EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

UCMERCED

Lecture 11

Overview

- Gram-Schmidt process
- QR decomposition
- Gram-Schmidt triangular orthogonalization

Reading

- Chapter 7 and Chapter 8 of *Numerical Linear Algebra* by Llyod Trefethen and David Bau
- Chapter 5 of *Matrix Computations* by Gene Golub and Charles Van Loan
- Chapter 5 of *Matrix Analysis and Applied Linear Algebra* by Carl Meyer

Geometric properties of orthonormal basis

- Let the columns of $U = [\mathbf{u}_1 \dots \mathbf{u}_n]$ are orthonormal, $U^\top U = I$
- If $\mathbf{w} = U\mathbf{x}$, then
  - mapping $\mathbf{w} = U\mathbf{x}$ is isometric: it preserves distance (as $\|\mathbf{w}\|^2 = \|\mathbf{z}\|^2$)
  - inner products are preserved: if $\mathbf{w}_1 = U\mathbf{x}_1$ and $\mathbf{w}_2 = U\mathbf{x}_2$, then $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$
  - angles are preserved: $\angle(\mathbf{w}_1, \mathbf{w}_2) = \angle(\mathbf{x}_1, \mathbf{x}_2)$
- Multiplication by U preserves inner products, angles, and distances
- It follows that $U^{-1} = U^\top$ and hence also $UU^\top = I$
- For any $\mathbf{x}$, $\mathbf{x} = UU^\top\mathbf{x}$, i.e.,

$$\mathbf{x} = \sum_{i=1}^{n} (\mathbf{u}_i^\top \mathbf{x})\mathbf{u}_i$$

  - $\mathbf{u}_i^\top \mathbf{x}$ is called the component of $\mathbf{x}$ in the direction of $\mathbf{u}_i$
  - $\mathbf{a} = U^\top\mathbf{x}$ resolves $\mathbf{x}$ into the vector of its $\mathbf{u}_i$ components
  - $\mathbf{x} = U\mathbf{a}$ reconstructs $\mathbf{x}$ from its $\mathbf{u}_i$ components
  - $\mathbf{x} = U\mathbf{a} = \sum_{i=1}^{n} a_i\mathbf{u}_i$ is called the expansion of $\mathbf{x}$

Examples

- Rotation by θ in $\mathbb{R}^2$ is given by

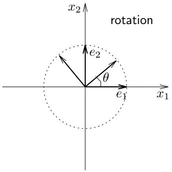
$$\mathbf{y} = U_\theta \mathbf{x},\, U_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

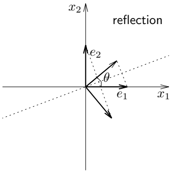
since $\mathbf{e}_1 = [\cos\theta, \sin\theta]^\top, \mathbf{e}_2 = [-\sin\theta, \cos\theta]^\top$

- Reflection across line $\mathbf{x}_2 = x_1 \tan(\theta/2)$ is given by

$$\mathbf{y} = U_\theta \mathbf{x},\, U_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

since $\mathbf{e}_1 = [\cos\theta, \sin\theta]^\top, \mathbf{e}_2 = [\sin\theta, -\cos\theta]^\top$





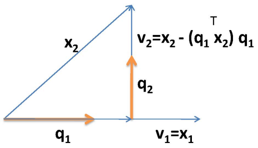
Gram-Schmidt process

- Given independent vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$, Gram-Schmidt process finds orthonormal vectors, $\mathbf{q}_1, \ldots, \mathbf{q}_n$ such that

$$\text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_r) = \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_r)$$

