Classical Numerical Analysis, Chapter 05

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu swise1@.utk.edu
University of Tennessee

Chapter 05, Part 2 of 2
Linear Least Squares Problem

QR Factorization

Theorem (QR factorization)

Suppose that $A \in \mathbb{C}^{m\times n}$, $m \geq n$. *There is an upper triangular matrix* $\hat{R} = [\hat{r}_{i,j}] \in \mathbb{C}^{n\times n}$ *with non–negative (real) diagonal elements and a matrix* $\hat{Q} \in \mathbb{C}^{m\times n}$ *satisfying* $\hat{Q}^H\hat{Q} = I_n$, *such that*

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

If $\mathsf{rank}(A) = n$, *then* $\hat{r}_{i,i} > 0$, *for all* $i = 1, \ldots, n$.

Proof.

The proof is by induction on the number of columns of A, $n \geq 1$.

$(n = 1)$: $A = \mathbf{a} \in \mathbb{C}^m$. If $\mathbf{a} \neq \mathbf{0}$, then

$$\hat{Q} = \frac{1}{\|\mathbf{a}\|_2}\mathbf{a}, \quad \hat{R} = [\|\mathbf{a}\|_2]$$

gives the result. If $\mathbf{a} = \mathbf{0}$, pick any $\mathbf{q} \in \mathbb{C}^m$ with the property that $\|\mathbf{q}\|_2 = 1$. Set

$$\hat{Q} = \mathbf{q}, \quad \hat{R} = [0],$$

and observe that the result is still satisfied.

Proof, Cont.

($n = k < m$): Suppose that the result is true for any $A \in \mathbb{C}^{m\times k}$, i.e., there is a matrix $\hat{Q} \in \mathbb{C}^{m\times k}$, with $\hat{Q}^H\hat{Q} = I_k$, and an upper triangular matrix $\hat{R} \in \mathbb{C}^{k\times k}$ such that

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

($n = k + 1 \leq m$): Suppose that $A \in \mathbb{C}^{m\times(k+1)}$. Write

$$A = [A_k \ \mathbf{a}], \quad A_k \in \mathbb{C}^{m\times k}, \quad \mathbf{a} \in \mathbb{C}^m.$$

There exist $\hat{Q}_k \in \mathbb{C}^{m\times k}$ and $\hat{R}_k \in \mathbb{C}^{k\times k}$, as above, such that $A_k = \hat{Q}_k\hat{R}_k$. Define

$$\hat{R} = \begin{bmatrix} \hat{R}_k & \mathbf{r} \\ \mathbf{0}^T & \alpha \end{bmatrix}, \quad \hat{Q} = \begin{bmatrix} \hat{Q}_k & \mathbf{q} \end{bmatrix},$$

where $\mathbf{r} \in \mathbb{C}^k$, $\mathbf{q} \in \mathbb{C}^m$, and $\alpha \in \mathbb{R}$ are to be determined.

Proof, Cont.

Now, we observe that

$$A = \hat{Q}\hat{R}, \quad \hat{Q}^H\hat{Q} = I_{k+1}, \tag{1}$$

has a solution iff the following are satified

$$A_k = \hat{Q}_k\hat{R}_k, \tag{2}$$

$$\mathbf{a} = \hat{Q}_k\mathbf{r} + \alpha\mathbf{q}, \tag{3}$$

$$\hat{Q}_k^H\hat{Q}_k = I_k, \tag{4}$$

$$\mathbf{q}^H\hat{Q}_k = \mathbf{0}^T, \tag{5}$$

$$\mathbf{q}^H\mathbf{q} = 1. \tag{6}$$

Equations (2) and (4) are satisfied as part of the induction hypothesis.

Proof, Cont.

Next, set

$$\alpha = \left\| \mathbf{a} - \hat{\mathrm{Q}}_k \hat{\mathrm{Q}}_k^{\mathrm{H}} \mathbf{a} \right\|_2 .$$

If $\alpha > 0$, equations (2)—(6) have a solution, namely

$$\mathbf{r} = \hat{\mathrm{Q}}_k^{\mathrm{H}} \mathbf{a} \in \mathbb{C}^k,$$

$$\mathbf{q} = \frac{1}{\alpha} \left(\mathbf{a} - \hat{\mathrm{Q}}_k \hat{\mathrm{Q}}_k^{\mathrm{H}} \mathbf{a} \right).$$

It is easy to see that $\mathbf{q}^{\mathrm{H}} \mathbf{q} = 1$, and, also, notice that

$$\mathbf{q}^{\mathrm{H}} \hat{\mathrm{Q}}_k = \frac{1}{\alpha} \left(\mathbf{a} - \hat{\mathrm{Q}}_k \hat{\mathrm{Q}}_k^{\mathrm{H}} \mathbf{a} \right)^{\mathrm{H}} \hat{\mathrm{Q}}_k = \frac{1}{\alpha} \left(\mathbf{a}^{\mathrm{H}} \hat{\mathrm{Q}}_k - \mathbf{a}^{\mathrm{H}} \hat{\mathrm{Q}}_k \hat{\mathrm{Q}}_k^{\mathrm{H}} \hat{\mathrm{Q}}_k \right) = \mathbf{0}^{\mathsf{T}}.$$

If $\alpha = 0$, which is possible, the construction fails. In this case, pick any $\mathbf{q} \in \mathbb{C}_\star^m$ that satisfies equations (5) and (6). Then, set $\alpha = 0$ and

$$\mathbf{r} = \hat{\mathrm{Q}}_k^{\mathrm{H}} \mathbf{a}.$$

The proof of the existence of the factorization is completed.

