Review of Linear Algebra

- Vector $\underline{v} \in \mathbb{R}^n$ is a column vector $\underline{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$

  $\underline{v}^T$ is a row vector $\underline{v}^T = [v_1 \cdots v_n]$

- Matrix $A \in \mathbb{R}^{m \times n}$ has m rows and n columns

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

- Vector norm : "length" of vector

  A norm $\|\cdot\|$ on $\mathbb{R}^n$ is a mapping from $\mathbb{R}^n$ to $\mathbb{R}$ satisfying

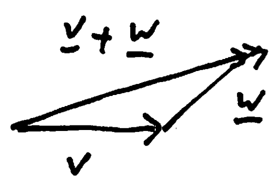
  (a) $\|\underline{v}\| \geq 0 \quad \forall \underline{x} \in \mathbb{R}^n$

  (b) $\|\alpha \underline{v}\| = |\alpha| \cdot \|\underline{v}\| \quad \forall \alpha \in \mathbb{R}$ and $\forall \underline{v} \in \mathbb{R}^n$

  (c) $\|\underline{v}\| = 0$ iff $\underline{v} = \underline{0}$

  (d) $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\| \quad \forall \underline{v}, \underline{w} \in \mathbb{R}^n$

  (triangle inequality)



Common norms:

$$\|\underline{v}\|_1 = \sum_{i=1}^{n} |v_i| \qquad l_1 \text{ norm}$$

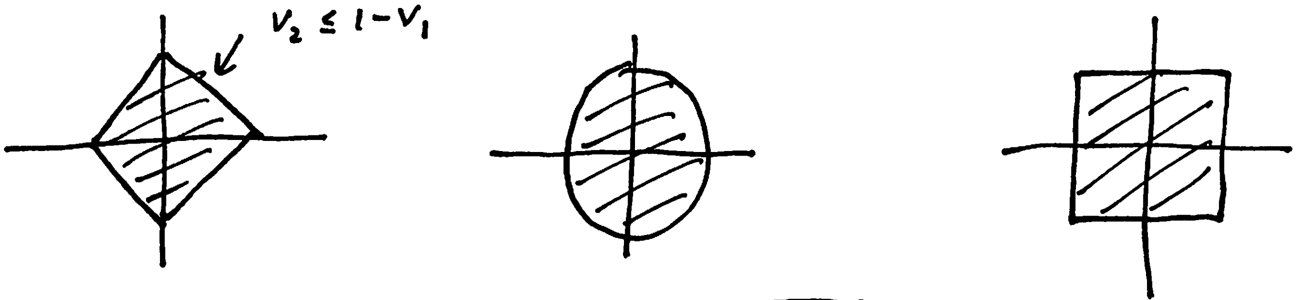
$\|\underline{v}\|_2 = \left(\sum_{i=1}^{n} v_i^2 \right)^{1/2}$ ℓ_2 norm

$\|\underline{v}\|_\infty = \max_i |v_i|$ ℓ_∞ norm

$\|\underline{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}, \quad p > 0$ ℓ_p norm

- Each norm defines a "unit sphere"

$$S_p = \{ \underline{v} \in \mathbb{R}^n : \|\underline{v}\|_p \le 1 \}$$



$\|\underline{v}_p\|_1 = |v_1| + |v_2| \le 1$ $\|\underline{v}\|_2 = \sqrt{v_1^2 + v_2^2} \le 1$ $\|\underline{v}\|_\infty \le 1$

- Inner products : measures "angle" bet. vectors

Inner product of $\underline{x}$ and $\underline{y} \in \mathbb{R}^n$: (dot product)

$$\langle \underline{x}, \underline{y} \rangle = \underline{x}^T \underline{y} = \sum_{i=1}^{n} x_i y_i = [x_1 \cdots x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}$$

$\underline{x}, \underline{y}$ are called orthogonal if $\underline{x}^T \underline{y} = 0$

If $\|\underline{x}\|_2 = \|\underline{y}\|_2 = 1$ and $\underline{x}^T \underline{y} = 0$, then $\underline{x}$ and $\underline{y}$ are called orthonormal.