- Thus, $\mathbf{q}_1, \ldots, \mathbf{q}_r$ are orthonormal basis for $\text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_r)$
- Idea: first orthogonalized each vector w.r.t. previous ones and then normalize result to have unit norm
  - $\mathbf{v}_1 = \mathbf{x}_1$
  - $\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$ (normalize)
  - $\mathbf{v}_2 = \mathbf{x}_2 - (\mathbf{q}_1^\top \mathbf{x}_2)\mathbf{q}_1$ (remove $\mathbf{q}_1$ component from $\mathbf{x}_2$)
  - $\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$ (normalize)
  - $\mathbf{v}_3 = \mathbf{x}_3 - (\mathbf{q}_1^\top \mathbf{x}_3)\mathbf{q}_1 - (\mathbf{q}_2^\top \mathbf{x}_3)\mathbf{q}_2$ (remove, $\mathbf{q}_1$, $\mathbf{q}_2$ components)
  - $\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|}$
  - etc.
- Find orthonormal basis incrementally

Geometric interpretation of Gram-Schmidt process



- Gram-Schmidt process:
  - $\mathbf{v}_1 = \mathbf{x}_1$
  - $\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$ (normalize)
  - $\mathbf{v}_2 = \mathbf{x}_2 - (\mathbf{q}_1^\top \mathbf{x}_2)\mathbf{q}_1$ (remove $\mathbf{q}_1$ component from $\mathbf{x}_2$)
  - $\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$ (normalize)
  - $\mathbf{v}_3 = \mathbf{x}_3 - (\mathbf{q}_1^\top \mathbf{x}_3)\mathbf{q}_1 - (\mathbf{q}_2^\top \mathbf{x}_3)\mathbf{q}_2$ (remove, $\mathbf{q}_1$, $\mathbf{q}_2$ components)
  - $\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|}$
  - etc.
- $\forall i$, we have

$$\begin{aligned}\mathbf{x}_i &= (\mathbf{q}_1^\top \mathbf{x}_i)\mathbf{q}_1 + (\mathbf{q}_2^\top \mathbf{x}_i)\mathbf{q}_2 + \cdots + (\mathbf{q}_{i-1}^\top \mathbf{x}_i)\mathbf{q}_{i-1} + \|\mathbf{v}_i\|\mathbf{q}_i \\ &= r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i\end{aligned}$$

QR decomposition

- In matrix form, $A = QR$ where $A \in \mathbb{R}^{m \times n}$, $Q \in \mathbb{R}^{m \times n}$, $R \in \mathbb{R}^{n \times n}$:

$$\underbrace{\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}}_{R}$$

- $Q^\top Q = I$, and R is upper triangular and invertible
- Usually computed using a variation of Gram-Schmidt process which is less sensitive to numerical errors
- Can also be computed by Householder transformation or Givens rotations
- Columns of Q are orthonormal basis in $\text{ran}(A)$

$$Q^\top A = R$$

General Gram-Schmidt process

- If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are dependent, we find $\mathbf{v}_j = 0$ for some j, which means $\mathbf{x}_j$ is linearly dependent on $\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}$
- Modified algorithm: when we have $\mathbf{v}_j = 0$, skip to the next vector $\mathbf{x}_{j+1}$ and continue

$$
\begin{aligned}
& k = 0 \\
& \text{for} \quad i = 1, \ldots, n \\
& \{ \\
& \quad \mathbf{v} = \mathbf{x}_i - \textstyle\sum_{j=1}^{k} \mathbf{q}_j \mathbf{q}_j^\top \mathbf{x}_i; \\
& \quad \text{if} \ \ \mathbf{v} \neq 0 \ \ \{k = k + 1; \mathbf{q}_k = \tfrac{\mathbf{v}}{\|\mathbf{v}\|}\}; \\
& \}
\end{aligned}
$$

Example

- Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$\mathbf{v}_1 = \mathbf{x}_1, \quad \mathbf{q}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$$

$$\mathbf{v}_2 = \mathbf{x}_2 - (\mathbf{q}_1^\top \mathbf{x}_2)\mathbf{q}_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - (\mathbf{q}_1^\top \mathbf{x}_3)\mathbf{q}_1 - (\mathbf{q}_2^\top \mathbf{x}_3)\mathbf{q}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Example (cont'd)