Proof, Cont.

Finally, suppose that $\mathbf{rank}(A) = n$. We want to prove that the diagonal elements of $\hat{R}$ must all be positive. To get a contradiction, suppose that $\hat{R}$ has a zero diagonal element and is, therefore, singular. In this case, there is a vector $\mathbf{x} \in \mathbb{C}^n_\star$ such that

$$\hat{R}\mathbf{x} = \mathbf{0} \in \mathbb{C}^n.$$

This implies that

$$A\mathbf{x} = \hat{Q}\hat{R}\mathbf{x} = \mathbf{0} \in \mathbb{C}^m,$$

which, in turn, implies that A is rank deficient. This is a contradiction. □

The Reduced QR Factorization

Definition

Suppose that $A \in \mathbb{C}^{m\times n}$, $m \geq n$. A factorization of the form

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

where $\hat{R} = [\hat{r}_{i,j}] \in \mathbb{C}^{n\times n}$ is an upper triangular matrix with non-negative (real) diagonal elements and $\hat{Q} \in \mathbb{C}^{m\times n}$ is a matrix satisfying $\hat{Q}^H\hat{Q} = I_n$, is called a **reduced QR factorization**, or sometimes a **reduced nonnegative QR factorization** to emphasize the fact that the diagonal entries are nonnegative.

Theorem (uniqueness)

Suppose that $A \in \mathbb{C}^{m\times n}$, $m \geq n$ *and* $\mathrm{rank}(A) = n$. *The reduced nonnegative QR factorization is unique.*

Proof Sketch.

Consider the Cholesky Factorization of A^HA. Is that unique? □

Lemma (reduced QR factorization via Gram–Schmidt)

Let $\mathrm{A} \in \mathbb{C}^{m \times n}$ *with* $m \geq n = \mathbf{rank}(\mathrm{A})$. *Then there exists a unique factorization*

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

where $\hat{\mathrm{Q}} \in \mathbb{C}^{m \times n}$ *has orthonormal columns, and* $\hat{\mathrm{R}} = [\hat{r}_{i,j}] \in \mathbb{C}^{n \times n}$ *is upper triangular with positive real diagonal entries.*

Proof.

Let $\mathrm{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ where $\mathbf{a}_i \in \mathbb{C}^m$ are the columns of A. We will inductively construct the columns of $\hat{\mathrm{Q}} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$ with $\mathbf{q}_i \in \mathbb{C}^m$ from the columns of A as follows:

Since $\mathbf{rank}(\mathrm{A}) = n$ we know that the columns of A are linearly independent. This, in particular, implies $\mathbf{a}_1 \neq 0$. Define

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{a}_1\|_2}\mathbf{a}_1 \quad \text{and} \quad \hat{r}_{j,1} = \begin{cases} \|\mathbf{a}_1\|_2 > 0, & j = 1, \\ 0, & j = 2, \ldots, n. \end{cases}$$

Proof, Cont.

Assume that, for $k \leq n-1$, we have found $\mathbf{q}_1, \ldots, \mathbf{q}_k$ that are orthonormal and, moreover, that

$$L_k = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \text{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}.$$

Since A is full rank, we know that $\mathbf{a}_{k+1} \neq \mathbf{0}$ and $\mathbf{a}_{k+1} \notin L_k$. Therefore, the vector

$$\mathbf{v}_{k+1} = \mathbf{a}_{k+1} - \sum_{j=1}^{k} (\mathbf{a}_{k+1}, \mathbf{q}_j)_2 \mathbf{q}_j \neq \mathbf{0}.$$

We define

$$\mathbf{q}_{k+1} = \frac{1}{\|\mathbf{v}_{k+1}\|_2} \mathbf{v}_{k+1} \quad \text{and} \quad \hat{r}_{j,k+1} = \begin{cases} (\mathbf{a}_{k+1}, \mathbf{q}_j)_2, & j = 1, \ldots, k, \\ \|\mathbf{v}_{k+1}\|_2 > 0, & j = k+1, \\ 0, & j = k+2, \ldots, n. \end{cases}$$

By construction, $\hat{\mathrm{Q}}\hat{\mathrm{R}} = \mathrm{A}$, as the reader should confirm.

Proof, Cont.

Furthermore, the matrix $\hat{\mathrm{Q}}$ has orthonormal columns and

$$\mathsf{span}\left(\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}\right) = \mathsf{span}\left(\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}\right), \quad k = 1, \cdots, n.$$

The matrix $\hat{\mathrm{R}}$ is upper triangular with positive diagonal entries. The uniqueness of this decomposition follows from a previous theorem. □

Remark (instability of Gram–Schmidt)

The classical Gram–Schmidt process used in the construction of the reduced QR factorization in the proof of Lemma suffers from numerical instabilities. For this reason, it is never used in practice. There exists a variant that is numerically stable, which we discuss below.

