A linear Algebra Review

Taken largely from a chapter written by Chung-Ming Kuan and published online in 2002

Basics

A matrix is an array of numbers

$$A = \begin{array}{r} \\ \\ \\ m \text{ rows} \end{array} \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{bmatrix}}_{n \text{ columns}}$$

a_{ij} is the entry in the i^{th} row and j^{th} column

an $n \times 1$ matrix is a column vector
an $1 \times n$ matrix is a row vector
For a matrix A, $\vec{a}_i$ is its i^{th} column

definitions

- a <u>square matrix</u> has an equal number of rows & columns.
- a <u>diagonal matrix</u> is all zeros except for the diagonal elements
- a diagonal matrix whose diagonal elements are all 1 is an <u>identity matrix</u> usually denoted I, I_n for the n-dimensional identity
- $\mathbf{0}$ is the matrix of all 0's

If f is a vector valued function $f : \mathbb{R}^m \to \mathbb{R}^n$, $\nabla_\theta f(\theta)$ is an $m \times n$ matrix of the first derivates of f w.r.t. the elements of θ

$$\nabla_\theta f(\theta) = \begin{bmatrix} \frac{\partial f_1(\theta)}{\partial \theta_1} & \frac{\partial f_2(\theta)}{\partial \theta_1} & \cdots & \frac{\partial f_n(\theta)}{\partial \theta_1} \\ \\ \frac{\partial f_1(\theta)}{\partial \theta_m} & \cdots & & \frac{\partial f_n(\theta)}{\partial \theta_m} \end{bmatrix}$$

if $n = 1$ then the gradient of f is a column vector. Also, if n is 1 the $m \times m$ Hessian matrix is

$$\nabla_\theta^2 f(\theta) = \nabla_\theta\left(\nabla_\theta f(\theta)\right) = \begin{bmatrix} \frac{\partial^2 f(\theta)}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 f(\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 f(\theta)}{\partial \theta_1 \partial \theta_m} \\ \\ \frac{\partial^2 f(\theta)}{\partial \theta_m \partial \theta_1} & & & \frac{\partial^2 f(\theta)}{\partial \theta_m \partial \theta_m} \end{bmatrix}$$

---

2 matrices are of equal size if they have the same # of rows & columns. Matrix equality is defined for matrices of the same size. (obvious)

The transpose of a matrix A is denoted A' or A^T

the i,j^{th} element of A^T is the j,i^{th} element of A

A matrix whose transpose is equal to itself is symmetric.

Two matrices of the same size can be added

$$C = A + B = B + A$$

Obviously $A + 0 = A$ and $(A+B)+C = A + (B+C)$

$(AB)' = B'A'$

proof

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

$$= \sum_{k=1}^{n} b'_{ik} a'_{kj}$$

$$\begin{array}{cc} & j \\ i & \begin{bmatrix} C \end{bmatrix} \end{array} = \begin{array}{cc} & k \\ i & \begin{bmatrix} B' \end{bmatrix} \end{array} \begin{array}{cc} & j \\ k & \begin{bmatrix} A' \end{bmatrix} \end{array}$$

$$= \begin{array}{cc} & k \\ i & \begin{bmatrix} A \end{bmatrix} \end{array} \begin{array}{cc} & j \\ k & \begin{bmatrix} B \end{bmatrix} \end{array}$$

$b'_{ik} = b_{kj}$ and $a'_{kj} = a_{ik}$ ■

A scalar c times A is written
$$cA = \{c a_{ij}\}_{ij}$$
obviously $cA = Ac$ and $-1A = -A$ and $A - A = 0$

Matrix multiplication

AB is only defined if the number of columns of A is the same as the number of rows of B

$$\begin{array}{cccc} & n & & p \\ m & \left[\; A \;\right] & n & \left[\; B \;\right] \end{array} = \begin{array}{cc} & p \\ m & \left[\; C \;\right] \end{array}$$

where
$$c_{mp} = \sum_{i=1}^{n} a_{mi} \cdot b_{ip}$$

Obviously $AB \neq BA$ (in general), however

$$A(BC) = (AB)C \leftarrow \text{associative}$$
$$A(B+C) = AB + AC \leftarrow \text{commutative}$$

