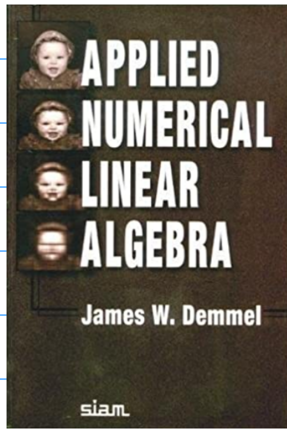
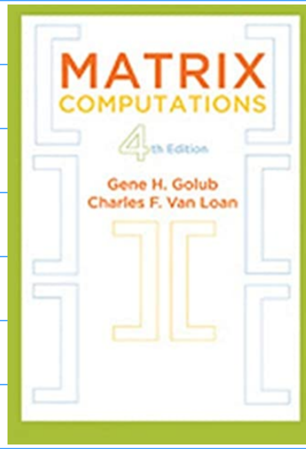
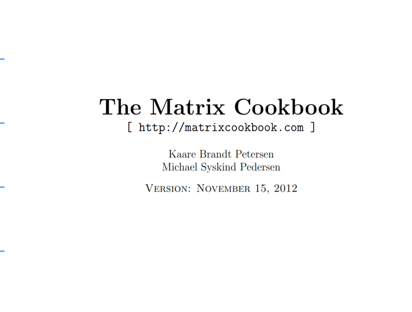
Lecture 9 – Review of Linear Algebra

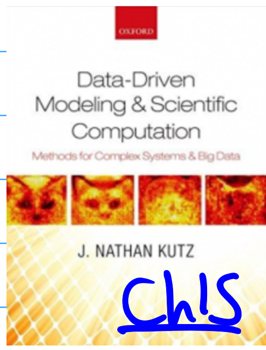
Starting this week, we will turn our attention away from signal processing to focus on "dimensionality reduction" techniques & in particular principal component analysis. These are techniques that rely heavily on linear algebra so we dedicate this lecture to a review of this topic.

Refs:









8.1 Basics & definitions

Def^n: A matrix $A \in \mathbb{R}^{n\times m}$ is a linear operator mapping vectors in $\mathbb{R}^m$ to $\mathbb{R}^n$,

$$\underline{x} \in \mathbb{R}^m, \quad A\underline{x} = \underline{b} \in \mathbb{R}^n$$

$$\underline{b} = A\underline{x} = \sum_{j=1}^{m} x_j \cdot \underline{a}_j \qquad A = \left[\underline{a}_1 \middle| \underline{a}_2 \middle| \cdots \middle| \underline{a}_m \right]$$

where $\underline{a}_j \in \mathbb{R}^n$ are the columns of A.

Similarly we define the action of a matrix $A \in \mathbb{R}^{n \times m}$ another matrix $B \in \mathbb{R}^{m \times d}$ as

$$AB = A\left[\underline{b}_1 \middle| \cdots \middle| \underline{b}_d \right] = \left[A\underline{b}_1 \middle| \cdots \middle| A\underline{b}_d \right] \in \mathbb{R}^{n \times d}$$

which is another matrix.

Def^n: The range of A is the set of vectors $\underline{b} \in \mathbb{R}^n$ that can be written as $A\underline{x}$ for some $\underline{x} \in \mathbb{R}^m$. ie

$$\mathrm{range}(A) := \mathrm{span}\{\underline{a}_1, \cdots, \underline{a}_m\}.$$

Def^n: The null space of A is the set of vectors $\underline{x} \in \mathbb{R}^m$ that satisfy $A\underline{x} = 0$.

$$\mathrm{null}(A) := \{\underline{x} \in \mathbb{R}^m \mid A\underline{x} = 0\}.$$

We define the column rank of A as the dimension of its column space, ie, number of linearly independent vectors in its column space. Analogously the row rank of A is the dimension of its row space. In fact, the row & column ranks are always equal! so we simply refer to it as rank of A. Notation $\mathrm{rank}(A)$.

