REVIEW OF LINEAR ALGEBRA INSTRUCTOR: K N RAGHAVAN IST @ MEPCO SIVAKASI, FIRST WEEK, 18–23 MAY 2026

1. Singular Value Decomposition (SVD) [s:svd]

Theorem 1.1. (SVD) *Let A be an $m \times n$ real matrix. Then there exists an orthogonal $m \times m$ real matrix U and an orthogonal $n \times n$ real matrix V such that*

- *$U^t AV$ is diagonal: with $D := U^t AV$ and $D = (d_{ij})$, we have $d_{ij} = 0$ for $i \neq j$.*
- *The diagonal entries of D are non-negative and weakly decreasing: putting $k = \min\{m, n\}$ and $\sigma_1 := d_{11}, \ldots, \sigma_k := d_{kk}$, we have $\sigma_1 \geqslant \ldots \geqslant \sigma_k \geqslant 0$.*

Remark 1.2. Some remarks about the above statement.

- A square matrix Z is called orthogonal if $ZZ^t = Z^t Z = I$, where Z^t denotes the transpose of Z and I the identity matrix.
- We can rewrite $U^t AV = D$ equivalently as $AV = UD$. Thinking of A as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by $x \mapsto Ax$ for $x \in \mathbb{R}^n$, the equality $AV = UD$ means that:
$$Av_1 = \sigma_1 u_1, \quad \ldots, \quad Av_k = \sigma_k u_k$$
where $v_1, \ldots, v_n$ denote the columns of V and $u_1, \ldots, u_m$ the columns of U.
- In case $m < n$, then $v_{m+1}, \ldots, v_n$ form an orthonormal set in the kernel of A.

Proof: The proof is based on this idea. As in the remark, we think of A as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by $x \mapsto Ax$, for $x \in \mathbb{R}^n$. Let v be a vector in $\mathbb{R}^n$ with $\|v\| = 1$ at which $\|Av\|$ attains a maximum (among vectors of unit length). Choose u in $\mathbb{R}^m$ such that $Av = \sigma_1 u$, with $\|u\| = 1$ and $\sigma_1 \geqslant 0$. The maximality of $\|Av\|$ now forces that the hyperplane $\{v\}^\perp$ maps to $\{u\}^\perp$ under A: a proof of this claim is outlined in item (2) in the exercises below.

A repeated application of the above idea—we could also use induction instead—gives a proof of the theorem. We choose the vector v as the first column of V and the vector u as the first column of U. □

1.1. Exercises. [ss:xsvd]

(1) A set $\{v_1, \ldots, v_m\}$ of non-zero vectors in $\mathbb{R}^n$ is said to be orthogonal if $v_i \cdot v_j = 0$ for $i \neq j$; it is said to be orthonormal if in addition $v_i \cdot v_i = 1$ for all i, $1 \leqslant i \leqslant m$.

(a) Let $\{v_1, \ldots, v_m\}$ be a set of vectors in $\mathbb{R}^n$ and let V be the $m \times n$ matrix whose columns are the vectors $v_1, \ldots, v_m$. Then the given set is orthonormal if and only if $V^t V = I$. In particular, the columns of an $n \times n$ orthogonal matrix form an orthonormal set of vectors in $\mathbb{R}^n$.

(b) An orthogonal set $\{v_1, \ldots, v_m\}$ of vectors in $\mathbb{R}^n$ is linearly independent. In particular, $m \leqslant n$. (Hint: Consider a general linear dependence relation and dot it successively with v_1, $\ldots$, v_m.)

(c) Let $\{v_1, \ldots, v_m\}$ be an orthonormal set of vectors in $\mathbb{R}^n$. We can find vectors v_{m+1}, $\ldots$, v_n in $\mathbb{R}^n$ such that $\{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$ is an orthonormal set. (Hint: Suppose $m < n$. Then we can find a non-zero vector w in $\mathbb{R}^n$ that is perpendicular to each of the vectors v_1, $\ldots$, v_m: this can be done by solving the homogeneous system $V^t x = 0$. Put $v_{m+1} := w / \|w\|$. Proceed by induction.)

(2) This exercise provides a proof of the claim made in the course of the proof above of the SVD.
   (a) Let f be a linear functional on $\mathbb{R}^n$ and v a unit vector in $\mathbb{R}^n$ at which $f(v)$ attains a maximum (or minimum). Then $\{v\}^\perp$ belongs to the kernel of f, and, if $f \neq 0$, $\{v\}^\perp$ equals the kernel of f.
   (b) Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let v be a unit vector in $\mathbb{R}^n$ at which $\|Ax\|$ attains a maximum (among vectors of unit length), say λ. Let u be such that $Ax = \lambda u$ with $\|u\| = 1$. Show that A maps $\{v\}^\perp$ to $\{u\}^\perp$. (Hint: Consider the functional $f : \mathbb{R}^n \to \mathbb{R}$ given by $x \mapsto u \cdot Ax$. Check that this attains a maximum at v and apply the previous item.)

