Lecture 6

One important type of matrices is called projector.

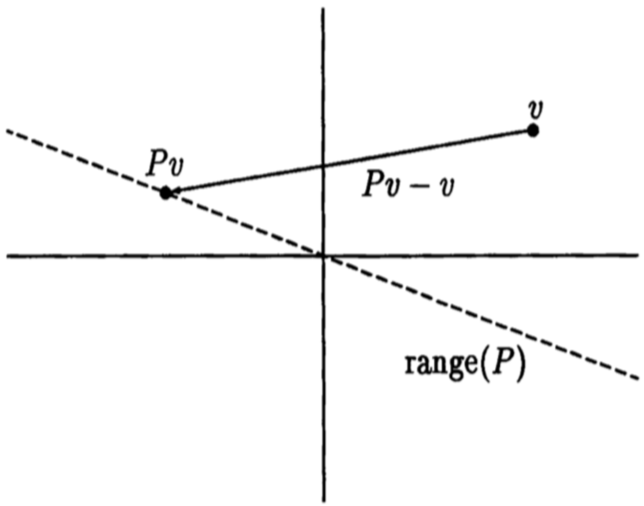
Definition: A projector (or oblique projector) is a square matrix P such that

$$P^2 = P$$

Elementary properties:

- if $v \in \text{range}(P)$, then $v = Px$, so that $Pv = P(Px) = Px = v$
- for any v, we have $P(Pv - v) = P^2v - Pv = 0$, so that $Pv - v \in \text{null}(P)$

Illustration:



Note that the projector P projects all vectors v to the space range(P).

Definition: Let P be a projector. Then $I - P$ is also a projector since

$$(I - P)^2 = I - 2P + P^2 = I - P$$

We call $I - P$ the complementary projector to P.

Elementary properties:

- $\text{range}(I - P) = \text{null}(P)$, namely, the projector $I - P$ projects all vectors to the space $\text{null}(P)$
- $\text{null}(I - P) = \text{range}(P)$
- $\text{range}(P) \cap \text{null}(P) = \{0\}$, namely, the projector separates $\mathbb{C}^m$ into two subspaces

To see the first property, if $v \in \text{null}(P)$, then $(I - P)v = v$, so $v \in \text{range}(I - P)$. On the other hand, if $v \in \text{range}(I - P)$, then $v = (I - P)x$, so that $Pv = P(I - P)x = Px - P^2x = 0$, which implies $v \in \text{null}(P)$.

To see the third property, if $v \in \text{range}(P) \cap \text{null}(P)$, then

$$\begin{aligned} v &= v - Pv && \text{since } v \in \text{null}(P) \\ &= (I - P)v \\ &= 0 && \text{since } v \in \text{range}(P) = \text{null}(I - P) \end{aligned}$$

We saw that a projector P separate $\mathbb{C}^m$ into two subspaces range(P) and null(P) with the property that range$(P) \cap$ null$(P) = \{0\}$.

Let S_1 and S_2 be two subspaces of $\mathbb{C}^m$ such that

- $S_1 \cap S_2 = \{0\}$.
- $S_1 + S_2 = \mathbb{C}^m$, where $S_1 + S_2$ denotes the set of vectors $s_1 + s_2$ with $s_i \in S_i$

