

Section 2.3 Symmetric Positive Definite Matrices and Orthogonal Matrices

We now move on to orthogonal matrices. We first review the concept of inner product and (vector) orthogonality.

Inner product. $\vec{u} \cdot \vec{v} = \sum_{j=1}^n x_j y_j$ if $\vec{u} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \vec{v} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$.
Also, use notation $\langle \vec{u}, \vec{v} \rangle = \vec{u} \cdot \vec{v} = \vec{u}^T \vec{v}$

Properties $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$
 $\langle \vec{u}, a\vec{v} + b\vec{w} \rangle = a\langle \vec{u}, \vec{v} \rangle + b\langle \vec{u}, \vec{w} \rangle$
 $\langle \vec{u}, \vec{u} \rangle \geq 0, = 0 \iff \vec{u} = \vec{0}$.

Length or norm $\|\vec{u}\| = \sqrt{\langle \vec{u}, \vec{u} \rangle} = \sqrt{x_1^2 + \dots + x_n^2}$
if $\vec{u} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

The Cauchy-Schwarz inequality

$$|\langle \vec{u}, \vec{v} \rangle| \leq \|\vec{u}\| \cdot \|\vec{v}\|$$

PF $g(\lambda) = \|\vec{u} + \lambda \vec{v}\|^2 = \|\vec{u}\|^2 + 2\lambda \langle \vec{u}, \vec{v} \rangle + \lambda^2 \|\vec{v}\|^2$
 $g(\lambda) \geq 0, \Delta = |2\langle \vec{u}, \vec{v} \rangle|^2 - 4\|\vec{u}\|^2 \|\vec{v}\|^2 \leq 0$.

Corollary $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|, \forall \vec{u}, \vec{v} \in \mathbb{R}^n$.

Easy to verify: $\|\vec{u} + \vec{v}\|^2 + \|\vec{u} - \vec{v}\|^2 = 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2$.
 $\forall \vec{u}, \vec{v} \in \mathbb{R}^n$.

Orthogonal: $\vec{u} \cdot \vec{v} = 0$.

A fact If $\vec{u}_1, \dots, \vec{u}_n$ are non zero and mutual orthogonal then they are linearly independent.

pf $\sum_{k=1}^n c_k \vec{u}_k = \vec{0} \quad \vec{u}_1 \cdot \sum_{k=1}^n c_k \vec{u}_k = 0$

$\Rightarrow \sum_{k=1}^n c_k \vec{u}_1 \cdot \vec{u}_k = 0$ But $\vec{u}_1 \cdot \vec{u}_1 \neq 0, \vec{u}_1 \cdot \vec{u}_k = 0$
if $k \neq 1$.

Hence $c_1 \vec{u}_1 \cdot \vec{u}_1 = 0 \quad c_1 = 0$.

Similarly, $c_2 = 0, \dots, c_n = 0$. Q.E.D.

Orthogonal basis: A group of n non zero, mutually orthogonal vectors form an orthogonal basis for \mathbb{R}^n .

Orthonormal basis An orthogonal basis for \mathbb{R}^n in which each vector has length 1.

If $\vec{u}_1, \dots, \vec{u}_n$ form an orthonormal basis for \mathbb{R}^n then for any $\vec{x} \in \mathbb{R}^n$

$$\vec{x} = \langle \vec{x}, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{x}, \vec{u}_2 \rangle \vec{u}_2 + \dots + \langle \vec{x}, \vec{u}_n \rangle \vec{u}_n$$

The Gram-Schmidt orthogonalization

Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$ be k linearly indep. vectors in \mathbb{R}^n . Then Define

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}$$

$$\begin{cases} \vec{w}_2 = \vec{v}_2 - \langle \vec{v}_2, \vec{u}_1 \rangle \vec{u}_1 \\ \vec{u}_2 = \frac{\vec{w}_2}{\|\vec{w}_2\|} \end{cases}$$

$$\begin{cases} \vec{w}_3 = \vec{v}_3 - \langle \vec{v}_3, \vec{u}_1 \rangle \vec{u}_1 - \langle \vec{v}_3, \vec{u}_2 \rangle \vec{u}_2 \\ \vec{u}_3 = \frac{\vec{w}_3}{\|\vec{w}_3\|} \end{cases}$$

$$\begin{cases} \vec{w}_k = \vec{v}_k - \langle \vec{v}_k, \vec{u}_1 \rangle \vec{u}_1 - \langle \vec{v}_k, \vec{u}_2 \rangle \vec{u}_2 - \dots - \langle \vec{v}_k, \vec{u}_{k-1} \rangle \vec{u}_{k-1} \\ \vec{u}_k = \frac{\vec{w}_k}{\|\vec{w}_k\|} \end{cases}$$