- Then we have

$$R = Q^\top A = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}^\top \begin{bmatrix} 1 & 1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \sqrt{2} & 2\sqrt{2} \\ 0 & -2 & -1 \\ 0 & 0 & \sqrt{3} \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \\ 0 & -1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & 2\sqrt{2} \\ 0 & -2 & -1 \\ 0 & 0 & \sqrt{3} \end{bmatrix}$$

Properties of QR decomposition

- Find orthonormal basis for ran(A) directly
- Let $A = BC$ with $B \in \mathbb{R}^{m \times p}$, $C \in \mathbb{R}^{p \times n}$, $p = \text{rank}(A)$
- To check whether $\mathbf{y} \in \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$: apply Gram-Schmidt procedure to $[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}]$
- If $A = Q_1 R_1$ with rank$(A) = p$, the full QR factorization is

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

  where $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ is orthogonal, i.e., columns of $Q_2 \in \mathbb{R}^{m \times (m-p)}$ are orthonormal, orthogonal to Q_1
- To find Q_2, one can use any matrix $\widetilde{A}$ s.t. $[A \;\; \widetilde{A}]$ is full rank (e.g., $\widetilde{A} = I$), and then apply general Gram-Schmidt process
- Q_1 are orthonormal vectors obtained form columns of A
- Q_2 are orthonormal vectors obtained from extra columns $\widetilde{A}$

Complementary subspaces

- $\text{ran}(Q_1)$ and $\text{ran}(Q_2)$ are complementary subspaces since
  - they are orthogonal (i.e., every vector in the first subspace is orthogonal to every vector in the second subspace)
  - the sum is $\mathbb{R}^m$ (i.e., every vector in $\mathbb{R}^m$ can be expressed as a sum of two vectors, one from each subspace)

$$\text{ran}(Q_1) \overset{\perp}{+} \text{ran}(Q_2) = \mathbb{R}^m \quad \text{ran}(Q_2) = \text{ran}(Q_1)^\perp \quad (\text{ran}(Q_1) = \text{ran}(Q_2)^\perp)$$

- We know $\text{ran}(Q_1) = \text{ran}(A)$; what is its orthogonal complement $\text{ran}(Q_2)$?

Complementary subspaces

- From $A^\top = \begin{bmatrix} R_1^\top & 0 \end{bmatrix} \begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix}$, we have

$$A^\top \mathbf{z} = 0 \iff Q_1^\top \mathbf{z} = 0 \iff \mathbf{z} \in \mathrm{ran}(Q_2)$$

  so $\mathrm{ran}(Q_2) = \mathrm{ran}(A^\top)$ (i.e., columns of Q_2 are an orthonormal basis for $\mathrm{ran}(A^\top)$)

- We conclude: $\mathrm{ran}(A)$ and $\mathrm{null}(A^\top)$ are complementary subspaces
  - $\mathrm{ran}(A) \overset{\perp}{+} \mathrm{null}(A^\top) = \mathbb{R}^m$ (recall $A \in \mathbb{R}^{m \times p}$
  - $\mathrm{ran}(A)^\perp = \mathrm{null}(A^\top)$ (and $\mathrm{null}(A^\top)^\perp = \mathrm{ran}(A)$)
  - called orthogonal decomposition induced by $A \in \mathbb{R}^{m \times p}$

Least squares via QR decomposition

- $A \in \mathbb{R}^{m \times n}$, $A = QR$ with $Q^\top Q = I$, $R \in \mathbb{R}^{n \times n}$
- The pseudo inverse is

$$(A^\top A)^{-1} A^\top = (R^\top Q^\top Q R)^{-1} R^\top Q^\top = R^{-1} Q^\top$$

- The projected point, $\mathbf{x}_{ls} = R^{-1} Q^\top \mathbf{y}$
- The projection matrix

$$A(A^\top A)^{-1} A^\top = A R^{-1} Q^\top = Q Q^\top$$

Least squares via full QR factorization

- Full QR factorization

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

where $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \in \mathbb{R}^{m \times m}$ is orthogonal, $R_1 \in \mathbb{R}^{n \times n}$ is upper triangular and invertible

