Linear Algebra Review

EECS 127 Spring 2026

(adapted by Phoenix Wilson from Axler's Linear Algebra Done Right and Olver's Applied Linear Algebra)

Overview

1. Vector Spaces
    a. Axioms
    b. Subspaces
    c. Sums/Direct Sums
    d. Span
    e. Linear Independence
    f. Bases
2. Inner Product Spaces
    a. Inner Products
    b. Norms
    c. Orthonormal Bases
    d. Projection
    e. Gram-Schmidt
    f. Orthogonal Complements
3. Matrices
    a. Operations
    b. Fundamental Subspaces
    c. Least Squares
    d. Minimum Norm Solution
    e. Eigenstuff
    f. Singular Value Decomposition
    g. Moore-Penrose Pseudoinverse
4. Classifications of Matrices
    a. Orthogonal Matrices
    b. Symmetric Matrices
5. Extra Resources

WARNING!!!

We intend for this presentation to provide a delineation of the fundamental ideas in your previous linear algebra classes. Please use this as an opportunity to review and highlight what you are still fuzzy about, and not as a comprehensive learning tool.

(also, not extensively reviewed, so there will be mistakes)

1. Vector Spaces

Axioms Vector Spaces

A **field** is an algebraic structure $(F, +, \cdot)$ containing a set F and binary operations $+ : F \times F \to F$ and $\cdot : F \times F \to F$.

F1. Associativity. $a + (b + c) = (a + b) + c$

F2. Commutativity. $a + b = b + a$

F3. Identities. $\exists 0 \in F : a + 0 = a \ \wedge \exists 1 \in F : a \cdot 1 = a$

F4. Additive Inverse. $\exists ! -a \in F : a + (-a) = 0$

F5. Multiplicative Inverse. $\exists ! a^{-1} \in F : a \cdot a^{-1} = 1 \quad a \neq 0$

F6. Distributivity. $a \cdot (b + c) = a \cdot b + a \cdot c$

Examples: $\mathbb{R}$, $\mathbb{C}$

Axioms (cont.)

A **vector space** over a field F is an algebraic structure $(V, +, \cdot)$ containing a set V and binary operations $+ : V \times V \to V$ and $\cdot : F \times V \to V$.

V1. Associativity. $u + (v + w) = (u + v) + w$ and $\lambda(\mu v) = (\lambda\mu)v$

V2. Commutativity. $u + v = v + u$

V3. Identities. $\exists! 0 \in V : u + 0 = u \wedge \exists! 1 \in F : 1v = v$

V4. Additive Inverse. $\exists! -u \in V : u + (-u) = 0$

V5. Distributivity. $(\lambda + \mu)u = \lambda u + \mu u$ and $\lambda(u + v) = \lambda u + \lambda v$

Main Idea: Matrix multiplication (addition and scalar multiplication of vectors) maps vector spaces to vector spaces.

Examples: $\mathbb{C}^n$, A^B (functions from a set B to a vector space A), $\mathbb{R}^{m\times n}$ (why we can treat block matrices like vectors)

Subspaces

Vector Spaces

A **subspace** U of a vector space V is a vector space (with functionally the same operations and over the same field) whose ground set is a subset of V.

S1. Additive Identity / Nonempty. $\exists ! 0 \in U$ or $U \neq \emptyset$

S2. Closure under Addition. $+ \restriction_{U \times U} : U \times U \to U$

S3. Closure under Scalar Multiplication. $\cdot \restriction_{F \times U} : F \times U \to U$

Why is this enough? You get associativity, commutativity, distributivity, and multiplicative identity for free. Closure gets you the additive inverse and additive identity (so long as U is nonempty).

