^TSx$. Then

$$q(x) = x^TSx = x^T(A^TA)x = (x^TA^T)(Ax)$$
$$= (Ax)^T(Ax) = (Ax)\cdot(Ax) = \|Ax\|^2$$

$\boxed{q(x) = \|Ax\|^2 \text{ is a quadratic form with } S = A^TA}$

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 $S = A^TA$

w/ orthonormal eigenbasis $\{u_1, \ldots, u_n\}$,

eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ ← A^TA is +ve-semidefinite

- The largest value is λ_1, achieved at any unit λ_1-eigenvector u_1.
- The smallest value is λ_n, achieved at any unit λ_n-eigenvector u_n.
- The second-largest value is λ_2, achieved at any unit λ_2-eigenvector $u_2 \perp u_1$. ... etc.

NB: these are eigenvectors/eigenvalues of $S = A^TA$, not of A (which need not be square).

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

So $\|A\| = \sqrt{\lambda_1}$ $\quad \lambda_1$ = largest eigenvalue of A^TA.

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

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

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

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

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

Check: $Au_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ -2 \\ -2 \\ -1 \end{pmatrix}$

has length $\frac{1}{\sqrt{2}} \cdot \sqrt{1^2+2^2+2^2+1^2} = \frac{\sqrt{10}}{\sqrt{2}} = \sqrt{5}$ ✓