SPD Matrices, QR Decompostion, Gram-Schmidt algorithm, SVD

Orthogonal Matrices: Definition and Properties

Definition

A square matrix $Q \in \mathbb{R}^{n \times n}$ is **orthogonal** if its columns (and rows) form an orthonormal basis for $\mathbb{R}^n$.

Equivalent Properties

The following statements are equivalent:

1. Q is orthogonal.
2. $Q^T Q = I$
3. $Q Q^T = I$
4. $Q^{-1} = Q^T$

Orthogonal Matrices: geometric property

Key Property

Orthogonal matrices preserve the Euclidean norm (length) and the dot product (angles). For any vector $x \in \mathbb{R}^n$:

$$\|Qx\|_2^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T I x = \|x\|_2^2$$

This means they represent rigid transformations, such as rotations and reflections.

Examples of Orthogonal Matrices in 2D

Rotation Matrix

A counter-clockwise rotation by an angle θ in $\mathbb{R}^2$ is represented by:

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

You can verify that $R(\theta)^T R(\theta) = I$.

Reflection Matrix

A reflection across a line passing through the origin with angle θ is:

$$H(\theta) = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}$$

This is also an orthogonal matrix. Note that $\det(R) = 1$ (rotation) while $\det(H) = -1$ (reflection).

Symmetric Matrices and Eigenvectors

Definition

A square matrix $A \in \mathbb{R}^{n \times n}$ is **symmetric** if $A = A^T$.

Geometric Meaning of Eigenvectors

For a matrix A, an eigenvector v is a special vector whose direction is unchanged by the linear transformation represented by A. It is only scaled by a factor λ, the eigenvalue.

$$A\mathsf{v} = \lambda\mathsf{v}$$

The eigenvectors define the axes along which the transformation acts as a simple stretching or compressing.

Properties of Symmetric Matrices

Theorem

For any real symmetric matrix A:

1. *All its eigenvalues are real.*
2. *Its eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof of Real Eigenvalues.

Let (λ, v) be an eigenpair, possibly complex. So $Av = \lambda v$. Taking the conjugate transpose gives $v^* A^* = \bar{\lambda} v^*$. Since A is real and symmetric, $A^* = A^T = A$. So $v^* A = \bar{\lambda} v^*$. Right-multiply by v: $v^* A v = \bar{\lambda} v^* v = \bar{\lambda} \|v\|_2^2$. Now, left-multiply $Av = \lambda v$ by v^*: $v^* A v = \lambda v^* v = \lambda \|v\|_2^2$. Comparing the two, we get $\lambda \|v\|_2^2 = \bar{\lambda} \|v\|_2^2$. Since $v \neq 0$, we must have $\lambda = \bar{\lambda}$, which means λ is real. $\qquad\square$

Proof of Orthogonal Eigenvectors.

Let $(\lambda_1, \mathsf{v}_1)$ and $(\lambda_2, \mathsf{v}_2)$ be two eigenpairs with $\lambda_1 \neq \lambda_2$. We know $A\mathsf{v}_1 = \lambda_1 \mathsf{v}_1$ and $A\mathsf{v}_2 = \lambda_2 \mathsf{v}_2$. Consider the expression $\mathsf{v}_1^T A \mathsf{v}_2$.

$$\mathsf{v}_1^T (A\mathsf{v}_2) = \mathsf{v}_1^T (\lambda_2 \mathsf{v}_2) = \lambda_2 \mathsf{v}_1^T \mathsf{v}_2$$

$$(\mathsf{v}_1^T A)\mathsf{v}_2 = (A^T \mathsf{v}_1)^T \mathsf{v}_2 = (A\mathsf{v}_1)^T \mathsf{v}_2 = (\lambda_1 \mathsf{v}_1)^T \mathsf{v}_2 = \lambda_1 \mathsf{v}_1^T \mathsf{v}_2$$

Equating the two expressions gives $\lambda_2 \mathsf{v}_1^T \mathsf{v}_2 = \lambda_1 \mathsf{v}_1^T \mathsf{v}_2$.

$$(\lambda_1 - \lambda_2)\mathsf{v}_1^T \mathsf{v}_2 = 0$$

Since $\lambda_1 \neq \lambda_2$, we must have $\mathsf{v}_1^T \mathsf{v}_2 = 0$, meaning the eigenvectors are orthogonal. □

The Spectral Theorem (Decomposition)

Theorem (Spectral Theorem for Symmetric Matrices)

If $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix, then it can be diagonalized by an orthogonal matrix Q.

$$A = Q \Lambda Q^T$$

where:

- Q is an orthogonal matrix whose columns are the orthonormal eigenvectors of A.
- Λ is a diagonal matrix whose entries are the corresponding real eigenvalues of A.