Solution of the Normal Equation via QR Factorization

Theorem

Let $\mathrm{A} \in \mathbb{C}^{m \times n}$ *be such that* $m \geq n = \mathsf{rank}(\mathrm{A})$. *Suppose* $\mathbf{f} \in \mathbb{C}^m$ *is given. The vector* $\tilde{\mathbf{x}} \in \mathbb{C}^n$ *is the unique least squares solution to* $\mathrm{A}\mathbf{x} = \mathbf{f}$ *iff*

$$\tilde{\mathbf{x}} = \hat{\mathrm{R}}^{-1}\hat{\mathrm{Q}}^{\mathrm{H}}\mathbf{f},$$

where $\hat{\mathrm{Q}}$ *and* $\hat{\mathrm{R}}$ *are the matrices from the reduced nonnegative QR factorization.*

Proof.

$\tilde{\mathbf{x}}$ is the unique least squares solution iff it satisfies the normal equation, $\mathrm{A}^{\mathrm{H}}\mathrm{A}\tilde{\mathbf{x}} = \mathrm{A}^{\mathrm{H}}\mathbf{f}$. But $\mathrm{A} = \hat{\mathrm{Q}}\hat{\mathrm{R}}$, so that the normal equation becomes

$$\hat{\mathrm{R}}^{\mathrm{H}}\mathrm{R}\tilde{\mathbf{x}} = \hat{\mathrm{R}}^{\mathrm{H}}\hat{\mathrm{Q}}^{\mathrm{H}}\mathbf{f}.$$

Theorem (full QR factorization)

Let $A \in \mathbb{C}^{m\times n}$, $m > n$. There exists a unitary matrix $Q \in \mathbb{C}^{m\times m}$ and an upper triangular matrix $R = [r_{i,j}] \in \mathbb{C}^{m\times n}$ (that is, $r_{i,j} = 0$, if $i > j$) with non-negative diagonal entries ($r_{i,i} \geq 0$, for each $i = 1, \ldots, n$) such that

$$A = QR.$$

Such a factorization is known as a full QR factorization.

Proof.

By a previous theorem, there is an upper triangular matrix $\hat{R} = [\hat{r}_{i,j}] \in \mathbb{C}^{n\times n}$ with nonnegative (real) diagonal elements and a matrix $\hat{Q} \in \mathbb{C}^{m\times n}$ satisfying $\hat{Q}^H\hat{Q} = I_n$, such that

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

Suppose $\hat{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$. The columns of $\hat{Q}$ form an orthonormal set.

Proof, Cont.

We create the square matrix

$$Q = \begin{bmatrix} \hat{Q} & \tilde{Q} \end{bmatrix} \in \mathbb{C}^{m \times m},$$

where the columns of $\tilde{Q} = [\mathbf{q}_{n+1}, \ldots, \mathbf{q}_m] \in \mathbb{C}^{m \times m-n}$ are orthonormal and chosen so that the set $\{\mathbf{q}_1, \ldots, \mathbf{q}_m\}$, that is, the set of columns of Q, is an orthonormal basis for $\mathbb{C}^n$. This can always be accomplished by a combination of the basis extension theorem and the Gram–Schmidt process. Next, we define

$$R = \begin{bmatrix} \hat{R} \\ O \end{bmatrix} \in \mathbb{C}^{m \times n},$$

where $O \in \mathbb{C}^{m-n \times n}$ is a zero matrix used for padding. Then, just note that $A = QR$ and the proof is complete. □

Sundry Results Using QR Factorization

Theorem (Hadamard inequality)

Suppose $\mathrm{A} \in \mathbb{C}^{n\times n}$, *and denote the columns of* A *by* $\mathbf{a}_j$, $j = 1, \ldots, n$. *Then*

$$|\det(\mathrm{A})| \leq \prod_{j=1}^{n} \|\mathbf{a}_j\|_2 .$$

Theorem (least squares and projection)

Let $\mathrm{A} \in \mathbb{C}^{m\times n}$, $m \geq n$, *and* $\mathbf{f} \in \mathbb{C}^m$. *The vector* $\mathbf{x}_o \in \mathbb{C}^n$ *is a least squares solution to* $\mathrm{A}\mathbf{x} = \mathbf{f}$ *iff* $\mathrm{A}\mathbf{x}_o = \mathrm{P}\mathbf{f}$, *where* $\mathrm{P} \in \mathbb{C}^{m\times m}$ *is the orthogonal projection onto* $\mathrm{im}(\mathrm{A})$.

The Moore–Penrose Pseudo–Inverse

Minimum Norm Least Squares Solution in the Rank-Deficient Case

Definition

Let $A \in \mathbb{C}^{m \times n}$ with $m \geq n > r = \mathbf{rank}(A)$ and $\mathbf{f} \in \mathbb{C}^m$. Define

$$\Phi(\mathbf{x}) = \|\mathbf{f} - A\mathbf{x}\|_2^2.$$

The **minimum norm least squares solution** of $A\mathbf{x} = \mathbf{f}$ is $\hat{\mathbf{x}} \in \mathbb{C}^n$ that satisfies

1. $\hat{\mathbf{x}}$ is a least squares solution, i.e., $\Phi(\hat{\mathbf{x}}) \leq \Phi(\mathbf{x})$, for all $\mathbf{x} \in \mathbb{C}^n$.
2. If $\Phi(\hat{\mathbf{x}}) = \Phi(\mathbf{x})$, then $\|\hat{\mathbf{x}}\|_2 \leq \|\mathbf{x}\|_2$.

