

## 1 Basics

### 1.1 Types of combinations

- **Affine:**  $\sum \lambda_i = 1$  (think infinite line  $\mu(u - v)$ )
- **Conic:**  $\lambda_i \geq 0$  (think positive subsection in direction of  $u \wedge v$ )
- **Convex:** Affine  $\wedge$  Conic (think intersection)

### 1.2 Norms

Assigns *non-negative* "sizes" to vectors.

- **1-Norm:**  $\sum |v_i|$  (measures travelled dist along axis)
- **2-Norm (Euclidian):**  $\sqrt{\sum v_i^2}$  (geometric distance)
- **p-Norm (Generalization):**  $\sqrt[p]{\sum v_i^p}$
- **Max-Norm:**  $\max\{v_i\}$

Other:

- $\|v\|^2 = v \cdot v$
- $\|1_n\| = \sqrt{n}$

### 1.3 Scalar Products

**Euclidian:**  $u \cdot v := u^T v$

Satisfy:

- $a \cdot (b + c) = a \cdot b + a \cdot c$  (linear in second factor)
  - $a \cdot (\lambda b) = \lambda(a \cdot b)$  (linear in second factor)
  - $a \cdot b = b \cdot a$  (symmetric for  $\mathbb{R}$ ) and  $a \cdot b = b^H \cdot a^H$  (hermitian for  $\mathbb{C}$ )
  - $\forall a \in V : a \cdot a (> 0) \vee (= 0 \text{ iff } a = 0)$  (positive definite)
- Other:
- $(x \cdot y)^2 \leq (x \cdot x)(y \cdot y)$

### 1.4 Angles

Given  $u, v \in \mathbb{R}^n$  and  $u' = \frac{u}{\|u\|}, v' = \frac{v}{\|v\|}$  unitized vectors:  
 $\cos(\alpha) = u' \cdot v'$ .

$$\sin : 0 \mapsto \frac{\sqrt{0}}{2}, 30 \mapsto \frac{\sqrt{1}}{2}, 45 \mapsto \frac{\sqrt{2}}{2}, 60 \mapsto \frac{\sqrt{3}}{2}, 90 \mapsto \frac{\sqrt{4}}{2}$$

### 1.5 Inequalities

#### 1.5.1 Cauchy-Schwarz

$$|u \cdot v| \leq \|u\| \|v\|, -1 \leq \frac{u \cdot v}{\|u\| \|v\|} \leq 1, -\|u\| \|v\| \leq u \cdot v \leq \|u\| \|v\|$$

#### 1.5.2 Triangle

$\|a + b\| \leq \|a\| + \|b\|$ , meaning the direct way is always  $\leq$  the indirect way.

### 1.6 Linear In/Dependence

**Linear Dependence Equivalent Definitions:**

1.  $\exists u \in V : u \in \text{span}\{V \setminus \{u\}\}$  (vector can be represented using others)
2.  $0 \in \text{span}(V)$  (0 combination)
3.  $\exists v_i \in V : v_i \in \text{span}\{V_{1..i-1}\}$  (vector can be represented by previous vectors)

### 1.7 CR Decomposition

$C$  : independent columns,  $R$  : combinations to get back to  $A$ . Basically run RREF on  $A$  and put identity columns into  $C$  and copy RREF without the ending zero-rows into  $R$ .

### 2 Matrices and Linear Transformations

Given a matrix in  $\mathbb{R}^{m \times n}$  ( $m$  rows,  $n$  columns), think of a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and how it possibly compresses information...

#### 2.1 Linear Transformations

- Definition:  $T(\lambda a) = \lambda T(a)$  and  $T(a + b) = T(a) + T(b)$
- Quick Checks:  $T(0) = 0$  and  $T(ax + by) = aT(x) + bT(y)$ . Basically check Homomorphism.