Def^n: $A \in \mathbb{R}^{n\times m}$ has full rank if $\text{rank}(A) = \min\{m,n\}$, ie, A has the maximal possible rank.

Thm: $A \in \mathbb{R}^{n\times m}$ with $n \geq m$ has full rank iff (if and only if) it maps no two distinct vectors to the same vector.

Given $\underline{x}, \underline{y} \in \mathbb{R}^m$, $\underline{x} \neq \underline{y} \iff A\underline{x} \neq A\underline{y}$.

Def^n: A nonsingular or invertible matrix is a square matrix $A \in \mathbb{R}^{n\times n}$ of full rank.

Note that by def^n the columns of an invertible matrix form a basis for $\mathbb{R}^n$.

Def^n: Any invertible matrix has a unique inverse $A^{-1} \in \mathbb{R}^{n\times n}$ satisfying

$$AA^{-1} = A^{-1}A = I$$ ← identity mat. in $\mathbb{R}^{n\times n}$.

Thm The following are equivalent for $A \in \mathbb{R}^{n\times n}$.

(a) A has an inverse A^{-1},
(b) $\text{rank}(A) = n$,
(c) $\text{range}(A) = \mathbb{R}^n$,
(d) $\text{null}(A) = \emptyset$,
(e) $\det(A) \neq 0$.

You might think of A^{-1} as a matrix and in turn $A^{-1}\underline{b}$ as a matrix vector product. But a more useful perspective is to think of A^{-1} as the inverse operator corresponding to A, ie, $A^{-1}\underline{b}$ is the unique vector $\underline{x}\in\mathbb{R}^n$ so that $A\underline{x}=\underline{b}$.

In this light, matrix inversion is simply a change of basis

$$\boxed{\underline{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{j=1}^{n} b_j \cdot \underline{e}_j} \quad \underset{A}{\overset{A^{-1}}{\rightleftarrows}} \quad \boxed{\begin{gathered} A^{-1}\underline{b} = \underline{x} \\ \underline{b} = \sum_{j=1}^{n} x_j \cdot \underline{a}_j \end{gathered}}$$

expansion of b in standard coordinate basis

expansion of b in column space of A.

8.2 Orthogonality

Given $\underline{x},\underline{y}\in\mathbb{R}^n$ we define their (Euclidean) inner product

$$\underline{x}^T\underline{y} = \sum_{j=1}^{n} x_j y_j$$

along with the Euclidean norm (length)

$$\|\underline{x}\| := (\underline{x}^T\underline{x})^{1/2}.$$

The angle between the two vector can then be defined as

$$\alpha = \cos^{-1}\left(\frac{\underline{x}^T\underline{y}}{\|\underline{x}\|\cdot\|\underline{y}\|}\right).$$

Def^n: $\underline{x}$ & $\underline{y}$ are orthogonal vectors if $\underline{x}^T\underline{y} = 0$. (ie they are at right angle).

A set of vectors $S := \{\underline{s}_1, \ldots, \underline{s}_k \in \mathbb{R}^n\}$ is orthogonal if its elements are pairwise orthogonal $\underline{s}_j^T \underline{s}_k = 0$ for $j \neq k$.
Furthermore, we say S is orthonormal if the $\underline{s}_k$ are also normalized, ie, $\|\underline{s}_k\| = 1$.

easy to check: Vectors in an orthogonal set are linearly independent.

Def^n: A matrix $Q \in \mathbb{R}^{n\times n}$ is called unitary or orthonormal if $Q^T = Q^{-1}$. ie,

$$Q = \begin{bmatrix} \underline{q}_1 & \cdots & \underline{q}_n \end{bmatrix}, \quad \begin{bmatrix} \underline{q}_1^T \\ \vdots \\ \underline{q}_n^T \end{bmatrix} \begin{bmatrix} \underline{q}_1 & \cdots & \underline{q}_n \end{bmatrix} = \begin{bmatrix} 1 & & \emptyset \\ & \ddots & \\ \emptyset & & 1 \end{bmatrix}$$

in other words, the colums of Q form an orthonormal basis for $\mathbb{R}^n$.

