$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}_{m\times n}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m\times 1}}_{b}$$

CX1104: **Linear Algebra for Computing**

Chap. No : **6.3.2**

Lecture : **Orthogonality**

Topic : **Gram–Schmidt Process**

Concept : **Gram-Schmidt Process for QR decomposition**

Instructor: **A/P Chng Eng Siong**

TAs: **Zhang Su, Vishal Choudhari**

Rev: 25 Sep 2021

Rev: 29th June 2020

QR Factorisation revisited

The QR decomposition can be performed by Gram–Schmidt. Given a matrix A (mxn sized),

$$A = QR$$

$$A = [q_1 \quad q_2 \quad \cdots \quad q_n] \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix}$$

The Q Factor (economy qr):

- Q is $m \times n$ with orthonormal columns and $Q^T Q = I$ dimension nxn
- If A is square $(m = n)$, then Q is orthogonal, i.e, $Q^T Q = QQ^T = I$
- If A is tall (m>n), then $QQ^T \neq I$,
  The matrix QQ^T is a projection matrix
  of dimension mxm, and it will project a vector R^m onto
  the columns space of A. In other words,
  $QQ^T y = \hat{y}$, where $\hat{y}$ is the least squares error approximation of y in the
  column space of A (see sec 7.1.4)

The R Factor:

- R is $n \times n$ upper triangular,
- If A has independent column, then R is invertible,
  else R is singular (not-invertible)

- Vectors $q_1, q_2, \dots, q_n$ are orthonormal m-dimensional vectors: $||q_i|| = 1 \quad and \quad q_i^T q_j = 0 \ if \ \ i \neq j$

NOTE:

Q is obtained by performing GS Process on A

The Gram Schmidt Process

What does Gram-Schmidt Process Do?

It orthogonalises a set of vectors!

$W = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$
$\boldsymbol{x}_i \in R^m$

Gram-Schmidt Process

$W' = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$
$\boldsymbol{v}_i \in R^m$ and
$\boldsymbol{v}_i . \boldsymbol{v}_j = 0$ when $i \neq j$

$$Span(W) = Span(W')$$

Typically, we are give a matrix A, and these x_i are columns of A.

The Gram Schmidt Process

In basic Gram-Schmidt, we assume that $\{x_1, x_2, ..\}$ are independent columns,