Then we can find a projector P such that

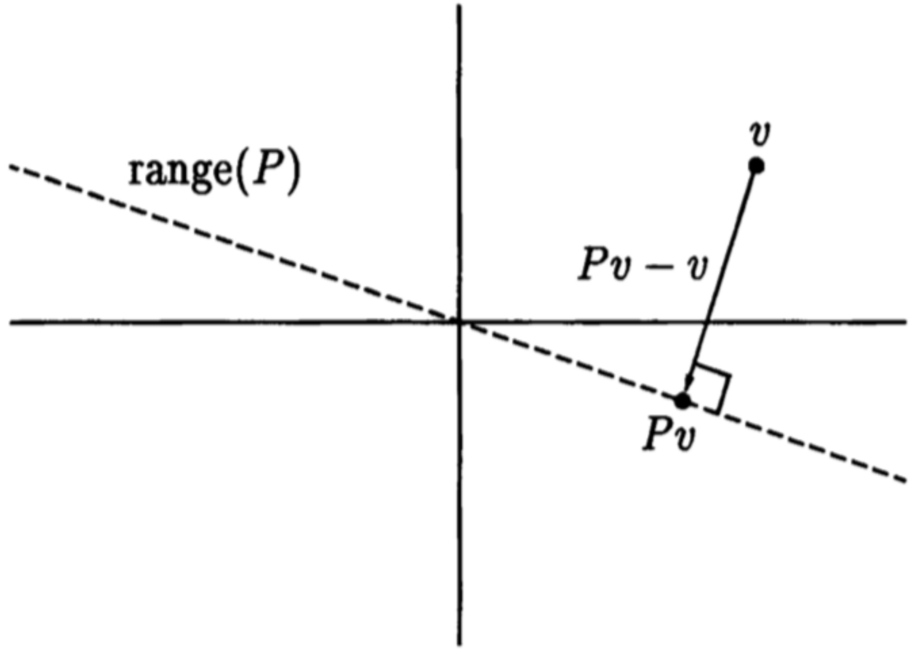
- range$(P) = S_1$
- null$(P) = S_2$

We say that P is the projector onto S_1 along S_2.

Orthogonal projectors

Definition: An orthogonal projector P is the one that projects onto a subspace S_1 along a subspace S_2 such that S_1 and S_2 are orthogonal.

Illustration:



Theorem: A projector P is orthogonal if and only if $P = P^*$ (that is, P is hermitian).

Proof: Assume that $P = P^*$.

Recall that P projects $\mathbb{C}^m$ onto $S_1 = \text{range}(P)$ along $S_2 = \text{null}(P) = \text{range}(I - P)$.

Let $s_1 \in S_1$ and $s_2 \in S_2$. Then $s_1 = Px$ and $s_2 = (I - P)y$.

Thus,

$$s_1^* s_2 = x^* P^* (I - P) y = x^* (P - P^2) y = 0, \qquad \text{since } P^2 = P$$

Hence, we see that S_1 and S_2 are orthogonal.

We prove the other direction next.

Theorem: A projector P is orthogonal if and only if $P = P^*$ (that is, P is hermitian).

Proof: Assume P is an orthogonal projector. Then P projects onto S_1 along S_2 with $S_1 \perp S_2$.

Let S_1 has dimension n.

Let $\{q_1, \cdots, q_m\}$ be an orthonormal basis for $\mathbb{C}^m$ where $\{q_1, \cdots, q_n\}$ is a basis for S_1 and $\{q_{n+1}, \cdots, q_m\}$ is a basis for S_2.

Observe that $Pq_j = q_j$ for $j \leq n$ and $Pq_j = 0$ for $j > n$.

Let Q be the unitary matrix whose columns are $\{q_1, \cdots, q_m\}$. Then

$$PQ = \left[\begin{array}{c|c|c|c|c} q_1 & \cdots & q_n & 0 & \cdots \end{array} \right]$$

which implies

$$Q^*PQ = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} = \Sigma$$

Thus, we obtain a SVD of P: $P = Q\Sigma Q^*$.

The above shows that $P^* = P$. The completes the proof.

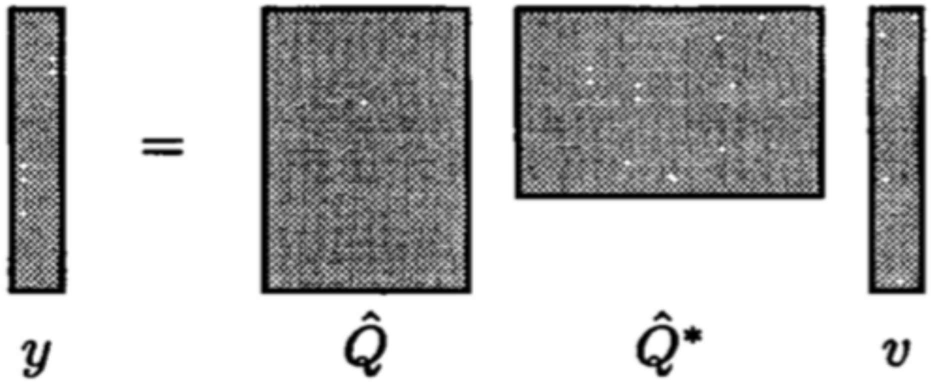
Projection with an orthonormal basis

Let $\{q_1, \cdots, q_n\}$ be a set of orthonormal vectors in $\mathbb{C}^m$ and $\hat{Q}$ be the $m \times n$ matrix having $\{q_1, \cdots, q_n\}$ as columns.