This is also called the eigendecomposition of A.

Symmetric Positive Definite (SPD) Matrices

Definition

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **Positive Definite** if for any non-zero vector $x \in \mathbb{R}^n$:

$$x^T A x > 0$$

Theorem (Equivalent Definitions for SPD)

The following are equivalent:

1. *A is SPD (i.e., $x^T A x > 0$ for all $x \neq 0$).*
2. *All eigenvalues of A are strictly positive ($\lambda_i > 0$).*
3. *All leading principal minors of A have positive determinants.*
4. *A can be written as $A = R^T R$ for some non-singular upper triangular matrix R (Cholesky decomposition).*

$(1 \implies 2)$: All eigenvalues are positive.

Let (λ, v) be an eigenpair of A. Since A is SPD, we have:

$$v^T A v > 0$$

But we also know $Av = \lambda v$. Substituting this in:

$$v^T(\lambda v) = \lambda(v^T v) = \lambda \|v\|_2^2 > 0$$

Since v is an eigenvector, $\|v\|_2^2 > 0$. Therefore, we must have $\lambda > 0$. $\qquad\square$

Proof of Equivalence: $(2) \implies (1)$

$(2 \implies 1)$: $x^T A x > 0$.

Since A is symmetric, we can use its spectral decomposition $A = Q \Lambda Q^T$. For any non-zero vector x, let $y = Q^T x$. Since Q is invertible, y is also non-zero.

$$x^T A x = x^T (Q \Lambda Q^T) x = (Q^T x)^T \Lambda (Q^T x)$$
$$= y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2$$

Since all $\lambda_i > 0$ and at least one $y_i \neq 0$, the sum $\sum \lambda_i y_i^2$ must be strictly positive. $\square$

QR Decomposition

Theorem

Any real matrix $A \in \mathbb{R}^{m \times n}$ can be factored into:

$$A = QR$$

where:

- *Q is an $m \times m$ orthogonal matrix.*
- *R is an $m \times n$ upper trapezoidal matrix.*

Meaning

The QR decomposition expresses the matrix A as a change of basis (rotation/reflection) followed by a scaling and shearing operation. The columns of Q form an orthonormal basis for $\mathbb{R}^m$, and R holds the coordinates of the columns of A in this new basis.

Full vs. Economy QR

Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

Full QR

$$A_{m \times n} = Q_{m \times m} R_{m \times n} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

Here Q is a square orthogonal matrix, and R has a block of zeros at the bottom. The columns of Q_1 form a basis for the column space of A, while the columns of Q_2 form a basis for its orthogonal complement.

Economy (or Thin or Reduced) QR

We can discard the parts that multiply by zero:

$$A_{m \times n} = \hat{Q}_{m \times n} \hat{R}_{n \times n}$$

Here $\hat{Q}$ has orthonormal columns (but is not square unless $m = n$), and $\hat{R}$ is a square upper triangular matrix. This is more memory efficient and is what is commonly used in practice.
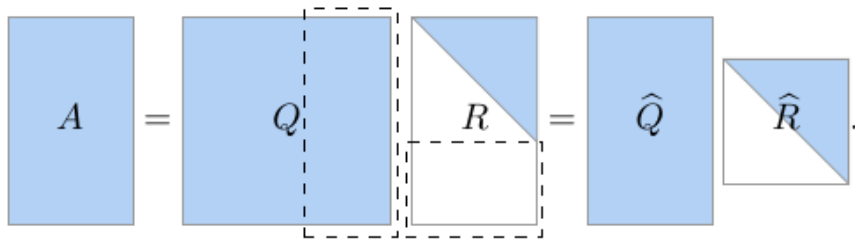
Figure: Full QR vs Economy QR

Gram-Schmidt Algorithm: Intuition

The Gram-Schmidt process is an algorithm for constructing an orthonormal basis from a set of linearly independent vectors.

Intuitive Idea

Imagine you have two vectors a_1 and a_2.

1. **Normalize the first vector:** Take a_1 and scale it to have unit length. This is your first basis vector, q_1.

2. **Remove projection:** Take the second vector a_2. Find its projection onto q_1 and subtract it from a_2. The result is a new vector that is orthogonal to q_1.

3. **Normalize the new vector:** Scale this new orthogonal vector to have unit length. This is your second basis vector, q_2.

Continue this process for all vectors.

Gram-Schmidt: Formulas for QR

Given the columns $a_1, \ldots, a_n$ of a matrix A.