- Multiplication by orthogonal matrix

$$\begin{aligned} \|A\mathbf{x} - \mathbf{y}\|^2 &= \left\| \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{x} - \mathbf{y} \right\|^2 \\ &= \left\| \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^\top \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{x} - \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^\top \mathbf{y} \right\|^2 \\ &= \left\| \begin{bmatrix} R_1\mathbf{x} - Q_1^\top \mathbf{y} \\ -Q_2^\top \mathbf{y} \end{bmatrix} \right\|^2 \\ &= \| R_1\mathbf{x} - Q_1^\top \mathbf{y} \|^2 + \| Q_2^\top \mathbf{y} \|^2 \end{aligned}$$

- Can be minimized by choosing $\mathbf{x}_{ls} = R_1^{-1} Q_1^\top \mathbf{y}$

Least squares via full QR factorization (cont'd)

- Least squares minimization

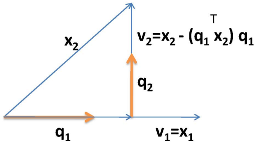
$$\|A\mathbf{x} - \mathbf{y}\|^2 \;=\; \left\| \; R_1\mathbf{x} - Q_1^\top \mathbf{y} \; \right\|^2 + \left\| \; Q_2^\top \mathbf{y} \; \right\|^2$$

- Optimal solution $\mathbf{x}_{ls} = R_1^{-1} Q_1^\top \mathbf{y}$
- Residual of the optimal $\mathbf{x}$ is

$$A\mathbf{x}_{ls} - \mathbf{y} = -Q_2 Q_2^\top \mathbf{y}$$

- $Q_1 Q_1^\top$ gives projection onto $\mathrm{ran}(A)$
- $Q_2 Q_2^\top$ gives projection onto $\mathrm{ran}(A)^\perp$

Gram-Schmidt as triangular orthogonalization



- Steps:
  - $\mathbf{v}_1 = \mathbf{x}_1$
  - $\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$ (normalize)
  - $\mathbf{v}_2 = \mathbf{x}_2 - (\mathbf{q}_1^\top \mathbf{x}_2)\mathbf{q}_1 = (I - \mathbf{q}_1\mathbf{q}_1^\top)\mathbf{x}_2$ (remove $\mathbf{q}_1$ component from $\mathbf{x}_2$)
  - $\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$ (normalize)
  - $\mathbf{v}_3 = \mathbf{x}_3 - (\mathbf{q}_1^\top \mathbf{x}_3)\mathbf{q}_1 - (\mathbf{q}_2^\top \mathbf{x}_3)\mathbf{q}_2 = (I - \mathbf{q}_1\mathbf{q}_1^\top - \mathbf{q}_2\mathbf{q}_2^\top)\mathbf{x}_3 = (I - \mathbf{q}_1\mathbf{q}_1^\top)(I - \mathbf{q}_2\mathbf{q}_2^\top)\mathbf{x}_3$ (remove, $\mathbf{q}_1$, $\mathbf{q}_2$ components)
  - $\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|}$
  - etc.

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}$$

Gram-Schmidt as triangular orthogonalization (cont'd)

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}$$

$$\begin{aligned}
\mathbf{x}_1 &= r_{11}\mathbf{q}_1 \\
\mathbf{x}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 = \tfrac{r_{12}}{r_{11}}\mathbf{x}_1 + r_{22}\mathbf{q}_2 = (\mathbf{q}_1^\top \mathbf{x}_2)\mathbf{q}_1 + r_{22}\mathbf{q}_2 \\
\mathbf{x}_3 &= r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3 = (\mathbf{q}_1^\top \mathbf{x}_3)\mathbf{q}_1 + (\mathbf{q}_2^\top \mathbf{x}_3)\mathbf{q}_2 + r_{33}\mathbf{q}_3 \\
&\vdots \\
\mathbf{x}_n &= r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + \cdots + r_{nn}\mathbf{q}_n
\end{aligned}$$