THEOREM 11

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

$$\vdots$$

$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \qquad (1)$$

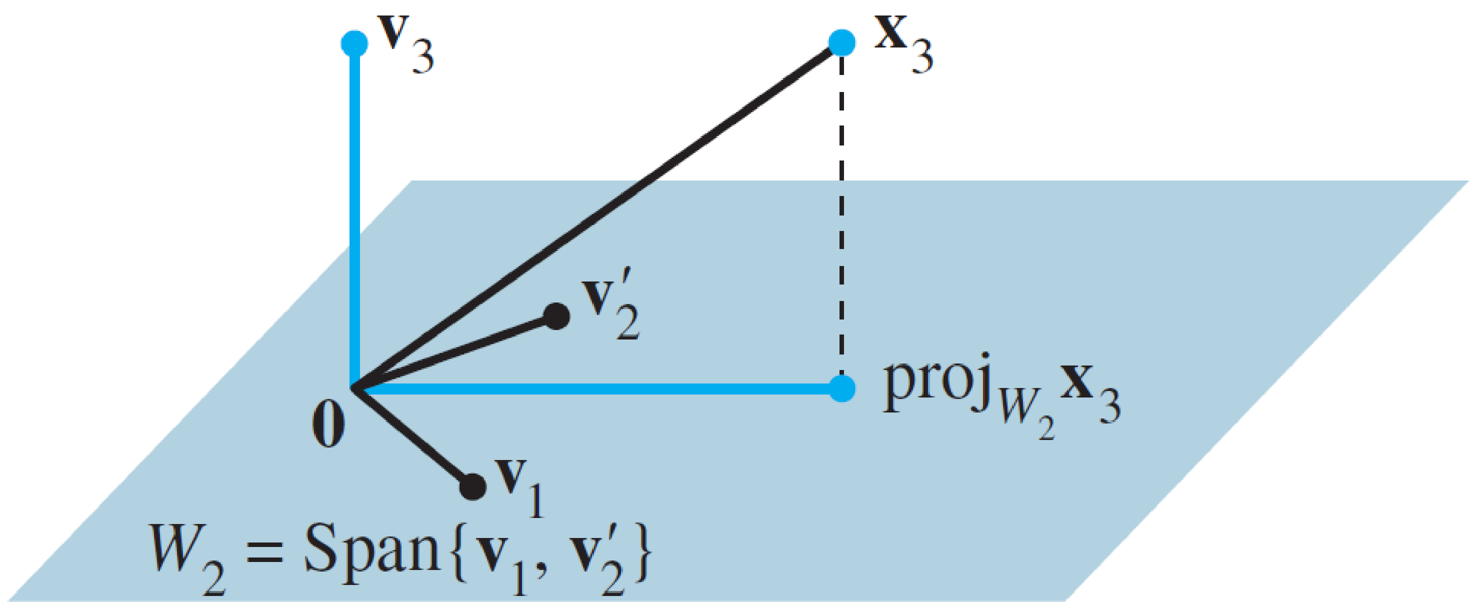


FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

Watch these worked out examples:
1. GramSchmidt: https://www.youtube.com/watch?v=Aslf3KGq2UE
2. QR: https://www.youtube.com/watch?v=6DybLNNkWyE
3. MIT Gram Schmidt: https://www.youtube.com/watch?v=TRktLuAktBQ&t=17s

MIT OpenCourseWare
2.46M subscribers

Gram-Schmidt Orthogonalization
Instructor: Ana Rita Pires
View the complete course: http://ocw.mit.edu/18-06SCF11

Worldwide Center of Mathematics
26.5K subscribers

Lay, Linear Algebra and its Applications (4th Edition)

354 CHAPTER 6 Orthogonality and Least Squares

Proof Theorem 11: Span of vectors generated by GS is same as original set of vectors

PROOF For $1 \leq k \leq p$, let $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. Set $\mathbf{v}_1 = \mathbf{x}_1$, so that $\text{Span}\{\mathbf{v}_1\} = \text{Span}\{\mathbf{x}_1\}$. Suppose, for some $k < p$, we have constructed $\mathbf{v}_1, \ldots, \mathbf{v}_k$ so that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W_k. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1} \tag{2}$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Note that $\text{proj}_{W_k} \mathbf{x}_{k+1}$ is in W_k and hence also in W_{k+1}. Since $\mathbf{x}_{k+1}$ is in W_{k+1}, so is $\mathbf{v}_{k+1}$ (because W_{k+1} is a subspace and is closed under subtraction). Furthermore, $\mathbf{v}_{k+1} \neq \mathbf{0}$ because $\mathbf{x}_{k+1}$ is not in $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. Hence $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal set of nonzero vectors in the $(k+1)$-dimensional space W_{k+1}. By the Basis Theorem in Section 4.5, this set is an orthogonal basis for W_{k+1}. Hence $W_{k+1} = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$. When $k + 1 = p$, the process stops. ■

Theorem 11 shows that any nonzero subspace W of $\mathbb{R}^n$ has an orthogonal basis, because an ordinary basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ is always available (by Theorem 11 in Section 4.5), and the Gram–Schmidt process depends only on the existence of orthogonal projections onto subspaces of W that already have orthogonal bases.

Example:

EXAMPLE 2 Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

SOLUTION

Step 1. Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{Span}\{\mathbf{x}_1\} = \text{Span}\{\mathbf{v}_1\}$.

Step 2. Let $\mathbf{v}_2$ be the vector produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is, let

$$\begin{aligned}
\mathbf{v}_2 &= \mathbf{x}_2 - \text{proj}_{W_1} \mathbf{x}_2 \\
&= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \qquad \text{Since } \mathbf{v}_1 = \mathbf{x}_1 \\
&= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}
\end{aligned}$$

As in Example 1, $\mathbf{v}_2$ is the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

Note: $v_2' = v_2 * 4$ to get rid of denominator in v_2

Step 3. Let $\mathbf{v}_3$ be the vector produced by subtracting from $\mathbf{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$ to compute this projection onto W_2:

$$\text{proj}_{W_2} \mathbf{x}_3 = \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_1} + \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2'}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_2'} = \frac{2}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

Then $\mathbf{v}_3$ is the component of $\mathbf{x}_3$ orthogonal to W_2, namely,

$$\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{W_2} \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

See Slide 3 of Chapter 6.2.5 for explanation.