Any linear transformation can be represented by a matrix:

$$A = \begin{pmatrix} | & \dots & | \\ T(e_1) & \dots & T(e_n) \\ | & \dots & | \end{pmatrix}.$$

#### 2.2 Spaces

For square we have: 1) Identity, 2) Diagonal 3) Upper/Lower 4) Symmetric ( $A^H = A$ )

- **Rank:**  $\text{rank}(A)$  = number of independent vectors. (Fullrank iff intertible for square matrices)
  - $\text{rank}(A) = \text{rank}(A^T) = \text{rank}(A^T A) = \text{rank}(A A^T)$
- **Column Space:**  $C(A) = \{Ax \mid x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$ , aka **Image**.  $\dim = r$
- **Row Space:**  $R(A) = C(A^T) = \{A^T x \mid x \in \mathbb{R}^m\} \subseteq \mathbb{R}^m$ .  $\dim = r$
- **Null Space:**  $N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$ . aka **Kernel**.  $\dim = n - r$ .
- **Left Null Space:**  $LN(A) = N(A^T) = \{x \in \mathbb{R}^m \mid x^T A = 0^T \text{ or } A^T x = 0\}$ .  $\dim = m - r$

A **basis** is defined as an independent set which spans your space. The dimension of a space is the cardinality of your basis for that space (which stays same independent of which basis represents that space).

#### 2.3 Don't Forget

- $AB \neq BA$
- $(AB)^{-1} = B^{-1} A^{-1}$
- $(AB)^T = B^T A^T$

### 3 Systems of Linear Equations

Basically  $Ax = b$ .

#### 3.1 LU Decomposition

Run REF on  $A \in \mathbb{R}^{m \times n}$  to generate  $U \in \mathbb{R}^{m \times n}$  and track coefficients in  $L \in \mathbb{R}^{n \times n}$ , where  $L$  contains 1's on the diagonal and the opposite values of the operations performed on the corresponding rows.

#### 3.2 Permutation Matrices

- Each row and column have exactly one 1.
- They are orthogonal, hence  $P^{-1} = P^T \wedge P P^T = P^T P = I$
- $\det(P) = \pm 1$

- $P = P_1 P_2$  is also a permutation matrix
- A permutation creates a bijection from  $[n] \rightarrow [n]$ .

### 3.3 LUP Decomposition

$$PA = LU. \text{ If } U = E_{m-1} P_{m-1} E_{m-2} P_{m-2} \dots E_1 P_1 A \Rightarrow P = P_{m-1} \cdot \dots \cdot P_1$$

### 4 Vector Spaces

A vector space is an algebra  $(V, +, \cdot)$ , where  $+$  :  $V \times V \rightarrow V$ ,  $\cdot$  :  $\mathbb{R} \cdot V \rightarrow V$  s.t. we have 1) commutativity 2) associativity 3) a zero vector 4) a negative vector 5) identity element  $\in \mathbb{R}$  6) compatibility of  $\cdot \in \mathbb{R} \wedge \cdot \in \mathbb{V}$  7) distributivity over  $+$   $\in \mathbb{V}$  and 8) distributivity over  $+$   $\in \mathbb{R}$

#### 4.1 Subspace

$U \subseteq V$  is a subset if we have 1) closure under  $+$  :  $U \times U \rightarrow U$  and 2) closure under  $\cdot$  :  $\mathbb{R} \times U \rightarrow U$ .

##### 4.1.1 Columns Space

See definition above. Construct by running RREF on  $A$  and select the columns of  $A$  based on the pivot columns of RREF. **Note:** R/REF changes the column space, make sure to pick from  $A$ .

##### 4.1.2 Row Space

See definition above. Construct by running RREF of  $A$  and selecting all non-zero rows of that RREF. **Note:** R/REF doesn't change row space, make sure to pick from R/REF.

**Lemma 4.27:** Given an invertible matrix  $M$  then  $R(A) = R(MA)$  (left multiplication only).

##### 4.1.3 Nullspace

See definition above.  $N(A) \subseteq \mathbb{R}^n$ . Construct by running RREF on  $A$ . For each non-pivot column set it's coefficient = 1 and find out what the coefficients of the pivot columns must be to get 0. This should yield  $n - r$  columns forming a basis of  $N(A)$ .