One can verify $(AB)' = B'A'$

Inner product of vectors

If $\vec{y}$ and $\vec{z}$ are d-dim vectors their <u>inner product</u> is
$$\vec{y}'\vec{z} = \sum_{i=1}^{d} y_i z_i$$

If $\vec{y}$ is m-dim and $\vec{z}$ is n-dim the outer <u>product</u> of them is

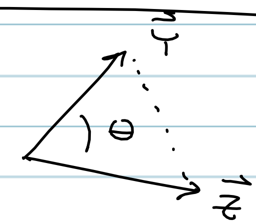
$\vec{y}\vec{z}'$ is a matrix with elements $\{y_i z_j\}_{i,j}$, $1 \leq i \leq m$, $1 \leq j \leq n$

The Euclidian norm of a vector $\vec{z}$ (it's "length") is

$$\|\vec{z}\| = (\vec{z}'\vec{z})^{1/2} = \left(\sum_{i=1}^{J} z_i^2\right)^{1/2}$$

$\vec{0}$ has Euclidian norm 0, a <u>unit vector</u> has Euclidian norm 1. Examples?

<u>orthogonality</u>



<u>Law of cosine</u>

$$\|\vec{y}-\vec{z}\|^2 = \|\vec{y}\|^2 + \|\vec{z}\|^2 - 2\|\vec{y}\|\|\vec{z}\| \cos\theta$$

expanding

$$\|(\vec{y}-\vec{z})'(\vec{y}-\vec{z})\| = \vec{y}'\vec{y} + \vec{z}'\vec{z} - 2\vec{y}'\vec{z}$$
$$= \|\vec{y}\|^2 + \|\vec{z}\|^2 - 2\vec{y}'\vec{z}$$

$$\Rightarrow \quad \vec{y}'\vec{z} = \|\vec{y}\|\,\|\vec{z}\| \cos\theta$$

which says that is $\theta = \frac{\pi}{2}$ (90°) then $\vec{y}'\vec{z} = 0$. In this case $\vec{y}$ and $\vec{z}$ are said to be <u>orthogonal</u>.

A square matrix A is orthogonal if

$$A'A = AA' = I$$

i.e. each row (and column) of A is a unit vector and orthogonal to the others

If $\vec{y} = c\vec{z}$ for some $c \neq 0$
$\vec{y}$ and $\vec{z}$ are said to be linearly dependent.

Consider differentiation w.r.t. vectors & matrices

let $\vec{a}$ and $\vec{\theta}$ be two D-dim vectors
then

$$\nabla_\theta (a'\theta) = a$$

Check: $\nabla_\theta (a'\theta) = \left[\frac{\partial a'\theta}{\partial \theta_1}, \frac{\partial a'\theta}{\partial \theta_2}, \cdots, \frac{\partial a'\theta}{\partial \theta_d} \right]$

$$= [\, a_1, a_2, \cdots, a_d \,] = a$$

$\nabla_\theta (\theta' A \theta) = 2A\theta$ where A is a $D \times D$ matrix

Check:

$$\nabla_\theta (\theta' A \theta) = \left[\frac{\partial \theta' A \theta}{\partial \theta_1}, \frac{\partial \theta' A \theta}{\partial \theta_2}, \cdots \frac{\partial \theta' A \theta}{\partial \theta_d} \right]$$

note $\theta' A \theta$ looks like

$$[\theta_1, \cdots, \theta_d] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ & & & \\ a_{d1} & \cdots & & a_{dd} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}$$