- can define more general inner products $\langle \cdot, \cdot \rangle$

Induced norm: $\|\underline{v}\| = \sqrt{\langle \underline{v}, \underline{v} \rangle}$

- Outer Product $\quad \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \quad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$$\underline{x}\,\underline{y}^T \in \mathbb{R}^{m\times n} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} [y_1 \cdots y_n] = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & & & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix}$$

write as $\underline{x} \otimes \underline{y}$

- Matrix product

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

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

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

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

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

- Special matrices:

Diagonal: $\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$ upper triangular: $\begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$

tridiagonal: $\begin{bmatrix} a & b & 0 & 0 \\ c & d & e & 0 \\ 0 & f & g & h \\ 0 & 0 & i & j \end{bmatrix}$ lower triangular. $\begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}$

Identity matrix: $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- Transpose
  - flip rows and columns
  - reflecting vector/matrix along line \

$$\begin{bmatrix} a \\ b \end{bmatrix}^T = [a \ b] \qquad (A^T)^T = A$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \qquad (AB)^T = B^T A^T \qquad (A+B)^T = A^T + B^T$$

- Linear Independence

vectors $\underline{v}_1, \ldots, \underline{v}_k$ are linearly indep. if $\sum_{i=1}^{k} c_i \underline{v}_i = \underline{0}$

$\Rightarrow c_1 = \ldots = c_k = 0.$

$$\begin{bmatrix} | & | & | \\ \underline{v}_1 & \underline{v}_2 & \underline{v}_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow c_1 = c_2 = c_3 = 0$$

Linear dependent vectors:

$$\underline{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \underline{x}_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \qquad \underline{x}_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

$\underline{x}_3 = -2\underline{x}_1 + \underline{x}_2$

$$\begin{bmatrix} | & | & | \\ \underline{x}_1 & \underline{x}_2 & \underline{x}_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ for } \begin{matrix} c_1 = 2 \\ c_2 = -1 \\ c_3 = 1 \end{matrix}$$

- <u>Span</u> of set of vectors $\underline{v}_1, \ldots, \underline{v}_k$ is a subspace

$$S = \{ c_1\underline{v}_1 + c_2\underline{v}_2 + \ldots + c_k\underline{v}_k \mid c_1 \in \mathbb{R}, \ldots, c_k \in \mathbb{R} \}$$

- A <u>basis</u> $\{\underline{v}_1, \ldots, \underline{v}_k\}$ for subspace S is a set of vectors in S s.t.

  (1) $\{\underline{v}_1, \ldots, \underline{v}_k\}$ are linearly indep.

  (2) $\{\underline{v}_1, \ldots, \underline{v}_k\}$ spans S, i.e. $S = \text{span}\{\underline{v}_1, \ldots, \underline{v}_k\}$

- Every basis for S has same # of elements = <u>dimension</u> of S

e.g. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is a basis for $\mathbb{R}^3$:

$$\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

- A $m \times n$ matrix. <u>Range space</u> $R(A)$ of A is set of $\underline{y} \in \mathbb{R}^m$ s.t. $\underline{y} = A\underline{x}$ for some $\underline{x} \in \mathbb{R}^n$.

  $R(A)$ is a subspace.

  <u>rank</u> of A = $\dim(R(A))$

  rank(A) = maximum number of linearly indep. rows
  = " " " columns

  $\leq \min\{m, n\}$

  A has full rank if rank$(A) = \min(m,n)$

A has full row rank if $\text{rank}(A) = m$

A " " column " if " $= n$.

- <u>Inverse</u> of a ~~square~~ nonsingular matrix A, denoted by A^{-1}, is the <u>unique</u> matrix $(n \times n)$ s.t.

$$AA^{-1} = A^{-1}A = I$$

(nonsingular, ~~or~~ invertible has full rank or $\det(A) \neq 0$)

If A^{-1} and B^{-1} exist, then $(AB)^{-1} = B^{-1}A^{-1}$.

$(A^T)^{-1} = (A^{-1})^T$.

If A is <u>orthonormal</u>, i.e. $A^TA = I$, then $A^{-1} = A^T$

$$A^TA = \begin{bmatrix} - \underline{a}_1 - \\ \vdots \\ - \underline{a}_n - \end{bmatrix} \begin{bmatrix} | & & | \\ \underline{a}_1 & \cdots & \underline{a}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

i.e. $\underline{a}_i^T \underline{a}_i = 1 \quad i = 1.., n$

$\underline{a}_i^T \underline{a}_j = 0 \quad i \neq j$.

If D is diagonal, i.e. $D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{bmatrix} = \text{diag}\{d_1, \ldots, d_n\}$,

$$D^{-1} = \begin{bmatrix} d_1^{-1} & & 0 \\ & \ddots & \\ 0 & & d_n^{-1} \end{bmatrix}$$

Eigendecomposition

- A $n \times n$ square matrix. $\underline{v} \neq \underline{0}$, $\underline{v} \in \mathbb{C}^n$, is an eigenvector of A if $A\underline{v} = \lambda \underline{v}$, where $\lambda \in \mathbb{C}$ is the eigenvalue corresponding to $\underline{v}$.