**Lemma 4.33:** Given an invertible matrix  $M$  then  $N(A) = N(MA)$ .





##### 4.1.4 Left Nullspace

See definition above.  $LN(A) := N(A^T) \subseteq \mathbb{R}^m$

### 4.2 Solution Space

For any  $Ax = b$  we have three options: 1) No solution 2) One solution 3) Infinite solutions.

- If  $A$  is not invertible and  $b \notin C(A)$  then no solution can exist.
- If  $A$  is invertible  $\Rightarrow N(A) = \{0\}$  then exactly one solution exist  $x = A^{-1}b$
- If  $A$  is not invertible but  $b \in C(A)$  then  $\exists x : Ax = b$  and  $\forall n \in N(A) : A(x + n) = b + 0 = b$ . This can happen when our transformation  $f$  is going from a higher dimensional space to a lower dimensional space, i.e  $n > m$ .

$R_0$	$r = n$ (full rank) invertible	$r < n$ (dependent columns) underdetermined	← free variables
	 1 solution	 $\infty$ many solutions	
$r = m$ (full rank)			
$r < m$ (zero rows)	overdetermined		← free variables
	 0 or 1 solution depending on c (if some $\star \neq 0$ , then 0)	 0 or $\infty$ many solutions	

**Inverse Theorem 3.11:** Let  $A \in \mathbb{R}^{m \times m}$ , then the following are equivalent:

1.  $\exists A^{-1}$
2.  $\forall b \in \mathbb{R}^m \exists x : Ax = b$ , and  $x$  is unique
3. The columns of  $A$  are independent

## 5 Orthogonality

**Definition:**  $u$  is orthogonal to  $v$  if  $u \cdot v = 0$ . Two subspaces  $U, V$  are orthogonal if  $\forall u \in U \forall v \in V : u \cdot v = 0$ . A basis can be used to check orthogonality.

**Theorem 5.1.7:** Let  $V, W$  be subspaces of  $\mathbb{R}^n$ , then the following are equivalent:

1.  $V = W^\perp$
2.  $\dim(V) + \dim(W) = n$
3.  $\forall u \in \mathbb{R}^n \exists$  unique  $v, w : u = v + w$

## 5.1 Four fundamental Subspaces

- $N(A) = R(A)^\perp$ 
  - Think how if  $Ax = 0$  then each row of  $A$  "dotted" by  $x = 0$ , which means these  $x$ 's are orthogonal to each row and hence the rowspace of  $A$ .
- $LN(A) = C(A)^\perp$ 
  - Argue with the same as above but just use  $A^T$  instead.

## 5.2 Properties

- $Q$  is orthogonal (more like orthonormal) iff  $Q^T Q = I$
- For square matrices  $Q Q^T = I$  and  $Q^T = Q^{-1}$
- For non-square matrices  $Q Q^T = I$  may *not* hold.
- Orthonormal matrices preserve **norm** (i.e.  $\det(Q) = \pm 1$  and  $\|Qx\| = \|x\|$ )
- Orthonormal matrices preserve **angle**.
- $A^{-1}$  is orthonormal.  $AB$  is orthonormal (since  $(AB)(AB)^T = ABB^T A^T = I$ )

## 5.3 Gram-Schmidt

We are given  $A$  a basis for some space and want to orthonormalize into  $Q$ . **Steps:**

1. Normalize  $v_1 \rightarrow q_1$
2. Subtract projection from previous vectors from current vector:
  1.  $v'_n = v_n - \sum_{i=1}^{n-1} \text{proj}_{q_i}(v_n) = v_n - \sum_{i=1}^{n-1} ((v_n \cdot q_i) q_i)$
  2.  $q_n = \frac{v'_n}{\|v'_n\|}$

## 5.4 QR Decomposition

$A = QR \Rightarrow Q^T A = R$ . Basically run Gram-Schmidt on  $A$  to generate  $Q$  and calculate  $R$ .

