First Year Refresher

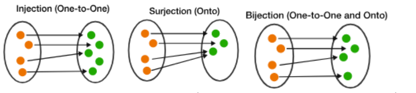
Properties of Transpose $(A^T)^T=A$; $(A+B)^T=A^T+B^T$; $(AC)^T=C^TA^T$
Symmetric Matrix $A \in IR^{n \times n}$ and $A^T = A$
Properties of Inverse $(A^{-1})^T = (A^T)^{-1}$; $(AB)^{-1} = B^{-1}A^{-1}$
Finding Inverse using Gaussian Elimination – $[A|I_n] \rightarrow \ldots \rightarrow [I_n|A]$
Determinant n=1: det(M) = a; n=2: det(M) = ad-bc;Diagonal/triangular matrix = product of diagonal elements.
Properties of Determinants $\det(A^T)=\det(A)$;$\det(AB)=\det(BA)= \det(A)\det(B)$; $\det(\lambda A) = \lambda^n\det(A)$; matrix is invertible ⇔ $\det(A) \neq 0$; $\det(A^{-1}) = 1/(\det(A))$.
Trace tr(A) sum of diagonal elements of A.
Proving linear independence/dependence Use Gaussian Elimination, all-zero column indicates linear dependence and reducing all rows to 0s 1s without any same indicates linear independence. Also, determinant = 0 indicates linear dependence.
Rank rk(A) is number of independent columns/rows of A, $rk(A) = rk(A^T)$. Invertible matrix $A \in IR^{n \times n}$ ⇔ rk(A) = n ⇔ columns are linearly independent ⇔ rows are linearly independent.
Non-singular linearly independent columns, rank is n, det != 0, invertible.
Vector space For u, v ∈ U, u+v ∈ U; u∈ U and c ∈ IR, cu ∈ U (closed under addition/scalar multiplication).
Generating set of subspace is if every vector of the subspace can be expressed as a linear combination of the vectors in the generating set.
Basis Minimal generating set (smallest set that spans subspace).
Finding basis from generating set Form matrix using vectors of generating set as columns, perform Gaussian Elimination and the columns associated with pivot columns form basis.
Finding simple basis from generating set Form matrix using vectors of generating set as rows, perform Gaussian Elimination, non-zero rows form simple basis.
Dimension number of vectors in the basis of a vector space.
Change of basis $I_{B'B}$ is such that $u_{B'} = I_{B'B}u_B$ (change-of-basis from B to B').
Property of change-of-basis $I_{B'B} = (I_{BB'})^{-1}$
Finding the change-of-basis matrix To find $I_{B'B}$, construct augmented matrix [B'|one vector from B]→ … → [I_n|one vector of X], put X vectors together to find change-of-basis matrix.
Finding coordinates of vector in new basis $u_{B'} = I_{B'B}u_B$.
Euclidean norm ||u|| Square root of sum of squares of elements of vector.
Parallelogram Law ∀u,v∈Rn,‖u+v‖^2 +‖u−v‖^2 =2‖u‖^2 +2‖v‖^2.
Scalar Product (Dot product) u.v = sum of all ui * vi = $u^Tv = v^Tu$
Euclidean norm is root of scalar product of u and u.
Property of scalar product u,v∈IR^n and θ∈IR, u.v=||u||v||* cos θ.
Orthogonal vectors u.v = 0, u⊥v; ***Orthogonal matrix*** $A^{-1}=A^T$.
Orthonormal vectors u.v = 0, ||u|| = 1, ||v|| = 1.
Columns of orthogonal matrix form orthonormal basis of IR^n.
Orthogonal subspaces all vectors in subspace are orthogonal, U⊥V.
Cauchy-Schwartz Inequality |u.v| < ||u||*||v||.
Linear Map A map f:V→W is linear if it preserves linear structure, i.e. it satisfies: For all u, v ∈ V, f(u+v) = f(u)+f(v) and u ∈ V and c ∈ IR, f(cu)=cf(u).
Linear map associated with matrix Multiply matrix with variables vector and write result. E.g., $\phi_A : IR^3 \rightarrow IR^2$, $[x\ y\ z]^T \mapsto [x+2y+3z\ 4x+5y+6z]^T$.
Transformation matrix of linear map in IR^n (LEFT)

$$\begin{bmatrix} y \\ x \\ y+x \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix}.$$

$$F_{D'B'} = I_{D'D}F_{DB}I_{BB'} \qquad \begin{array}{ccc} \text{coord in } B & \xrightarrow{F_{DB}} & \text{coord in } D \\ \uparrow I_{BB'} & & \uparrow I_{D'D} \\ \text{coord in } B' & \xrightarrow{F_{D'B'}} & \text{coord in } D' \end{array}$$

Matrix representation and basis change (RIGHT)
Equivalent matrices Two matrices A, A' $\in IR^{m \times n}$ are equivalent if there exist two invertible matrices $P \in IR^{m \times m}$ and $Q \in IR^{n \times n}$ such that $A=PA'Q^{-1}$.
Composition of linear maps g o f = g(f(x)), f, g, g o f are all linear maps.
Matrix representation of composition $(GF)_{DB} = G_{DC}F_{CB}$
Image space (Range space) im(f) = range(f) = {v ∈ IR^m|∃u ∈ IR^n: f(u) = v}
Kernel space (Null space) ker(f) = null(f) = {v ∈ IR^n : f(v) = 0}
Im(f) is a vector subspace of IR^m and ker(f) is a vector subspace of IR^n.
Finding basis of image Construct ***transformation matrix of linear map*** and apply ***Finding basis from generating set***.
Finding basis of kernel Apply Gaussian Elimination to ***transformation matrix of linear map (A)*** and use it to solve Ax=0.
Rank-Nullity Theorem For f:IR^n→IR^m linear map with matrix representation $A \in IR^{m \times n}$: dim(im(f)) + dim(ker(f)) = rk(A) + dim(ker(A)) = n
Kernel of injective linear map f: IR^n→IR^m injective ⇔ ker(f) = {**0**}.
Endomorphism is a linear map whose domain and codomain are same, e.g., f(x) = 2x that takes real number and returns real number.
Automorphism bijective endomorphism, e.g., f(n) = -n which takes integer and returns its negation preserves group operation of addition and inverse and is bijective.
[THEOREM] Kernel and Image of automorphism f: IR^n → IR^m automorphism ⇔ f is injective (ker(f) = {0}) ⇔ f is surjective (im(f) = IR^n).
Transition matrix of automorphism linear map is an automorphism iff the matrix representation $A \in IR^{n \times n}$ of f is invertible, and A^{-1} is the matrix representation of f^{-1}.

Injection (One-to-One) Surjection (Onto) Bijection (One-to-One and Onto)



Eigenvalues and Eigenvectors **Av = λv, v** is **eigenvector** with **eigenvalue λ**.
Spectrum of $A \in IR^{n \times n}$ is set of all eigenvalues of A, e.g., Sp(A) = {-1, 3}.
Characteristic Polynomial det(A-λI).
Eigenvalues are roots of characteristic polynomial.
Number of eigenvalues is at most n for $A \in IR^{n \times n}$ due to max no. of roots.
Eigenvalues, Trace, Determinant tr(A) = ∑i=1 n λ_i and det(A) = ∏i=1 n λ_i.
Algebraic multiplicity of λ (an eigenvalue) is how many times it is repeated as root of characteristic polynomial.
Geometric Multiplicity of λ (an eigenvalue) is dimension of eigenspace.
Eigenspace Set of eigenvectors associated with 1 λ denoted ***E_λ = ker(A-λI)***.
Finding eigenspace associated with eigenvalue (A-λI)v = 0, then write variables in terms of each other and construct eigenspace.
[THEOREM] Geometric multiplicity of eigenvalue is superior or equal to 1, and less or equal to algebraic multiplicity: 1<= dim E_λ<=algebraic mult of λ.
Diagonalisable endomorphism: Endomorphism of IR^n is diagonalisable if it has matrix representation in some bases if IR^n that is diagonal.
Diagonalisable matrix: Matrix is diagonalisable iff there exists diagonal matrix D and an invertible matrix P such that $A=PDP^{-1}$ (P is change-of-basis) from standard basis to basis where endomorphism associated to A is diagonal.
[THEOREM] Basis of eigenvectors $A \in IR^{n \times n}$ diagonalisable ⇔ there exist basis of IR^n of eigenvectors of A.
Proving matrix is diagonalisable According to ***[THEOREM] Basis of eigenvectors*** we can prove that A is diagonalisable by proving that eigenvectors form basis of IR^n (they span IR^n and are linearly independent).
[THEOREM] Algebraic and Geometric Multiplicity $A \in IR^{n \times n}$ with k eigenvalues, such that each eigenvalue has algebraic multiplicity r_i and geometric multiplicity g_i, A is diagonalisable iff for all I, r_i=g_i. ***Proving matrix is diagonalisable using algebraic geometric multiplicity*** use theorem above.
Diagonal form of diagonalisable matrix

Proposition 5.3 (Diagonal form of diagonalisable matrix)

If $A \in \mathbb{R}^{n\times n}$ is diagonalisable, the basis in which it is diagonal is a basis of its eigenvectors and the coefficient on the diagonal would be its eigenvalues. In other words, if $\lambda_1, \ldots \lambda_n \in \mathbb{C}$ are the eigenvalues of A, repeated as many time as they algebraic multiplicity, associated respectively with $v_1, \ldots v_n \in \mathbb{R}^n$, then:

$$A = [v_1 \ldots v_n]\begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}[v_1 \ldots v_n]^{-1}$$

Similar Matrices Two matrices (A, A') are similar if there exists an invertible matrix such that $A = PA'P^{-1}$.