Moore–Penrose Pseudo–Inverse

Definition

Let the matrix $A \in \mathbb{C}^{m \times n}$ with $\min(m, n) \geq r = \text{rank}(A)$. Assume that A has the SVD $A = U\Sigma V^H$ with

$$\Sigma = \underset{m \times n}{\text{diag}} \left(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0 \right),$$

where $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Define

$$\Sigma^\dagger = \underset{n \times m}{\text{diag}} \left(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0 \right).$$

Then the matrix

$$A^\dagger = V\Sigma^\dagger U^H$$

is called the **Moore–Penrose pseudo–inverse** of A.

Theorem (properties of $A^\dagger$)

In the notation of the previous definition, the Moore–Penrose pseudo–inverse satisfies:

1. *If $m = n$ and A^{-1} exists, then $\mathrm{A}^\dagger = \mathrm{A}^{-1}$.*
2. *If $m \geq n$ and A has full rank, then $\mathrm{A}^\dagger = (\mathrm{A}^{\mathrm{H}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{H}}$.*
3. $\mathrm{A}\mathrm{A}^\dagger\mathrm{A} = \mathrm{A}$.
4. $\mathrm{A}^\dagger\mathrm{A}\mathrm{A}^\dagger = \mathrm{A}^\dagger$.

Pseudo-Inverse and Full-Rank Least Squares

Proposition

Suppose that $A \in \mathbb{C}^{m\times n}$, *with* $m \geq n = \mathbf{rank}(A)$, *and* $\mathbf{f} \in \mathbb{C}^m$ *are given. Then* $\hat{\mathbf{x}} \in \mathbb{C}^n$ *is a least squares solution to* $A\mathbf{x} = \mathbf{f}$ *iff* $\hat{\mathbf{x}} = A^\dagger \mathbf{f}$.

Proof.

Let $A = U\Sigma V^H$ be an SVD for A and set $A^\dagger = V\Sigma^\dagger U^H$. Recall that $\hat{\mathbf{x}} \in \mathbb{C}^n$ is a least squares solution to $A\mathbf{x} = \mathbf{f}$ iff $A^H A\hat{\mathbf{x}} = A^H\mathbf{f}$. This is true iff

$$
\begin{aligned}
& V\Sigma^T\Sigma V^H\hat{\mathbf{x}} = V\Sigma^T U^H\mathbf{f} \\
\iff \quad & \Sigma^T\Sigma V^H\hat{\mathbf{x}} = \Sigma^T U^H\mathbf{f} \\
\iff \quad & V^H\hat{\mathbf{x}} = (\Sigma^T\Sigma)^{-1}\Sigma^T U^H\mathbf{f} \\
\iff \quad & \hat{\mathbf{x}} = V\Sigma^\dagger U^H\mathbf{f} \\
\iff \quad & \hat{\mathbf{x}} = A^\dagger\mathbf{f}.
\end{aligned}
$$

□

Theorem (minimal norm least squares solution)

Let $\mathrm{A} \in \mathbb{C}^{m \times n}$ *with* $m \geq n > r = \mathbf{rank}(\mathrm{A})$. *Then there is a unique minimum norm least squares solution* $\hat{\mathbf{x}}$, *which is given by* $\hat{\mathbf{x}} = \mathrm{A}^{\dagger}\mathbf{f}$.

Proof.

Let $\mathrm{A} = \mathrm{U}\boldsymbol{\Sigma}\mathrm{V}^{\mathrm{H}}$. For a fixed $\mathbf{x} \in \mathbb{C}^n$ set $\mathbf{w}_x = \mathrm{V}^{\mathrm{H}}\mathbf{x}$. Then

$$\boldsymbol{\Phi}(\mathbf{x}) = \|\mathbf{f} - \mathrm{U}\boldsymbol{\Sigma}\mathrm{V}^{\mathrm{H}}\mathbf{x}\|_2^2 = \|\mathbf{f} - \mathrm{U}\boldsymbol{\Sigma}\mathbf{w}_x\|_2^2 = \|\mathrm{U}^{\mathrm{H}}\mathbf{f} - \boldsymbol{\Sigma}\mathbf{w}_x\|_2^2.$$

In other words, finding a minimal norm least squares solution is equivalent to finding $\hat{\mathbf{w}} \in \mathbb{C}^n$ such that

$$\|\mathrm{U}^{\mathrm{H}}\mathbf{f} - \boldsymbol{\Sigma}\hat{\mathbf{w}}\|_2^2 \leq \|\mathrm{U}^{\mathrm{H}}\mathbf{f} - \boldsymbol{\Sigma}\mathbf{w}\|_2^2, \quad \forall\, \mathbf{w} \in \mathbb{C}^n,$$

which, since $\|\hat{\mathbf{x}}\|_2 = \|\hat{\mathbf{w}}\|_2$, also has minimal norm.

Given that $r = \mathbf{rank}(\mathrm{A})$, there are exactly r nonzero diagonal entries in $\boldsymbol{\Sigma}$.

