5. Orthogonal matrices

- matrices with orthonormal columns
- orthogonal matrices
- tall matrices with orthonormal columns
- complex matrices with orthonormal columns

Orthonormal vectors

a collection of real m-vectors $a_1, a_2, \ldots, a_n$ is *orthonormal* if

- the vectors have unit norm: $\|a_i\| = 1$
- they are mutually orthogonal: $a_i^T a_j = 0$ if $i \neq j$

Example

$$\left[\begin{array}{r} 0 \\ 0 \\ -1 \end{array}\right], \qquad \frac{1}{\sqrt{2}}\left[\begin{array}{r} 1 \\ 1 \\ 0 \end{array}\right], \qquad \frac{1}{\sqrt{2}}\left[\begin{array}{r} 1 \\ -1 \\ 0 \end{array}\right]$$

Matrix with orthonormal columns

$A \in \mathbf{R}^{m\times n}$ has orthonormal columns if its Gram matrix is the identity matrix:

$$
\begin{aligned}
A^TA &= \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array}\right]^T \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array}\right] \\
&= \left[\begin{array}{cccc} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{array}\right] \\
&= \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{array}\right]
\end{aligned}
$$

there is no standard short name for ‘matrix with orthonormal columns’

Matrix-vector product

if $A \in \mathbf{R}^{m \times n}$ has orthonormal columns, then the linear function $f(x) = Ax$

- preserves inner products:

$$(Ax)^T(Ay) = x^T A^T A y = x^T y$$

- preserves norms:

$$\|Ax\| = \left((Ax)^T(Ax)\right)^{1/2} = (x^T x)^{1/2} = \|x\|$$

- preserves distances: $\|Ax - Ay\| = \|x - y\|$

- preserves angles:

$$\angle(Ax, Ay) = \arccos\left(\frac{(Ax)^T(Ay)}{\|Ax\|\|Ay\|}\right) = \arccos\left(\frac{x^T y}{\|x\|\|y\|}\right) = \angle(x, y)$$

Left invertibility

if $A \in \mathbf{R}^{m \times n}$ has orthonormal columns, then

- A is left invertible with left inverse A^T: by definition

$$A^T A = I$$

- A has linearly independent columns (from p. 4-24 or p. 5-2):

$$Ax = 0 \qquad \Longrightarrow \qquad A^T Ax = x = 0$$

- A is tall or square: $m \geq n$ (see page 4-13)

Outline

- matrices with orthonormal columns
- **orthogonal matrices**
- tall matrices with orthonormal columns
- complex matrices with orthonormal columns

Orthogonal matrix

Orthogonal matrix

a *square* real matrix with orthonormal columns is called *orthogonal*

Nonsingularity (from equivalences on page 4-14): if A is orthogonal, then

- A is invertible, with inverse A^T:

$$\left.\begin{array}{l} A^TA = I \\ A \text{ is square} \end{array}\right\} \quad \Longrightarrow \quad AA^T = I$$

- A^T is also an orthogonal matrix
- rows of A are orthonormal (have norm one and are mutually orthogonal)

Note: if $A \in \mathbf{R}^{m\times n}$ has orthonormal columns and $m > n$, then $AA^T \neq I$

Permutation matrix

- let $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ be a permutation (reordering) of $(1, 2, \dots, n)$
- we associate with π the $n \times n$ *permutation matrix* A

$$A_{i\pi_i} = 1, \qquad A_{ij} = 0 \text{ if } j \neq \pi_i$$

- Ax is a permutation of the elements of x: $Ax = (x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_n})$
- A has exactly one element equal to 1 in each row and each column

Orthogonality: permutation matrices are orthogonal

- $A^T A = I$ because A has exactly one element equal to one in each row

$$(A^T A)_{ij} = \sum_{k=1}^{n} A_{ki} A_{kj} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

- $A^T = A^{-1}$ is the inverse permutation matrix

Example

- permutation on $\{1, 2, 3, 4\}$

$$(\pi_1, \pi_2, \pi_3, \pi_4) = (2, 4, 1, 3)$$

- corresponding permutation matrix and its inverse

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad A^{-1} = A^T = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