$$= [\theta_1, \cdots, \theta_d] \begin{bmatrix} \theta_1 a_{11} + \theta_2 a_{12} + \cdots \ \theta_d a_{1d} \\ \vdots \quad \ddots \quad \vdots \\ \theta_1 a_{d1} + \cdots + \theta_d a_{dd} \end{bmatrix}$$

$$= \begin{bmatrix} \theta_1^2 a_{11} + \theta_1 \theta_2 a_{12} + \theta_1 \theta_3 a_{13} + \cdots \ \theta_1 \theta_d a_{1d}, \\ \theta_2 \theta_1 a_{21} + \theta_2^2 a_{22} + \theta_2 \theta_3 a_{23} + \cdots \ \theta_2 \theta_d a_{2d}, \\ \vdots \\ \theta_d \theta_1 a_{d1} + \cdots \qquad \theta_d^2 a_{dd} \end{bmatrix}$$

Taking the derivative of this w.r.t. to θ_i repeatedly $\forall i$ yields

$$\frac{\partial \theta' A \theta}{\partial \theta_i} = [0, \cdots \quad 2\theta_i a_{ii} \quad \cdots, 0]$$

So ~~$\partial\theta' A \theta$~~ $\quad \nabla_\theta \theta' A \theta = 2A\theta$

Combining these two rules yields

$$\nabla_\theta^2 (\theta' A \theta) = 2A$$

Matrix Determinant

Let A be a square Matrix, and let A_{ij} be the sub-matrix of A after deleting row i and column j. The determinant of A is

$$|A| = \det(A) = \sum_{i=1}^{m} (-1)^{i+j} a_{ij} \det(A_{ij})$$

or equivalently $\qquad$ base case $\overset{2\times 2}{|A|} = a_{11}a_{22} - a_{12}a_{21}$

$$|A| = \det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

A square matrix with a non-zero determinant is called nonsingular, otherwise it is called singular

Fact: $\det(A) = \det(A')$

Sketch proof: base case is invariant to transpose, Since det is equivalent in both row and column expansion, 3×3 det. is clearly invariant to transpose. Induction from there.

Another super useful fact:

$$|cA| = c^n |A|$$

Pf: again, base case is 2×2 $|A| = ca_{11}ca_{22} - ca_{12}ca_{21}$
$= c^2 |A|$

from definition 3×3 will be $c^3|A|$, induction.

Without proof:

$$|AB| = |A||B| = |BA|$$

Many facts at once: if A is orthogonal then $AA' = I$. From definition of det it is easy to see that $|I| = 1$, so

$$|I| = |AA'| = (|A|)^2 \Rightarrow |A| = 1 \text{ or } -1$$

ie. orthogonal matrices must have det. 1 or -1.

Trace

The trace of a matrix is the sum of it's diagonal elements.

$$\text{trace}(I_n) = n$$
$$\text{trace}(A) = \text{trace}(A')$$
$$\text{trace}(cA + dB) = c\,\text{trace}(A) + d\,\text{trace}(B)$$

without proof:

$$\text{trace}(AB) = \text{trace}(BA)$$

if both products well defined

Matrix Inverses

A nonsingular matrix A has a unique inverse A^{-1} s.t. $AA^{-1} = A^{-1}A = I$.

Matrix inversion and transposition can be interchanged, i.e.

$$(A')^{-1} = (A^{-1})'$$

Because

$$ABB^{-1}A^{-1} = I$$ with A & B nonsingular and compatible

we have

$$(AB)^{-1} = B^{-1}A^{-1}$$

Most matrix inverses must be computed but some are easy. If A is diagonal then A^{-1} is diagonal with diagonal elements that are the reciprocal of the original. If A is orthogonal then because $AA' = I$ $A^{-1} = A'$.

Partitioned matrices can be inverted easily sometimes.

Matrix Rank (important in lin. alg.)

