Linear Algebra Review

Notation:

- Vectors $x \in \mathbb{R}^n$ are column vectors
$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- An $m \times n$ matrix A has m rows, n columns and entries A_{ij}.

- Matrix and vector addition defined componentwise:
$$(A+B)_{ij} = A_{ij} + B_{ij}; \qquad (x+y)_i = x_i + y_i$$

- If A is $m \times n$ and B is $n \times r$ then AB is the $m \times r$ matrix
$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

- The matrix I or sometimes $I_{n\times n}$ which is an $n \times n$ matrix with $I_{ii} = 1$ for all i and $I_{ij} = 0$ for any pair $i \neq j$ is called the $n \times n$ **identity matrix**.

- The **span** of a set of vectors $\{x_1, \ldots, x_p\}$ is the set of all vectors x of the form $x = \sum c_i x_i$. It is a vector space. The **column space** of a matrix, A, is the span of the set of columns of A. The **row space** is the span of the set of rows.

- A set of vectors $\{x_1, \ldots, x_p\}$ is **linearly independent** if $\sum c_i x_i = 0$ implies $c_i = 0$ for all i. The **dimension** of a vector space is the cardinality of the largest possible set of linearly independent vectors.

Defn: The **transpose**, A^T, of an $m \times n$ matrix A is the $n \times m$ matrix whose entries are given by

$$(A^T)_{ij} = A_{ji}$$

so that A^T is $n \times m$. We have

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

and

$$(AB)^T = B^T A^T .$$

Defn: **rank** of matrix A, rank(A): # of linearly independent columns of A. We have

$$\begin{aligned}\text{rank}(A) &= \dim(\text{column space of } A)\\ &= \dim(\text{row space of } A)\\ &= \text{rank}(A^T)\end{aligned}$$

If A is $m \times n$ then $\text{rank}(A) \le \min(m, n)$.

Matrix inverses

For now: all matrices square $n \times n$.

If there is a matrix B such that $BA = I_{n \times n}$ then we call B the inverse of A. If B exists it is unique and $AB = I$ and we write $B = A^{-1}$. The matrix A has an inverse if and only if rank$(A) = n$.

Inverses have the following properties:

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

(if one side exists then so does the other) and

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

Determinants

Again A is $n \times n$. The determinant if a function on the set of $n \times n$ matrices such that:

1. $\det(I) = 1$.

2. If A' is the matrix A with two columns interchanged then

   $$\det(A') = -\det(A)\,.$$

   (So: two equal columns implies $\det(A) = 0$.)

3. $\det(A)$ is a linear function of each column of A. If $A = (a_1, \ldots, a_n)$ with a_i denoting the ith column of the matrix then

   $$\begin{aligned}\det(a_1, \ldots, &a_i + b_i, \ldots, a_n) \\ &= \det(a_1, \ldots, a_i, \ldots, a_n) \\ &\quad + \det(a_1, \ldots, b_i, \ldots, a_n)\end{aligned}$$

Here are some properties of the determinant:

1. det(A^T) = det(A).

2. det(AB) = det(A)det(B).

3. det(A^{-1})=1/det(A).

4. A is invertible if and only if det(A) $\neq 0$ if and only if rank(A) $= n$.

5. Determinants can be computed (slowly) by expansion by minors.

Special Kinds of Matrices

1. A is symmetric if $A^T = A$.

2. A is orthogonal if $A^T = A^{-1}$ (or $AA^T = A^T A = I$).

3. A is idempotent if $AA \equiv A^2 = A$.

4. A is diagonal if $i \neq j$ implies $A_{ij} = 0$.

Inner Products, orthogonal, orthonormal vectors

Defn: Two vectors x and y are **orthogonal** if $x^T y = \sum x_i y_i = 0$.

Defn: The **inner product** or **dot product** of x and y is

$$< x, y >= x^T y = \sum x_i y_i$$

Defn: x and y are **orthogonal** if $x^T y = 0$.

Defn: The **norm** (or length) of x is $||x|| = (x^T x)^{1/2} = (\sum x_i^2)^{1/2}$

A is orthogonal if each column of A has length 1 and is orthogonal to each other column of A.

