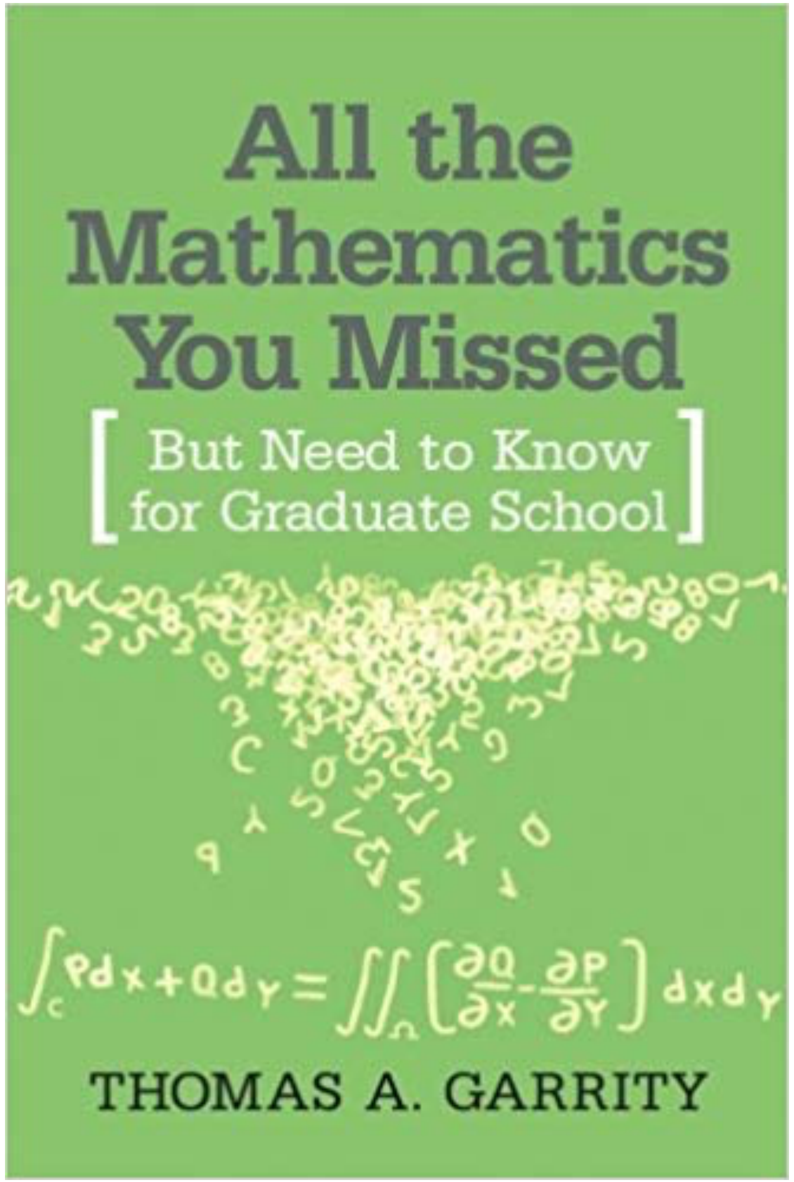
Math Review-1

For ENGN8535

Math for ENGN8535



Subjects

- Linear Algebra (Matrix computation, Matrix calculus)
- Probability theory (and statistical inference)
- Numerical Optimisation
- Python programming (Numpy)
- → Example applications in Data Analysis:
  Ranking and Recommendation algorithms.

Some Calculus

Derivatives

- For a single valued function $f\,(x_1,\ldots,x_n)$: $\mathbb{R}^n \rightarrow \mathbb{R}$
  - Partial derivative of x_i: $$\frac{\partial f}{\partial x_i} = \lim_{h \rightarrow 0} \frac{f(x + he_i) - f(x)}{h}$$
  - Gradient of f is $$\nabla f(x) = \begin{pmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_n \end{pmatrix}$$
  - Directional derivatives: for $||p||=1$

$$D(f(x), p) = \lim_{h \leftarrow 0} \frac{f(x + hp) - f(x)}{h} = \nabla f(x)^T p.$$

Hessian

- : the second order derivative of *f*

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix}$$

- If *f* is twice continuously differentiable, Hessian is symmetric.