A set of vectors $\vec{z}_1, \ldots, \vec{z}_n$ are linearly independent if $c_1, c_2, \cdots, c_n = 0$ is the only solution to

$$c_1\vec{z}_1 + c_2\vec{z}_2 + \cdots c_n\vec{z}_n = 0$$

otherwise they are linearly dependent. When 2 vectors are linearly dependent they lie on the same line; 3, plane or line; 4, plane, line, or volume.

Claim: The column rank and row rank of a matrix are equal.

Pf. The column rank of a matrix A is the max. number of linearly independent column vectors of A (row rank def. the same). If the col or row rank of A equals the corresponding dimensionality then A is said to be of full column or row rank.

The space spanned by a set of vectors $\{\vec{z}_1, \dots, \vec{z}_n\}$ is the collection of all linear combinations of those vectors and is denoted $\text{span}\{\vec{z}_1, \dots, \vec{z}_n\}$

The space spanned by the column vectors of A is $\text{span}(A)$ and is called the column space of A. $\text{span}(A')$ is the row space of A.

Let A be an $n \times k$ matrix with $k \leq n$ and suppose $r = \text{rowrank}(A) \leq n$ and

$c = \text{columnrank}(A) \leq k$

Assume w.l.o.g. the first r rows of A are lin independent then all rows of A can be expressed as

$$\vec{a}_i = q_{i1}\vec{a}_1 + q_{i2}\vec{a}_2 + \dots + q_{ir}\vec{a}_r$$

where the j^{th} element of $\vec{a}_i$ is

$$a_{ij} = q_{i1}a_{1j} + q_{i2}a_{2j} + \dots + q_{ir}a_{rj}$$

but from this it is clear that any <u>column</u> vector of A can be written as the linear combination of the r vectors

$$\vec{a}_j = \vec{q}_1 c_1 + \vec{q}_2 c_2 + \dots \vec{q}_r c_r$$

This means the column rank of A must also be less than or equal to r. The same arg. can be applied to the transpose of A, ie. starting with the column vectors yielding the result that the row rank of A must be less or equal to c.

$\Rightarrow$ rowrank (A) = columnrank(A) = rank(A)

Clearly: $\underline{\text{rank}(A) = \text{rank}(A')}$

if $\text{rank}(A) = n$ and A is $n \times n$ then A is <u>full rank</u>.

<u>Fact</u>: A full rank matrix is non singular and vice versa.

It can be shown that:

$$\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$$
$$\text{rank}(AB) \leq \min[\text{rank}(A), \text{rank}(B)]$$

Using these we have, when <u>A is non-singular</u>)

$$\text{rank}(AB) \leq \text{rank}(B) = \text{rank}(A^{-1}AB) \leq \text{rank}(AB)$$

$\Rightarrow$ $\text{rank}(AB) = \text{rank}(B)$ when A non-singular

the other direction works as well

$\text{rank}(BC) = \text{rank}(B)$ when C non-singular

Eigen values & Eigen vectors

Given a square matrix A if

$$A\vec{c} = \lambda\vec{c}$$

for some scalar λ and non-zero vector $\vec{c}$
then
$\vec{c}$ is an eigenvector of A corresponding to
λ an eigen value

Eigen vectors and eigen-values are particularly interpretable when A is a rotation or reflection matrix, the eigen vectors then are the axes of rotation and the eigen valuesign indicate reflection through some space

Given an eigen-value λ, let $\vec{c}_1, \cdots, \vec{c}_k$ be associated eigen vectors. Then,

$$A(a_1\vec{c}_1 + a_2\vec{c}_2 + \cdots + a_k\vec{c}_k) = \lambda(a_1\vec{c}_1 + a_2\vec{c}_2 + \cdots + a_k\vec{c}_k)$$

so any linear combination of the $\vec{c}$'s is again an eigen vector and the set of all such vectors is an eigen space associated with eigen-value λ.

If A $(n \times n)$ has n distinct eigenvalues, each eigenvalue must correspond to one eigen vector and the set of such eigen vectors must be linearly independent.