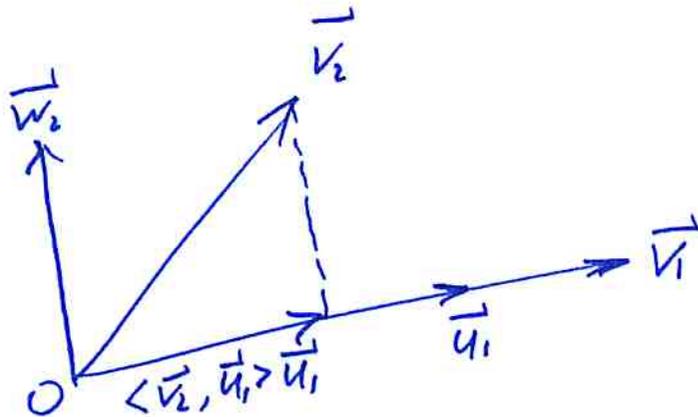
Then (1) $\vec{u}_1, \dots, \vec{u}_k$ are orthonormal

$$\text{i.e., } \langle \vec{u}_i, \vec{u}_j \rangle = \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

$1 \leq i, j \leq k.$

$$(2) \text{ span } \{ \vec{u}_1, \dots, \vec{u}_j \} = \text{span } \{ \vec{v}_1, \dots, \vec{v}_j \}$$

for $1 \leq j \leq k.$



Now, using $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_k$, we have

$$\begin{cases} \vec{v}_1 = \langle \vec{v}_1, \vec{u}_1 \rangle \vec{u}_1 \\ \vec{v}_2 = \langle \vec{v}_2, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v}_2, \vec{u}_2 \rangle \vec{u}_2 \\ \dots \\ \vec{v}_k = \langle \vec{v}_k, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v}_k, \vec{u}_2 \rangle \vec{u}_2 + \dots + \langle \vec{v}_k, \vec{u}_k \rangle \vec{u}_k \end{cases}$$

Equivalently,

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_k \end{bmatrix} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \dots & \vec{u}_k \end{bmatrix} \begin{bmatrix} \langle \vec{v}_1, \vec{u}_1 \rangle & \langle \vec{v}_2, \vec{u}_1 \rangle & \dots & \langle \vec{v}_k, \vec{u}_1 \rangle \\ & \langle \vec{v}_2, \vec{u}_2 \rangle & \dots & \langle \vec{v}_k, \vec{u}_2 \rangle \\ & & \ddots & \vdots \\ & & & \langle \vec{v}_k, \vec{u}_k \rangle \end{bmatrix}$$

$n \times k$ $n \times k$ $k \times k$
 $(n \geq k)$

$$A = QR$$

Theorem Suppose A is an $n \times k$ ($n \geq k$) real matrix with linearly indep. columns. (i.e., $\text{rank}(A) = k$.) Then

$A = QR$ [QR factorization]
 for an $n \times k$ matrix Q whose columns are orthonormal and R ~~an~~ upper triangular a $k \times k$ nonsingular matrix.

Example Least-squares.

How to solve

$$\begin{cases} 2x_1 - x_2 + x_3 = 1 \\ x_1 + 2x_2 - x_3 = 0 \\ x_1 - 2x_2 = 5 \end{cases} \quad ?$$

2 unknowns
 3 equations. No (classical) sol'n (x_1, x_2) .

In general $Ax = b$ $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $A: m \times n$

$m A_n x_i = m b_i$ $b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ $n < m$

The Least-squares sol'n

Called Normal equation. \longrightarrow $A^T A x = A^T b$

\uparrow \uparrow \uparrow
 $n \times n$ $n \times 1$ $n \times 1$

Assume A is $m \times n$ with $n < m$ and $\text{rank}(A) = n$. Then this eq. has a unique sol'n x

$Ax=b$ has a sol'n $x \iff b \in \text{col}(A)$. i.e.,
 b is a linear combination of columns of A .

Suppose, A is $m \times n$, $m > n$, $\text{rank } A = n$,
 and $b \in \mathbb{R}^m$, $b \notin \text{Col}(A)$.

Then $Ax=b$ has no sol'n.

