QR factorization

- Gram-Schmidt procedure, QR factorization
- orthogonal decomposition induced by a matrix

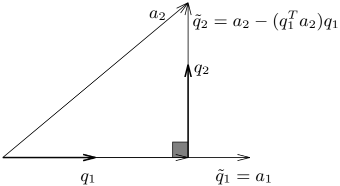
Gram-Schmidt procedure

given independent vectors $a_1, \ldots, a_n \in \mathbb{R}^m$, G-S procedure finds orthonormal vectors $q_1, \ldots, q_n$ s.t.

$$\mathbf{span}(a_1, \ldots, a_r) = \mathbf{span}(q_1, \ldots, q_r) \qquad \text{for } r \leq n$$

- thus, $q_1, \ldots, q_r$ is an orthonormal basis for $\mathbf{span}(a_1, \ldots, a_r)$
- rough idea of method: first *orthogonalize* each vector w.r.t. previous ones; then *normalize* result to have norm one

- step 1a. $\tilde{q}_1 := a_1$
- step 1b. $q_1 := \tilde{q}_1 / \|\tilde{q}_1\|$ *(normalize)*
- step 2a. $\tilde{q}_2 := a_2 - (q_1^\mathsf{T} a_2) q_1$ *(remove q_1 component from a_2)*
- step 2b. $q_2 := \tilde{q}_2 / \|\tilde{q}_2\|$ *(normalize)*
- step 3a. $\tilde{q}_3 := a_3 - (q_1^\mathsf{T} a_3) q_1 - (q_2^\mathsf{T} a_3) q_2$ *(remove q_1, q_2 components)*
- step 3b. $q_3 := \tilde{q}_3 / \|\tilde{q}_3\|$ *(normalize)*
- etc.



for $i = 1, 2, \ldots, n$ we have

$$\begin{aligned} a_i &= (q_1^\mathsf{T} a_i) q_1 + (q_2^\mathsf{T} a_i) q_2 + \cdots + (q_{i-1}^\mathsf{T} a_i) q_{i-1} + \|\tilde{q}_i\| q_i \\ &= r_{1i} q_1 + r_{2i} q_2 + \cdots + r_{ii} q_i \end{aligned}$$

(note that the r_{ij}'s come right out of the G-S procedure, and $r_{ii} \neq 0$)

QR decomposition

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

$$\underbrace{\left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array}\right]}_{A} = \underbrace{\left[\begin{array}{cccc} q_1 & q_2 & \cdots & q_n \end{array}\right]}_{Q} \underbrace{\left[\begin{array}{cccc} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{array}\right]}_{R}$$

- $Q^\mathsf{T} Q = I$, and R is upper triangular & invertible
- called QR **decomposition** (or factorization) of A
- usually computed using a variation on Gram-Schmidt procedure which is less sensitive to numerical (rounding) errors
- columns of Q are orthonormal basis for **range**(A)

General Gram-Schmidt procedure

- in basic G-S we assume $a_1, \ldots, a_n \in \mathbb{R}^m$ are independent
- if $a_1, \ldots, a_n$ are dependent, we find $\tilde{q}_j = 0$ for some j, which means a_j is linearly dependent on $a_1, \ldots, a_{j-1}$
- modified algorithm: when we encounter $\tilde{q}_j = 0$, skip to next vector a_{j+1} and continue:

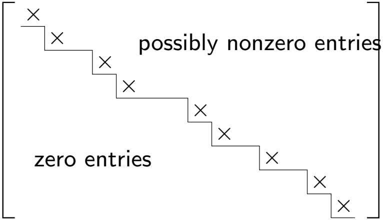
$$
\begin{array}{l}
r = 0 \\
\textbf{for } i = 1, \ldots, n \\
\quad \tilde{a} = a_i - \sum_{j=1}^{r} q_j q_j^T a_i \\
\quad \textbf{if } \tilde{a} \neq 0 \\
\quad\quad r = r + 1 \\
\quad\quad q_r = \tilde{a}/\|\tilde{a}\|
\end{array}
$$

Staircase form

on exit,

- $q_1, \ldots, q_r$ is an orthonormal basis for **range**(A) (hence $r = $ **Rank**(A))
- each a_i is linear combination of previously generated q_j's

in matrix notation we have $A = QR$ with $Q^T Q = I$ and $R \in \mathbb{R}^{r \times n}$ in *upper staircase form*