Taylor’s theorem

- Mean value theorem: for $\alpha \in (0,1)$

$$f(x+p) = f(x) + \nabla f(x+\alpha p)^T p$$

- Taylor’s theorem: for $\alpha \in (0,1)$

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+\alpha p) p$$

Vector valued function

- For a vector valued function $r: \mathbb{R}^n \to \mathbb{R}^m$

$$r(x) = \begin{pmatrix} f_1(x_1, x_2, \cdots, x_n) \\ f_2(x_1, x_2, \cdots, x_n) \\ \vdots \\ f_m(x_1, x_2, \cdots, x_n) \end{pmatrix}$$

- The Jacobian of r at x is

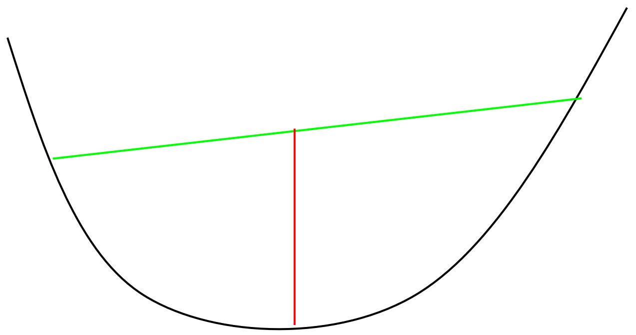
$$J(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \cdots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} \nabla f_1(x)^T \\ \nabla f_2(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{pmatrix}$$

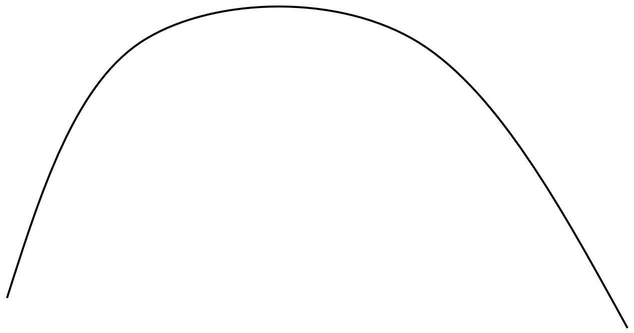
Convex function

- A function f is convex if for $\alpha \in [0,1]$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

- A function f is concave if $-f$ is convex





Lipschitz continuity

- A function f is said to be Lipschitz continuous on some set $\mathcal{N}$ if there is a constant L>0 such that

$$||f(x) - f(y)|| \leq L\ ||x - y||\ \text{for all } x,y \in \mathcal{N}$$

- If function f and g are Lipschitz continuous on $\mathcal{N}$, $f+g$ and fg are Lipschitz continuous on $\mathcal{N}$.

- ENGN8535 2021 S1: Math Review-1

Linear Algebra

Basic concepts

- Vector in R^n is an ordered set of n real numbers.
  - e.g. v = (1,6,3,4) is in R^4
  - A column vector: $\begin{pmatrix} 1 \\ 6 \\ 3 \\ 4 \end{pmatrix}$
  - A row vector: $\begin{pmatrix} 1 & 6 & 3 & 4 \end{pmatrix}$
- m-by-n matrix is an object in $R^{m \times n}$ with m rows and n columns, each entry filled with a (typically) real number:

$$\begin{pmatrix} 1 & 2 & 8 \\ 4 & 78 & 6 \\ 9 & 3 & 2 \end{pmatrix}$$

Basic concepts

Vector norms: A norm of a vector ||x|| is informally a measure of the "length" of the vector.