Examples: $u, v \in V, \operatorname{span}(u, v) \leq V$ $\qquad$ $A \in \mathbb{R}^{m \times n}, C(A) \leq \mathbb{R}^m$

Sums/Direct Sums

The **sum** of subspaces U and W is the subspace defined below

$$U + W := \{u + w \mid u \in U, w \in W\}$$

The sum is said to be a **direct sum** if each element in the resulting space can be written in exactly one way as the sum of elements from U and W

$$\begin{aligned} U + W = U \oplus W &\iff \forall v \in U + W, \exists! u \in U, w \in W : v = u + w \\ &\iff \forall u \in U, w \in W, u + w = 0 \implies u = w = 0 \\ &\iff U \cap W = \{0\} \end{aligned}$$

Examples: $\mathrm{span}(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}) \oplus \mathrm{span}(\begin{bmatrix}0\\0\\1\end{bmatrix}) = \mathbb{R}^3$

Span

Vector Spaces

The **span** of a list of vectors $v_1, \ldots, v_k \in V$ is the set of all possible linear combinations of those vectors and forms a subspace of V.

$$\operatorname{span}(v_1, \ldots, v_k) := \{\sum_{i=1}^{k} c_i v_i \mid c_1, \ldots, c_k \in F\}$$

This is the lowest dimensional subspace that intersects these vectors.

Examples: $\operatorname{span}(e_1, e_2) = \{c_1 \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix} \mid c_1, c_2 \in \mathbb{R}\} = \mathbb{R}^2$

Linear Independence

Vector Spaces

A list of vectors $v_1, \ldots, v_k \in V$ is **linearly independent** if each vector in its span can be written as exactly one linear combination of $v_1, \ldots, v_k$.

$$\forall v \in \operatorname{span}(v_1, \ldots, v_k), \exists! c_1, \ldots, c_k \in F : v = \sum_{i=1}^{k} c_i v_i$$

Or, equivalently, no vector in the list can be written as a linear combination of the others

$$\forall c_1, \ldots, c_k \in F, \sum_{i=1}^{k} c_i v_i = 0 \implies c_1 = \cdots = c_k = 0$$

Bases

Vector Spaces

A list of vectors $v_1, \ldots, v_n \in V$ is a **basis** of V if it is linearly independent and its span is equal to V.

Intuitively, a basis represents a vector space's choice of coordinates, since any vector in that vector space has $\underbrace{\text{exactly}}_{\text{(lin. ind.)}}$ $\underbrace{\text{one}}_{\text{(spanning)}}$ representation wrt a basis.

The length of any basis of a vector space is the same. For this reason, we define the **dimension** of a vector space to be the length of any basis for it.

Examples.

- e_1, e_2 is linearly independent and $\operatorname{span}(e_1, e_2) = \mathbb{R}^2$, so $\dim \mathbb{R}^2 = 2$
- $A \in \mathbb{R}^{n \times n}$ has full rank, what is a basis for $\mathbb{R}^n$?

2. Inner Product Spaces

Inner Products

An **inner product space** is a vector space V equipped with a binary function $\langle \cdot \mid \cdot \rangle : V \times V \to F$ known as an inner product.

I1. Positivity	$\langle v\|v\rangle \geq 0$
I2. Definiteness	$\langle v\|v\rangle = 0 \iff v = 0$
I3. Linearity in 1st Arg.	$\langle au + bw\|v\rangle = a\langle u\|v\rangle + b\langle w\|v\rangle$
I4. Conjugate Symmetry	$\langle u\|v\rangle = \overline{\langle v\|u\rangle}$

In this class, we will almost exclusively work with the Euclidean inner product over real vector spaces (dot product).

$$\langle u|v\rangle_{\text{euclidean}} := u^T v = \sum_{i=1}^{n} u_i v_i$$

Norms

Inner Product Spaces

An inner product induces a **norm** $||\cdot||: V \to F$.

N1. Positivity $||v|| \geq 0$

N2. Definiteness $||v|| = 0 \iff v = 0$

N3. Homogeneity $||cv|| = |c| \cdot ||v||$

$$\|v\|^2 := \langle v|v\rangle$$

Important Results:

Pythagorean Theorem.
If $\langle u \mid v\rangle = 0$, $||u+v||^2 = ||u||^2 + ||v||^2$.

Cauchy-Schwarz Inequality.
$|\langle u, v\rangle| \leq ||u|| \cdot ||v||$ w/ equality iff $u \in \text{span } v$ or $v \in \text{span } u$.