Since the choice of eigenvector is identifiable only up to a constant, often unit length eigenvectors are chosen.

$$E\left[(x-\mu)^T (x-\mu) \right]$$

Let C be the matrix consisting of these n, distinct, unit length, linearly independent eigenvectors. Clearly C is non-singular. That means we can write full rank

$$AC = C\Lambda$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \lambda_3 & \\ 0 & & & \ddots \quad \lambda_n \end{bmatrix}$$

and using the non-singularity of C we have

$$C^{-1}AC = \Lambda \quad \text{or} \quad \boxed{A = C\Lambda C^{-1}}$$

important, SVD

Again, when A has n distinct eigenvalues

$$\det(A) = \det(C\Lambda C^{-1}) = \det(\Lambda)\det(C)\det(C^{-1})$$
$$= \det(\Lambda) = \prod_{i=1}^{n} \lambda_i$$

$$\operatorname{trace}(A) = \operatorname{trace}(C\Lambda C^{-1}) = \operatorname{trace}(C^{-1}C\Lambda)$$
$$= \operatorname{trace}(\Lambda) = \sum_{i=1}^{n} \lambda_i$$

When $A = C\Lambda C^{-1}$ we have $A^{-1} = C\Lambda^{-1}C^{-1}$

so the eigenvectors of A^{-1} are the same as those of A. the eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A.

~~Because non-singular transformations preserve rank we have~~ too soon

$$\cancel{\boxed{\text{rank}(A) = \text{rank}(C\Lambda C') = \text{rank}(\Lambda) = \#\text{ non-zero eigenvalues of } A}}$$

Symmetric Matrices

Let $\vec{c}_1$ and $\vec{c}_2$ be two eigenvectors of A corresponding to distinct eigenvalues λ_1 & λ_2

If A is symmetric

$$\vec{c}_2{}' A \vec{c}_1 = \lambda_1 \vec{c}_2{}' \vec{c}_1 = \lambda_2 \vec{c}_2{}' \vec{c}_1$$

which because $\lambda_1 \neq \lambda_2 \Rightarrow \vec{c}_2{}' \vec{c}_1 = 0$

A symmetric matrix is <u>orthogonally diagonalizable</u> such that

$$C'AC = \Lambda \quad \text{or} \quad A = C\Lambda C'$$

where Λ is a diagonal matrix of eigenvalues and C is the orthogonal matrix of associated eigenvectors.

<u>Remember</u> non-singular transforms preserve rank so.

For a symmetric matrix A, $\text{rank}(\Lambda) = \text{rank}(A)$, i.e. the # of non-zero eigenvalues of A.

Lemma Let A be an $n \times n$ symmetric matrix. Then

$$\det(A) = \det(\Lambda) = \prod_{i=1}^{n} \lambda_i,$$

$$\operatorname{trace}(A) = \operatorname{trace}(\Lambda) = \sum_{i=1}^{n} \lambda_i$$

Obviously, a symmetric matrix is non-singular if all its eigen values are non-zero.

A symmetric matrix A is said to be <u>positive definite</u> if $\vec{b}'A\vec{b} > 0 \quad \forall \ \vec{b} \neq \vec{0}$. Pos. <u>semi-def.</u> if $\vec{b}'A\vec{b} \geq 0$

if A is pos. def $\Rightarrow$ A non-singular

A pos semi. def. $\not\Rightarrow$ " " , A may be sing.

If A is symmetric and orthogonally diagonalized as

$$C'AC = \Lambda$$

<u>and</u> if A is p.s.d. then for $\vec{b} \neq 0$

$$b'\Lambda b = b'(C'AC)b = \tilde{b}'A\hat{b} \geq 0$$

where $\tilde{b} = Cb$. This shows that Λ is also p.s.d. and Λ's diagonal elements must be positive.