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

– Common norms: L_1, L_2 (Euclidean)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

– $L_{infinity}$

$$\|x\|_\infty = \max_i |x_i|$$

Vector

- A column vector $x \in \mathbb{R}^n$ is denoted as $$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$
- The transpose of x is $$x^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$$
- The inner product of $x, y \in \mathbb{R}^n$ is $$x^T y = \sum_{i=1}^{n} x_i y_i$$
- Vector norm
  - 1-norm $$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$
  - 2-norm $$\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$$
  - ∞-norm $$\|x\|_\infty = \max_{i=1 \ldots n} |x_i|$$

Useful inequalities involving norms

Proposition 4.2. *The following inequalities hold for all $x \in \mathbb{R}^n$ (or $x \in \mathbb{C}^n$):*

$$\begin{aligned} \|x\|_\infty &\le \|x\|_1 \le n\|x\|_\infty, \\ \|x\|_\infty &\le \|x\|_2 \le \sqrt{n}\|x\|_\infty, \\ \|x\|_2 &\le \|x\|_1 \le \sqrt{n}\|x\|_2. \end{aligned}$$

Matrix

- A matrix $A \in \mathbb{R}^{m \times n}$ is $$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix}$$
- The transpose of A is $$A^T = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{m1} \\ A_{12} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & \ldots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{mn} \end{pmatrix}$$
- Matrix A is symmetric if
- Matrix norm: $||A||_p = \max ||Ax||_p$ for $||x||_p = 1$

- P=1 → L1 norm
- P=2 → Frobenius norm
- P= infinity → L-infinity norm.

Matrix norms

Matrix norms are functions $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ that satisfy the same properties as vector norms. Let $A \in \mathbb{R}^{m\times n}$. Here are a few examples of matrix norms:

- The Frobenius norm: $||A||_F = \sqrt{\mathrm{Tr}(A^T A)} = \sqrt{\sum_{i,j} A_{i,j}^2}$
- The sum-absolute-value norm: $||A||_{sav} = \sum_{i,j} |X_{i,j}|$
- The max-absolute-value norm: $||A||_{mav} = \max_{i,j} |A_{i,j}|$

Basic concepts

We will use lower case letters for vectors The elements are referred by x_i.

- Vector dot (inner) product:

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ x_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i .$$

If u•v=0, $||u||_2$!= 0, $||v||_2$!= 0 → *u and v are orthogonal*

If u•v=0, $||u||_2$ = 1, $||v||_2$ = 1 → *u and v are orthonormal*

- Vector outer product (direct product):

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix}$$

Basic concepts

We will use upper case letters for matrices. The elements are referred by A_{ij}.

- Matrix product:

$$A \in \mathbb{R}^{m\times n} \qquad B \in \mathbb{R}^{n\times p}$$

$$C = AB \in \mathbb{R}^{m\times p}$$

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

e.g.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

Special matrices

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$ diagonal

$$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$$ upper-triangular

$$\begin{pmatrix} a & b & 0 & 0 \\ c & d & e & 0 \\ 0 & f & g & h \\ 0 & 0 & i & j \end{pmatrix}$$ tri-diagonal

$$\begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}$$ lower-triangular

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ I (identity matrix)

Basic concepts

- Transpose

e.g. $$\begin{pmatrix} a \\ b \end{pmatrix}^T = \begin{pmatrix} a & b \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A + B)^T = A^T + B^T$

Linear independence

- A set of vectors is linearly independent if none of them can be written as a linear combination of the others.
- Vectors $v_1,\ldots,v_k$ are linearly independent if $c_1v_1+\ldots+c_kv_k = 0$ implies $c_1=\ldots=c_k=0$

$$\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

e.g.

$$\begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$(u,v)=(0,0)$, i.e. the columns are linearly independent.

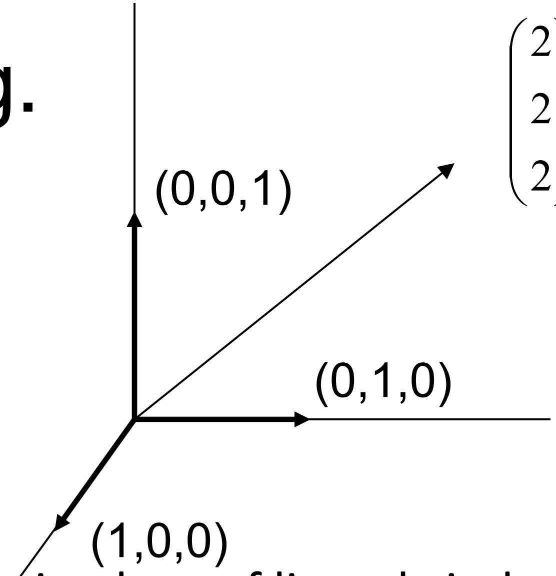
$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