Proof, Cont.

Therefore

$$\|\mathrm{U}^{\mathrm{H}}\mathbf{f} - \boldsymbol{\Sigma}\mathbf{w}\|_2^2 = \sum_{i=1}^{r} |\sigma_i w_i - (\mathrm{U}^{\mathrm{H}}\mathbf{f})_i|^2 + \sum_{j=r+1}^{m} |(\mathrm{U}^{\mathrm{H}}\mathbf{f})_i|^2.$$

The second sum does not depend on $\mathbf{w}$ so to minimize the norm of $\mathbf{w} = [w_i]$ we will set $w_i = 0$ for $i = r+1, \ldots, n$. In addition, since $\sigma_i > 0$ for $i \leq r$ we can set

$$w_i = \frac{1}{\sigma_i}(\mathrm{U}^{\mathrm{H}}\mathbf{f})_i, \quad i = 1, \ldots, r,$$

to make the first sum vanish. We have thus constructed the vector $\hat{\mathbf{w}} = \boldsymbol{\Sigma}^{\dagger}\mathrm{U}^{\mathrm{H}}\mathbf{f}$ that minimizes Φ and has minimal norm. As a consequence

$$\hat{\mathbf{x}} = \mathrm{V}\boldsymbol{\Sigma}^{\dagger}\mathrm{U}^{\mathrm{H}}\mathbf{f},$$

is the minimal norm least squares solution. □

The Modified Gram–Schmidt Process

The Standard Gram–Schmidt Process

Definition

Suppose that $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{C}^m_\star = \mathbb{C}^m \setminus \{\mathbf{0}\}$, $1 \leq n \leq m$ is given. The **Gram–Schmidt process** is an algorithm for generating the set of vectors $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\} \subset \mathbb{C}^m$ recursively as follows: for $j = 1$

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{a}_1\|_2}\mathbf{a}_1.$$

For $j = 2, \ldots, n$, provided $\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}$ have already been successfully computed, define

$$\mathbf{v}_j = P_j(\mathbf{a}_j) = \mathbf{a}_j - \sum_{k=1}^{j-1}(\mathbf{q}_k^{\mathsf{H}}\mathbf{a}_j)\mathbf{q}_k.$$

If $\mathbf{v}_j = \mathbf{0}$, the algorithm terminates without completion. Otherwise, the algorithm proceeds with

$$\mathbf{q}_j = \frac{1}{\|\mathbf{v}_j\|_2}\mathbf{v}_j.$$

Properties of the Gram–Schmidt Process

Theorem

Suppose that $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{C}^m_\star$, *with* $1 \leq n \leq m$, *is linearly independent. Then, the Gram–Schmidt process proceeds to completion to produce an orthonormal set* $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\} \subset \mathbb{C}^m$. *Furthermore,*

$$\operatorname{span}\left(\{\mathbf{a}_1, \ldots, \mathbf{a}_j\}\right) = \operatorname{span}\left(\left\{\mathbf{q}_1, \ldots, \mathbf{q}_j\right\}\right),$$

for each $j = 1, \ldots, n$.

Note that, we are writing the algorithm in this way, because we will obtain the vectors $\{\mathbf{a}_j\}_{j=1}^n \subset \mathbb{C}^m$ from the columns of $\mathrm{A} \in \mathbb{C}^{m \times n}$, $m \geq n$, that is,

$$\mathrm{A} = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix},$$

and we use the Gram–Schmidt process (standard or modified form) to obtain a QR factorization of A.

Gram-Schmidt as a Sequence of Projections

Let us recast the Gram–Schmidt process using the language of orthogonal projection matrices. To do so, for $j = 1, \ldots, n$, we define the matrix $\mathrm{P}_j \in \mathbb{C}^{m \times m}$ by its action on an arbitrary vector $\mathbf{w} \in \mathbb{C}^m$:

$$\mathrm{P}_j \, \mathbf{w} = \left(\mathrm{I}_m - \sum_{k=1}^{j-1} \mathbf{q}_k \mathbf{q}_k^{\mathrm{H}} \right) \mathbf{w} = \mathbf{w} - \sum_{k=1}^{j-1} (\mathbf{q}_k^{\mathrm{H}} \mathbf{w}) \mathbf{q}_k, \tag{7}$$

where $\{\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}\}$ is an orthonormal set.

The key step in the Gram–Schmidt process can be characterized by the action of the matrix P_j, and now we will show that P_j is an orthogonal projection matrix, that is $\mathrm{P}_j^2 = \mathrm{P}_j$ and $\mathrm{P}_j^{\mathrm{H}} = \mathrm{P}_j$.

Furthermore, we will show that P_j can be written as the product of simpler projection matrices.

P_j is the Product of Rank–$(m-1)$ Orthogonal Projections

Definition

For a given vector $\mathbf{q} \in \mathbb{C}^m$*, with* $\|\mathbf{q}\|_2 = 1$*, define* $P_{q^\perp} \in \mathbb{C}^{m\times m}$ *via*

$$P_{q^\perp} = I_m - \mathbf{q}\mathbf{q}^H,$$

so that, for an arbitrary $\mathbf{w} \in \mathbb{C}^m$*,* $P_{q^\perp}\mathbf{w} = \mathbf{w} - (\mathbf{q}^H\mathbf{w})\mathbf{q}$.