Quadratic Forms

Suppose A is an $n \times n$ matrix. The function

$$g(x) = x^T A x$$

is called a quadratic form. Now

$$\begin{aligned} g(x) &= \sum_{ij} A_{ij} x_i x_j \\ &= \sum_i A_{ii} x_i^2 + \sum_{i<j} (A_{ij} + A_{ji}) x_i x_j \end{aligned}$$

so that $g(x)$ depends only on the total $A_{ij}+A_{ji}$. In fact

$$x^T A x = x^T A^T x = x^T \left(\frac{A + A^T}{2} \right) x$$

Thus we will assume that A is symmetric.

Eigenvalues and eigenvectors

If A is $n \times n$ and $v \neq 0 \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ such that

$$Av = \lambda v$$

then λ is **eigenvalue** (or characteristic or latent value) of A; v is corresponding **eigenvector**. Since $Av - \lambda v = (A - \lambda I)v = 0$ matrix $A - \lambda I$ is singular.

Therefore $\det(A - \lambda I) = 0$.

Conversely: if $A - \lambda I$ singular then there is $v \neq 0$ such that $(A - \lambda I)v = 0$.

Fact: $\det(A - \lambda I)$ is polynomial in λ of degree n.

Each root is an eigenvalue.

General A the roots could be multiple roots or complex valued.

Diagonalization

Matrix A is **diagonalized** by a non-singular matrix P if $P^{-1}AP \equiv D$ is diagonal.

If so then $AP = PD$ so each column of P is eigenvector of A with the ith column having eigenvalue D_{ii}.

Thus to be diagonalizable A must have n linearly independent eigenvectors.

Symmetric Matrices

If A is symmetric then

1. Every eigenvalue of A is real (not complex).

2. A is diagonalizable; columns of P may be taken unit length, mutually orthogonal: A is diagonalizable by an orthogonal matrix P; in symbols $P^TAP = D$.

3. Diagonal entries in $D =$ eigenvalues of A.

4. If $\lambda_1 \neq \lambda_2$ are two eigenvalues of A and v_1 and v_2 are corresponding eigenvectors then
$$v_1^T A v_2 = v_1^T \lambda_2 v_2 = \lambda_2 v_1^T v_2$$
and
$$\begin{aligned}(v_1^T A v_2) = (v_1^T A v_2)^T &= v_2^T A^T v_1 \\ &= v_2^T A v_1 = v_2^T \lambda_1 v_1 = \lambda_1 v_2^T v_1\end{aligned}$$
Since $(\lambda_1 - \lambda_2) v_1^T v_2 = 0$ and $\lambda_1 \neq \lambda_2$ we see $v_1^T v_2 = 0$. Eigenvectors corresponding to distinct eigenvalues are orthogonal.

Positive Definite Matrices

Defn: A symmetric matrix $\mathbf{A}$ is non-negative definite if $x^T\mathbf{A}x \ge 0$ for all x. It is positive definite if in addition $x^T\mathbf{A}x = 0$ implies $x = 0$.

$\mathbf{A}$ is non-negative definite iff all its eigenvalues are non-negative.

$\mathbf{A}$ is positive definite iff all eigenvalues positive.

A non-negative definite matrix has a symmetric non-negative definite square root. If

$$\mathbf{PDP}^T = \mathbf{A}$$

for $\mathbf{P}$ orthogonal and $\mathbf{D}$ diagonal then

$$\mathbf{A}^{1/2} = \mathbf{PD}^{1/2}\mathbf{P}^T$$

is symmetric, non-negative definite and

$$\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$$

Here $\mathbf{D}^{1/2}$ is diagonal with

$$(\mathbf{D}^{1/2})_{ii} = (\mathbf{D}_{ii})^{1/2}.$$

Many other square roots possible. If $\mathbf{AA}^T = \mathbf{M}$ and $\mathbf{P}$ is orthogonal and $\mathbf{A}^* = \mathbf{AP}$ then $\mathbf{A}^*(\mathbf{A}^*)^T = \mathbf{M}$.