$x3 = -2x1 + x2$

Span of a vector space

- If all vectors in a vector space may be expressed as linear combinations of a set of vectors $v_1,\ldots,v_k$, then $v_1,\ldots,v_k$ spans the space.
- The cardinality of this set is the dimension of the vector space.

e.g.



$$\begin{pmatrix}2\\2\\2\end{pmatrix} = 2\begin{pmatrix}1\\0\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\\0\end{pmatrix} + 2\begin{pmatrix}0\\0\\1\end{pmatrix}$$

- A basis is a maximal set of linearly independent vectors and a minimal set of spanning vectors of a vector space

Rank of a Matrix

- rank(A) (the rank of a m-by-n matrix A) is

  The maximal number of linearly independent columns

  =The maximal number of linearly independent rows

  =The dimension of col(A)

  =The dimension of row(A)

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$$

- If A is n by m, then
  - rank(A)<= min(m,n)
  - If n=rank(A), then A has full row rank
  - If m=rank(A), then A has full column rank

Inverse of a matrix

- Inverse of a square matrix A, denoted by A^{-1} is the *unique* matrix s.t.
  - $AA^{-1} = A^{-1}A = I$ (identity matrix)
- If A^{-1} and B^{-1} exist, then
  - $(AB)^{-1} = B^{-1}A^{-1}$,
  - $(A^T)^{-1} = (A^{-1})^T$
- For orthonormal matrices
- For diagonal matrices

$$\mathbf{A}^{-1} = \mathbf{A}^\mathsf{T}$$

$$\mathbf{D}^{-1} = \mathrm{diag}\{d_1^{-1}, \ldots, d_n^{-1}\}$$

Matrix-vector multiplication

$$S = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

has eigenvalues 3, 2, 0 with corresponding eigenvectors

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

On each eigenvector, S acts as a multiple of the identity matrix: but as a different multiple on each.

Any vector (say $x = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$) can be viewed as a combination of the eigenvectors: $x = 2v_1 + 4v_2 + 6v_3$

Matrix-vector multiplication

- Thus a matrix-vector multiplication such as *Sx* (*S*, *x* as in the previous slide) can be rewritten in terms of the eigenvalues/vectors:

$$Sx = S(2v_1 + 4v_2 + 6v_3)$$

$$Sx = 2Sv_1 + 4Sv_2 + 6Sv_3 = 2\lambda_1 v_1 + 4\lambda_2 v_2 + 6\lambda_3 v_3$$

- Even though *x* is an arbitrary vector, the action of *S* on *x* is determined by the eigenvalues/vectors.
- Suggestion: the effect of “small” eigenvalues is small.

Matrix calculus

	Scalar	Vector	Matrix
Scalar	$\frac{\mathrm{d}y}{\mathrm{d}x}$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} = \left[\frac{\partial y_i}{\partial x}\right]$	$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}x} = \left[\frac{\partial y_{ij}}{\partial x}\right]$
Vector	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y}{\partial x_j}\right]$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j}\right]$	
Matrix	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{X}} = \left[\frac{\partial y}{\partial x_{ji}}\right]$		

By Thomas Minka. Old and New Matrix Algebra Useful for Statistics

Examples

$$\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T$$

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x}$$

http://matrixcookbook.com/

Eigenvalues & Eigenvectors

- A scalar λ is an eigenvalue of an $n \times n$ matrix A if there is a nonzero vector x such that $Ax = \lambda x$.
  - Vector x is called an eigenvector.
- Matrix A is symmetric positive definite (SPD) if $A^T = A$ and all its eigenvalues are positive.
- If A has n linearly independent eigenvectors, A can have the eigen-decomposition: $A = X \Lambda X^{-1}$.
  - Λ is diagonal with eigenvalues as its diagonal elements
  - Column vectors of X are corresponding eigenvectors