Vector and Matrix Norms

Vector Norms: Non Zero Vector, norm of x > 0, scalar multiplication, triangular inequality
Lp norms (||x||2 is also called Euclidean norm = root(x.x))

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \quad \|x\|_1 = \sum_{i=1}^n |x_i| \quad \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \quad \|x\|_\infty = \max_{1\le i\le n} |x_i|$$

Matrix Norms A real-valued map ||.||: $IR^{m \times n}$ → IR satisfying: For any nonzero matrix A: ||A|| > 0, For any scalar λ: ||λA|| = |λ|.||A||, and ||A+B|| <= ||A|| + ||B|| (Triangular inequality) [left->right: abs max col, max sing value, max row]
Sub-multiplicative matrix norm satisfies additional axiom: ||AB|| <= ||A|||B||.

$$\|A\|_1 = \max_j \|a_j\|_1 \quad \|A\|_2 = \sigma_1(A) \quad \|A\|_\infty = \max_i \|a^i\|_1$$

Consistent and compatible norms A matrix norm ||.|| on $IR^{m \times n}$ is **consistent** with vector norms $||.||_a$ on IR^n and $||.||_b$ on IR^m if for all $A \in IR^{m \times n}$ and $x \in IR^n$: $||Ax||_b <= ||A|||x||_a$. If a=b, then ||.|| is **compatible** with $||.||_a$.

Subordinate matrix norm $\forall A \in \mathbb{R}^{m\times n}, \|A\| = \max\{\|Ax\| : x \in \mathbb{R}^n, \|x\| = 1\}$

Proposition 1.3 (Subordinate matrix norm expressions)

$$\begin{aligned} \forall A \in \mathbb{R}^{m\times n}, \|A\| &= \max\{\|Ax\| : x \in \mathbb{R}^n, \|x\| = 1\} \\ &= \max\{\tfrac{\|Ax\|}{\|x\|} : x \in \mathbb{R}^n, x \neq 0\} \qquad (1) \\ &= \max\{\|Ax\| : x \in \mathbb{R}^n, \|x\| \le 1\} \qquad (2) \end{aligned}$$

Prove (1) set y = x / norm(x) since x is not 0. Then just sub it in to Ay. X / norm(x) = 1
Prove(2) y <=1, norm(Ayx) = y norm(Ax) <= norm(Ax). Since x = 1 is subset of <= 1, done.

[THEOREM] Compatibility of subordinate matrix norm A vector norm is compatible with its subordinate matrix norm ||.||.
Subordinate matrix norm of L1, L2, Linf For p=1, 2, inf, the matrix norm ||.||p is subordinate to vector norm ||.||p. p=p therefore they are also compatible.
Vector Product a × b = (a2b3 – a3b2)i + (a3b1 – a1b3)j + (a1b2 – a2b1)k.

Finally (3) can be proven using the Cauchy-Schwartz inequality: $\forall y \in \mathbb{R}^n, |x \cdot y| \le \|x\|_2\|y\|_2$. We want to choose $y \in \mathbb{R}^n$, such that $\|y\|_2 = 1$ and $|x \cdot y| = \frac{1}{\sqrt{n}}\|x\|_1$ so we can apply the inequality to find the result. We have $x \cdot y = \sum_{i=1}^n x_i y_i$. We choose $\forall 1 \le i \le n, y_i = \frac{\text{sign}(x_i)}{\sqrt{n}}$, so:

$$\|y\|_2 = \sqrt{\sum_{i=1}^n y_i^2} = \sqrt{\sum_{i=1}^n \frac{1}{n}} = \sqrt{1} = 1 \text{ and } |x \cdot y| = \sum_{i=1}^n \frac{|x_i|}{\sqrt{n}} = \frac{1}{\sqrt{n}}\|x\|_1 \Longrightarrow \boxed{\|x\|_1 \le \sqrt{n}\|x\|_2}$$

Matrix Norms

$$\|A\|_\infty = \max_{i=1\ldots n} \sum_{j=1}^n |a_{ij}| \quad \text{and} \quad \|x\|_\infty = \max_{i=1\ldots n} |x_i| \quad \text{and} \quad \|Ax\|_\infty = \max_{i=1\ldots n} \left|\sum_{j=1}^n a_{ij}x_j\right|$$

We know that $\forall \lambda_1, \lambda_2, \ldots \lambda_k, |\sum_{i=1}^k \lambda_i| \le \sum_{i=1}^k |\lambda_i|$ *and that* $|\lambda_1\lambda_2| = |\lambda_1||\lambda_2|$, *so* $\forall i = 1 \ldots m$:

$$\left|\sum_{j=1}^n a_{ij}x_j\right| \le \sum_{j=1}^n |a_{ij}||x_j| \le \sum_{j=1}^n |a_{ij}| \max_{i=1\ldots n}|x_i| \le \left(\max_{i=1\ldots n}|x_i|\right)\sum_{j=1}^n |a_{ij}| \le \left(\max_{i=1\ldots n}|x_i|\right)\left(\max_{i=1\ldots m}\sum_{j=1}^n |a_{ij}|\right)$$

$$\Longrightarrow \boxed{\|Ax\|_\infty \le \|x\|_\infty \|A\|_\infty}$$

Compatibility:

Proof. Using the result from Prop. 1.3: $\forall A \in \mathbb{R}^{m\times n}, \|A\| = \max\{\frac{\|Ax\|}{\|x\|} : x \in \mathbb{R}^n, x \neq 0\}$.
Thus, for all $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n, x \neq 0$, $\frac{\|Ax\|}{\|x\|} \le \|A\| \Leftrightarrow \|Ax\| \le \|A\|\|x\|$.
In addition, for all $A \in \mathbb{R}^{m\times n}$ and $x = 0$, $\|A\|\|0\| = 0 \ge \|0\|$ so the inequality still holds.

Extension of Linear Maps to C^n

Non-negativity of z $\bar{z}$ = a^2+b^2 >= 0; ==0 ⇔ z=0.
Condition for complex number being in IR conjugate must equal number, for vectors/matrices check every single entry.
Standard Inner Product <.,.> on C^n is <u, v> = $\bar{u}^Tv$ = ∑i=1 n $\bar{u}_iv_i$.
Properties z1=z2 ⇔ $\bar{z}1 = \bar{z}2$; conjugate of something in brackets is equal to conjugate of individual components (when multiplying), inner product is not symmetric, <u, u> is non-negative.

$$\bar{u}^T u = \sum_{i=1}^n \bar{u}_i u_i \ge 0$$

Standard Norm ||u|| = root(<u, u>).

Least Square Method

Orthogonal Projection on subspace If U is a subspace generated by ordered basis (**u1, …, un**) consider the matrix U = [**u1 … un**], the orthogonal projection π_U on U is π_U: IR^m → IR^m
v → $\pi_{U(v)} = U(U^TU)^{-1}U^Tv$.
im(A) ⊥ ker(A^T) – to prove simply set x = Ay, A^Tz = 0. Then x.z = x^Tz and sub in vals
Unique decomposition in terms of vectors Let $A \in IR^{m \times n}$, for all $b \in IR^m$, $b_i \in$ im(A), and unique $b_k \in \ker(A^T)$ such that $b = b_i + b_k$.
Least Square Method (LSM) If Ax=b has no solution, LSM finds x such that $||Ax-b||_2$, or equivalently $||Ax-b||_2^2$ is minimized.
$||Ax-b||_2$ is minimized ⇔ $||Ax-b_i||_2 = 0$ ⇔ $Ax=b_i$ ⇔ $A^TAx = A^Tb$
Normal equation $A^TAx = A^Tb$ gives solution to least square method. ***Linear Regression*** Find model of best fit such that the sum of errors squared is minimized. A is constructed by putting column of 1s for variables without any linked columns (e.g., h) and then data, z is vector of variables to optimize (in order of A), b is last column results vector.

$$f(t) = s_1 - s_2\frac{t^2}{2} \qquad f(t) = s_1f_1(t) + s_2f_2(t)$$

time	height measurement
t_1	y_1
t_2	y_2
$\vdots$	$\vdots$
t_m	y_m

We define a matrix A as follows: $A = \begin{bmatrix} f_1(t_1) & f_2(t_1) \\ \vdots & \vdots \\ f_1(t_m) & f_2(t_m) \end{bmatrix} = \begin{bmatrix} 1 & -\frac{t_1^2}{2} \\ \vdots & \vdots \\ 1 & -\frac{t_m^2}{2} \end{bmatrix}$.