Going back to the connection between inversion & change of basis.

$$\boxed{\underline{b} = \sum_{j=1}^{n} b_j \underline{e}_j} \quad \underset{Q}{\overset{Q^T}{\rightleftarrows}} \quad \boxed{Q^T\underline{b} = \underline{x}, \quad \underline{b} = \sum_{j=1}^{n} x_j \underline{q}_j}$$

Uninatry/Orthogonal matrices preserve Euclidean geometry

$$(Q\underline{x})^T(Q\underline{y}) = \underline{x}^T Q^T Q \underline{y} = \underline{x}^T\underline{y} \iff \|Q\underline{x}\| = \|\underline{x}\|.$$

ie, rigid rotation!

8.3 Norms

So far we saw the Euclidean norm (2-norm) defined as $\|\underline{x}\| = \underline{x}^T\underline{x}$. But this is not the only notion of norm (size, length etc) that we can use. In fact, other norms are very useful in data analysis.

Def^n: A function $\|\cdot\| : \mathbb{R}^n \longrightarrow \mathbb{R}$ is called a norm on $\mathbb{R}^n$ if it satisfies the following axioms:

(a) $\|\underline{x}\| \geq 0$ & $\|\underline{x}\| = 0$ only if $\underline{x} = \underline{0}$

(b) $\|\underline{x}+\underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$ (triangle inequality)

(c) $\|\alpha \underline{x}\| = |\alpha| \, \|\underline{x}\|$.

We will particularly focus on the p-norms in the course

$$x \in \mathbb{R}^n, \quad \|\underline{x}\|_p := \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}, \qquad 1 \leq p \leq +\infty$$

$$\|\underline{x}\|_1 = \sum_{j=1}^{n} |x_j| \quad \text{(widely used in image processing)}$$

$$\|\underline{x}\|_2 = (\underline{x}^T\underline{x})^{1/2} \quad \text{(Usual Euclidean norm)}$$

$$\|\underline{x}\|_\infty = \max_{1 \leq j \leq n} |x_j|$$

Vector norms can be used to define norms on matrices. This is done by viewing matrices as linear operators between vector spaces.

$A \in \mathbb{R}^{n \times m}$ & any norms $\|\cdot\|_{(n)}, \|\cdot\|_{(m)}$ on $\mathbb{R}^n$ & $\mathbb{R}^m$ respectively.

We define
$$\|A\|_{(m,n)} := \sup_{\substack{x \in \mathbb{R}^m \\ x \neq 0}} \frac{\|A\underline{x}\|_{(n)}}{\|\underline{x}\|_{(m)}} = \sup_{\|\underline{x}\|_{(m)}=1} \|A\underline{x}\|_{(n)}$$

In otherwords, the induced matrix norm $\|\cdot\|_{(m,n)}$ is the smallest constant $C \geq 0$ for which the following inequality holds.

$$\|A\underline{x}\|_{(n)} \leq C\|\underline{x}\|_{(m)}.$$

In the case of certain p-norms the induced matrix norms have simple interpretations.

$$\mathbb{R}^{n\times m} \ni A = \left[\underline{a}_1 \Big| \cdots \Big| \underline{a}_m\right] = \begin{bmatrix} \tilde{\underline{a}}_1^T \\ \vdots \\ \tilde{\underline{a}}_n^T \end{bmatrix}$$

induced by $\|\cdot\|_1$ on both $\mathbb{R}^n$ & $\mathbb{R}^m$ →

$$\|A\|_1 = \max_{1\leq j\leq m} \|\underline{a}_j\|_1$$

$$\|A\|_\infty = \max_{1\leq i\leq n} \|\tilde{\underline{a}}_i^T\|_1.$$

We can also define matrix norms directly. Using the norm axioms.