Orthogonal Projections

Suppose S vector subspace of $\mathbb{R}^n$, $a_1, \ldots, a_m$ basis for S. Given any $x \in \mathbb{R}^n$ there is a unique $y \in S$ which is closest to x; y minimizes

$$(x-y)^T(x-y)$$

over $y \in S$. Any y in S is of the form

$$y = c_1 a_1 + \cdots + c_m a_m = Ac$$

A, $n \times m$, columns $a_1, \ldots, a_m$; c column with ith entry c_i. Define

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

(A has rank m so $A^T A$ is invertible.) Then

$$\begin{aligned}
&(x-Ac)^T(x-Ac) \\
&\quad = (x - Qx + Qx - Ac)^T(x - Qx + Qx - Ac) \\
&\quad = (x-Qx)^T(x-Qx) + (Qx-Ac)^T(x-Qx) \\
&\qquad + (x-Qx)^T(Qx-Ac) \\
&\qquad + (Qx-Ac)^T(Qx-Ac)
\end{aligned}$$

Note that $x - Qx = (I - Q)x$ and that

$$QAc = A(A^T A)^{-1} A^T Ac = Ac$$

so that

$$Qx - Ac = Q(x - Ac)$$

Then

$$(Qx - Ac)^T (x - Qx) = (x - Ac)^T Q^T (I - Q)x$$

Since $Q^T = Q$ we see that

$$\begin{aligned} Q^T(I - Q) &= Q(I - Q) \\ &= Q - Q^2 \\ &= Q - A(A^T A)^{-1} A^T A(A^T A)^{-1} A^T \\ &= Q - Q = 0 \end{aligned}$$

This shows that

$$\begin{aligned} (x - Ac)^T (x - Ac) =& (x - Qx)^T (x - Qx) \\ &+ (Qx - Ac)^T (Qx - Ac) \end{aligned}$$

Choose Ac to minimize: minimize second term.

Achieved by making $Qx = Ac$.

Since $Qx = A(A^T A)^{-1} A^T x$ can take

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

Summary: closest point y in S is

$$y = Qx = A(A^T A)^{-1} A^T x$$

call y the orthogonal projection of x onto S.

Notice that the matrix Q is idempotent:

$$Q^2 = Q$$

We call Qx the orthogonal projection of x on S because Qx is perpendicular to the residual $x - Qx = (I - Q)x$.

Partitioned Matrices

Suppose A_{11} $p \times r$ matrix, $A_{1,2}$ $p \times s$, $A_{2,1}$ $q \times r$ and $A_{2,2}$ $q \times s$. Make $(p+q) \times (r+s)$ matrix by putting A_{ij} in 2 by 2 matrix:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

For instance if

$$A_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A_{12} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$A_{21} = \begin{bmatrix} 4 & 5 \end{bmatrix}$$

and

$$A_{22} = [6]$$

then

$$A = \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \\ \hline 4 & 5 & 6 \end{array}\right]$$

Lines indicate partitioning.

We can work with partitioned matrices just like ordinary matrices always making sure that in products we never change the order of multiplication of things.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{bmatrix}$$

and

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22} \\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{bmatrix}$$

Note partitioning of A and B must match.

Addition: dimensions of A_{ij} and B_{ij} must be the same.

Multiplication formula A_{12} must have as many columns as B_{21} has rows, etc.

In general: need $A_{ij}B_{jk}$ to make sense for each i, j, k.

Works with more than a 2 by 2 partitioning.

Defn: **block diagonal** matrix: partitioned matrix A for which $A_{ij} = 0$ if $i \neq j$. If

$$A = \left[\begin{array}{cc} A_{11} & 0 \\ 0 & A_{22} \end{array} \right]$$

then A is invertible iff each A_{ii} is invertible and then

$$A^{-1} = \left[\begin{array}{cc} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{array} \right]$$

Moreover $\det(A) = \det(A_{11})\det(A_{22})$. Similar formulas work for larger matrices.

Partitioned inverses. Suppose $\mathbf{A}$, $\mathbf{C}$ are symmetric positive definite. Look for inverse of

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix}$$

of form

$$\begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{F}^T & \mathbf{G} \end{bmatrix}$$

Multiply to get equations

$$\begin{aligned} \mathbf{AE} + \mathbf{BF}^T &= \mathbf{I} \\ \mathbf{AF} + \mathbf{BG} &= \mathbf{0} \\ \mathbf{B}^T\mathbf{E} + \mathbf{CF}^T &= \mathbf{0} \\ \mathbf{B}^T\mathbf{F} + \mathbf{CG} &= \mathbf{I} \end{aligned}$$

Solve to get

$$\begin{aligned} \mathbf{F}^T &= -\mathbf{C}^{-1}\mathbf{B}^T\mathbf{E} \\ \mathbf{AE} - \mathbf{BC}^{-1}\mathbf{B}^T\mathbf{E} &= \mathbf{I} \\ \mathbf{E} &= (\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}^T)^{-1} \\ \mathbf{G} &= (\mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1} \end{aligned}$$