'corner' entries (shown as $\times$) are nonzero

Applications

- directly yields orthonormal basis for **range**(A)
- yields factorization $A = BC$ with $B \in \mathbb{R}^{m \times r}$, $C \in \mathbb{R}^{r \times n}$, $r = \textbf{Rank}(A)$
- to check if $b \in \textbf{span}(a_1, \ldots, a_n)$, apply Gram-Schmidt to $\left[\begin{array}{cccc} a_1 & \cdots & a_n & b \end{array}\right]$
- staircase pattern in R shows which columns of A are dependent on previous ones

works incrementally: one G-S procedure yields QR factorizations of $\left[\begin{array}{ccc} a_1 & \cdots & a_p \end{array}\right]$ for $p = 1, \ldots, n$

$$\left[\begin{array}{ccc} a_1 & \cdots & a_p \end{array}\right] = \left[\begin{array}{ccc} q_1 & \cdots & q_s \end{array}\right] R_p$$

where $s = \textbf{Rank}(\left[\begin{array}{ccc} a_1 & \cdots & a_p \end{array}\right])$ and R_p is leading $s \times p$ submatrix of R

'Full' QR factorization

with $A = Q_1 R_1$ the QR factorization as above, write

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

where $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ is orthogonal, $i.e.$, columns of $Q_2 \in \mathbb{R}^{m \times (m-r)}$ are orthonormal, orthogonal to Q_1

to find Q_2:

- find any matrix $\tilde{A}$ s.t. $\begin{bmatrix} A & \tilde{A} \end{bmatrix}$ has rank m ($e.g.$, $\tilde{A} = I$)
- apply general Gram-Schmidt to $\begin{bmatrix} A & \tilde{A} \end{bmatrix}$
- Q_1 are orthonormal vectors obtained from columns of A
- Q_2 are orthonormal vectors obtained from extra columns ($\tilde{A}$)

$i.e.$, any set of orthonormal vectors can be *extended* to an orthonormal basis for $\mathbb{R}^m$

Permutation

can permute columns with $\times$ to front of matrix:

$$\boxed{A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} P}$$

where:

- $Q^{\mathsf{T}} Q = I$
- $R_{11} \in \mathbb{R}^{r \times r}$ is upper triangular and invertible
- $P \in \mathbb{R}^{n \times n}$ is a permutation matrix
  (which moves forward the columns of a which generated a new q)

Complementary subspaces

if $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ and Q is orthogonal then **range**(Q_1) and **range**(Q_2) are called *complementary subspaces*, because

$$\textbf{range}(Q_2) = \textbf{range}(Q_1)^{\perp}$$

- they are orthogonal $i.e.$, every vector in the first subspace is orthogonal to every vector in the second subspace
- every vector in $\mathbb{R}^m$ can be expressed as a sum of two vectors, one from each subspace
- each subspace is the *orthogonal complement* of the other

Orthogonal decomposition induced by A

$$\textbf{range}(A)^{\perp} = \textbf{null}(A^{\mathsf{T}})$$

- from $A^{\mathsf{T}} = \left[\begin{array}{cc} R_1^{\mathsf{T}} & 0 \end{array}\right] \left[\begin{array}{c} Q_1^{\mathsf{T}} \\ Q_2^{\mathsf{T}} \end{array}\right]$ we see that

$$A^{\mathsf{T}} z = 0 \quad \Longleftrightarrow \quad Q_1^{\mathsf{T}} z = 0 \quad \Longleftrightarrow \quad z \in \textbf{range}(Q_2)$$

  so $\textbf{range}(Q_2) = \textbf{null}(A^{\mathsf{T}})$
- the columns of Q_2 are an orthonormal basis for $\textbf{null}(A^{\mathsf{T}})$
- called *orthogonal decomposition (of $\mathbb{R}^m$) induced by $A \in \mathbb{R}^{m \times n}$*
- every $y \in \mathbb{R}^n$ can be written uniquely as $y = z + w$, with $z \in \textbf{range}(A)$, $w \in \textbf{null}(A^{\mathsf{T}})$ (we'll soon see what the vector z is ...)