Lay, Linear Algebra and its Applications (4th Edition)

Example:

$$\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{W_2}\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

See Fig. 2 for a diagram of this construction. Observe that $\mathbf{v}_3$ is in W, because $\mathbf{x}_3$ and $\text{proj}_{W_2}\mathbf{x}_3$ are both in W. Thus $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal set of nonzero vectors and hence a linearly independent set in W. Note that W is three-dimensional since it was defined by a basis of three vectors. Hence, by the Basis Theorem in Section 4.5, $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal basis for W. ■

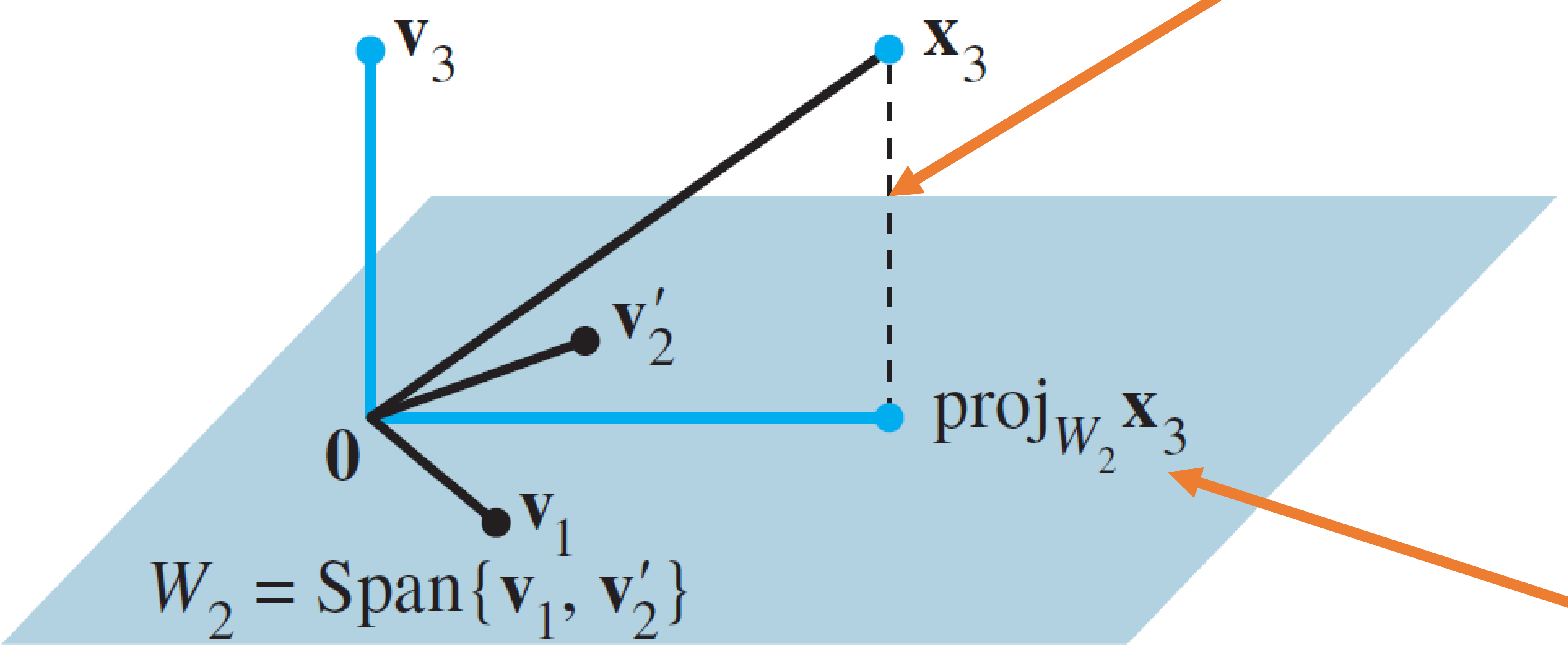


FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

$$\text{proj}_{W_2}\mathbf{x}_3 = \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1}_{\substack{\text{Projection of} \\ \mathbf{x}_3 \text{ onto } \mathbf{v}_1}} + \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2'}_{\substack{\text{Projection of} \\ \mathbf{x}_3 \text{ onto } \mathbf{v}_2'}} = \frac{2}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

The Basis Theorem

Let V be a p-dimensional vector space, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.