The algorithm computes the orthonormal columns $q_1, \ldots, q_n$ of Q and the entries of R as follows:

For $j = 1, 2, \ldots, n$:

1. Start with a temporary vector $v_j = a_j$.
2. For $i = 1, \ldots, j - 1$, subtract the projections:
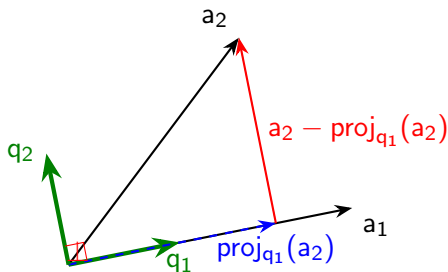
$$R_{ij} = q_i^T a_j$$

$$v_j = v_j - R_{ij} q_i$$

3. Compute the norm and normalize:

$$R_{jj} = \|v_j\|_2$$

If $R_{jj} = 0$, the column is linearly dependent. Stop or handle.

$$q_j = v_j / R_{jj}$$

Visualizing Gram-Schmidt for Two Vectors

Figure: The process starts with vectors a_1, a_2. q_1 is the normalized a_1. The projection of a_2 onto q_1 is subtracted from a_2 to create an orthogonal vector, which is then normalized to produce q_2.

Gram-Schmidt Example: Full Rank

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$. Columns are a_1, a_2.

Step 1 (j=1): $R_{11} = \|a_1\|_2 = \sqrt{1^2 + 0^2 + 1^2} = \sqrt{2}$. $q_1 = a_1/R_{11} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

Step 2 (j=2): $R_{12} = q_1^T a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}}$.

$v_2 = a_2 - R_{12}q_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \end{bmatrix}$.

$R_{22} = \|v_2\|_2 = \sqrt{1/4 + 1 + 1/4} = \sqrt{3/2}$. $q_2 = v_2/R_{22} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$.

$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ 0 & 2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{6} \end{bmatrix}$, $R = \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3/2} \end{bmatrix}$

Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 1 & 2 \end{bmatrix}$. Note $a_2 = 2a_1$.

Step 1 (j=1): $R_{11} = \|a_1\|_2 = \sqrt{2}$. $q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

Step 2 (j=2): $R_{12} = q_1^T a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = \frac{4}{\sqrt{2}} = 2\sqrt{2}$.

$v_2 = a_2 - R_{12}q_1 = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} - 2\sqrt{2} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

$R_{22} = \|v_2\|_2 = 0$. The algorithm breaks down. This indicates that a_2 is linearly dependent on a_1. The rank of the matrix is 1.

SVD: Generalizing Spectral Decomposition

Limitation of Eigendecomposition

The spectral decomposition $A = Q\Lambda Q^T$ is powerful, but it only works for **symmetric square** matrices.

The Question

Can we find a similar decomposition for *any* matrix $A \in \mathbb{R}^{m \times n}$, even rectangular ones?

The Answer: SVD

Yes! The Singular Value Decomposition (SVD) decomposes any matrix into two orthogonal matrices (U, V) and one diagonal matrix (Σ). It finds orthonormal bases for the four fundamental subspaces of the matrix.

Proof of Existence of the SVD I

Let $A \in \mathbb{R}^{m \times n}$. The proof is constructive.

1. Consider the matrix $A^T A$. This is an $n \times n$ symmetric, positive semi-definite matrix.

2. By the Spectral Theorem, we can find an orthonormal basis of eigenvectors for $A^T A$. Let these be $\{v_1, \ldots, v_n\}$ with corresponding real, non-negative eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

3. Define the **singular values** as $\sigma_i = \sqrt{\lambda_i}$. Let's say the first r singular values are non-zero (i.e., $\text{rank}(A^T A) = r$).

4. For $i = 1, \ldots, r$, define the vectors $u_i \in \mathbb{R}^m$ as:

$$u_i = \frac{1}{\sigma_i} A v_i$$

Proof of Existence of the SVD II

5. We must show these u_i vectors are orthonormal. Consider their dot product:

$$u_i^T u_j = \left(\frac{1}{\sigma_i} A v_i\right)^T \left(\frac{1}{\sigma_j} A v_j\right) = \frac{1}{\sigma_i \sigma_j} v_i^T A^T A v_j$$

$$= \frac{1}{\sigma_i \sigma_j} v_i^T (\lambda_j v_j) = \frac{\lambda_j}{\sigma_i \sigma_j} (v_i^T v_j)$$

Since the v's are orthonormal, $v_i^T v_j = \delta_{ij}$. If $i = j$, this gives $\frac{\lambda_i}{\sigma_i^2} = 1$. If $i \neq j$, it is 0. So the u_i are orthonormal.

6. We have r vectors $\{u_1, \ldots, u_r\}$. We can extend this set to an orthonormal basis for all of $\mathbb{R}^m$ by adding $m - r$ more vectors.