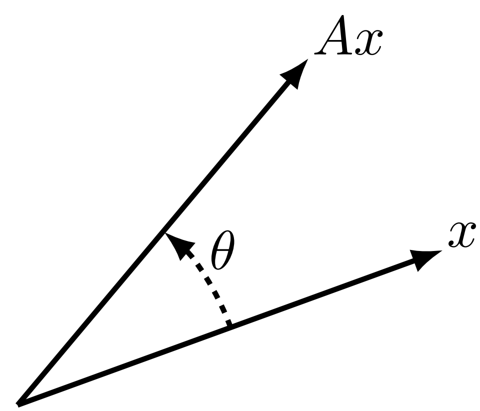
- A^T is permutation matrix associated with the permutation

$$(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4) = (3, 1, 4, 2)$$

Plane rotation

Rotation in a plane

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



Rotation in a coordinate plane in $\mathbf{R}^n$: for example,

$$A = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

describes a rotation in the (x_1, x_3) plane in $\mathbf{R}^3$

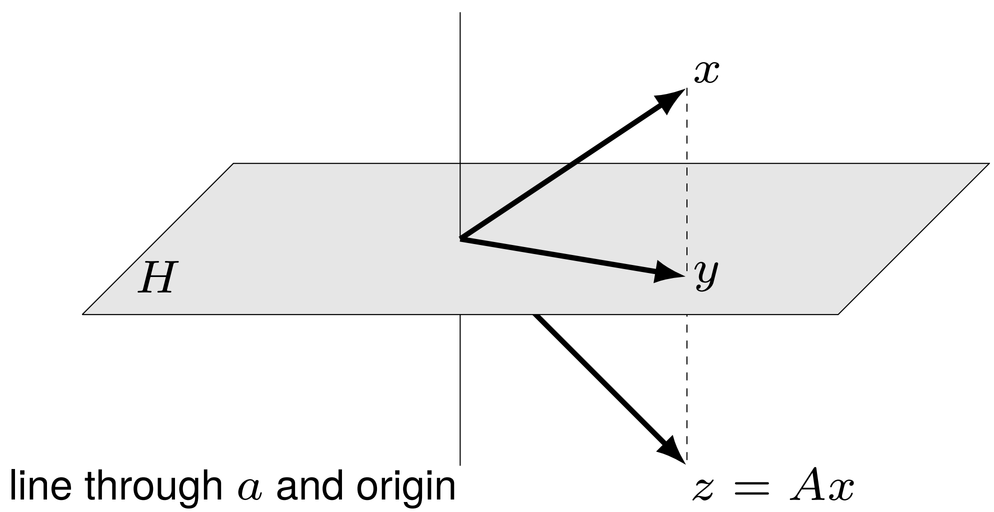
Reflector

$$A = I - 2aa^T \qquad \text{with } a \text{ a unit-norm vector } (\|a\| = 1)$$

- a reflector matrix is symmetric and orthogonal

$$A^T A = (I - 2aa^T)(I - 2aa^T) = I - 4aa^T + 4aa^T aa^T = I$$

- Ax is reflection of x through the hyperplane $H = \{u \mid a^T u = 0\}$



$$\begin{aligned} y &= x - (a^T x)a \\ &= (I - aa^T)x \\ z &= y + (y - x) \\ &= (I - 2aa^T)x \end{aligned}$$

(see page 2-44)

Product of orthogonal matrices

if $A_1, \ldots, A_k$ are orthogonal matrices and of equal size, then the product

$$A = A_1 A_2 \cdots A_k$$

is orthogonal:

$$\begin{aligned} A^T A &= (A_1 A_2 \cdots A_k)^T (A_1 A_2 \cdots A_k) \\ &= A_k^T \cdots A_2^T A_1^T A_1 A_2 \cdots A_k \\ &= I \end{aligned}$$

Linear equation with orthogonal matrix

linear equation with orthogonal coefficient matrix A of size $n \times n$

$$Ax = b$$

solution is

$$x = A^{-1}b = A^T b$$

- can be computed in $2n^2$ flops by matrix-vector multiplication
- cost is less than order n^2 if A has special properties; for example,

permutation matrix:	0 flops
reflector (given a):	order n flops
plane rotation:	order 1 flops

Outline

- matrices with orthonormal columns
- orthogonal matrices
- **tall matrices with orthonormal columns**
- complex matrices with orthonormal columns

Tall matrix with orthonormal columns

suppose $A \in \mathbf{R}^{m\times n}$ is tall ($m > n$) and has orthonormal columns

- A^T is a left inverse of A:

$$A^T A = I$$

- A has no right inverse; in particular

$$AA^T \neq I$$

on the next pages, we give a geometric interpretation to the matrix AA^T

Range

- the *span* of a collection of vectors is the set of all their linear combinations:

$$\operatorname{span}(a_1, a_2, \ldots, a_n) = \{x_1 a_1 + x_2 a_2 + \cdots + x_n a_n \mid x \in \mathbf{R}^n\}$$

