Quadratic Optimization: Continued

Last time: We discussed 2 quadratic optimization problems.

Let $q(x)$ be a quadratic form in n variables.

(1) Extremize $q(x)$ subject to $\|x\| = 1$

$\rightsquigarrow$ $q(x) = x^T S x$ for S symmetric

Eigenvalues of S: $\lambda_1 \le \cdots \le \lambda_n$

Maximum value is λ_n; achieved on a unit λ_n-eigenvector

Minimum value is λ_1; achieved on a unit λ_1-eigenvector

(2) Extremize $\|x\|^2$ subject to $q(x) = 1$

$\rightsquigarrow$ only works for q positive-definite

$q(x) = x^T S x$ for S positive-definite

Eigenvalues of S: $0 < \lambda_1 \le \cdots \le \lambda_n$

Minimum value of $\|x\|^2$ is $1/\lambda_n$; achieved on $1/\sqrt{\lambda_n}\, u_n$ u_n = unit λ_n-eigenvector

Maximum value of $\|x\|^2$ is $1/\lambda_1$; achieved on $1/\sqrt{\lambda_1}\, u_1$ u_1 = unit λ_1-eigenvector

Additional constraints: Let q be a quadratic form. $q(x) = x^T S x$, eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ of S.
Let u_1 be a λ_1-eigenvector.
What is the minimum value of $q(x)$ subject to
$$\|x\| = 1 \quad \text{and} \quad x \perp u_1 \, ?$$

NB: Without the "$x \perp u_1$" constraint, the answer is λ_1!
This comes up if you don't care about λ_1 (e.g. if $\lambda_1 = 0$ for a dumb reason).

Answer: The minimum value is λ_2, achieved at any unit λ_2-eigenvector that is orthogonal to u_1.
(This is automatic if $\lambda_2 \neq \lambda_1$.)

Why?

In the diagonal case, $q(x) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$ is minimized at $e_1 = (1, 0, \ldots, 0)$. Then $x \perp e_1 \iff x \cdot e_1 = 0$ $\iff x = (0, x_2, \ldots, x_n)$. So we're extremizing
$$q(0, x_2, \ldots, x_n) = \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2$$
which we know how to do.

In the non-diagonal case, we change variables to reduce to the diagonal case (exercise).

This also works for maximizing:

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

A: The maximum value is λ_{n-1}, achieved at any unit λ_{n-1}-eigenvector $\perp u_n$.

You can keep going:

Q: What is the maximum value of $q(x)$ subject to $\|x\|=1$ and $x \perp u_n$ and $x \perp u_{n-1}$?

A: The maximum value is λ_{n-2}, achieved at any unit λ_{n-2}-eigenvector in $\mathrm{Span}\{u_{n-1}, u_n\}^\perp$

This works for quadratic optimization problem #2:

Q: What is the maximum value of $\|x\|^2$ subject to $q(x)=1$ and $x \perp u_1$?

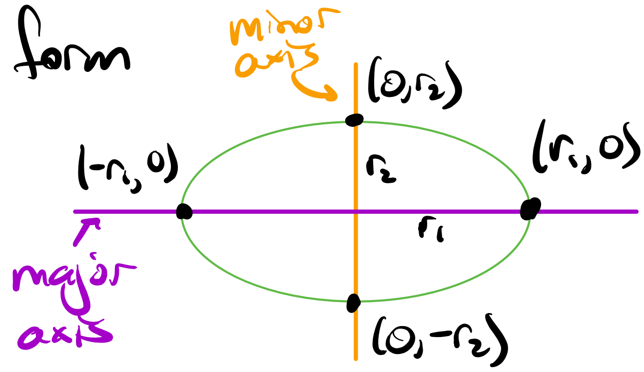
A: The maximum value is λ_2^{-1}, achieved at any unit λ_2-eigenvector $\perp u_1$.

Geometric Interpretation

Recall: An equation of the form

$$\left(\frac{x_1}{r_1}\right)^2 + \left(\frac{x_2}{r_2}\right)^2 = 1$$

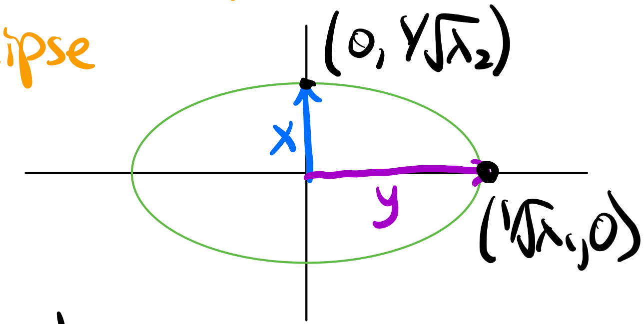
$(r_1 \geq r_2)$ defines an ellipse.