But, let's define x to min. $\|Ax-b\|^2$.
 that's why called least-squares sol'n.

$$F(x) = \|Ax - b\|^2 \quad (x \in \mathbb{R}^n)$$

$$= Ax \cdot Ax - 2Ax \cdot b + b \cdot b$$

$$= A^T A x \cdot x - 2x \cdot A^T b + b \cdot b$$

$$\nabla F(x) = 2A^T A x - 2A^T b = 0 \iff$$

$$\boxed{A^T A x = A^T b}$$

Sol'n of normal equation by QR factorization

$$A = QR$$

$\uparrow \quad \uparrow \quad \nwarrow$
 $m \times n \quad m \times n \quad n \times n$

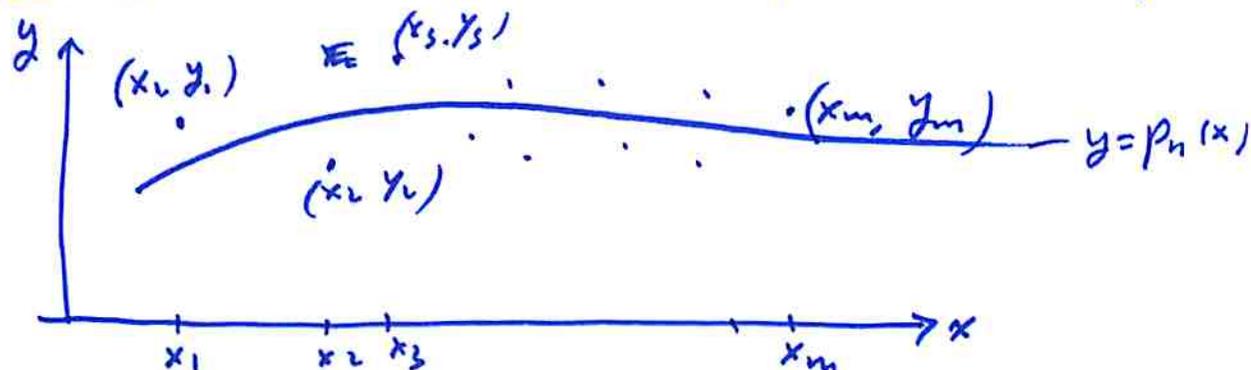
$$A^T A = (QR)^T Q A = R^T Q^T Q R = R^T R.$$

$$A^T A x = A^T b \iff R^T R x = R^T Q^T b \iff R x = Q^T b$$

$$\iff \boxed{x = R^{-1} Q^T b}$$

I have not been so organized here. I should have had the following example before. But, hope it is not too late.

Example (Data fitting, a least-squares prob.)



Measured data $(x_1, y_1), \dots, (x_m, y_m)$.

Fitting polynomial $y = p_n(x) = a_0 + a_1 x + \dots + a_n x^n$.

Here $n < m$. We need to find a_0, a_1, \dots, a_n .

Equations: $y_j = p_n(x_j) \quad j=1, 2, \dots, m$.

m equations
 $n+1$ unknowns (a_0, a_1, \dots, a_n) .

$$\begin{cases} a_0 + a_1 x_1 + a_2 x_1^2 + \dots + a_n x_1^n = y_1 \\ a_0 + a_1 x_2 + a_2 x_2^2 + \dots + a_n x_2^n = y_2 \\ \dots \\ a_0 + a_1 x_m + a_2 x_m^2 + \dots + a_n x_m^n = y_m \end{cases}$$

$$A \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

\uparrow \uparrow
 $m \times n$ $n \times 1$ $m \times 1$

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_m & x_m^2 & \dots & x_m^n \end{bmatrix} \quad m \times n$$

So far: Gram-Schmidt \Rightarrow QR factorization \Rightarrow LS prob.

Next: G-S \Rightarrow Schur's Lemma \Rightarrow Spectral Thm
for symmetric matrix

Definition An $n \times n$ real matrix A is orthogonal
(or unitary) if $A^T A = I$ (identity matrix).

$$A^T A = I \iff A^T = A^{-1}$$

\iff columns of A are orthonormal

\iff rows of A are orthonormal

$\iff \langle Au, Av \rangle = \langle u, v \rangle \quad \forall u, v \in \mathbb{R}^n$
(i.e., preserving inner product)

$\iff \|Au\| = \|u\| \quad \forall u \in \mathbb{R}^n$
(i.e., preserving length)

More properties

① If A is orthogonal then
 $\det A = \pm 1$ ($\det A = 1$: proper rotation,
 $A^T A = I$)

$$\text{PF } \det(A^T A) = \det I = 1 \implies (\det A)^2 = 1$$

$$\det A^T = \det A$$

$O(n) = \{ \text{all } n \times n \text{ real orthogonal matrices} \}$
 $SO(n) = \{ \text{all } n \times n \text{ real orthogonal matrices with determinants } 1 \}$

Both $O(n)$ and $SO(n)$ are groups
with ~~resp~~ resp. to matrix product.