(a) $\|A\| \geq 0$ & $\|A\|=0$ only if $A=\mathbb{0}$

(b) $\|A+B\| \leq \|A\| + \|B\|$

(c) $\|\alpha A\| = |\alpha|\cdot\|A\|$.

Notably useful example here is the Frobenius (Hilbert-Schmidt norm).

$$A\in\mathbb{R}^{n\times m} \quad \|A\|_F = \left(\sum_{i=1}^n\sum_{j=1}^m |a_{ij}|^2\right)^{1/2} = \left(\sum_{j=1}^m \|\underline{a}_j\|_2^2\right)^{1/2}$$

$$= \sqrt{\mathrm{Tr}(A^TA)} = \sqrt{\mathrm{Tr}(AA^T)},$$

where $\mathrm{Tr}(B)$ denotes trace of B, the sum of its diagonal entries $\mathrm{Tr}(B) = \sum_j b_{jj}$.

Unitary matrices also preserve certain matrix norms.

Thm $Q \in \mathbb{R}^{n \times n}$ (unitary) & $A \in \mathbb{R}^{n \times m}$ then

$$\|QA\|_2 = \|A\|_2 \quad \& \quad \|QA\|_F = \|A\|_F .$$

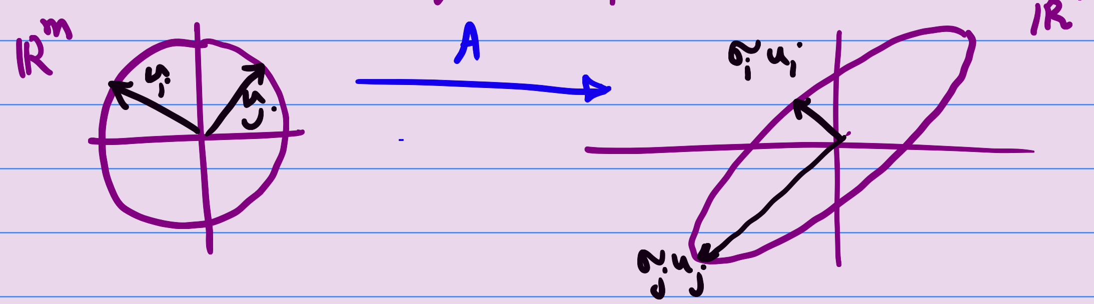
This thm is also true if Q is not square, $Q \in \mathbb{R}^{d \times n}$, $d > m$ but has orthonormal columns

8.4 The Singular Value Decomposition

We are now ready to present the Singular Value Decomposition (SVD), a remarkably powerful tool from linear algebra with profound applications in science, engineer, ML, etc.

Geometric Intuition

The image of the unit sphere under any $n \times m$ matrix is a hyper ellipse.



Thm Let $A \in \mathbb{R}^{n\times m}$ then there exist unitary matrices $U \in \mathbb{R}^{n\times n}$ & $V \in \mathbb{R}^{m\times m}$ along with a diagonal matrix $\Sigma \in \mathbb{R}^{n\times m}$, with positive entries, so that

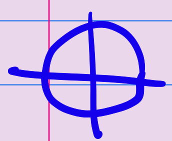
$$A = U\Sigma V^T$$

Note: This is true for any matrix!

$$\begin{bmatrix} | & & | \\ \underline{a}_1 & \cdots & \underline{a}_m \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \underline{u}_1 & \cdots & \underline{u}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_m \\ \hline & 0 & \end{bmatrix} \begin{bmatrix} \underline{v}_1^T \\ \vdots \\ \underline{v}_m^T \end{bmatrix}$$

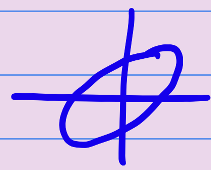
- The $\{\underline{u}_j\}$ the columns of U are called the left singular vectors of A & form a basis for $\mathbb{R}^n$.
- The $\{\underline{v}_j\}$ 〃 〃 V 〃 〃 〃 right singular vectors of A & 〃 〃 〃 $\mathbb{R}^m$
- The $\{\sigma_j\}$, are non-negative (typically in descending order) $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$ & are called the singular values of A.