We define the map

$$v \quad \mapsto \quad \sum_{i=1}^{n} (q_i q_i^*) v$$

It is a projector onto range($\hat{Q}$). In matrix form, we have $y = \hat{Q}\hat{Q}^* v$:



The projector $\hat{Q}\hat{Q}^*$ is an orthogonal projector since it is hermitian.

Let q be a unit vector. A special case of orthogonal projectors is the rank-one projector

$$P_q = qq^*$$

This is a building blocks of other higher-rank projectors (as seen in the previous page).

The complement is the rank $m-1$ orthogonal projector defined by

$$P_{\perp q} = I - qq^*$$

Lecture 7

An important numerical algorithm: QR factorization.

First, we consider the reduced QR factorization.

Let $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$ be of full rank n. Our aim is to find vectors $q_1, q_2, \cdots$, so that

$$\langle q_1, q_2, \cdots, q_j\rangle = \langle a_1, a_2, \cdots, a_j\rangle, \qquad j = 1, 2, \cdots, n$$

where a_j is the j-th column of A, and $\langle a_1, a_2, \cdots, a_j\rangle$ is the space spanned by $\{a_1, \cdots, a_j\}$.

We need the following result:

$$\left[\begin{array}{c|c|c|c} a_1 & a_2 & \cdots & a_n \end{array}\right] = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_n \end{array}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}$$

We need the following result:

$$\left[\begin{array}{c|c|c|c} a_1 & a_2 & \cdots & a_n \end{array}\right] = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_n \end{array}\right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}$$

The above is equivalent to

$$\begin{aligned} a_1 &= r_{11} q_1, \\ a_2 &= r_{12} q_1 + r_{22} q_2, \\ a_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3, \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \cdots + r_{nn} q_n. \end{aligned}$$

In matrix form, we have

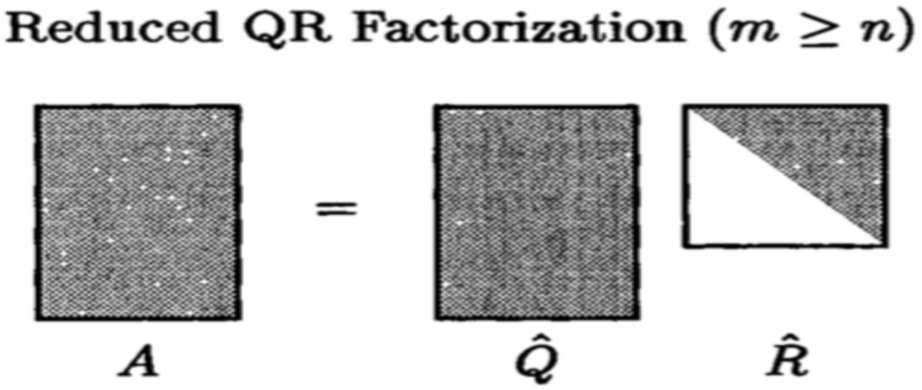
$$A = \hat{Q}\hat{R}$$

where $\hat{Q}$ is $m \times n$ with orthonormal columns, $\hat{R}$ is $n \times n$ upper-triangular. This is called the reduced QR factorization of A.