② If λ is an eigenvalue of an orthogonal matrix A , then $|\lambda| = 1$.

pf $Au = \lambda u \quad (u \neq 0)$

$\|u\| = \|Au\| = |\lambda| \|u\| \implies |\lambda| = 1. \quad \underline{Q.E.D.}$

Schur's Lemma For any $n \times n$ matrix A , there exists an orthogonal matrix U such that $U^{-1}AU = T$ is an upper triangular matrix with diagonal entries being eigenvalues of A .

pf. Use the Gram-Schmidt.

$AU_1 = U_1 \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix}$

1st col. in U_1 is an eigenvector for λ_1 .

$U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix}$

U_1 is unitary by G-S.

Similar, $U_2^{-1}U_1^{-1}AU_1U_2$

$= \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{bmatrix}$

Q.E.D.

If $A^T = A$, $n \times n$, real. then all λ 's are real. and eigenvectors can be chosen to be real. So,

$U^{-1}AU = T$

$U^{-1} = U^T$

$(U^{-1}AU)^T = T^T$

$T = U^{-1}AU = U^T A U = T^T$

So, $T = T^T = D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

So, A is diagonalizable by an orthogonal matrix!

Spectral Theorem for (Real) Symmetric Matrices

If A is an $n \times n$ real symmetric matrix, then A has n real eigenvalues (possibly ~~repeated~~ repeated) $\lambda_1, \dots, \lambda_n$ and a set of orthonormal eigenvectors $u_1, \dots, u_n \in \mathbb{R}^n$. If $U = [u_1 \ u_2 \ \dots \ u_n]$ then U is orthogonal. $U^{-1}AU = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$.

and $A = \sum_{j=1}^n \lambda_j u_j \otimes u_j$.

Spectral decomposition.

Corollary If A is SPD with spectral decomposition $A = \sum_{j=1}^n \lambda_j u_j \otimes u_j$, then define $\sqrt{A} = \sum_{j=1}^n \sqrt{\lambda_j} u_j \otimes u_j$. $\sqrt{A} \sqrt{A} = A$.
Symmetric and orthogonal matrices are normal matrices.

Definition A matrix is a normal matrix if it has a full set of orthonormal eigenvectors.

Thm. N is normal $\Leftrightarrow NN^* = N^*N$

$$N^* = \overline{N}^T$$

(transpose + conjugate)

Well, it's too much now!

Let's end this section (section 2.3), though Jordan canonical form can be included here.

Well, except SVD and polar decomposition.

Polar decomposition. Let $y(x) = Fx + b$ where $x \in \mathbb{R}^3$, F is a 3×3 matrix with $\det F > 0$. $y = y(x)$ is called a deformation. It consists a translation $(+b)$ and homogeneous part Fx . It turns out $x \mapsto Fx$ consists of rotation and stretching. This is due to polar decomposition of the matrix F .

Theorem. (Polar decomposition). Let F be an $n \times n$ real matrix such that $\det F > 0$. Then there exist unique rotational matrix R (i.e., orthogonal and $\det R = 1$) and symmetric positive definite matrix U such that

$$F = RU.$$

PF. Construction. $F^T F = (RU)^T RU = U^T R^T R U = U^T U = U^2$

So, define $U = \sqrt{F^T F}$. ($F^T F$ is SPD. So its square root is well defined.)

Define $R = F U^{-1}$. So, $F = R U$

Easy to verify that $R^T R = U^{-T} F^T F U^{-1} = U^{-1} U^2 U^{-1} = I$.

$$\det R = \det F \cdot \det U^{-1} = \cancel{\det R} \det \sqrt{\det F^T \cdot \det F} \det U^{-1} = \det U \cdot \det U^{-1} = 1$$

Note: $\det \sqrt{A} = \sqrt{\det A}$ if A is SPD

$$\det A = \lambda_1 \lambda_2 \dots \lambda_n$$

$$\det \sqrt{A} = \sqrt{\lambda_1} \sqrt{\lambda_2} \dots \sqrt{\lambda_n}$$

Q.E.D.