$$\begin{aligned}
\mathbf{x}_i &= (\mathbf{q}_1^\top \mathbf{x}_i)\mathbf{q}_1 + (\mathbf{q}_2^\top \mathbf{x}_i)\mathbf{q}_2 + \cdots + (\mathbf{q}_{i-1}^\top \mathbf{x}_i)\mathbf{q}_{i-1} + \|\widetilde{\mathbf{q}}_i\|\mathbf{q}_i \\
&= r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i
\end{aligned}$$

$$r_{ij} = \mathbf{q}_i^\top \mathbf{x}_j \quad (i \neq j)$$

Gram-Schmidt as triangular orthogonalization (cont'd)

- Gram-Schmidt process

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}$$

- At the j-th step, want to find a unit vector $\mathbf{q}_j \in \{\mathbf{x}_1, \ldots, \mathbf{x}_j\}$ that is orthogonal to $\mathbf{q}_1, \ldots \mathbf{q}_{j-1}$

$$\begin{aligned}
\mathbf{v}_j &= \mathbf{x}_j - (\mathbf{q}_1^\top \mathbf{x}_j)\mathbf{q}_1 - (\mathbf{q}_2^\top \mathbf{x}_j)\mathbf{q}_2 - \cdots - (\mathbf{q}_{j-1}^\top \mathbf{x}_j)\mathbf{q}_{j-1} \\
\mathbf{q}_1 &= \frac{\mathbf{x}_1}{r_{11}} \\
\mathbf{q}_2 &= \frac{\mathbf{x}_2 - r_{12}\mathbf{q}_1}{r_{22}} = \frac{(I - \mathbf{q}_1\mathbf{q}_1^\top)\mathbf{x}_2}{\|(I - \mathbf{q}_1\mathbf{q}_1^\top)\mathbf{x}_2\|} = \frac{P_2\mathbf{x}_2}{\|P_2\mathbf{x}_2\|} \\
&\vdots \\
\mathbf{q}_n &= \frac{\mathbf{x}_n - \sum_{i=1}^{n-1} r_{in}\mathbf{q}_i}{r_{nn}} = \frac{(I - Q_{n-1}Q_{n-1}^\top)\mathbf{x}_n}{\|(I - Q_{n-1}Q_{n-1}^\top)\mathbf{x}_n\|} = \frac{P_n\mathbf{x}_n}{\|P_n\mathbf{x}_n\|} \\
r_{ij} &= \mathbf{q}_i^\top \mathbf{x}_j \quad (i \neq j) \\
|r_{jj}| &= \|\mathbf{x}_j - \textstyle\sum_{i=1}^{j} r_{ij}\mathbf{q}_i\|_2 \quad \text{(often choose } r_{jj} > 0)
\end{aligned}$$

Gram-Schmidt as triangular orthogonalization (cont'd)

- Let $A \in \mathbb{R}^{m\times n}, m \geq n$ be a matrix of full rank with columns $\mathbf{x}_i$, consider the sequence of formulas

$$\mathbf{q}_1 = \frac{P_1\mathbf{x}_1}{\|P_1\mathbf{x}_1\|},\ \mathbf{q}_2 = \frac{P_2\mathbf{x}_2}{\|P_2\mathbf{x}_2\|},\ \ldots,\ \mathbf{q}_n = \frac{P_n\mathbf{x}_n}{\|P_n\mathbf{x}_n\|}$$

  where $P_j \in \mathbb{R}^{m\times m}$ of rank $m-(j-1)$ projects $\mathbf{x}$ onto the space orthogonal to $\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}$ ($P_1 = I$ when $j = 1$)
- Projection P_j can be represented explicitly
- Let Q_{j-1} denote the $m \times (j-1)$ matrix containing the first $j-1$ columns of Q

$$Q_{j-1} = \begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_{j-1}\end{bmatrix}$$

  then

$$P_j = I - Q_{j-1}Q_{j-1}^\top$$

Modified Gram-Schmidt process

- Recall for rank-one orthogonal projection with $\mathbf{q} \in \mathbb{R}^m$

$$P_{\mathbf{q}} = \frac{\mathbf{q}\mathbf{q}^\top}{\mathbf{q}^\top\mathbf{q}}$$

- The complements are rank $m-1$ orthogonal projections

$$P_{\perp\mathbf{q}} = I - \frac{\mathbf{q}\mathbf{q}^\top}{\mathbf{q}^\top\mathbf{q}}$$