- $R$  is upper triangular and invertible
- $C(Q) = C(A)$

## 6 Projections

The projection of  $b \in \mathbb{R}^m$  onto a subspace  $S \in \mathbb{R}^m$  is the point in  $S$  that's closest to  $b$ . i.e.  $\text{proj}_S(b) = \arg\min_{p \in S} \|b - p\|^2$  (yes error squared.)

- **1D Case:** Let  $a \in \mathbb{R}^m$  span  $S$ . Then  $\text{proj}_S(b) = \frac{aa^T}{a^T a} b$
- **ND Case:** Let  $S = C(A)$  and  $b \in \mathbb{R}^m$ . Then  $\text{proj}_S(b) = A\hat{x}$  s.t.  $A^T A\hat{x} = A^T b$ .
  - If  $b \in S$  iff  $Ax = b$  then  $\hat{x}$  preserves the  $x$ .
  - Otherwise  $\hat{x}$  minimizes the least square error.

**Theorem 5.2.6:** Let  $S = C(A)$ , then  $\text{proj}_S(b) = Pb$  s.t.  $P = A(A^T A)^{-1} A^T$ .

Other:

- $P^2 = P$  (projecting multiple times doesn't change the projection).
- If  $\text{proj}_S(b) = Pb$  then  $\text{proj}_{S^\perp}(b) = (I - P)b$
- $(I - P)^2 = I - P$  (since projecting onto the orthogonal complement multiple times doesn't change anything)

## 6.1 Least Squares

Assume  $Ax = b$  does not always have a solution, however we want to get the "best" solution according  $\min_{x' \in \mathbb{R}^n} \|Ax' - b\|^2$ . We can solve this using projections as follows:

- First write down the equation in form of e.g.  $b_i = \lambda_3 x^3 + \lambda_2 x^2 + \lambda_1 x + \lambda_0$
- Now write using matrices:  $\begin{pmatrix} | & \dots & | \\ x_i^3 & \dots & 1 \\ | & \dots & | \end{pmatrix} \begin{pmatrix} \lambda_3 \\ \vdots \\ \lambda_0 \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$
- Normal Equations:  $(A^T A)x' = (A^T b) \Rightarrow Nx' = y \Rightarrow x' = N^{-1}y$

## 7 Pseudoinverse

- **Left Pseudoinverse:**  $A^\dagger A = I$
- **Right Pseudoinverse:**  $AA^\dagger = I$

## 7.1 Left Pseudoinverse for Full Column Rank

Use a left pseudoinverse for  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  s.t.  $n < m$ , meaning we are transforming from a smaller space to a larger space. This means that we are not losing information from the input space but we cannot represent the whole output space, meaning  $b$  will probably not lie in  $C(A)$  ( $A$  is a basis and has full column rank), hence we basically do least squares since the system is **overdetermined**.

Hence  $A_{\text{left}}^\dagger = (A^T A)^{-1} A^T \Rightarrow A^\dagger A = I$

## 7.2 Right Pseudoinverse for Full Row Rank

Use right pseudoinverse for  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  s.t.  $n > m$ , meaning we are transforming from a larger space to a smaller space and hence losing information. This makes the system underdetermined (many possible solutions). This means that there exist a non-trivial nullspace. Here the right-pseudoinverse minimizes the norm of our solution.

Hence  $A_{\text{right}}^\dagger = A^T (AA^T)^{-1} \Rightarrow AA^\dagger = I$

## 7.3 Left Pseudoinverse for General Matrices

For general matrices  $A$  the left pseudoinverse cannot be defined as  $A^\dagger = (A^T A)^{-1} A^T$  because  $(A^T A)^{-1}$  might not be defined. Hence we need to use a different approach.

Basically we do a CR decomposition since  $C$  has full-column rank and  $R$  has full row rank.  $A = CR \Rightarrow A^\dagger = (CR)^\dagger = R^\dagger C^\dagger = R^T (RR^T)^{-1} (C^T C)^{-1} C^T$

This satisfies that for  $Ax = b \Rightarrow \hat{x} = A^\dagger b$  and  $\hat{x}$  is the unique solution satisfying  $\min_{x \in \mathbb{R}^n} \|x\|$  s.t.  $A^T Ax = A^T b$ .