Eigenvalues & Eigenvectors

- **Eigenvectors** (for a square $m \times m$ matrix $\mathbf{S}$)

$$\mathbf{S}\mathbf{v} = \lambda \mathbf{v}$$

(right) eigenvector $\mathbf{v} \in \mathbb{R}^m \neq \mathbf{0}$

eigenvalue $\lambda \in \mathbb{R}$

Example

$$\begin{pmatrix} 6 & -2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

- **How many eigenvalues** are there at most?

$$\mathbf{S}\mathbf{v} = \lambda \mathbf{v} \iff (\mathbf{S} - \lambda \mathbf{I})\,\mathbf{v} = \mathbf{0}$$

only has a non-zero solution if $|\mathbf{S} - \lambda \mathbf{I}| = 0$

this is a m-th order equation in λ which can have **at most m distinct solutions** (roots of the characteristic polynomial) – <u>can be complex even though S is real.</u>

Eigenvalues & Eigenvectors

For symmetric matrices, eigenvectors for distinct eigenvalues are **orthogonal**

$$Sv_{\{1,2\}} = \lambda_{\{1,2\}} v_{\{1,2\}}, \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow v_1 \bullet v_2 = 0$$

All eigenvalues of a real symmetric matrix are **real**.

$$\text{for complex } \lambda, \text{ if } |S - \lambda I| = 0 \text{ and } \mathrm{S} = \mathrm{S}^{\mathrm{T}} \Rightarrow \lambda \in \Re$$

All eigenvalues of a positive semidefinite matrix are **non-negative**

$$\overbrace{\forall w \in \Re^n, w^T S w \geq 0}, \text{ then if } \mathrm{Sv} = \lambda \mathrm{v} \Rightarrow \lambda \geq 0$$

Example

- Let

$$S = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Real, symmetric.

- Then

$$S - \lambda I = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \Rightarrow (2-\lambda)^2 - 1 = 0.$$

- The eigenvalues are 1 and 3 (nonnegative, real).
- The eigenvectors are orthogonal (and real):

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Plug in these values and solve for eigenvectors.

Eigen/diagonal Decomposition

- Let $\mathbf{S} \in \mathbb{R}^{m \times m}$ be a **square** matrix with ***m*** **linearly independent eigenvectors** (a "non-defective" matrix)
- **Theorem**: Exists an **eigen decomposition**

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$$

diagonal

  - (cf. matrix diagonalization theorem)

Unique for distinct eigen-values

- Columns of ***U*** are **eigenvectors** of ***S***
- Diagonal elements of $\mathbf{\Lambda}$ are **eigenvalues** of $\mathbf{S}$

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m), \quad \lambda_i \geq \lambda_{i+1}$$

Diagonal decomposition: why/how

Let ***U*** have the eigenvectors as columns: $U = \begin{bmatrix} v_1 & ... & v_n \end{bmatrix}$

Then, ***SU*** can be written

$$SU = S\begin{bmatrix} v_1 & ... & v_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & ... & \lambda_n v_n \end{bmatrix} = \begin{bmatrix} v_1 & ... & v_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & ... & \\ & & \lambda_n \end{bmatrix}$$

Thus $\boldsymbol{SU=U\Lambda}$, or $\boldsymbol{U^{-1}SU=\Lambda}$

And $\boldsymbol{S=U\Lambda U^{-1}}$.

Diagonal decomposition - example

Recall $$S = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}; \lambda_1 = 1, \lambda_2 = 3.$$

The eigenvectors $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ form $$U = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Inverting, we have $$U^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

Recall $UU^{-1} = 1$.

Then, $\boldsymbol{S=U\Lambda U^{-1}} =$ $$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

Example continued

Let's divide ***U*** (and multiply ***U***$^{-1}$) by $\sqrt{2}$

Then, ***S***= $$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

U $\quad$ Λ $\quad$ ***U***^{T}

Symmetric Eigen Decomposition

- If $\mathbf{S} \in \mathbb{R}^{m\times m}$ is a **symmetric** matrix:
- **Theorem**: Exists a (unique) **eigen decomposition** $S = Q\Lambda Q^T$
- where ***Q*** is **orthogonal:**
  - $Q^{-1} = Q^T$
  - Columns of ***Q*** are normalized eigenvectors
  - Columns are orthogonal.
  - (everything is real)

Spectral decomposition

- If A is real and symmetric, all its eigenvalues are real, and there are n orthogonal eigenvectors.
- The spectral decomposition of a symmetric matrix A is $A=Q\Lambda Q^T$.
  - Λ is diagonal with eigenvalues as its diagonal elements
  - Q is orthogonal, i.e. $Q^TQ = QQ^T = I$.
  - Column vectors of Q are corresponding eigenvectors.