Proposition

Suppose that $\{\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}\} \subset \mathbb{C}^m$ *is an orthonormal set and* $P_j \in \mathbb{C}^{m\times m}$ *is as defined in* (7)*. Then,*

$$P_j = P_{q_{j-1}^\perp} P_{q_{j-2}^\perp} \cdots P_{q_1^\perp} = I_m - \hat{Q}\hat{Q}^H,$$

where

$$\hat{Q} = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_{j-1} \\ | & & | \end{bmatrix} \in \mathbb{C}^{m\times(j-1)}.$$

The Modified Gram–Schmidt Process

Definition

Suppose that $S = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{C}^m_\star$, with $1 \leq n \leq m$. The **modified Gram–Schmidt Process** is an algorithm for generating the set of vectors $Q = \{\mathbf{q}_1, \ldots \mathbf{q}_n\} \subset \mathbb{C}^m_\star$ recursively as follows: for $j = 1$,

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{a}_1\|_2}\mathbf{a}_1.$$

For $2 \leq j \leq n$, suppose that $\{\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}\}$ have been computed. Set

$$\begin{aligned} \mathbf{v}_{j,1} &= \mathbf{a}_j \\ \mathbf{v}_{j,k+1} &= \mathrm{P}_{\mathrm{q}_k^\perp}\mathbf{v}_{j,k}, \quad k = 1, \ldots, j-1, \\ \mathbf{v}_j &= \mathbf{v}_{j,j}. \end{aligned}$$

If $\mathbf{v}_j = \mathbf{0}$ the process terminates. Otherwise, the process continues with

$$\mathbf{q}_j = \frac{1}{\|\mathbf{v}_j\|_2}\mathbf{v}_j.$$

Algorithm Standard Gram–Schmidt Pseudocode (Listing A.1)

```
for j = 1 to n do
    v_j = a_j
    for k = 1 to j − 1 do
        v_j = v_j − (q_k^H a_j) q_k
    end for
    q_j = v_j / ||v_j||_2
end for
```

Algorithm Modified Gram–Schmidt Pseudocode (Listing 5.1)

```
for j = 1 to n do
    v_j = a_j
end for
for j = 1 to n do
    q_j = v_j / ||v_j||_2
    for k = j + 1 to n do
        v_k = v_k − (q_j^H v_k) q_j
    end for
end for
```

Outcome and Complexity

In exact arithmetic, the results are the same.

Proposition (modified Gram–Schmidt)

Let $\{\mathbf{a}_j\}_{j=1}^n \subset \mathbb{C}^m$ with $m \geq n$ be linearly independent. The sequence $\{\mathbf{q}_j\}_{j=1}^n \subset \mathbb{C}^m$ obtained by the modified Gram–Schmidt process is orthonormal and the same as that obtained by the standard Gram–Schmidt process. Consequently, for every $j \in \{1, \ldots, n\}$ it holds that

$$\mathsf{span}\left(\{\mathbf{a}_1, \ldots, \mathbf{a}_j\}\right) = \mathsf{span}\left(\{\mathbf{q}_1, \ldots, \mathbf{q}_j\}\right).$$

Theorem (complexity of triangular orthogonalization)

Let $\mathrm{A} \in \mathbb{C}^{m \times n}$ with $m \geq n = \mathsf{rank}(\mathrm{A})$. Then the modified Gram–Schmidt process requires $\mathcal{O}(2mn^2)$ floating point operations to compute the reduced QR factorization of A.

Numerical Comparison Using a Random Matrix

```
>>  n = 200;
    A = rand(n);
    [Q,err] = ClassicalGramSchmidt(A);
    norm(eye(n)-Q'*Q)

ans =

    9.0046e-12

>>  [Q,R,err] = ModifiedGramSchmidt(A);
    norm(eye(n)-Q'*Q)

ans =

    1.5679e-13
```

Numerical Comparison Using the Hilbert Matrix

```
>>  n = 200;
    A = 0.00001*eye(n)+hilb(n);
    [Q,err] = ClassicalGramSchmidt(A);
    norm(eye(n)-Q'*Q)

ans =

    1.4320

>>  [Q,R,err] = ModifiedGramSchmidt(A);
    norm(eye(n)-Q'*Q)

ans =

    2.0814e-11
```

Householder Reflectors

Householder Reflectors

Definition

Suppose that $\mathbf{w} \in \mathbb{C}^n$ with $\|\mathbf{w}\|_2 = 1$. The **Householder reflector** with respect to $\mathbf{span}\{\mathbf{w}\}^\perp$ is the matrix