- By definition of P_j

$$P_j = P_{\perp\mathbf{q}_{j-1}} \cdots P_{\perp\mathbf{q}_2} P_{\perp\mathbf{q}_1}$$

where $P_1 = I$, and thus

$$\mathbf{v}_j = P_{\perp\mathbf{q}_{j-1}} \cdots P_{\perp\mathbf{q}_2} P_{\perp\mathbf{q}_1} \mathbf{x}_j$$

Modified Gram-Schmidt process (cont'd)

- For numerical stability, evaluating the following formulas in order (for now consider an algorithm is considered as stable if it is not too sensitive to the effects of rounding errors)

$$
\begin{aligned}
\mathbf{v}_j^{(1)} &= \mathbf{x}_j, \\
\mathbf{v}_j^{(2)} &= P_{\perp \mathbf{q}_1} \mathbf{v}_j^{(1)} &&= \mathbf{v}_j^{(1)} - \mathbf{q}_1 \mathbf{q}_1^\top \mathbf{v}_j^{(1)}, \\
\mathbf{v}_j^{(3)} &= P_{\perp \mathbf{q}_2} \mathbf{v}_j^{(2)} &&= \mathbf{v}_j^{(2)} - \mathbf{q}_2 \mathbf{q}_2^\top \mathbf{v}_j^{(2)}, \\
&\vdots &&\vdots \\
\mathbf{v}_j = \mathbf{v}_j^{(j)} &= P_{\perp \mathbf{q}_{j-1}} \mathbf{v}_j^{(j-1)} &&= \mathbf{v}_j^{(j-1)} - \mathbf{q}_{j-1} \mathbf{q}_{j-1}^\top \mathbf{v}_j^{(j-1)}
\end{aligned}
$$

- Projection matrix $P_{\perp \mathbf{q}_i}$ can be applied to $\mathbf{v}_j^{(i)}$ for each $j > i$ immediately after $\mathbf{q}_i$ is known
- $\mathbf{v}_1 = \mathbf{v}_1^{(1)} = \mathbf{x}_1$, $\mathbf{q}_1 = \frac{1}{r_{11}} \mathbf{v}_1^{(1)}$
- $\mathbf{v}_2 = \mathbf{v}_2^{(2)} = \mathbf{v}_2^{(1)} - \mathbf{q}_1 \mathbf{q}_1^\top \mathbf{v}_2^{(1)} = \mathbf{v}_2^{(1)} - (\frac{1}{r_{11}} \mathbf{v}_1^{(1)}) r_{12}$
- $\mathbf{v}_3 = \mathbf{v}_3^{(3)} = \mathbf{v}_3^{(2)} - \mathbf{q}_2 \mathbf{q}_2^\top \mathbf{v}_3^{(2)}$,
  $\mathbf{v}_3^{(2)} = \mathbf{v}_3^{(1)} - \mathbf{q}_1 \mathbf{q}_1^\top \mathbf{v}_3^{(1)} = \mathbf{v}_3^{(1)} - (\frac{1}{r_{11}} \mathbf{v}_1^{(1)}) r_{13}$

Modified Gram-Schmidt algorithm

- Steps

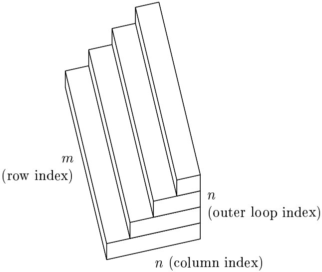
$$\begin{aligned} \mathbf{v}_j^{(1)} &= \mathbf{x}_j, \\ &\vdots \qquad\qquad \vdots \\ \mathbf{v}_j = \mathbf{v}_j^{(j)} &= P_{\perp \mathbf{q}_{j-1}} \mathbf{v}_j^{(j-1)} = \mathbf{v}_j^{(j-1)} - \mathbf{q}_{j-1}\mathbf{q}_{j-1}^\top \mathbf{v}_j^{(j-1)} \end{aligned}$$