Exercise

- Examine the symmetric eigen decomposition, if any, for each of the following matrices:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 \\ -2 & 3 \end{bmatrix} \quad \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$$

Singular Value Decomposition (SVD)

- The singular values of an $m \times n$ A are the square roots of the eigenvalues of $A^{\mathrm{T}}A$.
- Any matrix A can have the singular value decomposition (SVD): $A=U\Sigma V^{\mathrm{T}}$.
    - Σ is diagonal with singular values as its elements.
    - U and V are orthogonal matrices.
    - The column vectors of U are called left singular vectors of A; the column vectors of V is called the right singular vector of A.

Singular Value Decomposition

For an $m \times n$ matrix $\mathbf{A}$ of rank r there exists a factorization (Singular Value Decomposition = **SVD**) as follows:

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

$m \times m$ | $m \times n$ | V is $n \times n$

The columns of ***U*** are orthogonal eigenvectors of $\boldsymbol{AA^T}$.

The columns of ***V*** are orthogonal eigenvectors of $\boldsymbol{A^TA}$.

Eigenvalues $\lambda_1 \ldots \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of $\boldsymbol{A^TA}$.

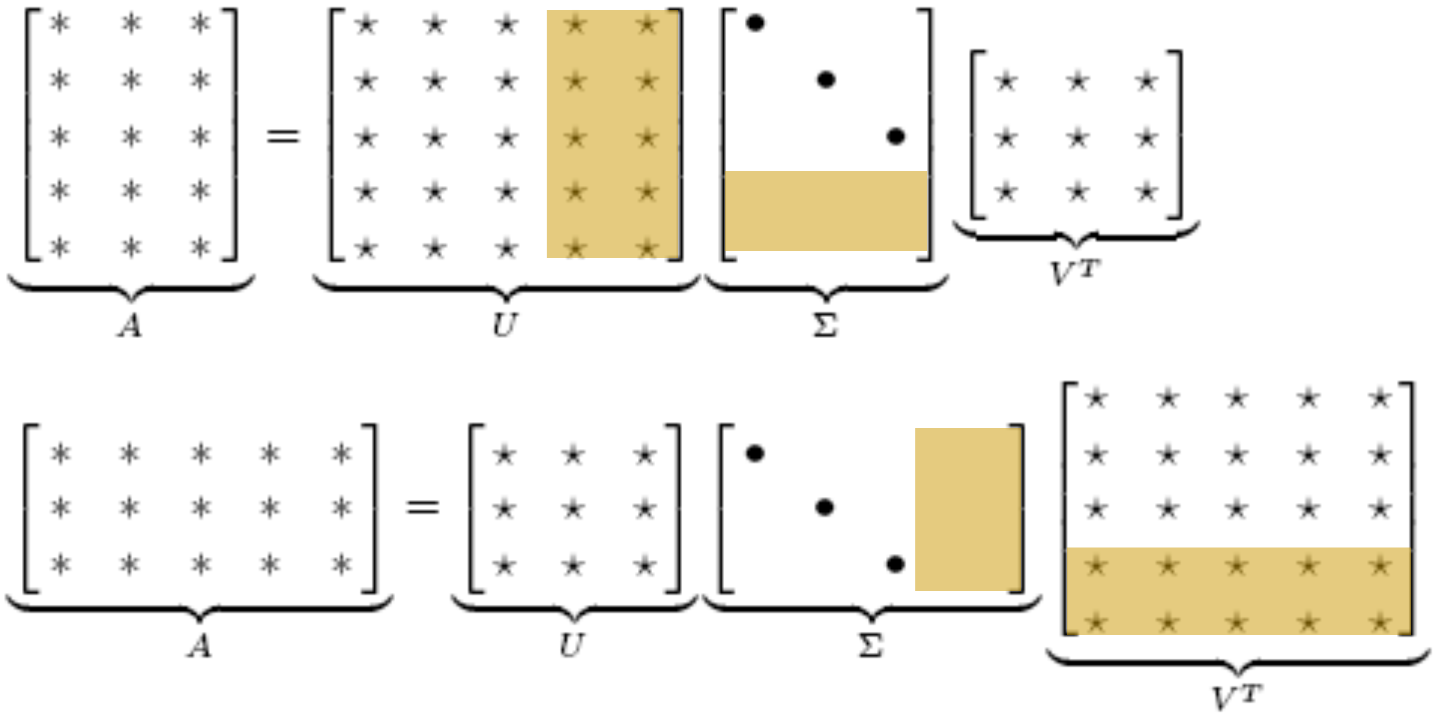
$$\sigma_i = \sqrt{\lambda_i}$$

$$\Sigma = diag(\sigma_1 \ldots \sigma_r)$$