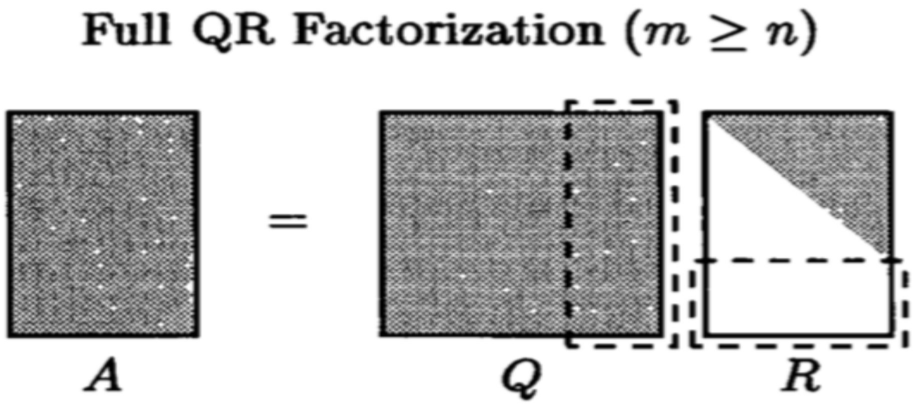
Full QR factorization

Let $A \in \mathbb{C}^{m \times n}$ $(m \geq n)$.

The full QR factorization of A is obtained from the reduced QR factorization $A = \hat{Q}\hat{R}$



by adding $m - n$ orthonormal columns of $\hat{Q}$:



The full QR factorization is $A = QR$, where Q is $m \times m$ unitary and R is $m \times n$ upper triangular.

Gram-Schmidt orthogonalization

Recall that our aim is to find orthonormal vectors $q_1, q_2, \cdots$, so that

$$\langle q_1, q_2, \cdots, q_j \rangle = \langle a_1, a_2, \cdots, a_j \rangle, \qquad j = 1, 2, \cdots, n$$

where a_j is the j-th column of A, and $\langle a_1, a_2, \cdots, a_j \rangle$ is the space spanned by $\{a_1, \cdots, a_j\}$.

Or, equivalently

$$\begin{aligned} a_1 &= r_{11} q_1, \\ a_2 &= r_{12} q_1 + r_{22} q_2, \\ a_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3, \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \cdots + r_{nn} q_n. \end{aligned}$$

For $j = 1$, we need to find q_1 so that

$$\langle q_1 \rangle = \langle a_1 \rangle$$

This is easy, we can define

$$v_1 = a_1, \quad \text{and} \quad q_1 = v_1 / \|v_1\|$$

If we define $r_{11} = \|v_1\|$, then we have

$$q_1 = a_1 / r_{11} \qquad \Longleftrightarrow \qquad a_1 = r_{11} q_1$$

For $j = 2$, we need to find q_2 so that

$$\langle q_1, q_2 \rangle = \langle a_1, a_2 \rangle$$

First, using the new column a_2, we define

$$v_2 = a_2 - (q_1^* a_2) q_1$$

Then we see that v_2 is orthogonal to q_1. To complete, we define

$$q_2 = v_2 / \|v_2\|$$

Let $r_{12} = q_1^* a_2$ and $r_{22} = \|v_2\|$. We have

$$a_2 = r_{12} q_1 + r_{22} q_2$$

For $j = 3$, we need to find q_3 so that

$$\langle q_1, q_2, q_3 \rangle = \langle a_1, a_2, a_3 \rangle$$

First, using the new column a_3, we define

$$v_3 = a_3 - (q_1^* a_3) q_1 - (q_2^* a_3) q_2$$

Then we see that v_3 is orthogonal to q_1 and q_2. To complete, we define

$$q_3 = v_3 / \|v_3\|$$

Let $r_{13} = q_1^* a_3$, $r_{23} = q_2^* a_3$ and $r_{33} = \|v_3\|$. We have

$$a_3 = r_{13} q_1 + r_{23} q_2 + r_{33} q_3$$

General step. Assume that we have found $q_1, q_2, \cdots, q_{j-1}$ such that

$$\langle q_1, q_2, \cdots, q_{j-1}\rangle = \langle a_1, a_2, \cdots, a_{j-1}\rangle$$

Next, we we need to find q_j so that

$$\langle q_1, q_2, \cdots, q_j\rangle = \langle a_1, a_2, \cdots, a_j\rangle$$

First, using the new column a_j, we define

$$v_j = a_j - (q_1^* a_j)q_1 - (q_2^* a_j)q_2 - \cdots - (q_{j-1}^* a_j)q_{j-1}$$

Then we see that v_j is orthogonal to $q_1, q_2, \cdots, q_{j-1}$. To complete, we define

$$q_j = v_j/\|v_j\|$$

Let $r_{ij} = q_i^* a_j$ for $i \neq j$ and $r_{jj} = \|v_j\|$. We have

$$a_j = r_{1j}q_1 + r_{2j}q_2 + \cdots + r_{j-1,j}q_{j-1} + r_{jj}q_j$$

This completes our construction. This is called the Gram-Schmidt iteration.