$A^\dagger$  can be defined (using SVD) as  $V \Sigma^\dagger U^T$  where  $\Sigma^\dagger$  is taking the reciprocal of non-zero singular values and then transposing the matrix.

## 8 Farkas Lemma

Farkas Lemma provides a way to determine if a system of linear inequalities is feasible. It essentially states that exactly one of two alternatives is true.

**Geometric Intuition:** Imagine a cone formed by the vectors representing the inequalities. Farkas Lemma helps determine if a given vector  $b$  is inside this cone (feasible system) or if there's a hyperplane separating  $b$  from the cone (infeasible system).

Given  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$  one and exactly one of these statements is true:

1. **Feasibility:**  $\exists x \in \mathbb{R}^n$  s.t.  $Ax \leq b \wedge x \geq 0$  (there exists a non-negative solution)
2. **Infeasibility Certificate:**  $\exists y \in \mathbb{R}^m$  s.t.  $A^T y \geq 0 \wedge y \geq 0 \wedge b \cdot y < 0$  (There's a non-negative linear combination of the inequalities that leads to a contradiction)

## 8.1 Fourier-Motzkin Elimination

Basically we want to go from  $m$  inequalities with  $n$  variables to possibly  $\frac{m^2}{4}$  inequalities with  $n - 1$  variables. Geometrically this is analogous to projecting the shadow of our "cone" from  $n$ -D to  $n - 1$ -D.

1. We separate the variable we want to eliminate onto say the LHS.

- We make sure the inequality direction is consistent for all equations.
- We normalize the equations so that the coefficients (of the variable we want to eliminate) are  $0 \vee \pm 1$
- We get a new set of equations by combining the  $+x_i$  equations with  $-x_i$  equations.
- Repeat until we get to a low dimension case
  - If we have an inconsistency, quit.
  - Otherwise backsubstitute values to get a possible  $x$  which satisfies the equation.

## 9 Determinants

For 2x2:  $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ . For NxN: (Cofactors:) Make  $+-+...$  grid. Pick a row/column and calculate  $\pm A_{i,j} \det(\dots)$  recursively.

**Quadratic Formula:** Either complete the square or

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

## 9.1 Properties

**Fundamental:**

- $\det(I) = 1$
- If we swap the rows of  $A \rightarrow B$  once, then  $\det(B) = -\det(A)$ .
- The determinant is a linear function of each row separately.
  - If a row of  $A$  is multiplied by a scalar  $t$ , then  $\det(A') = t \det(A)$ .
  - If a row of  $A$  is replaced by the sum of itself and a multiple of another row, the determinant is unchanged.

**Derived:**

- If any two rows are equal then  $\det(A) = 0$
- If  $A$  has a row of zeros then  $\det(A) = 0$
- Subtracting a multiple of one row from another row leaves the determinant unchanged.
- If  $A$  is triangular (upper or lower), the determinant is the product of the diagonal entries.
- $\det(A) = 0$  if and only if  $A$  is singular (not invertible)
- $\det(AB) = \det(A) \det(B)$
- $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det(A) = \det(A^T)$

## 10 Complex Numbers

Let  $z = (a + bi) \in \mathbb{C}$ .

**Conjugate:**

- $\bar{z} = a - bi$
- $z\bar{z} = \|z\| = a^2 + b^2$
- $\overline{x + y} = \bar{x} + \bar{y}$
- $\overline{xy} = \bar{x}\bar{y}$

**Norm:**

- $\|z\| = \sqrt{a^2 + b^2} \in \mathbb{R}$
- $\|xy\| = \|y\|\|x\|$
- $\|z^n\| = \|z\|^n$

**Hermitian of a matrix:**

Basically transpose and conjugate each entry.