And we write z our vector of parameters $z = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} h \\ g \end{bmatrix}$.

Then Solve Normal Eqn

Spectral Decomposition of Symmetric Matrices

Orthogonal matrices $Q^{-1} = Q^T$. ***Symmetric Matrix*** $A^T = A$.
Rotation of angle θ [cos θ -sinθ, sinθ cosθ].
Reflection through axis of angle θ/2 [cosθ sinθ, sinθ -cosθ].
[THEOREM] Orthogonal Transformations preserve lengths and angles norm(Qx) = norm(x). Square both sides, = (Qx)^T Qx and expand. To prove angle dot product Qx dot Qy and show equals x.y. SCALAR abs a x abs b x cos theta.
Determinant of orthogonal matrix det Q = 1 or det Q = -1.
Eigenvalues of orthogonal matrix all eigenvalues have modulus of 1: |λ| = 1.
[THEOREM] All eigenvalues of symmetric matrices are real. Assume complex and take complex conjugate of Alambda = lambda u.
[THEOREM] Geometric Multiplicity of eigenvalues in symmetric matrices for all eigenvalues of A we have, geometric multiplicity of λ = algebraic multiplicity of λ.
[THEOREM] Eigenvectors for distinct eigenvalues are orthogonal Two eigenvectors of A associated to two distinct eigenvalues are orthogonal.
[THEOREM] Spectral Theorem real symmetric matrix A can be diagonalized in the following way: $A = QDQ^T$ (= QDQ^{-1}), where Q is orthogonal, and D is a diagonal matrix construct using the (not necessarily distinct) eigenvalues of A.
Method to find Spectral Decomposition Find the characteristic polynomial of A and calculate roots to find eigenvalues. For each distinct eigenvalue find eigenspace. For each eigenspace find orthonormal basis. Combine all bases to form Q. Construct A with eigenvalues.

QR Decomposition

QR Decomposition using Gram-Schmidt (GS) Process Define $proj_u(v)$ = ((u.v)/(u.u))*u, then for u != 0 we have u.(v-$proj_u$(v)) = u.v – (u.v)(u.u)/(u.u) = 0.

$$\text{Let } A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

- $u_1 := a_1 = (1,1,0)^T$, $a_2 = (1,0,1)^T$, $a_3 = (0,1,1)^T$,
- $e_1 = (1,1,0)^T/\sqrt{2}$.
  $e_1 \cdot a_1 = \sqrt{2}$, $e_1 \cdot a_2 = 1/\sqrt{2}$, $e_1 \cdot a_3 = 1/\sqrt{2}$,
- $u_2 = a_2 - (e_1 \cdot a_2)e_1 = (1/2, -1/2, 1)^T$, $e_2 = (1,-1,2)^T/\sqrt{6}$. $e_2 \cdot a_2 = 3/\sqrt{6}$, $e_2 \cdot a_3 = 1/\sqrt{6}$,
- $u_3 = a_3 - (e_1 \cdot a_3)e_1 - (e_2 \cdot a_3)e_2 = (-1,1,1)^T/\sqrt{3}$ with $e_3 = (-1,1,1)^T/\sqrt{3}$. $e_3 \cdot a_3 = 1/\sqrt{6}$.

$$Q = [e_1, e_2, e_3] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix},$$

$$R = \begin{bmatrix} e_1 \cdot a_1 & e_1 \cdot a_2 & e_1 \cdot a_3 \\ 0 & e_2 \cdot a_2 & e_2 \cdot a_3 \\ 0 & 0 & e_3 \cdot a_3 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{3}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{2}{\sqrt{3}} \end{bmatrix}, \text{ with } A = QR$$

Householder Map Suppose hyper-plane P going through origin with unit normal **u** ∈ IR^m, i.e., P = {x ∈ IR^m: **u.x** = 0}. The householder matrix defined by $H_u = I - 2uu^T$ induces reflection wrt P.
Properties H_u is involutory: $H_u = H_u^{-1}$, H_u is orthogonal (as all reflections): $H_u^T = H_u^{-1}$, H_u preserves Euclidean length of vectors $||H_u(x)|| = ||x||$ and angles between vectors, the orthogonal projection Q on hyperplane P is given by $Q = I - uu^T$ with $Q^2 = Q$ and $Q = Q^T$, all reflections and rotations are orthogonal transformations.

Singular Value Decomposition

$$x \neq 0, \ x^TAx = \sum_{i=1}^n \lambda_i y_i^2 > 0.$$

Positive matrix is symmetric and for all x in IR^n – {0}, $x^TAx > 0$.
Positive semi-definite matrix is summetric and for all x in IR^n, $x^TAx >= 0$.
Negative matrix is symmetric and for all x in IR^n – {0}, $x^TAx < 0$.
Negative semi-definite matrix is summetric and for all x in IR^n, $x^TAx <= 0$.
Characterising Positive (semi-) definiteness with eigenvalues a symmetric matrix A is positive-definite ⇔ all its eigenvalues are strictly positive, if it is positive semi-definite ⇔ all its eigenvalues are non-negative.
$A^TA \in IR^{n \times n}$ and $AA^T \in IR^{m \times m}$ are symmetric and also positive semi-definite.
Singular Value Decomposition (SVD) of A is $A = USV^T$ where U and V are orthogonal, S is a diagonal matrix = diag(σ_1, σ_2, …, σ_p) where p = min(m, n) and the σs are singular values of A.
More cols than rows: VS^T = A^T U
Rank of A = number of positive singular values in S.
$||A||_2 = \sigma_1$ (Largest singular value of A).
Positive singular values of A are the positive square roots of eigenvalues of AA^T or A^TA.
SVD Method for more rows than columns Obtain eigenvalues σ_1^2 >= … >= σ_n^2 and the orthonormal eigenvectors v_1, …, v_n of A^TA. Construct orthogonal matrix V = [v1, …, vn]. Use formula to calculate U. US = AV
Principal Component Analysis (PCA) $A \in IR^{m \times n}$ represents m samples of n dimensional data. The principal axes of A are columns of V. The first principal axis of collection of m samples is defined as: $w_{(1)}$ = arg max (||w|| = 1) w^TA^TAw.
First principal axis Same assumption of representing m samples of n dimensional data. $A = USV^T$ and V = [v1, …, vn]. v1 is the first principal axis of A, corresponding to the column vector of V associated with the largest singular value of A.
Principal components (scores) of a collection of samples The columns of US are principal components or scores of A.

Generalized Eigenvectors

Definition Given a square matrix, a non-zero vector v is a generalised eigenvector of rank m associated with eigenvalue λ for A if $(A-\lambda I)^mv = 0$ and $(A-\lambda I)^{m-1}v$!= 0 (eigenvector of λ is generalised of rank 1).
Number of generalized eigenvectors associated with λ - λ with algebraic multiplicity k has k linearly independent generalized eigenvectors.
Number of generalized eigenvectors has n linearly independent generalised eigenvectors. In other words, there exist basis of C^n of generalised eigenvectors of A.
Jordan Normal Form (JNF) A matrix is in JNF if it is a diagonal matrix where each diagonal element is a Jordan block of size k_i with diagonal coefficient λ_i (λ are not necessarily distinct).
[THEOREM] Existence of Jordan Normal Form Consider the basis of generalised eigenvectors of A such that for each eigenvector v1 associated with eigenvalue λ, the associated generalised eigenvectors in the basis vk are chosen as $(A - \lambda I)v^k = v^{k-1}$. Then, in this basis, A is reduced to JNF. Coefficients on the diagonal of the Jordan blocks are corresponding eigenvalues of A. Different Jordan blocks can have the same eigenvalue on the diagonal.
Blocks in JNF and multiplicities Algebraic multiplicity of eigenvalue λ is the sum of sizes of blocks with λ on diagonal. Geometric multiplicity of λ is number of blocks with λ on diagonal.