Lay, Linear Algebra and its Applications (4th Edition)

Orthogonal basis vs Orthonormal basis

The columns of Q
are orthogonal,
as well as orthonormal!
$Q^T Q = I$

Orthonormal == orthogonal +
(length of vector ==1)

Orthonormal Bases

An orthonormal basis is constructed easily from an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$: simply normalize (i.e., "scale") all the $\mathbf{v}_k$. When working problems by hand, this is easier than normalizing each $\mathbf{v}_k$ as soon as it is found (because it avoids unnecessary writing of square roots).

EXAMPLE 3 Example 1 constructed the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

An orthonormal basis is

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{45}}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \blacksquare$$

The entries of R matrix during GS

Consider A has independent col (mxn matrix)

$$A = [\, x_1, x_2, x_3, \dots, x_n]$$

Gram–Schmidt Process

$$\mathrm{Proj}_v x = \frac{v.x}{v.v}\, v$$

$$v_1 = x_1$$

$$v_2 = x_2 - \mathrm{Proj}_{v_1} x_2$$

$$v_3 = x_3 - \mathrm{Proj}_{v_1} x_3 - \mathrm{Proj}_{v_2} x_3$$

$$v_i = x_i - \sum_{j=1}^{i-1} \mathrm{Proj}_{v_j} x_i$$

Ortho-normalization

$$u_1 = \frac{v_1}{||v_1||}$$

$$u_2 = \frac{v_2}{||v_2||}$$

$$u_i = \frac{v_i}{||v_i||}$$

We can now express the x_i over our newly computed orthonormal basis: e_j

$$x_1 = u_1.x_1\ u_1$$

$$x_2 = u_1.x_2\ u_1 + u_2.x_2\ u_2$$

$$x_3 = u_1.x_3\ u_1 + u_2.x_3\ u_2 + u_3.x_3\ u_3$$

$$\vdots$$

$$x_n = \sum_{j=1}^{n} u_j\,.x_n\ \ u_j$$

This can be written in matrix form:

$$A = QR$$

where:

$$Q = [\, u_1, u_2, u_3, \dots, u_n]$$

and

$$R = \begin{pmatrix} u_1.x_1 & u_1.x_2 & u_1.x_3 & \cdots \\ 0 & u_2\,.x_2 & u_2\,.x_3 & \cdots \\ 0 & 0 & u_3\,.x_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

https://www.math.ucla.edu/~yanovsky/Teaching/Math151B/handouts/GramSchmidt.pdf

Example: Gram–Schmidt on a 3x3 matrix

Example [edit]

Consider the decomposition of

$$A = \begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix}.$$

Recall that an orthonormal matrix Q has the property

$$Q^{\mathsf{T}} Q = I.$$

Then, we can calculate Q by means of Gram–Schmidt as follows:

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix} = \begin{pmatrix} 12 & -69 & -58/5 \\ 6 & 158 & 6/5 \\ -4 & 30 & -33 \end{pmatrix};$$

$$Q = \begin{pmatrix} \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} & \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} & \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \end{pmatrix} = \begin{pmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{pmatrix}$$

Ref: https://en.wikipedia.org/wiki/QR_decomposition#Using_the_Gram–Schmidt_process

Thus, we have

$$Q^{\mathsf{T}} A = Q^{\mathsf{T}} Q\, R = R;$$

$$R = Q^{\mathsf{T}} A = \begin{pmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{pmatrix}$$

Hence,

$$A = \begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix} =$$

$$\begin{pmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{pmatrix} \times \begin{pmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{pmatrix}$$

Q: Orthogonal Matrix $\quad$ R: Upper Triangular Matrix

Warning: modify QR when A has dependent column!