(This is a circle horizontally stretched by r_1 & vertically stretched by r_2.)

If $q(x_1, x_2) = \lambda_1 x_1^2 + \lambda_2 x_2^2$ is diagonal & positive-definite then $q(x_1, x_2) = 1$ defines the ellipse

$$\left(\frac{x_1}{1/\sqrt{\lambda_1}}\right)^2 + \left(\frac{x_2}{1/\sqrt{\lambda_2}}\right)^2 = 1 \quad \begin{pmatrix} r_1 = 1/\sqrt{\lambda_1} \\ r_2 = 1/\sqrt{\lambda_2} \end{pmatrix}$$



and extremizing $\|x\|^2 = I$ subject to $q(x) = 1$ amounts to finding the shortest & longest vectors on the ellipse.

$\|x\|^2 = 1/\lambda_2$

$\|y\|^2 = 1/\lambda_1$

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.

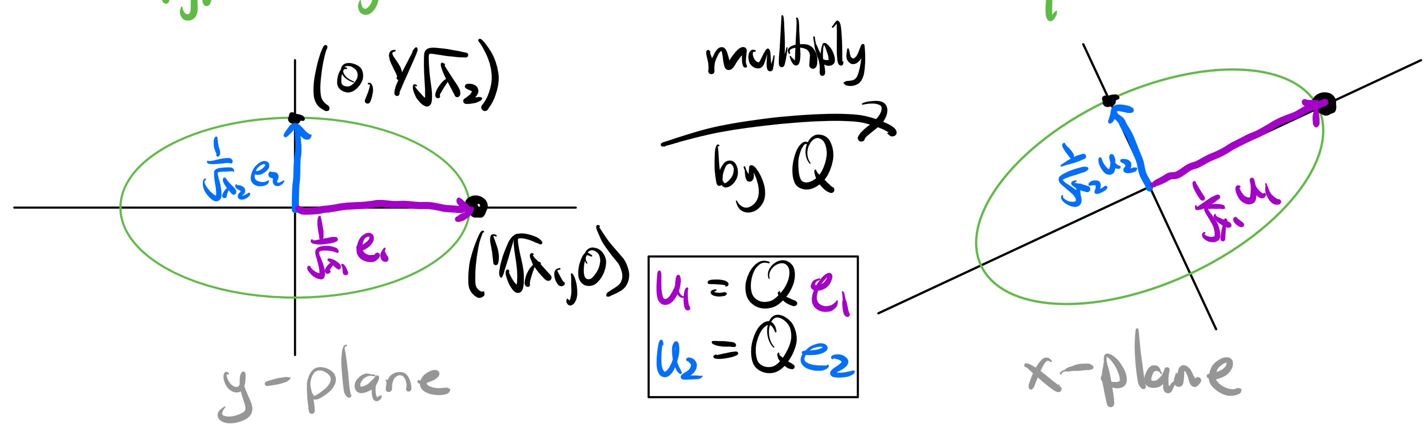
Non-diagonal case:

$q(x) = x^T S x$ for S positive-definite.

Let $\lambda_1 \leq \lambda_2$ be the eigenvalues, u_1, u_2 orthonormal eigenvectors.

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

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



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

Upshot: $q(x) = 1$ defines a (rotated) ellipse

The major axis is in the u_1-direction.

→ The longest vector is $\pm \frac{1}{\sqrt{\lambda_1}} u_1$

The minor axis is in the u_2-direction.

→ The shortest vector is $\pm \frac{1}{\sqrt{\lambda_2}} u_2$.

So we've drawn a picture of quadratic optimization problem #2.

Everything works in higher dimensions; just get rotated ellipsoids.

LDL^T & Cholesky

This is an IOU about positive-definite symmetric matrices. It amounts to an LU decomposition that's 2x as fast to compute.

Thm: A positive-definite symmetric matrix S can be uniquely decomposed as

$$S = LDL^T \quad \text{and} \quad S = \boxed{L_1 L_1^T} \leftarrow \text{Cholesky}$$

where:

- D: diagonal w/ positive diagonal entries
- L: lower-unitriangular
- L_1: lower-triangular with positive diagonal entries.

Proof: [supplement]

NB: Let $U = DL^T$.

(scales the rows of L^T by the diagonal entries of D)

Then U is upper-Δ with positive diagonal entries

$\Rightarrow$ in REF, so $S = LU$ is the LU decomposition!