$A \in \mathbb{R}^{n\times n}$, one can find its JNF form with the following steps:

1. Compute the eigenvalues of A: $\lambda_1 \ldots \lambda_m$, for example using the characteristic polynomial of A. We note a_i the algebraic multiplicity of eigenvalue λ_i.
2. For each eigenvalue λ_i, compute the associated eigenspace E_{λ_i}, we note g_i the geometric multiplicity of eigenvalue λ_i and $v_{i1}^1, v_{i2}^1, \ldots, v_{ig_i}^1$ the associated eigenvectors.
   Note: in the notation v_{ij}^1, the exponent 1 indicates that it is a generalised eigenvector of rank 1, i indicates that it is associated with eigenvalue λ_i, and the j is just an index to differentiate it from the others ($j = 1, 2, \ldots g_i$).
3. if $g_i < a_i$, find the missing $a_i - g_i$ generalised eigenvectors: for each eigenvector $v_{ij}^1 \in \mathbb{C}^n$ associated with λ_i, try to find all the $v_{ij}^k \in \mathbb{C}^n$ such that:
   $$(A - \lambda_i I)v_{ij}^k = v_{ij}^{k-1}$$
   This can be solved using Gaussian Elimination if needed.
   ⚠ You would need to do so for all eigenvectors associated with λ_i as each of them might generate part of the $a_i - g_i$ generalised eigenvectors you are looking for. See Example 6.1.
4. Write these vectors in a change-of-basis matrix as followed:
   $$B = \left[v_{11}^1, \ldots, v_{11}^{k_{11}}, \ldots v_{1g_1}^1, \ldots, v_{1g_1}^{k_{1g_1}}, \ldots v_{m1}^1, \ldots, v_{m1}^{k_{m1}}, \ldots v_{mg_m}^1, \ldots, v_{mg_m}^{k_{mg_m}},\right]$$
5. Write the JNF of A as:
   $$J = \begin{bmatrix} J_{k_{11}}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{k_{mg_m}}(\lambda_m) \end{bmatrix}$$
   where each $J_{k_{ij}}(\lambda_i)$ is a Jordan block of size k_{ij} with diagonal coefficient λ_i:
   $$J_{k_{ij}}(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_i & 1 & \ddots & 0 & 0 \\ 0 & 0 & \lambda_i & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_i & 1 \\ 0 & 0 & \ldots & 0 & 0 & \lambda_i \end{bmatrix}$$
   In other word, each block $J_{k_{ij}}(\lambda_i)$ is associated with eigenvalue λ_i and corresponds to the transformation of each generalised eigenvector v_{ij}^k obtained from the eigenvector v_{ij}^1.
   We have $B^{-1}AB = J$

Cholesky Decomposition

Lower Triangular Matrix has 0 for all i<j A_{ij} (diagonal has elements, below and left also has).
Upper Triangular Matrix has 0 for all i>j A_{ij} (diagonal has elements, above and right also has).
Diagonal coefficients of positive (semi-) definite matrix A is symmetric, if A is positive definite, all diagonal elements are strictly positive, if A is positive semi-definite, all diagonal elements are non-negative.
Largest coefficients of positive (semi-) definite matrix A is symmetric, if A is positive definite, all diagonal elements are strictly positive; if positive semi-definite, all diagonal elements are non-negative.
Largest coefficients of positive (semi-) definite matrices A is symmetric, A is positive definite, then for all indexes i, j, max (A_{ii}, A_{jj}) > |A_{ij}|; if positive semi-definite, then for all indexes i, j, max (A_{ii}, A_{jj}) >= |A_{ij}|. Consequently, in both cases, largest coefficient of A is on diagonal.
If A is symmetric and positive (semi-)definite, then the 1x1, 2x2, …, mxm matrices in upper left corner of A are positive (semi-) definite.
Cholesky Decomposition $A \in IR^{n \times n}$, $A = LL^T$ where L is a lower triangular matrix.
Existence of Cholesky Decomposition A is symmetric and square, A is positive semi-definite ⇔ there exists a matrix $L \in IR^{n \times n}$ such that L is lower triangular and $A = LL^T$.
Existence of Unique Cholesky Decomposition A is symmetric and square, A is positive definite ⇔ there exists a unique matrix $L \in IR^{n \times n}$ such that: L is lower triangular, $A = LL^T$, the diagonal elements of L are positive.
Finding Cholesky Decomposition Compute LL^T using general lower triangular matrix, equate to A.
Using Cholesky Decomposition to solve equation Ax = b → LL^Tx = b → Ly = b (y = L^Tx) use forward substitution → Solve L^Tx = y using backward substitution.

SPECTRAL DECOMP: S = QDQ^T
S performs a scaling operation in the direction of its eigenvectors.

More specifically, given $\lambda_1, \lambda_2 \neq 0$, the circle of equation $\|u\|_2 = 1$ is mapped to the ellipse of equation $\left(\frac{u_1'}{\lambda_1}\right)^2 + \left(\frac{u_2'}{\lambda_2}\right)^2 = 1$, with $u' = (u_1', u_2')^T = Su$.

PROOF TIP:
If Symmetric = can be diagonalised

$$LL^T = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

8.2.1 Householder Map

Theorem 8.1

Suppose we have a hyper-plane P going through the origin with unit normal $u \in \mathbb{R}^m$, i.e., $P = \{x \in \mathbb{R}^m : u \cdot x = 0\}$.
The **Householder** matrix defined by $H_u = I - 2uu^T$ induces reflection wrt P.

Exercise 8.1. *You can verify that:*

- H_u *is* ***involutory***: $H_u = H_u^{-1}$
- H_u *is* ***orthogonal*** *(as all reflections)*: $H_u^T = H_u^{-1}$.
- H_u *preserves the euclidian length of vectors*: $\|H_u(x)\| = \|x\|$.

$\lim_{n\to\infty} a_n = l$ if and only if:

$$\boxed{\forall \epsilon > 0, \ \exists N \in \mathbb{N} \text{ such that } \forall n > N, \ |a_n - l| < \epsilon}$$

Example 9.2 (A not so easy convergence). *Let (a_n) be the sequence defined by $\forall n > 0$, $a_n = \sum_{i=1}^n \frac{1}{i(i+1)}$. Show that it is converging.*
Let $n, m > 0$, then:

$$\begin{aligned} a_n - a_m &= \sum_{i=1}^n \frac{1}{i(i+1)} - \sum_{i=1}^m \frac{1}{i(i+1)} \\ &= \sum_{i=m+1}^n \frac{1}{i(i+1)} \\ &= \sum_{i=m+1}^n \frac{1}{i} - \frac{1}{i+1} \\ &= \frac{1}{m+1} - \frac{1}{n+1} \end{aligned}$$

Let $\epsilon > 0$ and $N = \lceil \frac{2}{\epsilon} \rceil$ then for $n > N$: $\frac{1}{n+1} < \frac{2}{\epsilon}$. It comes that for $n, m > N$ we have:

$$|a_n - a_m| = \left|\frac{1}{m+1} - \frac{1}{n+1}\right| \le \left|\frac{1}{m+1}\right| + \left|\frac{1}{n+1}\right| < \epsilon$$

So (a_n) is a Cauchy sequence of real numbers so it converges.
Actually, we can simplify the expression of a_n to $a_n = 1 - \frac{1}{n+1}$ so it is easy to show that it converges and its limit is 1.

QR Decomposition

- For $u, v \in \mathbb{R}^m$, we defined the **projection** of v onto $u \neq 0$ as:

$$\text{proj}_u(v) = \frac{u \cdot v}{u \cdot u} u \quad (2)$$

- There is a square matrix P such that $Pv = \text{proj}_u(v)$.
- The matrix P must satisfy: $P = P^2$ (why?).
- The matrix $(I - P)$ is also a projector (why?).
- Projector P defines complementary subspaces:

$$\begin{aligned}\text{range}(P) &= \text{null}(I-P) \Rightarrow S_1 \\ \text{range}(I-P) &= \text{null}(P) \Rightarrow S_2\end{aligned} \quad (3)$$

- If S_1 and S_2 are orthogonal, we call P an orthogonal projector (orthogonal projector $\neq$ orthogonal matrix).

P is orthogonal projector iff $P = P^*$

① if $P = P^* \rightarrow Px \in S_1$, $(I-P)y \in S_2$ $\quad S_1 \perp S_2$

$(Px)^*(I-P)y = 0$

$x^*P^*(I-P)y = 0$

$x^*(P - \cancel{P^2})y = 0$

$0 = 0$

② S_1, S_2 as before, $S_1 \perp S_2 \rightarrow P = P^*$

choose orthonormal basis for $\mathbb{C}^m$

$Q = [q_1 \; q_2 \cdots q_m]$ s.t. $\{q_1, \ldots, q_n\}$ a basis for S_1

$\{q_{n+1}, \ldots, q_m\}$ a basis for S_2

$PQ = [Pq_1 \; Pq_2 \cdots Pq_m] = [q_1 \cdots q_n \; 0 \cdots 0]$

$Q^*PQ = [q_1^* \cdots q_m^*][q_1 \cdots q_n \; 0 \cdots 0] = \Sigma$

$Q^*PQ = \Sigma$

$QQ^*PQQ^* = Q\Sigma Q^*$

$P = Q\Sigma Q^*$ $\quad P^* = (Q\Sigma Q^*)^* = Q\Sigma^* Q^* = Q\Sigma Q^*$

$\hat{Q} = [q_1 \cdots q_n]$

$P = \hat{Q}\hat{Q}^* \rightarrow q_1 q_1^*$

SYSTEM OF EQUATIONS

Try to set A = LU. To this write L stub with diagonals = 1 and then e.g for row 3 in the first box write how many negative R1's you needed and in second box how many negative R2's you needed.

If A's diagonal elements approx. 0 then gaussian elim fails. So use partial pivoting. Create permutation matrix P, only 1 row in every row and column. PA = swap rows, AP = swap columns. Permutation can change condition number!