Recall that multiplication by unitary matrices is equivalent to rigid rotation. Hence, the SVD tells us that the action of any matrix A is equivalent to the following:



$$x \xrightarrow{\quad\quad\quad A \quad\quad\quad} b = Ax$$

$$x \xrightarrow{V^T} V^T x \xrightarrow{\Sigma} \Sigma V^T x \xrightarrow{U} U\Sigma V^T x$$



rotation   scaling   New basis

Thm Every $A \in \mathbb{R}^{n\times m}$ has an SVD, the singular values σ_j unique & if $n=m$ & the σ_j are distinct then the left & right singular vectors $\underline{u}_j, \underline{v}_j$ are unique up to a sign (ie, can be multiplied by ± 1).

The SVD reveals many of the fundamental properties of a matrix:

(a) $\mathrm{rank}(A) = \#$ of non-zero singular values.

(b) suppose $A \in \mathbb{R}^{n\times m}$, $n \geq m$ is rank $r < m$, ie,

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_m = 0$$

then $v_{r+1}, \ldots, v_m$ span $\mathrm{null}(A)$ and $u_1, \ldots, u_r$ span $\mathrm{range}(A)$.

(c) $$\|A\|_F = \left(\sum_{j=1}^{\min(m,n)} \sigma_j^2\right)^{1/2}.$$

(d) Nuclear norm $$\|A\|_* := \sum_{j=1}^{\min(m,n)} \sigma_j$$

(e) $A \in \mathbb{R}^{n\times n}$, $$\det(A) = \prod_{j=1}^{n} \sigma_j$$

8.5 Eigenvalue decomposition

Yet another way to look at SVD is that the action of any matrix A is equivalent to that of a diagonal matrix after appropriate unitary transformations

$$A\underline{x} = \underline{b}$$

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

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

This is call diagonalization of A. But SVD is not the only possible way.

Def^n Suppose $A \in \mathbb{R}^{n\times n}$ is square. Then a number $\lambda \in \mathbb{C}$ & a vector $\underline{q} \in \mathbb{C}^n$ are called an eigenvalue & eigenvector (respectively) for A if

$$A\underline{q} = \lambda\underline{q}.$$

The span of eigenvectors of A is called the eigenspace of A & the set of eigenvalues is called the spectrum.

Let Q denote the matrix with columns q_j the eigen vectors of A & Λ the digonal matrix of corresponding eigenvalues λ_j.

Then the eigenvalue decomposition (eigendecomposition) of A is a factorization of the form

$$A = Q \Lambda Q^{-1}$$

Such decompositions do not always exist! if they do then we say A is non-defective.

If A is Hermitian (self-adjoint/symmetric) then it is non-defective (has eigendecomp) & all of its eigenvalues are real.

Further, we say A is non-negative definite symm. (NDS) if its eigenvalues are non-negative, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$.

Equivalently $x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$. (Note A is assumed symmetric automatically).

NDS matrices are also called positive definite, or Hermitian by authors.

Note: (a) There are fundamental differences between eigendecomp. & SVD.

(a.1) SVD always exist & σ_j are real valued. but eigendecomp. does not always exist & λ_j may be complex valued.

(a.2) SVD uses different bases for domain & range of matrix while eigendecomp assumes the same basis for both.

(b) if matrix A is Hermitian then the singular values of A coincide with the absolute value of its eigenvalues. (check)