To sum up, we have found the following

$$\begin{aligned} a_1 &= r_{11} q_1, \\ a_2 &= r_{12} q_1 + r_{22} q_2, \\ a_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3, \\ &\vdots \\ a_n &= r_{1n} q_1 + r_{2n} q_2 + \cdots + r_{nn} q_n. \end{aligned}$$

This is equivalent to the following reduced QR factorization

$$\left[\begin{array}{c|c|c|c} a_1 & a_2 & \cdots & a_n \end{array} \right] = \left[\begin{array}{c|c|c|c} q_1 & q_2 & \cdots & q_n \end{array} \right] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ & & \ddots & \\ & & & r_{nn} \end{bmatrix}$$

Remark: $A = \hat{Q}\hat{R}$.

Note that we have $r_{ii} > 0$, so that the matrix $\hat{R}$ is nonsingular.

Here is the summary of the steps for computer implemenations:

Algorithm 7.1. Classical Gram–Schmidt (unstable)

for $j = 1$ **to** n

$\quad v_j = a_j$

$\quad$ **for** $i = 1$ **to** $j-1$

$\quad\quad r_{ij} = q_i^* a_j$

$\quad\quad v_j = v_j - r_{ij} q_i$

$\quad r_{jj} = \|v_j\|_2$

$\quad q_j = v_j / r_{jj}$

Remark: this algorithm is not stable due to rounding errors.

Exercise

Let

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

(a) Find the reduced QR factorization of A.

(b) Find the orthogonal projector P onto range(A).

For

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

we have

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad a_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

We define

$$v_1 = a_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad r_{11} = \|v_1\| = \sqrt{2}$$

So,

$$q_1 = v_1/\|v_1\| = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}$$

and that

$$a_1 = r_{11} q_1$$

Next, we define

$$v_2 = a_2 - (q_1^* a_2) q_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - 2/\sqrt{2} \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$$

We can define

$$r_{12} = (q_1^* a_2) = 2/\sqrt{2} \quad \text{and} \quad r_{22} = \|v_2\| = \sqrt{3} \quad \text{and} \quad q_2 = v_2/\|v_2\|$$

Then we have

$$q_2 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{pmatrix} \quad \text{and} \quad a_2 = r_{12} q_1 + r_{22} q_2$$

Finally, the reduced QR factorization is

$$A = \hat{Q}\hat{R} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & \sqrt{3} \end{pmatrix}$$

Recall

$$A = \hat{Q}\hat{R} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & \sqrt{3} \end{pmatrix}$$

Notice that q_1 and q_2 span range(A). The orthogonal projector is

$$\begin{aligned} P &= \hat{Q}\hat{Q}^* = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \end{pmatrix} \\ &= \begin{pmatrix} 5/6 & 1/3 & 1/6 \\ 1/3 & 1/3 & -1/3 \\ 1/6 & -1/3 & 5/6 \end{pmatrix} \end{aligned}$$

Note that $P = P^*$.

Solving linear systems

One can use QR factorization to solve linear systems $Ax = b$, where $A \in \mathbb{C}^{m \times m}$ nonsingular.

First, we obtain the (full) QR factorization $A = QR$. Then

$$Ax = b \quad \Longleftrightarrow \quad QRx = b$$

Now we define y such that $Qy = b$. Then

$$Qy = b \quad \Longleftrightarrow \quad y = Q^* b$$

Finally, we can solve

$$Rx = y$$

to obtain x. Note that this is a triangular system, and is hence easy to solve.

Exercise

Let

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Solve $Ax = b$.

