SVD

1) \( A = Q_1 \Sigma Q_2^T \) given \( m \times n \)

\[
AA^T = Q_1 \Sigma Q_2^T Q_2 \Sigma^T Q_1^T = Q_1 \Sigma \Sigma^T Q_1^T
\]

\( m \times n \) matrix

\( Q_1 \) eigenvector matrix for \( A A^T \)

\( \Sigma \) eigenvalue matrix for \( A A^T \)

\( \Sigma^T \Sigma \) with \( \sigma_1^2, ..., \sigma_p^2 \) on diagonal.

Likewise

\[
A^T A = Q_2 \Sigma^T Q_1^T \Sigma \Sigma^T Q_2^T = Q_2 \Sigma \Sigma^T Q_2^T
\]

\( Q_2 \) is the eigenvector matrix for \( A^T A \)

\( \Sigma^T \Sigma \) has the same \( \sigma_1^2, ..., \sigma_p^2 \) but

\( n \times n \)

2) \( A^T A \) is symmetric, so it has a complete set of eigenvectors
let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors.

These get into the calculation of $VQ_2$:

$$A^T A \mathbf{x}_j = \lambda_j \mathbf{x}_j \quad \text{with} \quad \mathbf{x}_j^T \mathbf{x}_j = 1.$$

Eigenvalues are nonnegative:

$$(\langle \mathbf{x}_j, \mathbf{x}_k \rangle^T \mathbf{x}_j = (A^T A \mathbf{x}_j)^T \mathbf{x}_j = (A \mathbf{x}_j)^T (A \mathbf{x}_j) \geq 0)$$

$$\lambda_j = \| A \mathbf{x}_j \|_2^2 \geq 0.$$

Suppose $\lambda_1, \ldots, \lambda_r$ are positive and remaining $n-r$ of $A \mathbf{x}_j$ and $\lambda_j$ are zero. ($\sum \lambda_j = \| A \mathbf{x}_j \|_2^2$)

For $1 \leq j \leq r$, let $\sigma_j = \sqrt{\lambda_j}$ and

$$q_j = \frac{1}{\sigma_j} A \mathbf{x}_j.$$

The $q_j$'s are pairwise orthogonal and have norm 1.
\[
\frac{q^T q}{\lambda e} = \frac{1}{\sigma_x} \frac{1}{\sigma_e} (A x^j)^T (A x_e)
\]

\[
= \frac{1}{\sigma_x \sigma_e} \frac{j}{x^j} (A^T A x_e)
\]

\[
= \frac{\lambda e}{\sigma_x \sigma_e} x^j x_e = S_\lambda e
\]

By Gram–Schmidt, extend the \( q_j \)s to an orthonormal basis of \( R^m \).

\( q_1, q_2, \ldots, q_m \) become the columns of the matrix \( Q \).

Then \( (Q^T A Q_2)_{ij} \) is \( q_i^T A x_j \)

(row \( i \) times matrix \( x \) column \( j \) vector)

\( q_i^T A x_j = 0 \) if \( j > r \) because \( A x_j = 0 \).
\[ q^T Ax_i = q^T (0, 0, q_i) \text{ if } i \leq r \]

since \( Ax_i = 0 \).

But \( q^T q = 5 \), so only nonzero
in \( \Sigma = Q_1^T A Q_2 \) pick the
first \( r \) diagonal entries \( q_{11}, \ldots, q_{rr} \).

Then \( A = Q_1 \Sigma Q_2^T \)

**SVD decomposition.**

Textbook \( A = \begin{bmatrix} -2 & 11 \\ -10 & 5 \end{bmatrix} \)

5.3

page 37

A sym. \( g(x) = \langle Ax, x \rangle \)

Then \( g(x) \) is real, there exist
numbers \( m \in M \) such that
\[
m \| x \|^2 \leq \langle Ax, x \rangle \leq M \| x \|^2
\]

\[
m \leq \frac{\langle Ax, x \rangle}{\| x \|^2} \leq M
\]
Here the variational formulation or minimization idea again enters.

Back to solving \( Ax = b \) with the added assumption that \( A \) is SPD. One of the most important methods for solving this system for very large \( n \) is called the conjugate gradient method.

Consider the quadratic functional \( f: \mathbb{R}^n \to \mathbb{R} \) given by

\[
f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + C
\]

Claim: minimizing the functional \( f \) is equivalent to solving \( Ax = b \).

Note the gradient of \( f \) at \( x \) is

\[
\nabla f(x) = A x - b
\]
from calculating a necessary condition for \( \varepsilon \) to minimize \( f \) is that
\[
\nabla f(\varepsilon) = A \varepsilon - b = 0
\]
This is also sufficient if \( A \) is SPD for then \( WA \) is nonsingular and
\[
Q(x, y) = \langle Ax, y \rangle = \langle x, Ay \rangle
\]
define a new inner product with where \( N(x) = Q(x, x)^{1/2} \) in which
\[
f(x) = \frac{1}{2} N(x - A^{-1}b)^2
\]
\[
N(x - A^{-1}b)^2 = Q(x - A^{-1}b, x - A^{-1}b)
\]
\[
= \langle A(x - A^{-1}b), x - A^{-1}b \rangle
\]
\[
= \langle Ax - b, x - A^{-1}b \rangle
\]
\[
= \langle Ax, x \rangle - \langle Ax, A^{-1}b \rangle - \langle b, x \rangle + \langle b, A^{-1}b \rangle
\]
\[
= \langle Ax, x \rangle - 2 \langle b, x \rangle + \langle b, A^{-1}b \rangle
\]
\[
= 2f(x)
\]
where $c$ is $\frac{1}{2} < b$, $A^{-1}b >$.

$E = A^{-1}b$ is the unique minimizer of $f$.

Computationally, how do we minimize such an $f$? We can try “steepest descent”.
Build an infinite sequence.

1. Pick $x_0$.

2. For $n \geq 1$, define

$$x_{n+1} = x_n - \alpha_n \nabla f(x_n)$$

where $\alpha_n$ is a “small” stepsize.

$$
\begin{array}{c}
\xrightarrow{\alpha_n} \\
\xrightarrow{\nabla f(x_n)} \\
\xrightarrow{x_n+1}
\end{array}
$$