Singular values.

Singular Value Decomposition

- Illustration of SVD dimensions and sparseness



SVD example

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

Thus m=3, n=2. Its SVD is

$$\begin{bmatrix} 0 & 2/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Typically, the singular values arranged in decreasing order.

Low-rank Approximation

- SVD can be used to compute optimal **low-rank approximations**.
- Approximation problem: Find $\boldsymbol{A_k}$ of rank $\boldsymbol{k}$ such that

$$A_k = \min_{X:rank(X)=k} \|A - X\|_F$$ *Frobenius norm*

$$\|A\|_F \equiv \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}\,.$$

A_k and X are both $m \times n$ matrices.

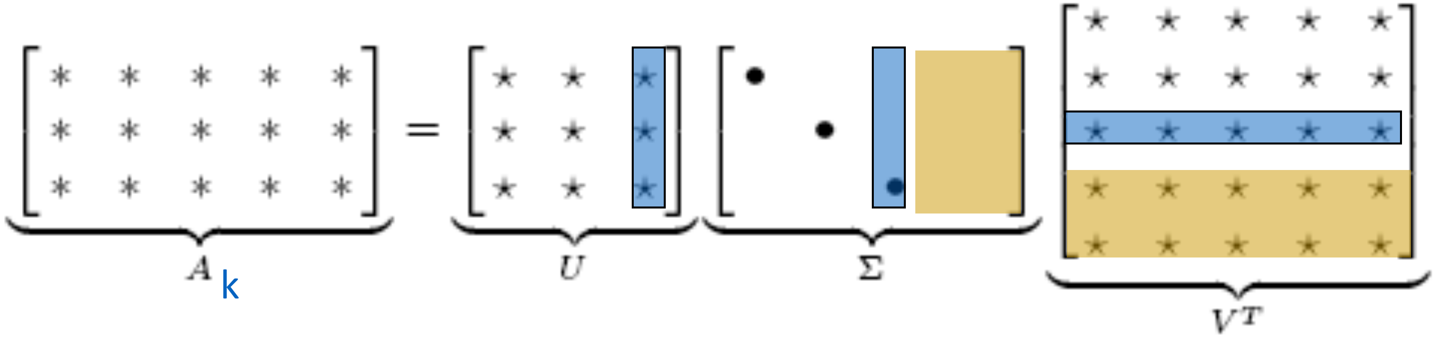
Typically, want $k << r$.

Low-rank Approximation

- Solution via SVD

$$A_k = U \,\mathrm{diag}(\sigma_1, \ldots, \sigma_k, \underbrace{0, \ldots, 0}_{})V^T$$

set smallest r-k
singular values to zero



$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

column notation: sum of rank 1 matrices

Approximation error

- How good (bad) is this approximation?
- It's the best possible, measured by the Frobenius norm of the error:

$$\min_{X:rank(X)=k} \|A - X\|_F = \|A - A_k\|_F = \sigma_{k+1}$$

where the σ_i are ordered such that $\sigma_i \geq \sigma_{i+1}$.

Suggests why Frobenius error drops as *k* increased.

