5. Orthogonality

- matrices with orthonormal columns
- orthogonal matrices
- Gram-Schmidt orthogonalization

Orthonormal vectors

the list m-vectors $a_1, a_2, \ldots, a_n$ is *orthonormal* if

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

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

Examples

- standard unit n-vectors $e_1, \ldots, e_n$
- the three vectors

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

Orthonormal expansion

orthonormal n-vectors $a_1, \ldots, a_n$ are linearly independent, hence basis for $\mathbb{R}^n$

- therefore, for any n-vector x,

$$x = \beta_1 a_1 + \cdots + \beta_n a_n \quad \text{for some unique } \beta_i$$

  this is called *orthonormal expansion* of x (in the orthonormal basis)

- multiplying by a_i^T on left, we have $\beta_i = a_i^T x$ and hence

$$x = (a_1^T x) a_1 + \cdots + (a_n^T x) a_n$$

Matrix with orthonormal columns

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

$$A^TA = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & \vdots & & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

there is no standard short name for a “matrix with orthonormal columns”

- A is left-invertible with left inverse A^T
- A has linearly independent columns: $Ax = 0 \Longrightarrow A^TAx = x = 0$
- A is tall or square: $m \geq n$
- if A is tall $m > n$, then A has no right inverse; in particular

$$AA^T \neq I$$

Matrix-vector product

if $A \in \mathbb{R}^{m\times n}$ 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:

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

Projection on range of matrix with orthonormal columns

suppose $Q \in \mathbb{R}^{m\times n}$ has orthonormal columns $q_1, \ldots, q_n$

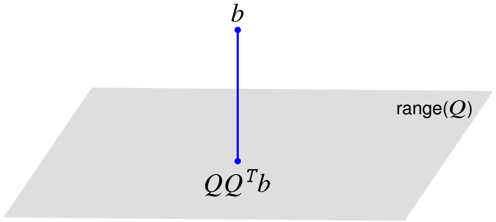
- $Q\hat{x}$ is called the *orthogonal projection* of $b \in \mathbb{R}^m$ onto $\mathrm{range}(Q)$ if

$$\|Q\hat{x} - b\| < \|Qx - b\| \quad \text{for all} \quad x \neq \hat{x}$$

  i.e., it is the vector on $\mathrm{range}(Q)$ closest to b

- we next show $\hat{x} = Q^Tb$ so that the orthogonal projection of b onto $\mathrm{range}(Q)$ is

$$QQ^Tb$$



Proof: the squared distance of b to an arbitrary point Qx in $\mathrm{range}(Q)$ is

$$\begin{aligned}
\|Qx - b\|^2 &= \|Q(x - \hat{x}) + Q\hat{x} - b\|^2 \quad \text{(where } \hat{x} = Q^T b\text{)} \\
&= \|Q(x - \hat{x})\|^2 + \|Q\hat{x} - b\|^2 + 2(x - \hat{x})^T Q^T (Q\hat{x} - b) \\
&= \|Q(x - \hat{x})\|^2 + \|Q\hat{x} - b\|^2 \\
&= \|x - \hat{x}\|^2 + \|Q\hat{x} - b\|^2 \\
&\geq \|Q\hat{x} - b\|^2
\end{aligned}$$

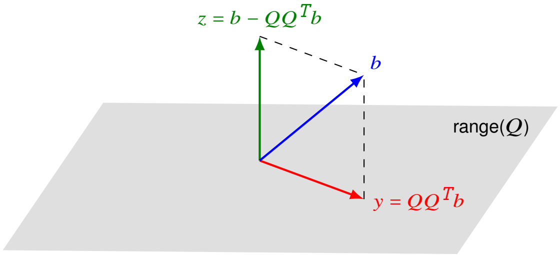
with equality only if $x = \hat{x}$

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

Orthogonal decomposition

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

$$y \in \operatorname{range}(Q), \quad z \in \operatorname{range}(Q)^\perp$$



- decomposition exists and unique for every b:

$$b = Qx + z, \quad Q^Tz = 0 \quad \Longleftrightarrow \quad x = Q^Tb, \quad z = b - QQ^Tb$$

- z is orthogonal projection on $\operatorname{range}(Q)^\perp = \operatorname{null}(Q^T) = \{u \mid Q^Tu = 0\}$

Outline

- matrices with orthonormal columns
- **orthogonal matrices**
- Gram-Schmidt orthogonalization

Orthogonal matrix

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

Nonsingularity: if A is orthogonal, then

- A is invertible, with inverse A^T:

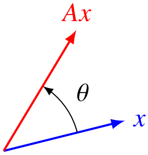
$$\left.\begin{array}{l} A^T A = I \\ A \text{ is square} \end{array}\right\} \quad \Longrightarrow \quad AA^T = I$$

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

Example: rotation in a plane

rotation matrices are orthogonal

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



Rotation in a coordinate plane in $\mathbb{R}^n$: 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 $\mathbb{R}^3$