(3) Let A, U, V, D, $\underline{\sigma}$, and k be as in the statement of the SVD.
   (a) Then $V^tA^tD = D^t$ and $A^tU = VD^t$. Thus $A^tu_1 = \sigma_1 v_1, \ldots, A^tu_k = \sigma_k v_k$.
   (b) $A^tAV = VD^tD$ and $AA^tU = UDD^t$: thus, $A^tAv_i = \sigma_i^2 v_i$ for and $AA^tu_i = \sigma_i^2 u_i$ for $1 \leqslant i \leqslant k$.
   (c) It follows from the previous item that AA^t and A^tA have the same eigenvalues, all of which are non-negative, and each admits an orthonormal basis of eigenvectors. If v is an eigenvector for A^tA, then Av is an eigenvector for AA^t with the same eigenvalue. The expressions $AA^t = U(DD^t)U^t$ and $A^tA = V(D^tD)V^t$ are SVDs.

(4) Let A be a real $n \times n$ matrix of rank k (for instance, $n = 9$ and $k = 7$). What is the range of possible values of the rank of A^2? Now let A be symmetric in addition. What about the same question? More generally, if A is a real $m \times n$ matrix of rank k, what is the rank of A^tA (or AA^t)?

2. Householder QR Decomposition

[s:hhqr]

[x:hhqr]

2.1. **Exercises.**

(1) Follow the Householder QR algorithm described in class to write the following matrix A as a product QR where Q is 3×3 orthogonal and R is 3×2 upper triangular. Rather than describe Q directly, it is enough to record v_1 and v_2 such that $Q = H_{v_2}H_{v_1}$, where H_v denotes the Householder matrix corresponding to the vector v.

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$$

(2) (Definition of Thin QR Factorisation) Let $A = QR$ be a QR factorisation of an $m \times n$ matrix A. Suppose that $m \geqslant n$. Let Q_1 be the $m \times n$ matrix consisting of the first n columns of Q; and let R_1 be the $n \times n$ matrix consisting of the first n rows of R.
   (a) Observe that $A = Q_1R_1$. This is called a thin QR factorisation of A.
   (b) Suppose in addition that the columns of A are linearly independent. Then the diagonal elements of R_1 can be chosen to be positive and under that condition the thin factorisation of A is unique.

(3) (Routine verification) Given a non-vector x and a unit vector u in $\mathbb{R}^m$, show that $H_v(x)$ is along u for $v = x \pm \|x\|\, u$. (Here, as elsewhere, $H_v = I - 2vv^t/v^tv$ is the reflection in the hyperplane orthogonal to v.)

(4) Let v and w be non-zero vectors in $\mathbb{R}^2$. Describe the composition H_wH_v geometrically in words.

(5) Show that every orthogonal real 2×2 matrix is one of the following two kinds:
- It is a rotation, by an angle θ in the counter-clockwise direction, in which case its matrix with respect to the standard basis is:

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

- It is a reflection, in the line that is at an angle $\theta/2$ counter-clockwise from the positive x-axis, in which case its matrix with respect to the standard basis is:

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

(6) We have seen that, given a non-zero vector v in $\mathbb{R}^m$, the orthogonal projection P_v onto the line through v of $\mathbb{R}^m$ is given by vv^t/v^tv, which can be written also as $v(v^tv)^{-1}v^t$. Generalizing this, prove the following: if A is an $m \times n$ real matrix with linearly independent columns (so that $m \geqslant n$), the orthogonal projection to the column space of A is given by $A(A^tA)^{-1}A^t$.

(7) Prove the following result, called the Cartan-Dieudonné Theorem: Any $n \times n$ real orthogonal matrix M can be written as a product of at most n Householder transformations H_v: $M = H_{v_1} \cdots H_{v_k}$ for some $1 \leqslant k \leqslant n$ and some sequence of non-zero vectors $v_1, \ldots, v_k$ in $\mathbb{R}^n$. (Hint: This follows easily from our procedure for the Householder QR decomposition. Observe that if, in an orthogonal matrix, all entries in the first column other than the first are zero, then so are all entries in the first row other than the first.)

(8) Let M be a real orthogonal 3×3 matrix with determinant 1. By the Cartan-Dieudonné theorem, it is the product of two Householder transformations. Use this and item (4) above to show that each such matrix represents a rotation in a certain axis—a certain line in three space passing through the origin—by a certain angle. In terms of the Householder transformation factors, describe geometrically in words the line and the angle.

(9) The classical Gram-Schmidt algorithm leads to a thin QR factorization, as we now observe. Let A be an $m \times n$ matrix with linearly independent columns (which in particular means that $m \geqslant n$). Let us modify these columns by the standard Gram-Schmidt process to make them an orthonormal set of n vectors. Let Q be the $m \times n$ matrix whose columns are these orthonormal vectors. Then $A = QR$ where R is an upper triangular matrix with positive diagonal entries, and we have a thin QR factorisation of A.