Iterative Methods

$$x^{(k+1)} = Bx^{(k)} + d, k = 0, 1, 2, \ldots \quad (1)$$

Stopping criterion usually defined by relative residual:

$$\frac{||b - Ax^{(k)}||}{||b||} \leq \epsilon$$

- Jacobi iteration (easy to parallelise):

$$x_i^{(k+1)} = \frac{1}{a_{i,i}}\left(b_i - \sum_{j=1, i\neq j}^{n} a_{i,j}x_j^{(k)}\right) \quad (2)$$

- Gauss-Seidel iteration (lower storage requirements):

$$x_i^{(k+1)} = \frac{1}{a_{i,i}}\left(b_i - \sum_{j=1}^{i-1} a_{i,j}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{i,j}x_j^{(k)}\right) \quad (3)$$

- Successive overrelaxation (SOR) iteration:

$$x_i^{(k+1)} = \frac{\omega}{a_{i,i}}\left(b_i - \sum_{j=1}^{i-1} a_{i,j}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{i,j}x_j^{(k)}\right) + (1-\omega)x_i^{(k)} \quad (4)$$

- These schemes "split" matrix A into $A = G + R$
- Assuming we define this split so G is non-singular:

$$\begin{aligned} Ax = b &\Leftrightarrow (G + R)x = b \\ &\Leftrightarrow x = G^{-1}b - G^{-1}Rx \\ &\Leftrightarrow x = Bx + d \end{aligned}$$

- Jacobi iteration: $G = D$, $R = L + U$
- Gauss-Seidel iteration: $G = D + L$, $R = U$
- SOR iteration: $G = (\omega^{-1}D + L)$, $R = ((1 - \omega^{-1})D + U)$

If abs(B) < 1 then converges for any starting x0:

Proof: We have a complete metric space $(\mathbb{R}^m, ||\cdot||)$.

$$\begin{aligned}||f(x) - f(y)|| &= ||Bx + d - (By + d)|| \\ &= ||B(x - y)|| \\ &\leq ||B|| \; ||x - y||\end{aligned}$$

Thus f is continuous and a contraction with contracting factor $||B|| < 1$. The *Banach fixed point theorem* tells us it has a unique fixed point x_*.

Modified Gram-Schmidt

At each iteration, first normalise and then subtract that component from all the vectors that come after it.

More precisely, we are interested in small perturbations:

$$\text{cond}(x) = \lim_{\delta \to 0} \max_{||\delta x|| \leq \delta} \frac{||f(x) - f(x + \delta x)||}{||\delta x||} \quad (2)$$

If f is differentiable:

$$\text{cond}(x) = ||J(x)|| \quad (3)$$

$\text{cond}(x)$ depends on the norm chosen.

We may care about the **relative** perturbation $||\delta x||/||x||$ and **relative** output change $||\delta f||/||f(x)||$ (why?):

$$\kappa(x) = \max_{\delta x}\left(\frac{||\delta f||}{||f(x)||} \Big/ \frac{||\delta x||}{||x||}\right) \quad (4)$$

Again, if f is differentiable:

$$\kappa(x) = \frac{||J(x)||}{||f(x)||/||x||} \quad (5)$$

Where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$. As $\mathbf{A}$ is non-square, it can not be inverted and the cond number as defined in Definition 1.2 can not be calculated. But:

$$\mathbf{Ax} = \mathbf{b} \Rightarrow \mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$$

Assuming the columns of $\mathbf{A}$ are linearly independent then $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ can be inverted and we en with:

$$x = (A^TA)^{-1}A^Tb$$

The matrix $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is called the pseudo-inverse of $\mathbf{A}$ and is denoted $\mathbf{A}^\dagger$.

Definition 1.3: Condition Number of a Non-Square Matrix
Let $\mathbf{A}$ be a non-square matrix, then its condition number is given by:

$$\boxed{\kappa(\mathbf{A}) = ||\mathbf{A}^\dagger||\,||\mathbf{A}||}$$

Consider computing Ax for a matrix $A \in \mathbb{C}^{m\times n}$:

$$\kappa(x) = ||A||\frac{||x||}{||Ax||} \leq ||A|| \; ||A^{-1}|| \quad (6)$$

Because it shows up often, we define the condition number for a matrix A as:

$$\kappa(A) = ||A|| \; ||A^{-1}||.$$

This gives a bound on relative change in A^{-1} given a perturbation on A.

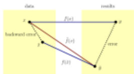
Machine epsilon: $\epsilon_{\text{machine}} = \frac{1}{2}\beta^{1-t}$.

- Half the distance between 1 and the next larger floating point number (size of relative gap).

For any x, there exists $\texttt{fl}(x)$ s.t. $|x - \texttt{fl}(x)| \leq \epsilon_{\text{machine}}|x|$.

- Relative error of the computation: $||\tilde{f}(x) - f(x)||/||f(x)||$.
- An algorithm $\tilde{f}$ for problem f is stable if:

$$\frac{||\tilde{f}(x) - f(\tilde{x})||}{||f(\tilde{x})||} = \mathcal{O}(\epsilon_{\text{machine}}); \; \frac{||\tilde{x} - x||}{||x||} = \mathcal{O}(\epsilon_{\text{machine}}) \quad (7)$$



"A stable algorithm gives *nearly* the right answer to *nearly* the right question."

Non-Linear Sys. Of Eqns

If A is positive definite, solving $Ax = b$ is equivalent to solving $\min_x f(x) = \frac{1}{2}x^TAx - x^Tb$. At a local minimum:

$$\begin{aligned}\nabla f(x) &= \frac{1}{2}2Ax - b = Ax - b = 0 \\ \nabla^2 f(x) &= A \quad \text{(PD by definition)}\end{aligned}$$

HESSIAN

$f(x, y) = \frac{1}{3}x^3 + \frac{1}{3}y^3 - x^2 - y^2$

$f_x = x^2 - 2x = 0 \quad x(x-2) = 0$ when $x = 0, 2$

$f_y = y^2 - 2y = 0 \quad y(y-2) = 0$ when $y = 0, 2$

Stationary points at $(0,0)$, $(0,2)$, $(2,0)$, $(2,2)$

d) ii)

$f_{xx} = 2x - 2$, $f_{xy} = 0$, $f_{yx} = 0$, $f_{yy} = 2y - 2$ $\quad$ Therefore Hessian matrix $H = \begin{pmatrix} 2x-2 & 0 \\ 0 & 2y-2\end{pmatrix}$

For maxima, $\det H > 0$ and $f_{xx}(a,b) < 0$

For minima, $\det H > 0$ and $f_{xx}(a,b) > 0$

For saddle, $\det H < 0$

d) iii)

$(0,0)$ $\begin{vmatrix}-2 & 0 \\ 0 & -2\end{vmatrix} = 4 > 0$ $\quad f_{xx}(0,0) = 2(0) - 2 = -2 < 0$ Hence maxima

$(0,2)$ $\begin{vmatrix}-2 & 0 \\ 0 & 2\end{vmatrix} = -4 < 0$ Hence saddle

$(2,0)$ $\begin{vmatrix}2 & 0 \\ 0 & -2\end{vmatrix} = -4 < 0$ Hence saddle

$(2,2)$ $\begin{vmatrix}2 & 0 \\ 0 & 2\end{vmatrix} = 4 > 0$ $\quad f_{xx}(2,2) = 2(2) - 2 = 2 > 0$ Hence minima

QR ALGORITHM

2. Compute for $i = 0, 1, 2, \ldots$ the QR decomposition $\tilde{\mathbf{Q}}_{i+1}\mathbf{R}_{i+1} = \mathbf{A}_i$
Swap the two factors: $\mathbf{A}_{i+1} := \mathbf{R}_{i+1}\tilde{\mathbf{Q}}_{i+1}$

A_k is similar to A.

$A_k = Q_k^TAQ_k$ so A_k and A have same eigenvalues and v is an eigenvector of A_k iff Q_kv is an eigenvector of A.

Sequence A_k converges to an upper triangular matrix under certain conditions.

Eigenvalue of upper triangular matrix are simply its diagonal elements.

If A is symmetric, so are all A_k.

If A is symmetric, the algorithm converges, under certain conditions, to a diagonal matrix, hence the eigenvectors of A are in effect the columns of Q_k for large enough k.

LU Decomposition A non-singular matrix can be factorised as A = LU where L is lower triangular matrix and U is upper triangular matrix iff A can be reduced to its row echelon form without swapping any two rows (the latter condition holds iff all principal minors of A are non-singular).

If A is non-singular and A = LU with diagonal elements of L being all one, the decomposition is unique.

$A^n = Q_1 \ldots Q_n R_1 \ldots R_n$

A symmetric positive definite matrix with distinct eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_n > 0$ with eigen decomposition $A = Q\Lambda Q^T$. Suppose $Q^T = LU$ with lower triangular L and the diagonal elements of U are positive. Then $A_k \rightarrow \Lambda$.