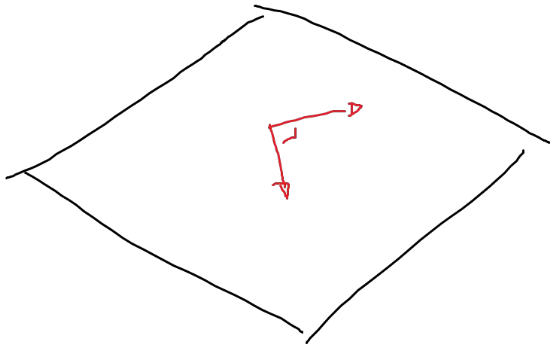
Triangle Inequality.
$||u+v|| \leq ||u|| + ||v||$ w/ equality iff one is non-neg. multiple of other

Orthonormal Bases

A basis $v_1, \ldots, v_n \in V$ is said to be an **orthonormal basis** of V if each vector is orthogonal to one another and has unit norm.

$$\langle v_i | v_j \rangle = \delta[i-j] = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Orthonormal bases are nice to work in!



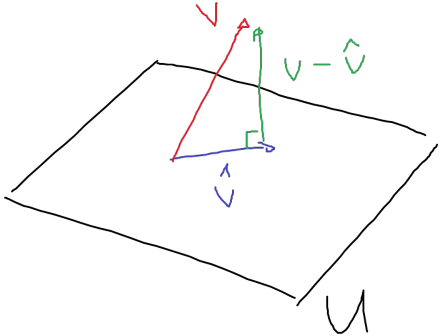
$$\|v\|^2 = \sum_{i=1}^{n} |\langle v | v_i \rangle|^2$$

$$\mathrm{proj}_{\mathrm{span}(v_1, \ldots, v_k)} w = \sum_{i=1}^{k} \langle w | v_i \rangle v_i$$

Projection

The (orthogonal) **projection** of a vector $v \in V$ onto a subspace $U \leq V$ is the vector $\hat{v} \in U$ which makes the error orthogonal to all of U.

$$\forall u \in U, \langle u \,|\, v - \hat{v} \rangle = 0$$



Given a basis $u_1, \ldots, u_k \in U$ of U, we can compute $\hat{v}$ by solving the following system of equations for the coefficients:

$$\langle u_j \,|\, v - \sum_{i=1}^{k} c_i u_i \rangle = 0$$

Projection (cont.)

The projection $\hat{v}$ is not only unique, but also minimizes the norm of the error vector.

$$\text{proj}\,_U v := \hat{v} = \arg\min_{u \in U} ||v - u||$$

Proof.

Let $\hat{v} \in U$ be the projection of v onto U and fix $u \in U$.

$$\|v-u\|^2 = \|(v-\hat{v})+(\hat{v}-u)\|^2 = \|v-\hat{v}\|^2 + \|\hat{v}-u\|^2 + 2\langle v-\hat{v} \mid \hat{v}-u\rangle$$

By the closure property of subspaces, $\hat{v} - u \in U$, thus by def. of $\hat{v}$:

$$= ||v - \hat{v}||^2 + ||\hat{v} - u||^2 + 0$$

$$\geq ||v - \hat{v}||^2$$

Gram-Schmidt

Inner Product Spaces

Projecting onto the span of orthonormal vectors is really easy! However, not every list of vectors is orthonormal.

Gram-Schmidt is an algorithm that converts a list of linearly independent vectors $v_1, \ldots, v_k \in V$ to a list of orthonormal vectors $u_1, \ldots, u_k \in V$ with the same span.

$$\text{for } i = 1, 2, \ldots, k$$

$$z_i \leftarrow v_i - \sum_{j=1}^{i-1} \langle v_i \mid u_j \rangle u_j = v_i - \mathbf{proj}_{\mathrm{span}(u_1, \ldots, u_{i-1})} v_i$$

$$u_i \leftarrow \frac{z_i}{||z_i||}$$

Orthogonal Complements

The **orthogonal complement** of a subspace $U \leq V$ is the set of all vectors orthogonal to every vector in U, which also forms a subspace of V.

$$U^{\perp} := \{ w \in V \,|\, \forall\, u \in U, \langle w \,|\, u \rangle = 0 \}$$