- suppose A has n linearly indep. eigenvectors

$$[A\underline{v}_1 \cdots A\underline{v}_n] = [\lambda_1 \underline{v}_1 \cdots \lambda_n \underline{v}_n]$$

$$AV = V\Lambda \quad \text{where} \quad V = \begin{bmatrix} | & & | \\ \underline{v}_1 & \cdots & \underline{v}_n \\ | & & | \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

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

$$A^{-1} = V \Lambda^{-1} V^{-1} \qquad \Lambda^{-1} = \begin{bmatrix} \lambda_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{-1} \end{bmatrix}$$

- Fact: If A is square, symmetric ($A^T = A$), then eigenvalues of A are real and the (real) eigenvectors of A are orthogonal (can be made orthonormal) i.e.

$$A = U \Lambda U^T \quad \text{where } U \text{ is orthonormal matrix:}$$

$$= \sum_{i=1}^{n} \lambda_i \underline{u}_i \underline{u}_i^T \qquad U^T U = U U^T = I$$

- Solving for eigenvalues:

$$A\underline{v} = \lambda \underline{v}$$

or $(A - \lambda I)\underline{v} = \underline{0}$

has $\underline{v} \neq \underline{0}$ solution iff $\underbrace{\det(A - \lambda I)}_{\text{characteristic polynomial}} = 0$

- Singular Value Decomposition (SVD)

A $m \times n$ real matrix

$A = U \Sigma V^T$

U $m \times m$ orthogonal matrix i.e. $U^T U = I_{m \times m}$

Σ $m \times n$ rectangular matrix with ≥ 0 real numbers on diagonal

V $n \times n$ orthogonal matrix i.e. $V^T V = I_{n \times n}$

- diagonal entries $\sigma_i = \Sigma_{ii}$ of Σ are uniquely determined by A : singular values (real)

  \# non-zero singular values = rank(A)

- columns of V are eigenvectors of $A^T A$

  " U are " of $A A^T$

- singular values are eigenvalues of $\sqrt{}$ of eigenvalues of $A A^T$ or $A^T A$.

Applications:

- matrix pseudoinverse
- low-rank matrix approximation
- suppose order $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$

  $\sigma_{r+1} = 0 = \ldots = \sigma_{\min\{m,n\}} = 0$

Then $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

<u>Low rank matrix inversion</u>:

- In linear regression, Gaussian model, need to calculate inverse of covariance matrix X^TX (where each row of $n \times m$ matrix X is a data sample)
- If # features m >> # samples n inverting X^TX is complex: $O(m^3)$.
- Use SVD to help:

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

$$X^TX = (V\Sigma^T U^T)(U\Sigma V^T)$$

$$= V\Sigma^T\Sigma V^T$$

Now $(V(\Sigma^T\Sigma)^{-1}V^T)V\Sigma^T\Sigma V^T = I$

Thus, $(X^TX)^{-1} = V(\underbrace{\Sigma^T\Sigma}_{\text{diagonal matrix inverting is easy.}})^{-1}V^T$

Derivatives

$f : \mathbb{R}^n \to \mathbb{R}$, $\underline{x} \in \mathbb{R}^n$, ith partial derivative of f at $\underline{x}$:

$$\frac{\partial f(\underline{x})}{\partial x_i} = \lim_{\alpha \to 0} \frac{f(\underline{x} + \alpha \underline{e}_i) - f(\underline{x})}{\alpha} \quad \text{if exists}$$

$$\underline{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{ith position}$$

suppose all partial derivatives exist.

Gradient of f at $\underline{x}$:

$$\nabla f(\underline{x}) = \begin{bmatrix} \frac{\partial f(\underline{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\underline{x})}{\partial x_n} \end{bmatrix}$$

Vector-valued functions

$f : \mathbb{R}^n \to \mathbb{R}^m \qquad f = (f_1, \ldots, f_m)$

Gradient matrix of f is $n \times m$ matrix

$$\nabla f(\underline{x}) = [\, \nabla f_1(\underline{x}) \; \cdots \; \nabla f_m(\underline{x}) \,]$$

Transpose of ∇f is Jacobian of f.

Chain rule for differentiation

$f : \mathbb{R}^k \to \mathbb{R}^m$, $\quad g : \mathbb{R}^m \to \mathbb{R}^n$ differentiable fns

$h(\underline{x}) = g(f(\underline{x}))$ composition $\quad h : \mathbb{R}^k \to \mathbb{R}^n$

Chain Rule: $\underbrace{\nabla h(\underline{x})}_{k \times n} = \underbrace{\nabla f(\underline{x})}_{k \times m} \underbrace{\nabla g(f(\underline{x}))}_{m \times n}$