Fixed Point Equations

Definition 10.7 (Fixed point)

Let S be a non-empty set and $f : S \to S$ a function form S to itself. Then $p \in S$ is called a fixed point if:

$$\boxed{f(p) = p}$$

Definition 10.8 (Contraction)

Let (S, d) a metric space and $f : S \to S$. f is called a contraction of S (or a contracting map) if there exists $0 \leq \alpha < 1$ called the contraction constant such that:

$$\boxed{\forall x, y \in S, \; d(f(x), f(y)) \leq \alpha d(x, y)}$$

Theorem 10.4 (Fixed point theorem)

Let (S, d) a complete metric space and f a contraction of S. Then f has a unique fixed point.

Theorem 10.5

The differential equation

$$\begin{cases} \frac{dx}{dt} = \dot{x}(t) &= f(x, t) \\ x(t_0) &= x_0 \end{cases}$$

has a unique solution: there exists a unique function $x : (t_0 - d, t_0 + d) \to \mathbb{R}$ *satisfying*

$$\frac{dx}{dt} = f(x(t), t) \text{ with } x(t_0) = x_0$$

where $d = \min(a, b/(B + bc))$ *with*

$$B = \sup\{|f(x, t)| : |x - x_0| < b, \quad |t - t_0| < a\} > 0.$$

Iterative Techniques to compute Eigenvalues/Eigenvectors

Definition 3.1:
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. A dominant eigenvalue of $\mathbf{A}$ is an eigenvalue with the largest modulus. A dominant eigenvector is an eigenvector corresponding to a dominant eigenvalue.

3.2.1 Power Iteration

Theorem 3.1:
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ a diagonalisable matrix with eigenvalues of distinct modulus. Let $\lambda \in \mathbb{R}$ be the dominant eigenvalue. We consider the sequence $(\mathbf{x}_k)$ defined by $\mathbf{x}_{k+1} = \frac{\mathbf{Ax}_k}{||\mathbf{Ax}_k||}$ and $x_0 \in \mathbb{R}^n \setminus \{0\}$. Then:

$$\boxed{\mathbf{x}_k \xrightarrow[k\to\infty]{} \mathbf{v}} \text{ and } \boxed{||\mathbf{Ax}_k|| \xrightarrow[k\to\infty]{} |\lambda|}$$

Where $\mathbf{v} \in \mathbb{R}^n$ is a normalised dominant eigenvector. In this case, the notion of convergence that we use for $(\mathbf{x}_k)$ is not exactly rigorous, as you can see in Example 3.1, if the dominant eigenvalue is negative, it can end up oscillating between $\mathbf{v}$ and $-\mathbf{v}$. So here it has to be understood as "the sequence converges to the corresponding eigenspace"

Although quite simple to apply, the power iteration has some limitations:

1. x_0 being chosen at random, it is possible for it to be such that $\alpha_1 = \ldots = \alpha_p = 0$. In this case, the iteration will yield the second dominant eigenvalue and eigenvector.
   - Need to make sure that there is at least one non zero component in the corresponding eigenspace.
   - Usually not an issue in practice as rounding errors during the normalisations will introduce a non zero component.
2. We assumed that all eigenvalues have distinct modulus which is not always the case. It is possible for a matrix to have multiple eigenvalue of maximum modulus. In this case, the power iteration will converge to a linear combination of the corresponding eigenvectors.
3. Convergence may be slow if the dominant eigenvalue is not "very dominant".

3.2.2 Inverse Power Iteration

With the power iteration, we have seen that it is possible to approximate the dominant eigenvalue of a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$. As previously, suppose that $\mathbf{A}$ has eigenvalue with different modulus and in addition, suppose that it is non-singular. This means that none of its eigenvalues are zero and μ is an eigenvalue of $\mathbf{A}$ if and only if $\frac{1}{\mu}$ is an eigenvalue of $\mathbf{A}^{-1}$ with the same eigenvectors. So if we denote λ to be the eigenvalue of $\mathbf{A}$ with the smallest modulus, then $\frac{1}{\lambda}$ is a dominant eigenvalue of $\mathbf{A}^{-1}$.

Example 3.2:
Considering the matrix from Example 3.1, it is non-singular and its inverse is:

$$\mathbf{A}^{-1} = \frac{1}{96}\begin{bmatrix} 19 & -13 & 17 & 1 \\ -13 & 19 & 1 & 17 \\ 17 & 1 & 19 & -13 \\ 1 & 17 & -13 & 19 \end{bmatrix}$$

The eigenvalues of $\mathbf{A}^{-1}$ are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{6}$ and $-\frac{1}{8}$ with respective eigenspaces:

Go to page 15

$$E_{\frac{1}{2}} = \text{span}\left\{\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}\right\}, E_{\frac{1}{4}} = \text{span}\left\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right\}, E_{\frac{1}{6}} = \text{span}\left\{\begin{bmatrix}-1\\1\\1\\-1\end{bmatrix}\right\} \text{ and } E_{-\frac{1}{8}} = \text{span}\left\{\begin{bmatrix}-1\\-1\\1\\1\end{bmatrix}\right\}.$$

Starting from $\mathbf{x}_0 = \begin{bmatrix}1\\0\end{bmatrix}$, similar iterations to Example 3.1 would show that the process converges to the vector $\begin{bmatrix}1\\-1\end{bmatrix}$ and $||\mathbf{A}^{-1}\mathbf{x}_k||_\infty \xrightarrow[k\to\infty]{} \frac{1}{2}$.

From this example we can write the Theorem, that is similar to Theorem 3.1.

Theorem 3.2:
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ a diagonalisable non-singular matrix with eigenvalues of distinct modulus. Let $\lambda \in \mathbb{R}$ be the eigenvalue with the smallest modulus. We consider the sequence $(\mathbf{x}_k)$ defined by $\mathbf{x}_{k+1} = \frac{\mathbf{A}^{-1}\mathbf{x}_k}{||\mathbf{A}^{-1}\mathbf{x}_k||}$ and $x_0 \in \mathbb{R}^n \setminus \{0\}$. Then:

$$\boxed{\mathbf{x}_k \xrightarrow[k\to\infty]{} \mathbf{v}} \text{ and } \boxed{||\mathbf{A}^{-1}\mathbf{x}_k|| \xrightarrow[k\to\infty]{} \left|\frac{1}{\lambda}\right|}$$

Where $\mathbf{v} \in \mathbb{R}^n$ is a normalised eigenvector corresponding to λ.

3.2.3 Shifts

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $s \in \mathbb{R}$. We call the matrix $\mathbf{A} - s\mathbf{I}$ a shifted matrix and we have the following property:

Property 3.1:
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $s \in \mathbb{R}$. $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ if and only if $\lambda - s$ is an eigenvalue of $\mathbf{A} - s\mathbf{I}$ with the same eigenvectors.

Thanks to the shifted matrix, it is possible to focus on a particular eigenvalue of the matrix A.

Theorem 3.3:
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ a diagonalisable matrix with eigenvalues of distinct modulus and $s \in \mathbb{R}$. Suppose that the shifted matrix $\mathbf{A} - s\mathbf{I}$ is non-singular (it is non-singular as long as s is not an eigenvalue of A), then by performing inverse power iterations, we can find the eigenvalue of $\mathbf{A}$ that is the closest to s.

3.3.1 Rayleigh Quotient

Definition 3.2:
Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$, the Rayleigh quotient $R(\mathbf{A}, \mathbf{x})$ is given by:

$$\boxed{R(\mathbf{A}, \mathbf{x}) = \frac{\mathbf{x}^T\mathbf{Ax}}{\mathbf{x}^T\mathbf{x}}}$$

When using an iterative technique to find the eigenvalue/eigenvector of a matrix, you can use the Rayleigh quotient to monitor the convergence to the eigenvalue as it gives access to an approximation of the eigenvalue directly and not its modulus.

3.3.2 Deflation

Suppose that thanks to the power iteration, you have found the dominant eigenvalue of $\mathbf{A} \in \mathbb{R}^{n\times n}$, how do you find the second dominant eigenvalue? One way would be to shift but you would have to pick a s that is close to the this second dominant eigenvalue. Problem: you don't know its value. Another way is to deflate the matrix $\mathbf{A}$ into a matrix $\mathbf{B} \in \mathbb{R}^{(n-1)\times(n-1)}$ that has the same eigenvalues than $\mathbf{A}$ except for the dominant which is absent. How does deflation works?