From: pg 4-6 http://ee263.stanford.edu/lectures/qr.pdf

Note: here the columns of Q are denoted as q_i

QR decomposition

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

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

- 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:

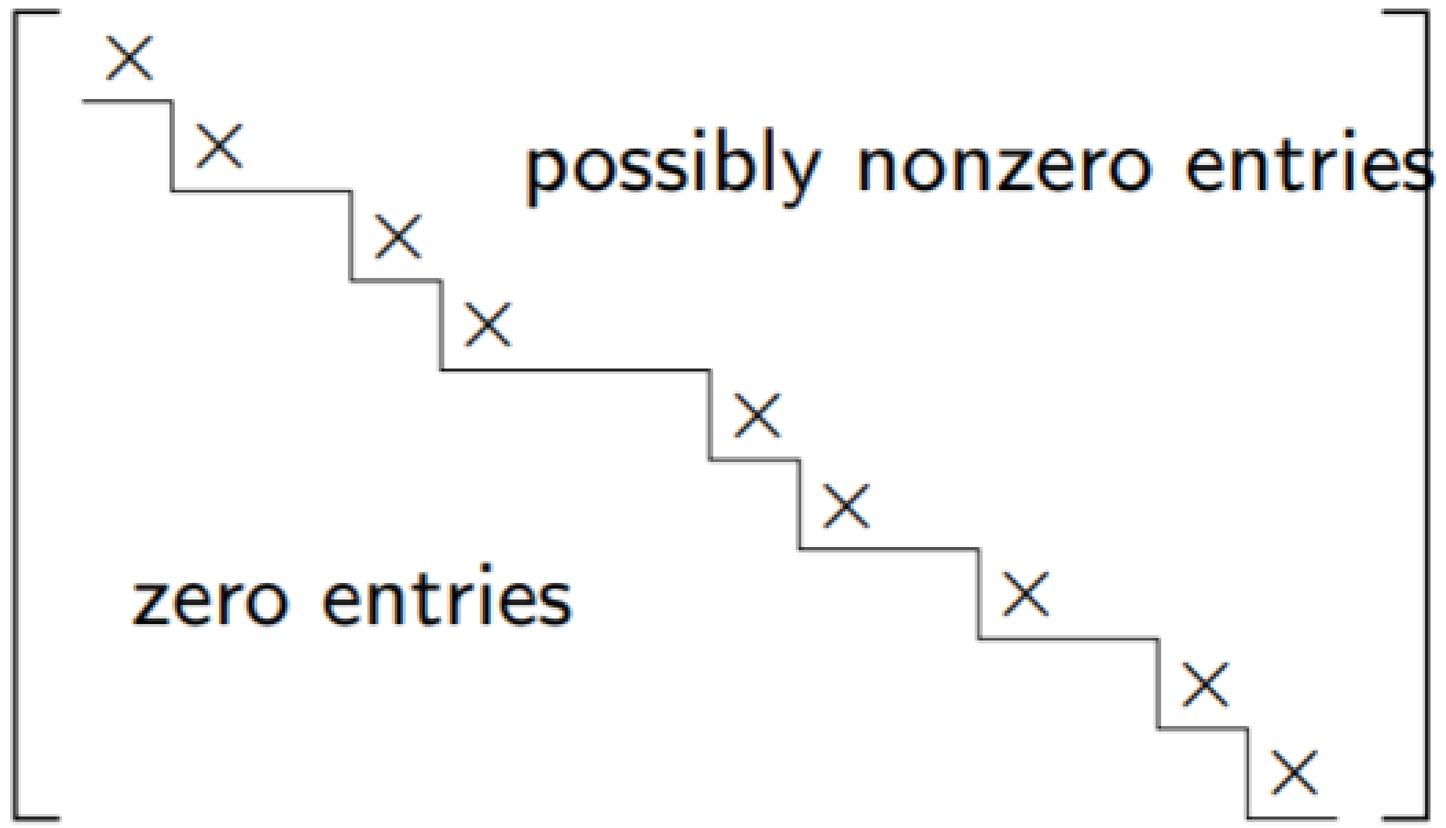
$$
\begin{aligned}
& r = 0 \\
& \textbf{for } i = 1, \ldots, n \\
& \qquad \tilde{a} = a_i - \textstyle\sum_{j=1}^{r} q_j q_j^\top a_i \\
& \qquad \textbf{if } \tilde{a} \neq 0 \\
& \qquad\qquad r = r + 1 \\
& \qquad\qquad q_r = \tilde{a}/\|\tilde{a}\|
\end{aligned}
$$

Warning: modify QR when A has dependent column!

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 $Q^\mathsf{T} Q = I$ and $R \in \mathbb{R}^{r \times n}$ in *upper staircase form*



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

From: pg 4-6 http://ee263.stanford.edu/lectures/qr.pdf

How to get full QR decomposition when A is tall and skinny?