The ambient vector space V is partitioned by any subspace and its orthogonal complement. In this way, their sum is direct:

$$U \oplus U^{\perp} = V$$

Examples: $(\mathrm{span}(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}))^{\perp} = \mathrm{span}(\begin{bmatrix}0\\0\\1\end{bmatrix})$

3. Matrices

Operations

Matrices

A **matrix** $A \in F^{m \times n}$ is an m by n rectangular grid of elements from F.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} = \begin{bmatrix} \alpha_1^T \\ \vdots \\ \alpha_m^T \end{bmatrix}$$

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} = \begin{bmatrix} \alpha_1 & \dots & \alpha_m \end{bmatrix}$$

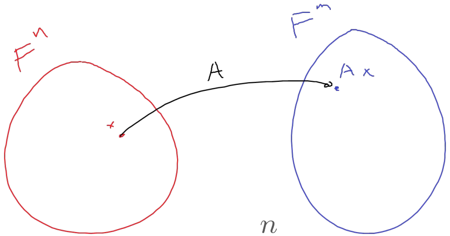
Operations (cont.)

We define **matrix-vector multiplication** between a vector $x \in F^n$ and a matrix $A \in F^{m \times n}$ over the same field in the following equivalent ways:

$$Ax = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i a_i$$

$$Ax = \begin{bmatrix} \alpha_1^T \\ \vdots \\ \alpha_m^T \end{bmatrix} x = \begin{bmatrix} \alpha_1^T x \\ \vdots \\ \alpha_m^T x \end{bmatrix}$$

Operations (cont.) Matrices



Think about matrix-vector multiplication like a function $f : F^n \to F^m, x \mapsto \sum_{i=1}^{n} x_i a_i$

When f is surjective (onto), everything in F^m must be mapped to.

$$f \text{ is surjective} \iff \dim \operatorname{span}(a_1, \ldots, a_n) = m \; (m \text{ lin. ind. columns})$$

When f is injective, each mapping is unique.

$$f \text{ is injective} \iff a_1, \ldots, a_n \text{ is linearly independent}$$

When f is bijective, everything in F^m has a unique mapping to it.

$$f \text{ is bijective} \iff a_1, \ldots, a_n \text{ is linearly independent and } m = n$$

Fundamental Subspaces

A matrix $A \in F^{m \times n}$ has **4 fundamental subspaces**

1. **Column Space** $C(A) := \{Ax \mid x \in F^n\} = \text{span}(a_1, \ldots, a_n)$
2. **Null Space** $\mathcal{N}(A) := \{x \in F^n \mid Ax = 0\}$
3. **Row Space** $C(A^T) := \{A^T b \mid b \in F^m\} = \text{span}(\alpha_1, \ldots, \alpha_m)$
4. **Left Null Space** $\mathcal{N}(A^T) := \{b \in F^m \mid A^T b = 0\}$

Fundamental Subspaces (cont.)

Important Results

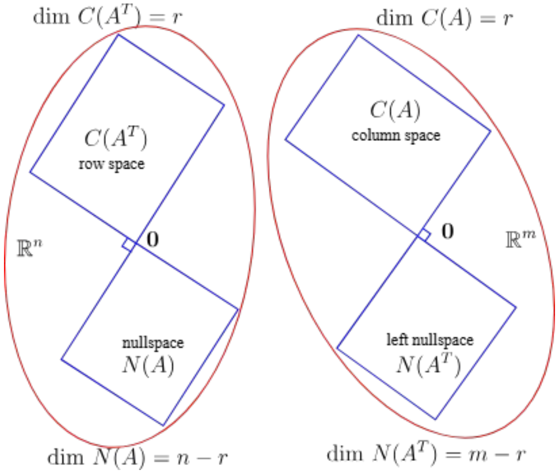
$$Ax = 0 \iff \alpha_1^T x = \cdots = \alpha_m^T x = 0 \iff \boxed{\mathcal{N}(A) = (C(A^T))^\perp} \implies \boxed{\mathcal{N}(A) \oplus C(A^T) = F^n}$$

$$A^T b = 0 \iff a_1^T b = \cdots = a_n^T b = 0 \iff \boxed{\mathcal{N}(A^T) = (C(A))^\perp} \implies \boxed{\mathcal{N}(A^T) \oplus C(A) = F^m}$$

$$\mathcal{N}(A) \oplus C(A^T) = F^n \implies \boxed{\dim \mathcal{N}(A) + \dim C(A^T) = n}$$