7. By construction, we have $A v_i = \sigma_i u_i$. This can be written in matrix form as $AV = U\Sigma$, leading to $A = U\Sigma V^T$. □

The Three Geometric Transformations of SVD
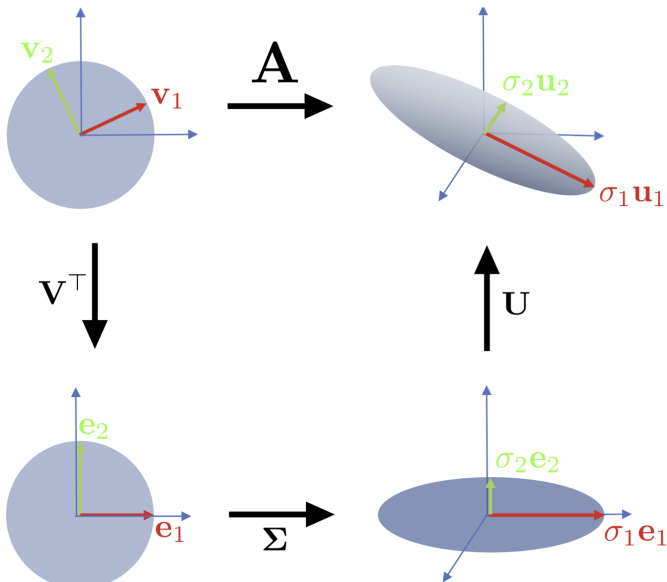
The Three Geometric Transformations of SVD

- 1. **Initial Rotation (V^T):**
  - The matrix V^T rotates (or reflects) the original coordinate system.
  - It aligns the input basis vectors with the principal directions along which the transformation will occur.
- 2. **Scaling (Σ):**
  - The diagonal matrix Σ scales these aligned axes.
  - Each singular value σ_i stretches or shrinks the corresponding principal direction.
  - A unit circle (or sphere) is transformed into an ellipse (or ellipsoid).
- 3. **Final Rotation (U):**
  - The matrix U applies a final rotation (or reflection) to the scaled axes.
  - This aligns the stretched ellipse/ellipsoid with the final output coordinate system.

Full SVD

For $A \in \mathbb{R}^{m \times n}$:

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$$

- U: Orthogonal matrix with columns = left singular vectors.
- V: Orthogonal matrix with columns = right singular vectors.
- Σ: Diagonal matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$.

Reduced SVD (Economy)

If $m \geq n$ and rank$(A) = r$:

$$A_{m \times n} = U_{m \times r} \Sigma_{r \times r} V_{r \times n}^T$$

This form is more compact and only uses the parts of the matrices that contribute to reconstructing A. It's the outer product form:

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Full SVD

Reduced SVD

Meaning of U and V

The columns of U and V provide orthonormal bases for the four fundamental subspaces of A. Let rank$(A) = r$.

- First r columns of U: Orthonormal basis for the **Column Space** $C(A)$.
- Last $m - r$ columns of U: Orthonormal basis for the **Left Null Space** $N(A^T)$.
- First r columns of V: Orthonormal basis for the **Row Space** $C(A^T)$.
- Last $n - r$ columns of V: Orthonormal basis for the **Null Space** $N(A)$.

Importance in Data Science / ML

- **Dimensionality Reduction (PCA):** SVD is the engine behind Principal Component Analysis. It finds the directions of greatest variance in the data.
- **Low-Rank Approximation:** The Eckart-Young theorem states that truncating the SVD gives the best low-rank approximation of a matrix, which is used for data compression and denoising.
- **Recommender Systems:** Used in collaborative filtering (e.g., Netflix prize) to find latent factors in user-item rating matrices.
- **Solving Linear Systems:** Provides a robust way to solve least-squares problems, especially for ill-conditioned or rank-deficient matrices.

Polar Decomposition I

Theorem

Any square matrix $A \in \mathbb{R}^{n \times n}$ can be factored into:

$$A = UP$$

where:

- *U is an orthogonal matrix.*
- *P is a symmetric positive semi-definite matrix.*

Polar Decomposition II

Derivation from SVD

Start with the SVD of A: $A = U_{svd}\Sigma V^T$. We can rewrite this as:

$$A = (U_{svd}V^T)(V\Sigma V^T)$$

Let $U = U_{svd}V^T$ and $P = V\Sigma V^T$.

- U is a product of orthogonal matrices, so it is orthogonal.
- P is symmetric since $(V\Sigma V^T)^T = V\Sigma^T V^T = P$. It is also positive semi-definite because its eigenvalues are the singular values of A, which are non-negative.

Polar Decomposition III

Use in Mechanics

This decomposition is analogous to writing a complex number as $z = e^{i\theta}r$. In continuum mechanics, the deformation gradient tensor F is decomposed as $F = RU$, where R is a pure rotation and U is a pure stretch. This separates the rigid body motion from the strain (deformation) of the material.