- Algorithm:

for $i = 1$ to n **do**
  $\mathbf{v}_i = \mathbf{x}_i$
end for
for $i = 1$ to n **do**
  $r_{ii} = \|\mathbf{v}_i\|$
  $\mathbf{q}_i = \frac{\mathbf{v}_i}{r_{ii}}$
  for $j = i + 1$ to n **do**
    $r_{ij} = \mathbf{q}_i^\top \mathbf{v}_j$
    $\mathbf{v}_j = \mathbf{v}_j - r_{ij}\mathbf{q}_i$
  end for
end for

Complexity of Gram-Schmidt process

- Algorithmically, complexity of general or modified Gram-Schmidt process requires $\sim 2mn^2$ flops for an $m \times n$ matrix
- Geometrically, it can be easily seen that the complexity is $\sim 2mn^2$ flops



- As $m, n \to \infty$, the figure converges to a right triangular prism, with volume $mn^2/2$, multiply four flops per unit volume gives $2mn^2$ flops

Gram-Schmidt as triangular orthogonalization

- The projection matrix $P_j = I - Q_{j-1}Q_{j-1}^\top$ can be represented explicitly

$$
\begin{aligned}
\mathbf{x}_1 &= r_{11}\mathbf{q}_1 \ (\mathbf{q}_1 = \tfrac{1}{r_{11}}\mathbf{x}_1) \\
\mathbf{x}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 = \tfrac{r_{12}}{r_{11}}\mathbf{x}_1 + r_{22}\mathbf{q}_2 \ (\mathbf{q}_2 = \tfrac{1}{r_{22}}(\mathbf{x}_2 - \tfrac{r_{12}}{r_{11}}\mathbf{x}_1)) \\
\mathbf{x}_3 &= r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3 = \tfrac{r_{13}}{r_{11}}\mathbf{x}_1 + r_{22}\mathbf{q}_2 + r_{33}\mathbf{q}_3 \\
&\vdots \\
\mathbf{x}_i &= r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i
\end{aligned}
$$

- The first iteration multiply the first column $\mathbf{x}_1$ by $1/r_{11}$ and then subtract r_{1j} times the result from each of the remaining columns $\mathbf{x}_j$
- Equivalent to right-multiplication by a matrix R_1

$$
\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}
\begin{bmatrix} \frac{1}{r_{11}} & \frac{-r_{12}}{r_{11}} & \frac{-r_{13}}{r_{11}} & \cdots \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{bmatrix}
= \begin{bmatrix} \mathbf{q}_1 & \mathbf{v}_2^{(2)} & \cdots & \mathbf{v}_n^{(2)} \end{bmatrix}
$$

Gram-Schmidt triangular orthogonalization (cont'd)

- At step i the modified Gram-Schmidt algorithm subtracts r_{ij}/r_{ii} times column i of the current A from columns $j > i$ and replaces column i by $1/r_{ii}$ times itself
- This corresponds to multiplication by an upper triangular matrix R_i

$$R_2 = \begin{bmatrix} 1 & & & \\ & \frac{1}{r_{22}} & \frac{-r_{23}}{r_{22}} & \cdots \\ & & 1 & \\ & & & \ddots \end{bmatrix}, \quad R_3 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \frac{1}{r_{33}} & \frac{-r_{34}}{r_{33}} & \cdots \\ & & & & \ddots \end{bmatrix}, \quad \cdots$$

- At the end, we have

$$A \underbrace{R_1 R_2 \cdots R_n}_{R^{-1}} = Q$$

- This shows that the Gram-Schmidt algorithm is a method of triangular orthogonalization