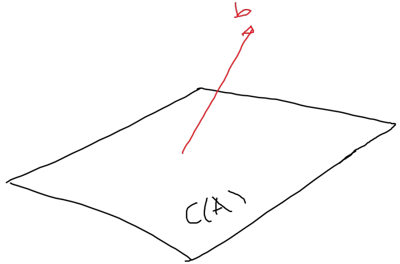
$$\mathcal{N}(A^T) \oplus C(A) = F^m \implies \boxed{\dim \mathcal{N}(A^T) + \dim C(A) = m}$$

$$\boxed{\dim C(A) = \dim C(A^T)} \implies \boxed{\underbrace{\dim \mathcal{N}(A)}_{\text{nullity } A} + \underbrace{\dim C(A)}_{\text{rank } A} = n}$$



Least Squares

Matrices



Given a system $Ax = b$, it may not be the case that $b \in C(A)$.

Least squares is an optimization criterion which finds the vector which maps to the vector closest to b wrt the Euclidean norm.

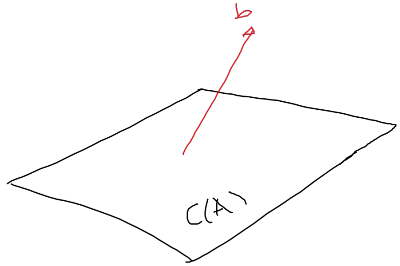
$$\underset{x \in \mathbb{R}^n}{\arg\min} \, ||Ax - b||_2^2$$

From our study of projections, we know that the unique element of $C(A)$ that minimizes the error is $\text{proj}_{C(A)} b$

$$a_1^T(Ax - b) = \cdots = a_n^T(Ax^* - b) = 0 \implies A^T(Ax^* - b) = 0$$

$$A^T A x^* = A^T b$$

Least Squares (cont.)



Note that even though the projection is unique, it may not always be the case that the vector which maps to it is itself unique.

In fact, for any solution x^*, we have that for any $z \in \mathcal{N}(A)$, $z + x^*$ is also a solution.

$$A^T A(x^* + z) = A^T A x^* + A^T(0) = A^T A x^* = A^T b$$

Hence, we typically fix an arbitrary solution x^* and write the solution set as the following affine set (not a subspace)

$$S_{lstq} = x^* + \mathcal{N}(A) := \{x^* + z \mid z \in \mathcal{N}(A)\}$$

Minimum Norm Solution

Assuming $b \in C(A)$, the system $Ax = b$ may have infinitely many solutions. What if we wanted a “canonical” one?

The **minimum norm solution** to $Ax = b$ is the solution with the least norm.

$$\arg\min_{x \in \mathbb{R}^n} ||x||_2$$
$$s.t.\ Ax - b = 0$$

Every solution takes the form $x^* + z$ for some $z \in \mathcal{N}(A)$ and a fixed $x^* \in \mathbb{R}^n$.

$$||x||^2 = ||x^* + z||^2 = ||x^*_{\mathcal{N}(A)} + x^*_{\mathcal{N}(A)^\perp} + z||^2 = \underbrace{||x^*_{\mathcal{N}(A)^\perp}||^2}_{\text{fixed}} + ||x^*_{\mathcal{N}(A)} + z||^2$$

To minimize this, we take $z = -x^*_{\mathcal{N}(A)}$, thus $\underline{x} \in \mathcal{N}(A)^\perp$

$$\underline{x} \in \mathcal{N}(A)^\perp = C(A^T) \implies \underline{x} = A^T u \quad \exists u \in \mathbb{R}^m \implies b = A(A^T u)$$

Minimum Norm Solution (cont.)