**Properties:**

- $z + \bar{z} = 2\Re(z) = 2a$
- $z - \bar{z} = 2i\Im(z) = 2ib$
- $\|z\| = \|\bar{z}\|$
- $z^{-1} = \frac{\bar{z}}{\|z\|^2}$  (multiplicative inverse)
- Triangle Inequality:**  $\|x + y\| \leq \|x\| + \|y\|$

**Eulers Formula:**

- $e^{i\theta} = \cos \theta + i \sin \theta$
- $\theta = \arctan\left(\frac{\Im(z)}{\Re(z)}\right) = \arctan\left(\frac{b}{a}\right)$

## 11 Change of Basis

To transform a linear transformation  $M_A$  in basis  $A$  to basis  $B$ :

$$M_B = P_{A \rightarrow B}^{-1} M_A P_{B \rightarrow A}$$

Here,  $P$  is calculated as:

- Express each  $b_i$  (basis  $B$ ) in terms of basis  $A$ :  $[b_i]_A = x_i$ , where  $Ax_i = b_i$ .
- Construct  $P = ([b_1]_A \dots [b_n]_A)$ .

**Intuition:**

- $e_1$  in basis  $B$  equals  $b_1$ , written as  $[b_1]_A = x_1$  such that  $Ax_1 = b_1$ .
- Transform in basis  $A$ , then "undo" the change of basis.

**Example:** Given  $A = (e_1 \ e_2 \ e_3)$  and  $B = (b_1 \ b_2 \ b_3)$ :

- Compute  $[b_1]_A, [b_2]_A, [b_3]_A$  to find  $P$ .
- Use  $M_B = P^{-1} M_A P$ .

## 12 Eigenvalues and Eigenvectors

Basically we want to find the Eigenvalues  $\lambda$  s.t.  $Ax = \lambda x \Rightarrow (A - \lambda I)x = 0 \Rightarrow \det(A - \lambda I) = 0$ , where the  $x$  which satisfy this for their given  $\lambda$  are called Eigenvectors.

Since  $Av_i = \lambda_i v_i = v_i \lambda_i \Rightarrow AV = V\Lambda \Rightarrow A = V\Lambda V^{-1} \Rightarrow A^k = V\Lambda^k V^{-1}$ .

### 12.1 Terms

- The set of Eigenvectors is called the **spectrum**.
- The **characteristic polynomial** is  $\det(A - \lambda I) = 0$
- The set of vectors corresponding to a  $\lambda$  s.t.  $Av = \lambda v$  are called an **Eigenspace**.
- Multiplicities:**
  - The number of times an eigenvalue appears as a root of the characteristic polynomial is called **algebraic multiplicity**.
  - The **geometric multiplicity** of  $\lambda$  is the dimension of the Eigenspace of  $\lambda$ . Calculate as  $\dim(N(A - \lambda I))$
  - Key rule: Geometric multiplicity  $\leq$  Algebraic multiplicity

### 12.2 Observations

- If  $\lambda$  is real, then it has a corresponding real Eigenvectors

- If for a real matrix  $(\lambda, v)$  is a complex Eval/EVec pair, then  $(\bar{\lambda}, \bar{v})$  is too.
- For orthonormal matrices  $\lambda \in \mathbb{C} \wedge |\lambda| = 1$ .
- $A^k v = \lambda^k v$
- $\det(A - \lambda I)$  is a polynomial in  $\lambda$  with degree  $n$ .
  - The coefficient of  $\lambda^n$  is  $(-1)^n$ .
- For  $k$  distinct Eigenvalues, there exist  $k$  independent Eigenvectors.
- The characteristic polynomial can be factored as  $0 = \det(A - xI) = (-1)^n (x - \lambda_1) \cdot \dots \cdot (x - \lambda_n)$ .
- $\det(A) = \prod \lambda_i$  because  $\det(A) = \det(A - 0I) = (-1)^n \cdot (\lambda_1) \cdot \dots \cdot (-\lambda_n)$
- $\text{Tr}(A) = \sum \lambda_i$ . (Also  $\text{Tr}(AB) = \text{Tr}(BA) \wedge \text{Tr}(A(BC)) = \text{Tr}((BC)A)$ )
- A projection matrix  $P$  projecting onto  $U \in \mathbb{R}^n$  has two Eigenvalues of 0, 1.