- the *range* of a matrix $A \in \mathbf{R}^{m \times n}$ is the span of its column vectors:

$$\operatorname{range}(A) = \{Ax \mid x \in \mathbf{R}^n\}$$

Example

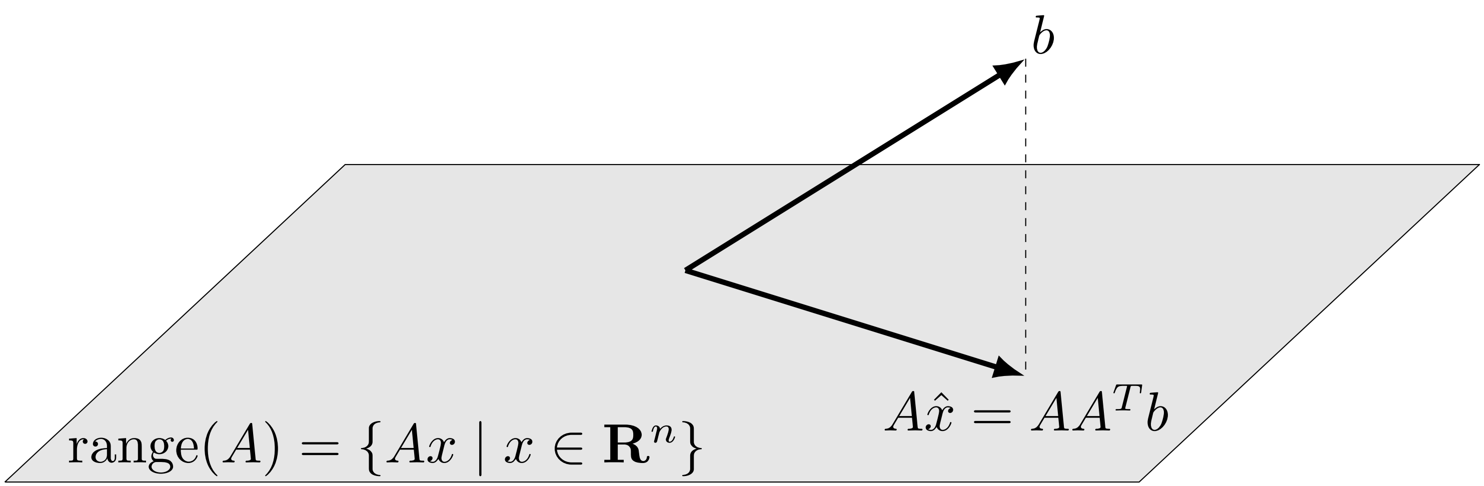
$$\operatorname{range}(\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & -1 & 1 \end{bmatrix}) = \left\{ \begin{bmatrix} x_1 + x_3 \\ x_1 + x_2 + 2x_3 \\ -x_2 + x_3 \end{bmatrix} \mid x_1, x_2, x_3 \in \mathbf{R} \right\}$$

Projection on range of matrix with orthonormal columns

suppose $A \in \mathbf{R}^{m \times n}$ has orthonormal columns; we show that the vector

$$AA^T b$$

is the orthogonal projection of an m-vector b on $\mathrm{range}(A)$



- $\hat{x} = A^T b$ satisfies $\|A\hat{x} - b\| < \|Ax - b\|$ for all $x \neq \hat{x}$
- this extends the result on page 2-12 (where $A = (1/\|a\|)a$)

Proof

the squared distance of b to an arbitrary point Ax in $\mathrm{range}(A)$ is

$$\begin{aligned}
\|Ax - b\|^2 &= \|A(x - \hat{x}) + A\hat{x} - b\|^2 \qquad \text{(where } \hat{x} = A^T b\text{)} \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 + 2(x - \hat{x})^T A^T (A\hat{x} - b) \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \\
&= \|x - \hat{x}\|^2 + \|A\hat{x} - b\|^2 \\
&\geq \|A\hat{x} - b\|^2
\end{aligned}$$