As it turns out, even if there are multiple solutions to the system $b = AA^T u$, the minimum norm solution is always unique.

$$AA^T u = b$$

$$\underline{x} = A^T u$$

Eigenstuff

Matrices

For a square matrix $A \in F^{n \times n}$, the **eigenvalue problem** is

$$Av = \lambda v \qquad v \neq 0$$

$$(A - I\lambda)v = 0 \implies \det(A - I\lambda) = 0$$

We call $p(\lambda) := \det(A - \lambda I)$ the characteristic polynomial.

Each root λ_i is called an eigenvalue and the solutions to $(A - \lambda_i I)v_i = 0$ are called eigenvectors corresponding to λ_i.

Eigenvectors corresponding to distinct eigenvalues are linearly independent.

Eigenstuff (cont.) Matrices

For each eigenvalue $\lambda_i \in \sigma(A)$, we define its **algebraic multiplicity** as the multiplicity of λ_i as a root to $p(\lambda)$ and its **geometric multiplicity** as the number of linearly independent eigenvectors corresponding to it.

$$(\lambda_i)_{alg} \geq (\lambda_i)_{geo}$$

A matrix $A \in F^{n\times n}$ is said to be **diagonalizable** if for each of its eigenvalues, the algebraic multiplicity is equal to the geometric multiplicity.

$$AV = [Av_1 \quad \dots \quad Av_n] = [\lambda_1 v_1 \quad \dots \quad \lambda_n v_n] = [v_1 \quad \dots \quad v_n]\mathrm{diag}(\lambda_1, \dots, \lambda_n) = V\Lambda$$

$$A \text{ is diagonalizable} \implies A \text{ has } n \text{ linearly independent eigenvectors} \implies A = V\Lambda V^{-1}$$

Singular Value Decomposition

Any matrix $A \in F^{m \times n}$ has a **singular value decomposition**.

$$A = \underbrace{[U]_{m \times m}}_{\text{orthogonal}} [\Sigma]_{m \times n} \underbrace{[V^T]_{n \times n}}_{\text{orthogonal}}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} [\Sigma_r]_{r \times r} & [0]_{r \times (n-r)} \\ [0]_{(m-r) \times r} & [0]_{(m-r) \times (n-r)} \end{bmatrix}$$

We call the diagonal entries of Σ_r the **singular values** of A, where $r := \mathrm{rank}(A)$ and conventionally order them in the following way:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$$

Singular Value Decomposition (cont.)

$$AA^T = U\Sigma V^T V \Sigma^T U^T = U \begin{bmatrix} [\Sigma_r]_{r\times r} & [0]_{r\times(n-r)} \\ [0]_{(m-r)\times r} & [0]_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} [\Sigma_r]_{r\times r} & [0]_{r\times(m-r)} \\ [0]_{(n-r)\times r} & [0]_{(n-r)\times(m-r)} \end{bmatrix} U^T$$

$$= U \begin{bmatrix} [\Sigma_r^2]_{r\times r} + [0]_{r\times(n-r)}[0]_{(n-r)\times r} & [\Sigma_r]_{r\times r}[0]_{r\times(m-r)} + [0]_{r\times(n-r)}[0]_{(n-r)\times(m-r)} \\ [0]_{(m-r)\times r}[\Sigma_r]_{r\times r} + [0]_{(m-r)\times(n-r)}[0]_{(n-r)\times r} & [0]_{(m-r)\times r}[0]_{r\times(m-r)} + [0]_{(m-r)\times(n-r)}[0]_{(n-r)\times(m-r)} \end{bmatrix} U^T$$

$$= U \begin{bmatrix} [\Sigma_r^2]_{r\times r} & [0]_{r\times(m-r)} \\ [0]_{(m-r)\times r} & [0]_{(m-r)\times(m-r)} \end{bmatrix} U^T$$

$$\implies AA^T U = [AA^T u_1 \quad \dots \quad AA^T u_m] = U \begin{bmatrix} [\Sigma_r^2]_{r\times r} & [0]_{r\times(m-r)} \\ [0]_{(m-r)\times r} & [0]_{(m-r)\times(m-r)} \end{bmatrix} = [u_1 \quad \dots \quad u_m] \begin{bmatrix} \sigma_1^2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r^2 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} = [\sigma_1^2 u_1 \quad \dots \quad \sigma_r^2 u_r \quad 0u_{r+1} \quad \dots \quad 0u_m]$$