**Gotchas:**

- Even though the Eigenvalues of  $A, A^T$  are same, their Eigenvectors differ.
- The Eigenvalues of  $A + B$  cannot be trivially determined.
- The Eigenvalues of  $AB$  or  $BA$  are not trivially determined. (Unless  $A, B$  have equal dimensional square matrices, then they share the non-zero Eigenvalues, but might have different multiplicities.)
- Gauss Elimination doesn't preserve Eigenvalues and Eigenvectors.

## 12.3 Dynamic Systems

Write down equation in the form of  $\vec{g}_n = M\vec{g}_{n-1}$  with  $g_0$  being the base case. Let  $g \in \mathbb{R}^m$ . Since  $g_n = M^n g_0$  we have that  $M \in \mathbb{R}^{m \times m}$ , hence quadratic. Let  $v_1, \dots, v_m$  be the Eigenvectors of  $M$ .

- Check dimensions:** If  $\text{span}\{v_1, \dots, v_m\} \neq \mathbb{R}^m$  quit.
- Eigenbasis:** Let  $V = (v_1 \dots v_m)$  form the new basis of  $\mathbb{R}^m$ .
- Exponentiation:** We have  $g_n = M^n g_0 = V\Lambda^n V^{-1} g_0$ . Extract your solution from  $g_n$ .

## 13 Similar Matrices and Spectral Theorem

$A, B$  are called similar matrices if  $\exists S$  s.t.  $B = S^{-1}AS$ . Similar matrices are equal dimensional square matrices. Similar matrices share Eigenvalues.

- Spectral Theorem:** Any symmetric matrix has  $n$  Eigenvalues and an orthonormal basis made out of Eigenvectors of  $A$ .
- Symmetric matrices can be diagonalized as  $S = V\Lambda V^{-1} = V\Lambda V^T$ .
- The rank of a symmetric matrix is the number of non-zero Eigenvalues.
- $S = \sum_{i=1}^n \lambda_i v_i v_i^T$ .
- Symmetric matrices only have real Eigenvalues.

### 13.1 Rayleigh Quotient

$$Av = \lambda v \Rightarrow v^T Av = \lambda v^T v \Rightarrow \lambda = R(v) = \frac{v^T Av}{v^T v}.$$

$$\lambda_{\min} \leq R(v) \leq \lambda_{\max}$$

### 14 Definiteness

- **Positive Semidefinite (PSD):**  $\lambda_i \geq 0$
- **Positive Definite (PD):**  $\lambda_i > 0$

**Intuition:** Look at the quadratic form  $q(x) = x^T Ax$ . If it always makes a positive ellipsoid it's PD and it's positive Eigenvalues show that growth. If it touches 0 (except for origin) it's PSD.

- If  $A, B$  are PSD/PD then  $A + B$  is also PSD/PD, because  $x^T Ax + x^T Bx \geq 0 \Rightarrow x^T (A + B)x$

### 15 Gram Matrices

$G = V^T V$ ,  $G$  is called a Gram matrix.

#### Properties:

- $A^T A \in \mathbb{R}^{n \times n}$  and  $AA^T \in \mathbb{R}^{m \times m}$  have the same non-zero Eigenvalues.

### 16 SVD

Any matrix  $A$  can be factored as  $A = U \Sigma V^T$ .

- $U$  has the **left-singular vectors** and is orthonormal.
- $V$  has the **right-singular vectors** and is orthonormal.
- $\Sigma$  has the **singular values** and contains non-negative values only.

#### Construction:

- $A^T A = U \Lambda_1 U^T$ . Here we have that  $\Lambda_1 = \Sigma^T \Sigma$ .  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$  s.t.  $k = \min(n, m)$
- $AA^T = V \Lambda_2 V^T$ . Here we have that  $\Lambda_2 = \Sigma \Sigma^T$ .  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$  s.t.  $k = \min(n, m)$
- $\sigma_i = \sqrt{\lambda_i}$ .
- For both:  $\Sigma$  is constructed s.t.  $\sigma_1 \geq \dots \geq \sigma_k \geq 0$ . Rank: number of non-zero singular values.