<u>Lemma</u> A symmetric matrix is P(S)D iff. it's eigenvalues are all positive (non-negative)

For a symmetric pos. def. matrix A, $A^{-1/2}$ is such that $A^{-1/2\prime} A^{-1/2} = A^{-1}$. This can be arrived at via

$$A^{-1} = C \Lambda^{-1} C' = (C \Lambda^{-1/2} C')(C \Lambda^{-1/2} C')$$

so we may choose $A^{-1/2} = C \Lambda^{-1/2} C'$

Orthogonal Projection

A matrix A is idempotent if $A^2 = A$. Given a vector $\vec{y}$ in Euclidian space V a projection of $\vec{y}$ onto a subspace S of V is a linear transformation of $\vec{y}$ to S. The projection can be written $P\vec{y}$ where P is a transformation matrix. Projecting a projection should not effect the projection, i.e.

$$P(P\vec{y}) = P^2\vec{y} = P\vec{y}$$

The matrix P is called a projection matrix if it is idempotent.

A projection of $\vec{y}$ onto S is orthogonal if $P\vec{y}$ is orthogonal to the difference between $\vec{y}$ and $P\vec{y}$. Algebraically

$$\begin{aligned}(\vec{y} - P\vec{y})' P\vec{y} &= ((I-P)\vec{y})' P\vec{y} \\ &= \vec{y}'(I-P)'P\vec{y}\end{aligned}$$

This can only be zero if $(I-P)'P = 0$. This can only happen if $P = P'P$. This shows that P must be symmetric.

Conclusion: P is an orthogonal projection matrix iff. P is symmetric and idempotent.

If P is an orthogonal projection matrix, it can clearly be seen that $I-P$ is idempotent

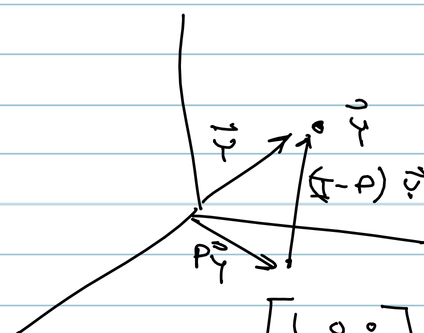
$$(I-P)(I-P) = I - 2P - PP = I - P$$

Since $I-P$ is symmetric it is also an orthogonal projection matrix.

Since $(I-P)P = 0$ the projections Py and $(I-P)y$ must be orthogonal.

So, any vector $\vec{y}$ can be uniquely decomposed into two orthogonal components:

$$\vec{y} = P\vec{y} + (I-P)\vec{y}$$



$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$I - P = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

If A is symmetric & idempotent and C is orthogonal Then

$$\Lambda = C'AC = C'ACC'AC = \Lambda^2$$

This can only happen if the entries of Λ are 0 or 1

This implies ~~Since~~ that the rank of $A'A$ and AA' and A are all the same, ie.

$$\text{rank}(A) = \text{rank}(A'A) = \text{rank}(AA')$$

If A is full column rank $\Rightarrow$ $\text{rank}(A'A) = k$ w/ k <u>pos.</u> eigv's

Facts:- A symmetric & idempotent matrix is positive semi-definite with eigenvalues 0 & 1

- Trace(Λ) is the # of non-zero eigenvalues of A and hence rank(Λ) = trace(Λ).

Remember: if A is symmetric rank(A) = rank(Λ)
and trace(A) = trace(Λ)

Combining these two yields:

Fact: For a symmetric & idempotent matrix A, rank(A) = trace(A) the number of non-zero eigenvalues of A.

Let A be an $n \times k$ matrix. Clearly $A'A$ and AA' are symmetric

If A is full column rank $k<n$,
$P = A(A'A)^{-1}A'$ is symmetric & idempotent and as such an orthogonal projection matrix.
As

$$\text{trace}(P) = \text{trace}(A'A(A'A)^{-1}) = \text{trace}(I_k) = k$$

we see that P as k eigenvalues equal to 1 and as such $\text{rank}(P) = k$. Similarly, $\text{rank}(I-P) = n-k$

↑ homework