Singular Value Decomposition (cont.)

$$\Longrightarrow \begin{cases} AA^T u_i = \underbrace{\sigma_i^2}_{\lambda_i(AA^T)} u_i & \text{for } i = 1, 2, \ldots, r \\ AA^T u_i = \underbrace{0}_{\lambda_i(AA^T)} u_i & \text{for } i > r \end{cases}$$

One can prove something very similar for $A^T A$.

Singular Value Decomposition (cont.)

The eigenvectors of AA^T form the columns of U (**left singular vectors**), the eigenvectors of A^TA form the columns of V (**right singular vectors**), and the positive square roots of the eigenvalues of both form the entries of Σ_r.

Spectral theorem (later) provides some justification for why U and V are orthogonal, and why Σ is real.

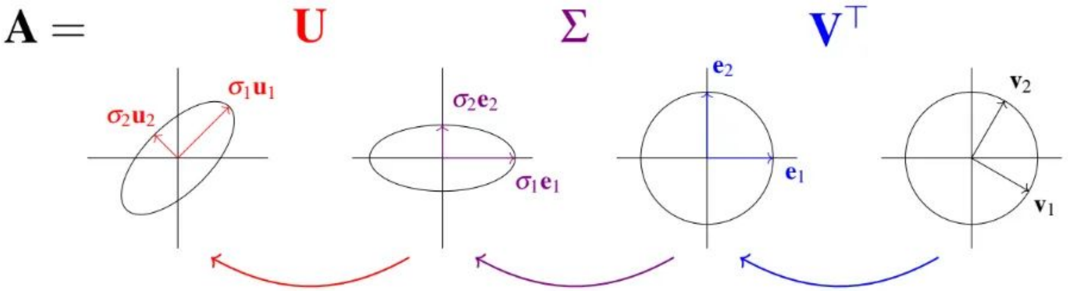
Algorithm for computing SVD:

1. Find eigenvalues/eigenvectors of A^TA or AA^T.
2. Compute U or V from $A = U\Sigma V^T$ vector-wise.

Singular Value Decomposition (cont.)

Main Corollaries.

1. Any linear transformation is a rotation followed by a stretch/compression along right singular vectors, followed by another rotation.



2. Any matrix can be decomposed into a series of rank 1 updates.

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Moore-Penrose Pseudoinverse

Any matrix $A \in F^{m\times n}$ has a unique **pseudoinverse**.

$$Ax = b \implies U\Sigma V^T x = b \implies \Sigma V^T x = U^T b$$

If we define $\Sigma^\dagger := \begin{bmatrix} [\Sigma_r^{-1}]_{r\times r} & [0]_{r\times(m-r)} \\ [0]_{(n-r)\times r} & [0]_{(n-r)\times(m-r)} \end{bmatrix}$

$$\Sigma^\dagger \Sigma V^T x = \begin{bmatrix} [I]_{r\times r} & [0]_{r\times(n-r)} \\ [0]_{(n-r)\times r} & [0]_{(n-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} [V_r^T]_{r\times n} \\ [V_{n-r}^T]_{(n-r)\times n} \end{bmatrix} x = \begin{bmatrix} [V_r^T]_{r\times n} \\ [0]_{(n-r)\times n} \end{bmatrix} x = \Sigma^\dagger U^T b$$

$$V \begin{bmatrix} [V_r^T]_{r\times n} \\ [0]_{(n-r)\times n} \end{bmatrix} x = \begin{bmatrix} [V_r]_{n\times r} & [V_{n-r}]_{n\times(n-r)} \end{bmatrix} \begin{bmatrix} [V_r^T]_{r\times n} \\ [0]_{(n-r)\times n} \end{bmatrix} x = V_r V_r^T x = V \Sigma^\dagger U^T b$$

Moore-Penrose Pseudoinverse (cont.)

$$Ax = b \implies \underbrace{V_r V_r^T x}_{\text{proj}_{C(V_r)} x} = V\Sigma^\dagger U^T b$$

One can show that $C(V_r) = C(A^T)$, so the left-hand side becomes:

$$\tilde{x} = V\Sigma^\dagger U^T b \qquad \tilde{x} \in C(A^T)$$

However, from previous study, we know that the minimum norm solution is the unique solution which lies entirely in $C(A^T)$.

$$\underline{x} = \tilde{x} = V\Sigma^\dagger U^T b = A^\dagger b$$

Moore-Penrose Pseudoinverse (cont.)