We have

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad a_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad a_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

We have already known that

$$q_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} \quad \text{and} \quad q_2 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{pmatrix}$$

Next, we define

$$\begin{aligned} v_3 &= a_3 - (q_1^* a_3)q_1 - (q_2^* a_3)q_2 \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - 1/\sqrt{2} \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} + 1/\sqrt{3} \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{pmatrix} \\ &= \begin{pmatrix} -1/6 \\ 1/3 \\ 1/6 \end{pmatrix} \end{aligned}$$

We define

$$r_{33} = \|v_3\| = 1/\sqrt{6}, \quad r_{13} = q_1^* a_3 = 1/\sqrt{2}, \quad r_{23} = (q_2^* a_3) = -1/\sqrt{3}$$

So, the QR factorization of A is

$$A = QR = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} & -1/\sqrt{6} \\ 0 & 1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{3} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 2\sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3} & -1/\sqrt{3} \\ 0 & 0 & 1/\sqrt{6} \end{pmatrix}$$

Next, we have

$$y = Q^* b = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix}$$

Finally, we solve $Rx = y$, namely

$$\begin{pmatrix} \sqrt{2} & 2\sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3} & -1/\sqrt{3} \\ 0 & 0 & 1/\sqrt{6} \end{pmatrix} x = \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix}$$

which gives

$$x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Lecture 8

We will re-formulate the Gram-Schmidt iteration using projectors.

Let $A \in \mathbb{C}^{m \times n}$ of full rank, and $\{a_j\}$ be the columns of A. We consider

$$q_1 = \frac{P_1 a_1}{\|P_1 a_1\|}, \quad q_2 = \frac{P_2 a_2}{\|P_2 a_2\|}, \quad \cdots \quad q_n = \frac{P_n a_n}{\|P_n a_n\|},$$

where each P_j is an orthogonal projector.

More precise, each P_j

- is an $m \times m$ matrix
- projects vectors onto the space orthogonal to $\langle q_1, \cdots, q_{j-1} \rangle$

Remark that, we choose $P_1 = I$, the identity matrix.

We notice that the vectors $\{q_j\}$ are the same vectors obtained from the Gram-Schmidt iteration.

Recall that we need P_j such that

- it is an $m \times m$ matrix
- it projects vectors onto the space orthogonal to $\langle q_1, \cdots, q_{j-1} \rangle$

The orthogonal projector P_j can be constructed as follows. First we define a matrix $\hat{Q}_{j-1}$ by

$$\hat{Q}_{j-1} = \left[\begin{array}{c|c|c|c} & & & \\ q_1 & q_2 & \cdots & q_{j-1} \\ & & & \end{array} \right]$$

Then we define

$$P_j = I - \hat{Q}_{j-1} \hat{Q}_{j-1}^*$$

(why?)

We remark that this formula is also not used in practice as it is unstable.

Modified Gram-Schmidt iteration

From the above discussions, we have

$$v_j = P_j a_j \quad \text{and} \quad q_j = \frac{v_j}{\|v_j\|}$$

To define P_j in a more numerically stable way, we do the following.

Recall that $P_{\perp q}$ is the orthogonal projector onto the space orthogonal to q.

Using this, we can define P_j as

$$P_j = P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1}$$

Recall that $P_1 = I$.

We need to compute

$$v_j = P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1} a_j$$

We perform the following.

$$\begin{aligned}
v_j^{(1)} &= a_j, \\
v_j^{(2)} &= P_{\perp q_1} v_j^{(1)} &&= v_j^{(1)} - q_1 q_1^* v_j^{(1)}, \\
v_j^{(3)} &= P_{\perp q_2} v_j^{(2)} &&= v_j^{(2)} - q_2 q_2^* v_j^{(2)}, \\
&\vdots &&\vdots \\
v_j = v_j^{(j)} &= P_{\perp q_{j-1}} v_j^{(j-1)} &&= v_j^{(j-1)} - q_{j-1} q_{j-1}^* v_j^{(j-1)}.
\end{aligned}$$

This is the modified Gram-Schmidt iteration.

We summarize the modified Gram-Schmidt iteration as follows.

Algorithm 8.1. Modified Gram–Schmidt

for $i = 1$ to n

$\quad v_i = a_i$

for $i = 1$ to n

$\quad r_{ii} = \|v_i\|$

$\quad q_i = v_i / r_{ii}$

$\quad$ for $j = i+1$ to n

$\quad\quad r_{ij} = q_i^* v_j$

$\quad\quad v_j = v_j - r_{ij} q_i$

Note that, the projector $P_{\perp q_i}$ is applied to $v_j^{(i)}$ for each $j > i$ immediately after q_i is known.