$$H_w = I_n - 2\mathbf{w}\mathbf{w}^H.$$

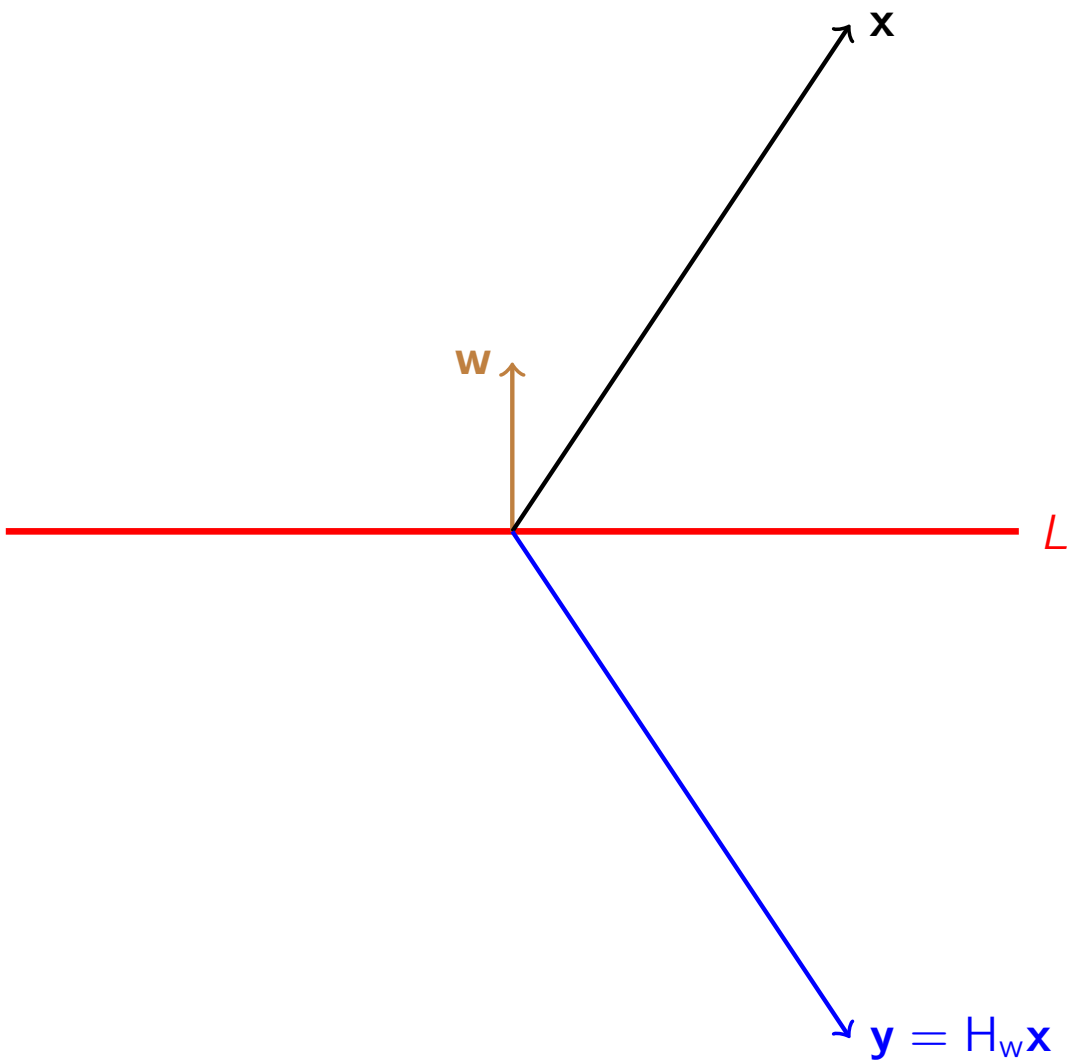

Figure: The reflection of the point x about the subspace $L = \mathbf{span}\{w\}^{\perp}$.

Proposition (properties of H_w)

Suppose that $\mathbf{w} \in \mathbb{C}^n$ *with* $\|\mathbf{w}\|_2 = 1$. *The reflector* H_w *satisfies the following properties:*

1. H_w *is Hermitian.*
2. H_w *is unitary.*
3. H_w *is involutory, i.e.,* $H_w^2 = I_n$.
4. *For any* $\mathbf{x} \in \mathbb{C}^n$, $\|H_w\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.

Proof.

Clearly,

$$H_w^H = I_n^H - 2\left(\mathbf{w}\mathbf{w}^H\right)^H = I_n - 2\mathbf{w}\mathbf{w}^H = H_w$$

which shows that H_w is Hermitian. To see the other properties, observe that

$$H_w^H H_w = H_w^2 = \left(I_n - 2\mathbf{w}\mathbf{w}^H\right)\left(I_n - 2\mathbf{w}\mathbf{w}^H\right) = I_n - 4\mathbf{w}\mathbf{w}^H + 4\mathbf{w}\mathbf{w}^H\mathbf{w}\mathbf{w}^H = I_n.$$

□

Lemma (action of a reflector)

Let $\mathbf{x} = [x_j] \in \mathbb{C}^n_\star$, *with*

$$x_j = |x_j| e^{\mathrm{i}\alpha_j}, \quad \alpha_j \in [0, 2\pi).$$

Suppose that $\mathbf{w} \in \mathbb{C}^n$, $\|\mathbf{w}\|_2 = 1$, *has the property that*

$$\mathrm{H_w}\mathbf{x} = k \|\mathbf{x}\|_2 \mathbf{e}_j,$$

where $k \in \mathbb{C}$ *and* $\mathbf{e}_j$ *is the* j*–th canonical basis vector. Then it must be that*

$$k = \pm e^{\mathrm{i}\alpha_j}.$$

Proof.