3. Full Rank Least Squares (LS) Problem [s:frls]

3.1. **Exercises.** [x:frls]

(1) (Normal Form of a System of Linear Equations) Let $Ax = b$ be a system of linear equations with A being a real matrix of size $m \times n$, b a vector in $\mathbb{R}^m$, and x an $n \times 1$ matrix of variables. Then the system $A^tAx = A^tb$ is called the normal form of $Ax = b$. Show the following:
   (a) The normal form always has a solution, even if the original system does not.
   (b) The solution set of the normal form coincides with that of $Ax = b^\star$, where $b^\star$ is the orthogonal projection of b on the column space of A. In particular, in case $Ax = b$ has a solution, the solution set coincides with that of the normal form.

4. On the Pseudo-inverse

[s:pseudo]

Let T be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. Observe that the restriction of T to $\ker(T)^\perp$ (the orthogonal complement to the kernel of T) is an isomorphism onto $\mathrm{Image}(T)$. The pseudo-inverse of T, denoted T^+, is the unique linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ that is determined by the following conditions:

- T^+ restricted to the $\mathrm{Image}(T)$ is the inverse of the restriction of T to $\ker(T)^\perp$.
- T^+ vanishes on $\mathrm{Image}(T)^\perp$.

Let A be an $m \times n$ matrix. The pseudo-inverse, denoted A^+, is defined to be the $n \times m$ matrix corresponding to the linear transformation T^+, where T is the linear transformation corresponding to A. Correspondence between matrices and linear transformations in this context are all with respect to standard bases.

We will use the language of matrices in what follows. The translation to the language of linear transformations is obvious.

[p:a+]

Proposition 4.1. *Let A be a real $m \times n$ matrix. Then:*

1. *$(A^+)^+ = A$; $A^+ = (A^tA)^{-1}A^t$ if A has linearly independent columns; in particular, $A^+ = A^{-1}$ if A is square and invertible.*
2. *$AA^+A = A$, $A^+AA^+ = A^+$, $(AA^+)^t = AA^+$, and $(A^+A)^t = A^+A$. These are called the Moore-Penrose Equations.*
3. *AA^+ is the orthogonal projection to the column space of A; and A^+A is the orthogonal projection to the orthogonal complement of the null space of A.* (Thinking of A as a linear transformation in the usual way, the column space is just the image and the null space is the kernel.)

We leave the proof as an exercise to the reader, with the following hint: to check that two $m \times n$ matrices are equal, it is enough to check that they coincide on the null space of A and also on its the orthogonal complement; similarly, to check that two $n \times m$ matrices are equal, it is enough that they coincide on the image of A and on its orthogonal complement.

[x:sfls]

4.1. Exercises.

1. The Moore-Penrose equations characterize A^+. More precisely, if X is any $n \times m$ matrix such that $AXA = A$, $XAX = X$, $(XA)^t = XA$, and $(AX)^t = AX$, then $X = A^+$.
2. (a) For a diagonal $m \times n$ matrix $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$, with σ_1, …, σ_r being all non-zero, Σ^+ is the diagonal $n \times m$ matrix $\mathrm{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)$.
   (b) Let $A = U\Sigma V^t$ be any SVD of an $m \times n$ real matrix A, with Σ being diagonal: $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$, with $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0$. Then:
      (i) $A^+ = V\Sigma^+U^t$, and this expression is an SVD of A^+.
      (ii) $A^t = V\Sigma^tU^t$, and this expression is an SVD of A^t.
3. Let $Ax = b$ be a system of linear equations, where A is an $m \times n$ real matrix, b is a vector in $\mathbb{R}^m$, and x is an $n \times 1$ matrix of variables.
   (a) Suppose that the system admits a solution. Show that there is a unique solution x_0 such that x_0 is orthogonal to the null space of A. Show moreover that this unique solution x_0 is also the unique solution for which $\|x\|$ is minimized among all solutions x. We call x_0 the least norm solution.
   (b) In general (even when the system $Ax = b$ does not admit a solution), prove the following:
      (i) There is a unique vector $x^\star$ in $\mathbb{R}^n$ such that the following conditions are satisfied:

- $\|Ax^\star - b\| \leqslant \|Ax - b\|$ for all $x \in \mathbb{R}^n$.
- For $x \in \mathbb{R}^n$ with $\|Ax - b\| = \|Ax^\star - b\|$, we have $\|x\| \geqslant \|x^\star\|$.

This vector $x^\star$ is called the least norm least squares solution of the system $Ax = b$. (Hint: Replace b by its orthogonal projection $b^\star$ on the column space of A. Now, invoke the previous item and let $x^\star$ be the unique least-norm solution to $Ax = b^\star$.)

(ii) The least norm least squares solution is given by $x^\star = A^+ b$, where A^+ is the pseudo-inverse of A.

(iii) The least norm least squares solution $x^\star = A^+ b$ is the least norm solution of the normal form $A^t A x = A^t b$ (see Exercise (1) of §3 for the definition and properties of the normal form).

(4) Let A be an $m \times n$ real matrix.

(a) Observe that both A^+ and A^t have Image$(A)^\perp$ as their kernel, both are isomorphisms when restricted to Image(A), and both have image equal to Kernel$(A)^\perp$.

(b) Identify the subspace of $\mathbb{R}^m$ on which A^t and A^+ coincide.