'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 $\mathbb{R}^m$

look ahead: QQ^T relationship to Least Squares

Ref: relating QQ^T to least squares solution (see Sec 7.1.4)
and Boyd's lecture:
https://see.stanford.edu/materials/lsoeldsee263/05-ls.pdf Pg 5-8

Least-squares via QR factorization

- $A \in \mathbf{R}^{m\times n}$ skinny, full rank
- factor as $A = QR$ with $Q^TQ = I_n$, $R \in \mathbf{R}^{n\times n}$ upper triangular, invertible
- pseudo-inverse is

$$(A^TA)^{-1}A^T = (R^TQ^TQR)^{-1}R^TQ^T = R^{-1}Q^T$$

so $x_{\mathrm{ls}} = R^{-1}Q^Ty$

- projection on $\mathcal{R}(A)$ given by matrix

$$A(A^TA)^{-1}A^T = AR^{-1}Q^T = QQ^T$$

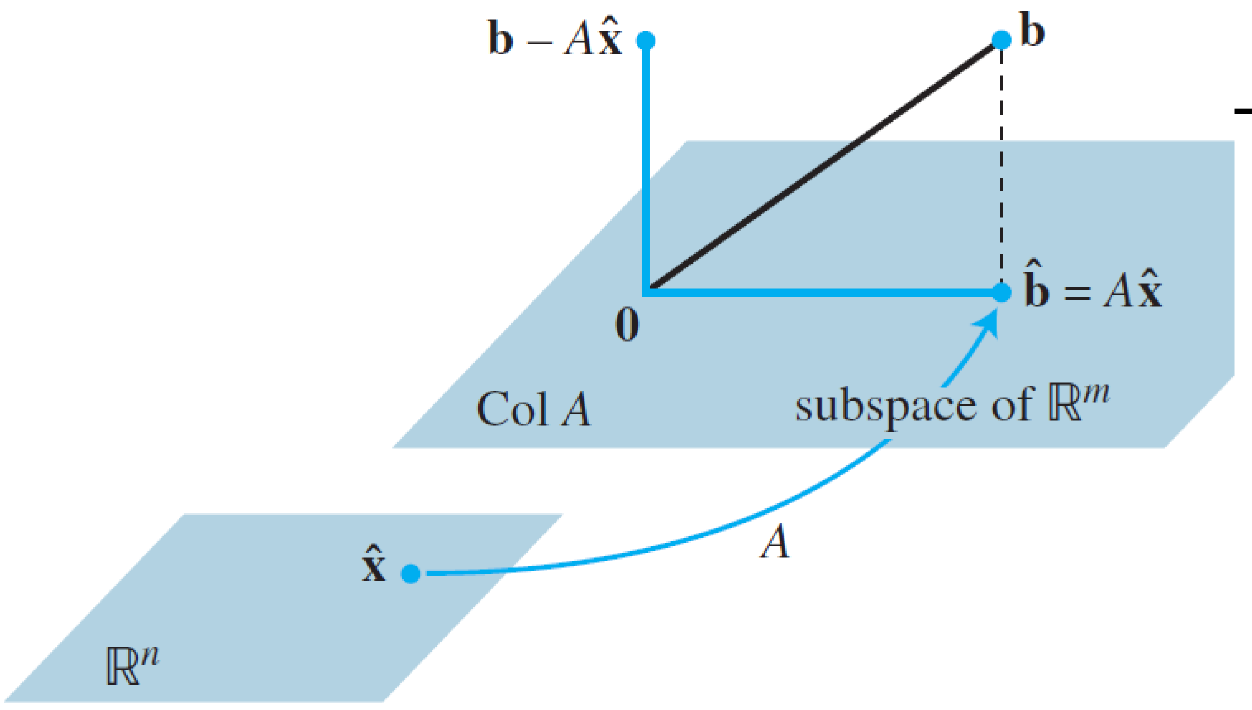


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

THEOREM 14

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.
b. The columns of A are linearly indpendent.
c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

Therefore if $\hat{\mathrm{x}} = (A^TA)^{-1}A^Tb$, then
$\hat{b} = A\hat{x}$, means
$\hat{b} = A(A^TA)^{-1}A^T\ b$

And this matrix $A(A^TA)^{-1}A^T$ is a projection matrix, projecting b into the column space of A.

See 7.1.3 (pg 4) : projection matrix of least squares