Example: permutation matrices

- permutation matrix is square with exactly one entry of each row/column is one
  - an identity matrix with rows and columns interchanged
- let $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ be a *permutation* (reordering) of $(1, 2, \ldots, n)$
- permutation matrix A,

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

  is orthogonal

- Ax is a permutation of the elements of x: $Ax = (x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n})$

Proof

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

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

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

Example: permutation matrices

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

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

- A^T is permutation matrix associated with the permutation

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

Matrix multiplication with permutation matrix

- left multiplying a matrix by a permutation matrix, will switch the corresponding rows
- right multiplying will switch the corresponding columns

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
plane rotation:	order 1 flops

Outline

- matrices with orthonormal columns
- orthogonal matrices
- **Gram-Schmidt orthogonalization**

Gram-Schmidt (G-S) procedure

given vectors $a_1, \ldots, a_n \in \mathbb{R}^m$

step 1a. $\tilde{q}_1 := a_1$

step 1b. $q_1 := \tilde{q}_1 / \|\tilde{q}_1\|$ (normalize)

step 2a. $\tilde{q}_2 := a_2 - (q_1^T a_2) q_1$ (remove q_1 component from a_2)

step 2b. $q_2 := \tilde{q}_2 / \|\tilde{q}_2\|$ (normalize)

step 3a. $\tilde{q}_3 := a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2$ (remove q_1, q_2 components)

step 3b. $q_3 := \tilde{q}_3 / \|\tilde{q}_3\|$ (normalize)

etc.

- $\tilde{q}_k$ and q_k are orthogonal to $q_1, \ldots, q_{k-1}$
- when $a_1, \ldots, a_n$ are independent, G-S produces orthonormal basis $q_1, \ldots, q_n$:

$$\text{span}(a_1, \ldots, a_k) = \text{span}(q_1, \ldots, q_k), \quad k \leq n$$

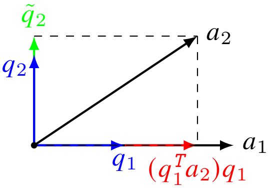
- if $a_1, \ldots, a_{k-1}$ are independent, but $a_1, \ldots, a_k$ are dependent, then $\tilde{q}_k = 0$

Interpretation

$$\begin{aligned}\tilde{q}_k &= a_k - q_1(q_1^T a_k) - q_2(q_2^T a_k) - \cdots - q_{k-1}(q_{k-1}^T a_k) \\ &= (I - q_1 q_1^T - q_2 q_2^T - \cdots - q_{k-1} q_{k-1}^T) a_k \\ &= (I - \begin{bmatrix} q_1 & q_2 & \cdots & q_{k-1} \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \cdots & q_{k-1} \end{bmatrix}^T) a_k\end{aligned}$$

this is the residual of a_k after subtracting its orthogonal projection on

$$\begin{aligned}\text{span}(a_1, a_2, \ldots, a_{k-1}) &= \text{span}(q_1, q_2, \ldots, q_{k-1}) \\ &= \mathrm{range}\left(\begin{bmatrix} q_1 & q_2 & \cdots & q_{k-1} \end{bmatrix}\right)\end{aligned}$$



Gram-Schmidt (G-S) algorithm

---

given vectors $a_1, \ldots, a_n \in \mathbb{R}^m$

set $q_1 = a_1/\|a_1\|$

for $k = 2, \ldots, n$

1. *orthogonalization:*

$$\tilde{q}_k = a_k - (q_1^T a_k)q_1 - \cdots - (q_{k-1}^T a_k)q_{k-1}$$

2. *test for linear dependence:* if $\tilde{q}_k = 0$ quit
3. *normalization:* $q_k = \tilde{q}_k/\|\tilde{q}_k\|$

Example

$$a_1 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 3 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix}$$

- $k = 1$, $\|a_1\| = 2$ and

$$q_1 = a_1/\|a_1\| = (-1/2, 1/2, -1/2, 1/2)$$

- $k = 2$, we have $q_1^T a_2 = 4$, and

$$\tilde{q}_2 = a_2 - (q_1^T a_2) q_1 = (1, 1, 1, 1)$$

  normalizing, we get

$$q_2 = \tilde{q}_2/\|\tilde{q}_2\| = (1/2, 1/2, 1/2, 1/2)$$

- $k = 3$; we have $q_1^T a_3 = 2$ and $q_2^T a_3 = 8$, so

$$\tilde{q}_3 = a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2 = (-2, -2, 2, 2)$$

normalizing, we get

$$q_3 = \tilde{q}_3 / \|\tilde{q}_3\| = (-1/2, -1/2, 1/2, 1/2)$$

- since no vector $\tilde{q}_i$ is zero, the vectors a_1, a_2, a_3 are linearly independent

References and further readings

- S. Boyd and L. Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares.* Cambridge University Press, 2018.
- L. Vandenberghe, *EE133A Lecture Notes*, University of California, Los Angeles.
- M. T. Heath. *Scientific Computing: An Introductory Survey* (revised second edition). Society for Industrial and Applied Mathematics, 2018.