Let $\lambda_1, \ldots, \lambda_n$ the eigenvalues of $\mathbf{A}$, ordered according to their magnitude with λ_1 being the dominant one with corresponding eigenvector $\mathbf{x}_1$. Define $\mathbf{H} \in \mathbb{R}^{n\times n}$ a non-singular matrix such that:

$$\mathbf{Hx}_1 = \alpha\mathbf{e}_1$$

Where $\alpha \in \mathbb{R}\setminus\{0\}$ and $\mathbf{e}_1 = [1, 0, \ldots, 0]^T$ is the first vector in the standard basis. We then have

$$\mathbf{HAH}^{-1}\mathbf{e}_1 = \mathbf{HA}\frac{\mathbf{x}_1}{\alpha} = \mathbf{H}\frac{\lambda_1}{\alpha}\mathbf{x}_1 = \lambda_1\mathbf{e}_1$$

So the first column of $\mathbf{HAH}^{-1}$ is $[\lambda_1, 0, \ldots, 0]^T$ thus we can write:

$$\mathbf{HAH}^{-1} = \begin{bmatrix}\lambda_1 & \mathbf{b}^T \\ \mathbf{0} & \mathbf{B}\end{bmatrix}$$

Where $\mathbf{B} \in \mathbb{R}^{(n-1)\times(n-1)}$. Now let λ be an eigenvalue of $\mathbf{A}$ and $\mathbf{x}$ a corresponding eigenvector. We have:

$$\mathbf{Ax} = \lambda\mathbf{x} \iff \mathbf{HAx} = \lambda\mathbf{Hx} \iff \mathbf{HAH}^{-1}(\mathbf{Hx}) = \lambda\mathbf{Hx}$$

So λ is eigenvalue of $\mathbf{A}$ if and only if it is also eigenvalue of $\mathbf{HAH}^{-1}$. This means that $\mathbf{B}$ has eigenvalues $\lambda_2, \ldots, \lambda_n$.

Property 3.2:
Let λ_2 be the second dominant eigenvector of $\mathbf{A}$ with $\lambda_2 \neq \lambda_1$. An eigenvector $\mathbf{x}_2$ of $\mathbf{A}$ corresponding to eigenvalue λ_2 is:

$$x_2 = \mathbf{H}^{-1}\begin{bmatrix}\beta \\ \mathbf{z}_2\end{bmatrix}$$

With $\beta = \frac{\mathbf{b}^T\mathbf{z}_2}{\lambda_2 - \lambda_1}$ and $\mathbf{z}_2$ is a dominant eigenvector of $\mathbf{B}$

Remark. Here we have assumed that $\lambda_1 \neq \lambda_2$ or equivalently that the dominant eigenvalue λ_1 had geometric multiplicity 1. Suppose it is no longer the case and the dominant eigenvalue has multiplicity p, that is $\lambda_1 = \lambda_2 = \ldots = \lambda_p$. Then the deflation would work similarly but with blocks instead of vectors and B would be a $(n-p)\times(n-p)$ matrix.

One integral part of deflation is the matrix $\mathbf{H}$. How can you build such matrix? One possibility is using the Householder transformation that represents the reflection through a hyperplane with normal vector $\mathbf{u}$:

$$\mathbf{H} = \mathbf{I} - \frac{2\mathbf{uu}^T}{\mathbf{u}^T\mathbf{u}}$$

$\mathbf{H}$ is symmetric and orthogonal, that is: $\mathbf{H} = \mathbf{H}^T = \mathbf{H}^{-1}$. Now remember that we want $\mathbf{H}$ to be such that $\mathbf{Hx}_1 = \alpha\mathbf{e}_1$ with $\alpha \in \mathbb{R}\setminus\{0\}$. To do that, define $\mathbf{u} = \mathbf{x}_1 + ||\mathbf{x}_1||_2\mathbf{e}_1$. We then have:

$$\mathbf{Hx}_1 = \mathbf{x}_1 - \frac{2\mathbf{u}(\mathbf{u}^T\mathbf{x}_1)}{\mathbf{u}^T\mathbf{u}}$$

But

$$\begin{aligned}\mathbf{u}^T\mathbf{x}_1 &= (\mathbf{x}_1 + ||\mathbf{x}_1||_2\mathbf{e}_1)^T\mathbf{x}_1 \\ &= \mathbf{x}_1^T\mathbf{x}_1 + ||\mathbf{x}_1||_2\mathbf{e}_1^T\mathbf{x}_1 \\ &= ||\mathbf{x}_1||_2(||\mathbf{x}_1||_2 + x_{1,1})\end{aligned}$$

Where $x_{1,1}$ is the first element of $\mathbf{x}_1$. We also have:

$$\begin{aligned}\mathbf{u}^T\mathbf{u} &= (\mathbf{x}_1 + ||\mathbf{x}_1||\mathbf{e}_1)^T(\mathbf{x}_1 + ||\mathbf{x}_1||\mathbf{e}_1) \\ &= \mathbf{x}_1^T\mathbf{x}_1 + 2||\mathbf{x}_1||_2\mathbf{e}_1^T\mathbf{x}_1 + ||\mathbf{x}_1||_2^2\mathbf{e}_1^T\mathbf{e}_1 \\ &= 2||\mathbf{x}_1||_2(||\mathbf{x}_1||_2 + x_{1,1})\end{aligned}$$

So all in all:

$$\mathbf{Hx}_1 = \mathbf{x}_1 - \mathbf{u} = -||\mathbf{x}_1||_2\mathbf{e}_1$$

So by setting $\alpha = -||\mathbf{x}_1||$, we do have $\mathbf{Hx}_1 = \alpha\mathbf{e}_1$ as intended.

Example 3.4:
Let's apply deflation to the matrix from Example 3.1 using the Householder matrix. We take $\mathbf{x}_1 = [1 \;\; 1 \;\; -1 \;\; -1]^T$, then $\mathbf{u} = [3 \;\; 1 \;\; -1 \;\; -1]^T$ and

$$\mathbf{H} = \frac{1}{6}\begin{bmatrix} -3 & -3 & 3 & 3 \\ -3 & 5 & 1 & 1 \\ 3 & 1 & 5 & -1 \\ 3 & 1 & -1 & 5 \end{bmatrix}$$

$\mathbf{H}$ being a Householder transformation, it is its own inverse and we have:

$$\mathbf{HAH}^{-1} = \mathbf{HAH} = \begin{bmatrix} -8 & 0 & 0 & 0 \\ 0 & 4 & 4/3 & -4/3 \\ 0 & 4/3 & 10/3 & 0 \\ 0 & -4/3 & 0 & 14/3 \end{bmatrix}$$

So

$$\mathbf{B} = \begin{bmatrix} 4 & 4/3 & -4/3 \\ 4/3 & 10/3 & 0 \\ -4/3 & 0 & 14/3 \end{bmatrix}$$

And $\mathbf{b}^T = [0 \;\; 0 \;\; 0]$. Performing power iterations on $\mathbf{B}$, the algorithm converges to $\mathbf{z}_2 = [1 \;\; 1/2 \;\; -1]^T$ with corresponding eigenvalue $\lambda_2 = 6$. In this case, we have $\beta = \frac{\mathbf{b}^T\mathbf{z}_2}{\lambda_2 - \lambda_1} = 0$ and

$$\mathbf{x}_2 = \mathbf{H}^{-1}\begin{bmatrix}0\\1\\1/2\\-1\end{bmatrix} = \mathbf{H}\begin{bmatrix}0\\1\\1/2\\-1\end{bmatrix} = \frac{3}{4}\begin{bmatrix}-1\\1\\1\\-1\end{bmatrix}$$

Which is indeed an eigenvector of $\mathbf{A}$.

Fixed points and contracting mapping theorem

Convergence of a sequence of real numbers Let $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ be a sequence of real numbers and $l \in \mathbb{R}$. The sequence (a_n) is said to converge to its limit l, noted $a_n \longrightarrow$ (n → inf written under arrow) l or $\lim_{n\to\inf} a_n = l$ iff $\forall\varepsilon > 0, \exists N \in \mathbb{N}$ such that $\forall n > N$, $|a_n - l| < \varepsilon$.

Steps to show sequence of real numbers converging Find limit l, Take $\varepsilon > 0$, Find $N \in \mathbb{N}$ such that $|a_n - l| < \varepsilon$ for n > N, the value of N will usually depend on and decrease with ε.

Cauchy Sequence Let $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ a sequence of real numbers. Then (a_n) is said to be a Cauchy sequence iff $\forall\varepsilon > 0, \exists N \in \mathbb{N}$ such that $\forall n, m > N$, $|a_n - a_m| < \varepsilon$.

Cauchy Test (a_n) is convergent iff it is a Cauchy sequence.

Definition 10.3 (Metric Space)

A metric space is a tuple (S, d) where S is an non-empty set and d is a metric over S meaning:
A function $d : S \times S \to \mathbb{R}$ such that:

1. $\forall x, y \in S, \; d(x, y) \geq 0$
2. $\forall x, y \in S, \; d(x, y) = 0 \iff x = y$
3. $\forall x, y \in S, \; d(x, y) = d(y, x)$
4. $\forall x, y, z \in S, \; d(x, y) \leq d(x, z) + d(z, y)$

Remark 10.3. *The first property can be derived from the 3 others: let $x, y \in S$, then*

$$\begin{aligned} &d(x, x) \leq d(x, y) + d(y, x) \text{ (property 4)} \\ \Rightarrow &0 \leq d(x, y) + d(y, x) \text{ (property 2)} \\ \Rightarrow &0 \leq d(x, y) + d(x, y) \text{ (property 3)} \\ \Rightarrow &0 \leq d(x, y) \end{aligned}$$

Proposition 10.1 (Metric in a normed vector space)

Let V be a vector space equipped with the norm $||.||$. Let the d be the function defined by:

$$d : \begin{cases} V \times V & \to \mathbb{R} \\ (\mathbf{x}, \mathbf{y}) & \mapsto \|\mathbf{x} - \mathbf{y}\| \end{cases}$$

Then (V, d) is a metric space.