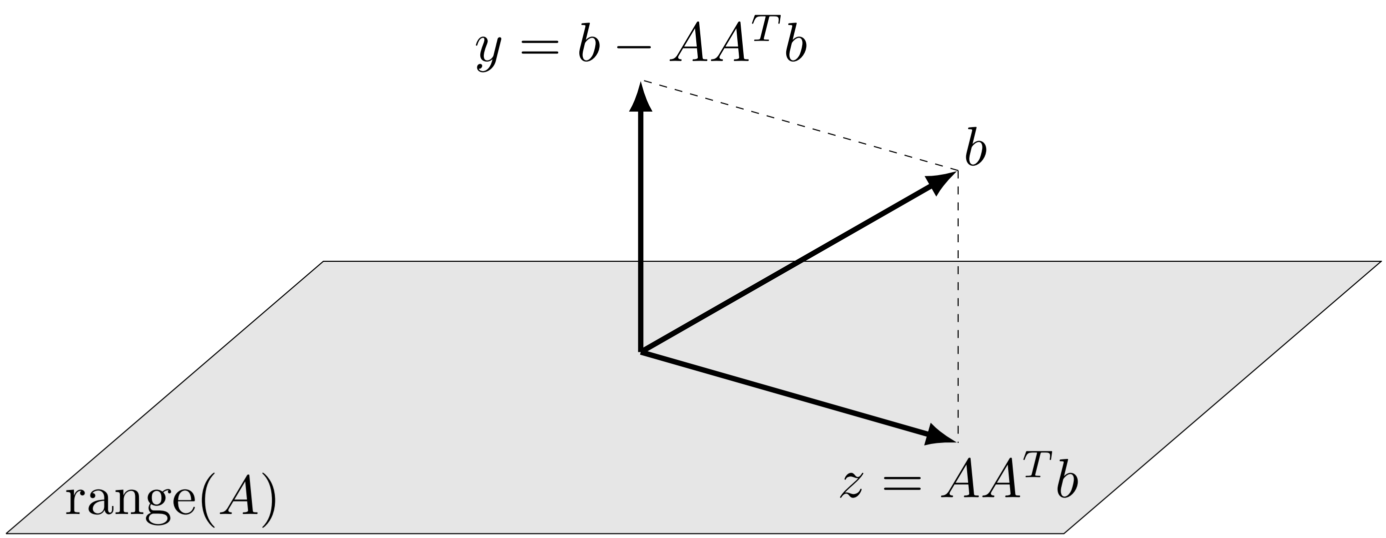
with equality only if $x = \hat{x}$

- line 3 follows because $A^T(A\hat{x} - b) = \hat{x} - A^T b = 0$
- line 4 follows from $A^T A = I$

Orthogonal decomposition

the vector b is decomposed as a sum $b = z + y$ with

$$z \in \mathrm{range}(A), \qquad y \perp \mathrm{range}(A)$$



such a decomposition exists and is unique for every b

$$b = Ax + y, \quad A^T y = 0 \qquad \Longleftrightarrow \qquad A^T b = x, \quad y = b - AA^T b$$

Outline

- matrices with orthonormal columns
- orthogonal matrices
- tall matrices with orthonormal columns
- **complex matrices with orthonormal columns**

Gram matrix

$A \in \mathbf{C}^{m\times n}$ has orthonormal columns if its Gram matrix is the identity matrix:

$$
\begin{aligned}
A^H A &= \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array}\right]^H \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array}\right] \\
&= \left[\begin{array}{cccc} a_1^H a_1 & a_1^H a_2 & \cdots & a_1^H a_n \\ a_2^H a_1 & a_2^H a_2 & \cdots & a_2^H a_n \\ \vdots & \vdots & & \vdots \\ a_n^H a_1 & a_n^H a_2 & \cdots & a_n^H a_n \end{array}\right] \\
&= \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{array}\right]
\end{aligned}
$$

- columns have unit norm: $\|a_i\|^2 = a_i^H a_i = 1$
- columns are mutually orthogonal: $a_i^H a_j = 0$ for $i \neq j$

Unitary matrix

Unitary matrix

a *square* complex matrix with orthonormal columns is called *unitary*

Inverse

$$\left.\begin{array}{l} A^H A = I \\ A \text{ is square} \end{array}\right\} \quad \Longrightarrow \quad AA^H = I$$

- a unitary matrix is nonsingular with inverse A^H
- if A is unitary, then A^H is unitary

Discrete Fourier transform matrix

recall definition from page 3-37 (with $\omega = e^{2\pi \mathrm{j}/n}$ and $\mathrm{j} = \sqrt{-1}$)

$$W = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \cdots & \omega^{-(n-1)} \\ 1 & \omega^{-2} & \omega^{-4} & \cdots & \omega^{-2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{-(n-1)} & \omega^{-2(n-1)} & \cdots & \omega^{-(n-1)(n-1)} \end{bmatrix}$$

the matrix $(1/\sqrt{n})W$ is unitary (proof on next page):

$$\frac{1}{n}W^H W = \frac{1}{n}WW^H = I$$

- inverse of W is $W^{-1} = (1/n)W^H$
- inverse discrete Fourier transform of n-vector x is $W^{-1}x = (1/n)W^H x$

Gram matrix of DFT matrix

we show that $W^H W = nI$

- conjugate transpose of W is

$$W^H = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

- i, j element of Gram matrix is

$$(W^H W)_{ij} = 1 + \omega^{i-j} + \omega^{2(i-j)} + \cdots + \omega^{(n-1)(i-j)}$$

$$(W^H W)_{ii} = n, \qquad (W^H W)_{ij} = \frac{\omega^{n(i-j)} - 1}{\omega^{i-j} - 1} = 0 \quad \text{if } i \neq j$$

(last step follows from $\omega^n = 1$)