You'll notice that nowhere in our derivation did we assume $b \in C(A)$!

As it turns out, the transformations we applied to b actually projected it onto the column space of A.

$\Sigma^\dagger U^T b$

$$A^T(Ax^\dagger - b) = V\Sigma^T U^T (U\Sigma V^T A^\dagger b - b)$$

$$= V\Sigma^T \Sigma\Sigma^\dagger U^T b - V\Sigma^T U^T b$$

$$(*) \quad \Sigma^T\Sigma\Sigma^\dagger = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} = \Sigma^T$$

$$= V\Sigma^T U b - V\Sigma^T U b = 0$$

So the pseudoinverse also gets us the least-squares solution!

Moore-Penrose Pseudoinverse (cont.)

To summarize, the pseudoinverse gets you the least-squares solution of minimum norm:

$$x^\dagger = A^\dagger b = \arg\min_{x \in \mathcal{S}_{lstq}} ||x||_2$$

4. Types of Matrices

Orthogonal Matrices

An orthogonal matrix $Q \in F^{n \times n}$ is a square matrix with orthonormal columns.

$$Q^T Q = \underbrace{QQ^T}_{\text{(rows are orthonormal too!)}} = I$$

Orthogonal matrices correspond to linear isometries (distance/angle-preserving):

$$||Qx||_2 = ||x||_2$$

If a matrix $Q \in F^{m \times n}$ has orthonormal columns but is not square, then:

$$Q^T Q = I$$

$$QQ^T x = \text{proj}_{\,C(Q)} x$$

Orthogonal Matrices (cont.)

A further note on the orthogonality property:

If $Q \in F^{m \times n}$ (m > n) has orthogonal columns:

$$[Q^T]_{n\times m}[Q]_{m\times n} = [Q^T q_1 \quad Q^T q_2 \quad \dots \quad Q^T q_n] = [e_1 \quad e_2 \quad \dots \quad e_n] = I_n$$

$$[Q]_{m\times n}[Q^T]_{n\times m} = [q_1 \quad \dots \quad q_n]\begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} = [q_1 q_1^T + \dots + q_n q_n^T]_{m\times m}$$

$$\operatorname{rank}(QQ^T) \leq \min\{\operatorname{rank}(Q), \operatorname{rank}(Q^T)\} \implies \operatorname{rank}(QQ^T) \leq n < m = \operatorname{rank}(I_m) \implies QQ^T \neq I_m$$

$$\text{(aside: } \sum_{k=1}^{n} q_k q_k^T v = \sum_{k=1}^{n} \langle q_k \mid v \rangle q_k = \mathbf{proj}_{C(Q)} v)$$

Symmetric Matrices

A matrix $S \in F^{n \times n}$ is said to be **symmetric** if $S = S^T$.

The **spectral theorem** says that all symmetric matrices have real eigenvalues and a full set orthonormal eigenvectors (are orthonormally diagonalizable).

The fact that AA^T and A^TA are symmetric guarantee existence of an SVD.

These matrices are the workhorse of 127!

5. Extra Resources

Extra Resources

What we didn't cover (non-exhaustive):

- Projection Matrices
- QR Decomposition
- LU Decomposition
- Low Rank Factorization
- Change of Basis
- Cholesky Decomposition

Here are some extra resources to dive deeper:

Applied Linear Algebra (Olver and Shakiban)

Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares (Boyd and Vandenberghe)

3Blue1Brown's "Essence of Linear Algebra"

EECS 127 Course Reader

Feedback Form

https://forms.gle/aHMEonWeTe7pX6KV7