C. Eckart, G. Young, *The approximation of a matrix by another of lower rank.* Psychometrika, 1, 211-218, 1936.

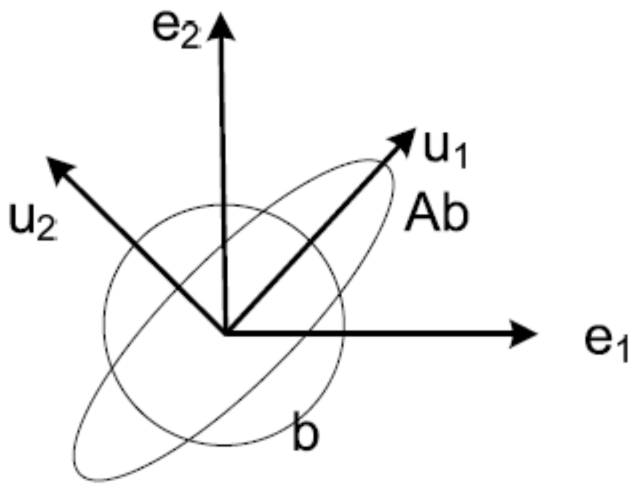
SVD Low-rank approximation

- Given a term-doc matrix A of m =50000, n=10 million (and rank close to 50000).
- We can construct an approximation A_{100} with rank 100.
  - Of all rank 100 matrices, it would have the lowest Frobenius error.
- Great … but why would we??
- Answer: *Latent Semantic Document Indexing, Movie Recommendation,…*

Eigen Value Decomposition

- Any **symmetric** matrix A can be decomposed as $A=UDU^T$, where
  - D is diagonal, with d=rank(A) non-zero elements
  - The fist d rows of U are orthogonal basis for col(A)=row(A)

- Re-interpreting Ab
  - First stretch b along the direction of u_1 by d_1 times
  - Then further stretch it along the direction of u_2 by d_2 times



U's column space R^2

Summary of SVD

- Any general matrix A can be decomposed as $A=UDV^T$, where

  - D is a diagonal matrix, with d=rank(A) non-zero elements
  - The fist d rows of U are orthogonal basis for col(A)
  - The fist d rows of V are orthogonal basis for row(A)

- Applications of the SVD
  - Matrix Pseudoinverse
  - Low-rank matrix approximation

LU decomposition

- The LU decomposition with pivoting of matrix A is $PA=LU$
  - P is a permutation matrix
  - L is lower triangular; U is upper triangular.
- The linear system $Ax=b$ can be solved by

1. Perform LU decompose $PA=LU$
2. Solve $Ly=Pb$
3. Solve $Ux=y$

Cholesky decomposition

- For an SPD matrix A, there exists the Cholesky decomposition $P^{\mathrm{T}}AP = LL^{\mathrm{T}}$
  - P is a permutation matrix
  - L is a lower triangular matrix
- If A is not SPD, the LBL decomposition can be used: $P^{\mathrm{T}}AP = LBL^{\mathrm{T}}$
  - B is a block diagonal matrix with blocks of dimension 1 or 2.

Subspaces, QR decomposition

- The null space of an $m \times n$ matrix A is

$$\mathbf{Null}(A)=\{w \mid Aw=0, w \neq 0\}$$

- The range of A is $\mathbf{Range}(A)=\{w \mid w=Av, \forall v\}$.
- Fundamental of linear algebra:

$$\mathbf{Null}(A) \oplus \mathbf{Range}(A^{\mathsf{T}}) = \mathbb{R}^n$$

- Matrix A has the QR decomposition $AP = QR$
  - P is permutation matrix; Q is an orthogonal matrix; R is an upper triangular matrix.

Recap: linear algebra

- Vectors and matrix
- Eigenvalue and eigenvector
- Singular value decomposition
- LU decomposition and Cholesky decomposition
- Subspaces and QR decomposition

End of Math Review-1

- Read the 3 Stanford University (CS229) lecture notes provide on Wattle.

- Read an old (2019) ENGN8535 lecture notes on Wattle.

Tomorrow’s tutorial (zoom based, at lecture time)

- Python Tutorial
- Numpy
- Github and GitLab (for CLab submission).
- Please start to practice Python programming. You will need it for CLab-1.

End of today’s lecture.