Definition 10.4

Convergence in a metric space] Let (S, d) be a metric space and (a_n) a sequence in S. (a_n) is said to converge to a limit $l \in S$ if and only if:

$$\boxed{\forall\epsilon > 0, \; \exists N \in \mathbb{N} \text{ such that } \forall n > N, \; d(a_n, l) < \epsilon}$$

Proposition 10.2 (Uniqueness of the limit)

Let (S, d) a metric space and (a_n) a sequence in S. If (a_n) is converging, then its limit is unique.

Cauchy Sequence

Definition 10.5 (Cauchy sequence in a metric space)

Let (S, d) a metric space and (a_n) a sequence in S. Then (a_n) is said to be a Cauchy sequence if and only if:

$$\boxed{\forall\epsilon > 0, \; \exists N \in \mathbb{N} \text{ such that } \forall n, m > N, \; d(a_n, a_m) < \epsilon}$$

As for the convergence, the definition of a Cauchy sequence in a metric space is similar to the one in $\mathbb{R}$. But that does not mean that everything that is valid in $\mathbb{R}$ is also valid in any metric space. For example the Cauchy test is no longer valid, instead we have:

Theorem 10.2

Let (S, d) a metric space and (a_n) a sequence in S: if (a_n) is convergent, then it is a Cauchy sequence

Remark 10.4. *While in $\mathbb{R}$ we have an equivalence, in any metric space we are only assured of one implication, this implication being proven in a similar fashion as in $\mathbb{R}$. The bast way to show that it is an implication is to have a counter example to the converse way.*

Definition 10.6 (Complete space)

Let (S, d) be a metric space. Then it is said to be a complete space if and only if every Cauchy sequence in S is also converging in S.

Proposition 10.3

For any $k > 0$, $\mathbb{R}^k$ equipped with any of the three metrics induced by ℓ_1, ℓ_2 or ℓ_∞ is complete.

Proposition 10.4

The space $(C[a, b], \|\cdot\|_\infty)$ is complete.

Theorem 10.3

Let (S, d) be a complete metric space and (a_n) a sequence in S. Then (a_n) is converging if and only if it is a Cauchy sequence.

Data: x ∈ R. Solution: 2x, computed as x ⊕ x

a) We have $f(x) = 2x$, $\tilde{f}(x) = \text{fl}(x) \oplus \text{fl}(x)$. Accounting for rounding of x using ϵ_1:

$$\tilde{f}(x) = x(1 + \epsilon_1) \oplus x(1 + \epsilon_1), |\epsilon_1| \leq \epsilon_{\text{machine}}$$

Accounting for floating point addition $\oplus$ using ϵ_2:

$$\tilde{f}(x) = 2x(1 + \epsilon_1)(1 + \epsilon_2), |\epsilon_2| \leq \epsilon_{\text{machine}}$$

Let $\tilde{x} = x(1 + \epsilon_1)(1 + \epsilon_2)$:

$$\frac{||\tilde{x} - x||}{x} = \frac{||x(\epsilon_1 + \epsilon_2 + \epsilon_1\epsilon_2)||}{||x||} \leq 2\epsilon_{\text{machine}} + \mathcal{O}(\epsilon_{\text{machine}}^2)$$

Therefore, this algorithm is *backward stable*:

$$\frac{||\tilde{x} - x||}{x} = \mathcal{O}(\epsilon_{\text{machine}}); \; f(\tilde{x}) = \tilde{f}(x)$$

Data: none. Solution: e, computed as P∞ k=0 1 k! from left to right using ⊕ and ⊗ and stopping when a summand is reached of magnitude < Emachine

$$\begin{aligned}\tilde{f}(x) &= 1 \oplus \frac{1}{2} \oplus \frac{1}{6} \oplus \ldots \\ &= \left((1 + \frac{1}{2})(1 + \epsilon_1) + \frac{1}{6}\right)(1 + \epsilon_2) \oplus \ldots\end{aligned}$$

where $|\epsilon_i| \leq \epsilon_{\text{machine}}, \forall\, i = 1, 2, \ldots$. Examining the leading powers of the ϵ_i:

$$\begin{aligned} f(x) - \tilde{f}(x) &= \frac{3}{2}\epsilon_1 + \left(\frac{3}{2} + \frac{1}{6}\right)\epsilon_2 + \ldots \\ &= \frac{3}{2}\epsilon_1 + \frac{10}{6}\epsilon_2 + \ldots \end{aligned}$$

This reveals that the coefficients of the ϵ_i terms are identical to the original series of factorials, which converges to e. The number of terms n added satisfies $n! \approx 1/\epsilon_{\text{machine}}$, so the number of terms n increases as $1/\epsilon_{\text{machine}}$ increases. The coefficients of ϵ_i terms are all $\mathcal{O}(1)$, and there will be n terms, so the overall error is:

$$f(x) - \tilde{f}(x) = \mathcal{O}(n\epsilon_{\text{machine}})$$

The key point is that this error scales with $n \times \epsilon_{\text{machine}}$, which is not $\mathcal{O}(\epsilon_{\text{machine}})$ since we have shown that n increases as $\epsilon_{\text{machine}}$ gets smaller. Therefore this algorithm is *unstable*. Note that this analysis does not even account for the error of truncating the series.

$A^TU = VS^T$

$\begin{bmatrix}3 & 2 \\ 2 & 3 \\ 2 & -2\end{bmatrix} \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1 \\ 1 & 1\end{bmatrix} = [v_1 \; v_2 \; v_3]\begin{bmatrix}5 & 0 \\ 0 & 3 \\ 0 & 0\end{bmatrix}$

$3\times2 \quad 2\times2 \quad 3\times3 \quad 3\times2$

$\frac{1}{\sqrt{2}}\begin{bmatrix}5 & -1 \\ 5 & 1 \\ 0 & -4\end{bmatrix} = [5v_1 \;\; 3v_2]$

$\Rightarrow \frac{1}{\sqrt{2}}\begin{bmatrix}5\\5\\0\end{bmatrix} = 5v_1$

$\Rightarrow v_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix}$

$\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\-4\end{bmatrix} = 3v_2$

$\Rightarrow v_2 = \frac{1}{3\sqrt{2}}\begin{bmatrix}-1\\1\\-4\end{bmatrix}$

$v_3 = v_1 \times v_2$

Example 9.4 (The SNCF distance). *Let V be a normed vector space equipped with the norm $||.||$. We define the function $d : V \times V \to \mathbb{R}$ by:*

- $\forall \mathbf{x}, \mathbf{y} \in V, \mathbf{x} \neq \mathbf{y} \Rightarrow d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| + \|\mathbf{y}\|$
- $\forall \mathbf{x} \in V, d(\mathbf{x}, \mathbf{x}) = 0$

Then (V, d) is a metric space. This metric is inspired by the French railway system where most of the train lines go through Paris. Let's show that d is indeed a metric:

Proof. Let $\mathbf{x}, \mathbf{y} \in V$ then if $\mathbf{x} \neq \mathbf{y}$ we have $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| + \|\mathbf{y}\| \geq 0$ by property of the norm. Now if $\mathbf{x} = \mathbf{y}$ then $d(\mathbf{x}, \mathbf{y}) = 0$ so all in all: $\forall \mathbf{x}, \mathbf{y} \in V, d(\mathbf{x}, \mathbf{y}) \geq 0$
Let $\mathbf{x}, \mathbf{y} \in V$, if $\mathbf{x} = \mathbf{y}$ then $d(\mathbf{x}, \mathbf{y}) = 0$. Now suppose that $d(\mathbf{x}, \mathbf{y}) = 0$ and $\mathbf{x} \neq \mathbf{y}$. Then $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| + \|\mathbf{y}\| = 0$ but the norms being non-negative, it comes that: $\|\mathbf{x}\| = \|\mathbf{y}\| = 0$ thus $\mathbf{x} = \mathbf{y} = 0$ which is a contradiction. Hence: $\forall \mathbf{x}, \mathbf{y} \in V, \; d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$
For $\mathbf{x}, \mathbf{y} \in V$ we have $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| + \|\mathbf{y}\| = \|\mathbf{y}\| + \|\mathbf{x}\| = d(\mathbf{y}, \mathbf{x})$: $\forall \mathbf{x}, \mathbf{y} \in V, \; d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.
Finally, let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, if $\mathbf{x} = \mathbf{y}$ then $d(\mathbf{x}, \mathbf{y}) = 0 \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$. If $\mathbf{x} = \mathbf{z}$ then $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{z}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$, same if $\mathbf{y} = \mathbf{z}$. Now if none of these equalities is true, we have $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| + \|\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| + 2\|\mathbf{z}\| = \|\mathbf{x}\| + \|\mathbf{z}\| + \|\mathbf{z}\| + \|\mathbf{y}\| = d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$. So $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V, \; d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$.
And we have shown that (V, d) is a metric space ∎