This tells us how to compute an LDL^T decomposition.

Procedure to compute $S=LDL^T$:

Let S be a symmetric matrix.

(1) Compute the LU decomposition $S=LU$.

→ If you have to do a row swap then stop: S is not positive-definite.

→ If the diagonal entries of U are not all positive then stop: S is not positive-definite.

(2) Let $D=$ the matrix of diagonal entries of U (set the off-diagonal entries $=0$). Then

$$S=LDL^T.$$

NB: An LDL^T decomposition can be computed in $\sim\frac{1}{3}n^3$ flops (as opposed to $2/3\,n^3$ for LU). This requires a slightly more clever algorithm. See the supplement — it's also faster by hand!

NB: This is still an LU decomposition — lets you solve $Sx=b$ quickly.

Eg: Find the LDL^T decomposition of

$$S = \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -1 \\ -2 & -1 & 14 \end{pmatrix}$$

2-column method:

$$L \qquad\qquad\qquad U$$

$$\begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \qquad \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -1 \\ -2 & -1 & 14 \end{pmatrix}$$

$R_2 -= 2R_1$, $R_3 += R_1$

$$\begin{pmatrix} 1 & & \\ 2 & & \\ -1 & & \end{pmatrix} \qquad \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 3 \\ 0 & 3 & 12 \end{pmatrix}$$

$R_3 -= 3R_2$

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix}$$

So $S = LDL^T$ for

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix} \qquad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Check:

$$DL^T = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix} = U \checkmark$$

Cholesky from LDL^T:

If S is positive-definite then $S = LDL^T$ where D is diagonal with positive diagonal entries.

If $D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$ set $\sqrt{D} = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}$

Then $\sqrt{D} \cdot \sqrt{D} = D$ and $\sqrt{D}^T = \sqrt{D}$, so

$$LDL^T = L\sqrt{D}\sqrt{D}L^T = (L\sqrt{D})(L\sqrt{D})^T$$

So just set

$$\boxed{L_1 = L\sqrt{D} \rightsquigarrow S = L_1 L_1^T}$$

Strang:

"$S = A^TA$ is how a positive-definite symmetric matrix is put together.

$S = L_1L_1^T$ is how you pull it apart"

Eg:

$\begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -1 \\ -2 & -1 & 14 \end{pmatrix} = L_1 L_1^T$ for

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{1} & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 2\sqrt{2} & 1 & 0 \\ -\sqrt{2} & 3 & \sqrt{3} \end{pmatrix}$$

Toward the SVD

We'll discuss the SVD next time.
Today let's prove the crucial ingredients from symmetric matrices.

Recall: If A is any matrix then $S = A^TA$ is positive-semidefinite: it has nonnegative eigenvalues.

Facts: Let v be an eigenvector of A^TA with eigenvalue λ and let $u = Av$.

(1) $\|u\|^2 = \lambda \|v\|^2$

(2) If $\lambda > 0$ then u is an eigenvector of AA^T with eigenvalue λ

(3) Let v' be another eigenvector of A^TA with eigenvalue λ', and let $u' = Av'$. If $v \cdot v' = 0$ then $u \cdot u' = 0$.

(This is automatic when $\lambda \neq \lambda'$)

Proof: (1)
$$\|u\|^2 = (Av)\cdot(Av) = (Av)^T(Av)$$
$$= v^TA^TAv = v^T(A^TAv) = v^T(\lambda v)$$
$$= \lambda v^Tv = \lambda v \cdot v = \lambda \|v\|^2.$$

(2) If $\lambda > 0$ then $u \neq 0$ by (1). Now compute:

$$AA^T u = AA^T(Av) = A(A^T A v) = A(\lambda v)$$
$$= \lambda A v = \lambda u.$$

(3) $$u \cdot u' = (Av) \cdot (Av') = (Av)^T (Av')$$
$$= v^T A^T A v' = v^T (A^T A v') = v^T (\lambda' v')$$
$$= \lambda' v^T v' = \lambda' v \cdot v' = 0.$$

We'll use the Facts to prove:

Theorem (SVD, Vector Form):

Let A be an $m \times n$ matrix of rank r.
Then there exist orthonormal sets
$\{v_1, \dots, v_r\}$ in $\mathbb{R}^n$ and
$\{u_1, \dots, u_r\}$ in $\mathbb{R}^m$
such that

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

for numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Here $\{v_1, \dots, v_r\}$ is an orthonormal eigenbasis of $A^T A$ for the nonzero eigenspaces, $\sigma_i = \sqrt{\lambda_i}$, and $u_i = \frac{1}{\sigma_i} A v_i$.