From the properties of the Householder reflector $\mathrm{H_w}$ we have that

$$\|\mathbf{x}\|_2 = \|\mathrm{H_w}\mathbf{x}\|_2 = |k| \, \|\mathbf{x}\|_2 \, \|\mathbf{e}_j\| ,$$

which necessarily implies that $|k| = 1$. In other words, there is $\beta \in \mathbb{R}$ such that

$$k = e^{\mathrm{i}\beta}.$$

Proof. Cont.

Next, observe that, since $\mathrm{H_w}$ is Hermitian,

$$\mathbf{x}^{\mathrm{H}}\mathrm{H_w}\mathbf{x} = \mathbf{x}^{\mathrm{H}} k \left\|\mathbf{x}\right\|_2 \mathbf{e}_j = k\bar{x}_j \left\|\mathbf{x}\right\|_2 \in \mathbb{R}.$$

As a consequence, we find

$$k\bar{x}_j \left\|\mathbf{x}\right\|_2 \in \mathbb{R}, \iff k\bar{x}_j \in \mathbb{R} \iff e^{\mathrm{i}\beta}|x_j|e^{-\mathrm{i}\alpha_j} \in \mathbb{R} \iff e^{\mathrm{i}(\alpha_j-\beta)} \in \mathbb{R},$$

but this is only possible if there is $m \in \mathbb{Z}$ for which

$$\beta = \alpha_j + m\pi,$$

or, equivalently, $k = \pm e^{\mathrm{i}\alpha_j}$ as we had claimed. □

Theorem (existence of reflector)

Let $\mathbf{x} = [x_j] \in \mathbb{C}^n_\star$, *with*

$$x_j = |x_j| e^{\mathrm{i}\alpha_j}, \quad \alpha_j \in [0, 2\pi).$$

Then, provided

$$\mathbf{x} \neq \pm e^{\mathrm{i}\alpha_j} \|\mathbf{x}\|_2 \mathbf{e}_j,$$

there is a vector $\mathbf{w} \in \mathbb{C}^n$, $\|\mathbf{w}\|_2 = 1$, *such that*

$$\mathrm{H}_\mathrm{w}\mathbf{x} = \pm e^{\mathrm{i}\alpha_j} \|\mathbf{x}\|_2 \mathbf{e}_j,$$

Proof.

Let $\sigma = \pm 1$. Define

$$\mathbf{v} = \mathbf{x} - \sigma e^{\mathrm{i}\alpha_j} \|\mathbf{x}\|_2 \mathbf{e}_j,$$

and, since by assumption $\mathbf{v} \neq \mathbf{0}$, we can define

$$\mathbf{w} = \frac{1}{\|\mathbf{v}\|_2}\mathbf{v}.$$

Proof, Cont.

Let us now show that

$$\mathrm{H_w}\mathbf{x} = \sigma e^{\mathrm{i}\alpha_j} \|\mathbf{x}\|_2 \mathbf{e}_j.$$

A straightforward computation reveals that

$$\mathbf{v}^{\mathrm{H}}\mathbf{x} = \|\mathbf{x}\|_2^2 - \sigma e^{-\mathrm{i}\alpha_j} \|\mathbf{x}\|_2 x_j = \|\mathbf{x}\|_2^2 - \sigma|x_j| \|\mathbf{x}\|_2 \in \mathbb{R},$$

and

$$\mathbf{v}^{\mathrm{H}}\mathbf{v} = \|\mathbf{x}\|_2^2 - 2\sigma|x_j| \|\mathbf{x}\|_2 + \sigma^2 \|\mathbf{x}\|_2^2 = 2\left(\|\mathbf{x}\|_2^2 - \sigma|x_j| \|\mathbf{x}\|_2\right) = 2\mathbf{v}^{\mathrm{H}}\mathbf{x}.$$

With these computations, it follows that

$$\mathrm{H_w}\mathbf{x} = \mathbf{x} - 2\mathbf{v}\frac{\mathbf{v}^{\mathrm{H}}\mathbf{x}}{\mathbf{v}^{\mathrm{H}}\mathbf{v}} = \mathbf{x} - \mathbf{v} = \sigma e^{\mathrm{i}\alpha_j} \|\mathbf{x}\|_2 \mathbf{e}_j,$$

as claimed. □

Lemma (construction of reflectors)

Assume that $m > k$. Suppose that $\mathbf{x} = [x_j] \in \mathbb{C}^k_\star$, with

$$x_1 = |x_1| e^{\mathrm{i}\alpha_1}, \quad \exists\, \alpha_1 \in [0, 2\pi),$$

satisfies the property that

$$\mathbf{x} \neq \pm e^{\mathrm{i}\alpha_1} \, \|\mathbf{x}\|_{\ell^2(\mathbb{C}^k)} \, \widehat{\mathbf{e}}_1,$$

where $\widehat{\mathbf{e}}_1 \in \mathbb{C}^k$ is the first canonical basis vector. Let $\mathrm{H_w} \in \mathbb{C}^{k \times k}$ be the Householder reflector that satisfies

$$\mathrm{H_w}\mathbf{x} = \sigma e^{\mathrm{i}\alpha_1} \, \|\mathbf{x}\|_{\ell^2(\mathbb{C}^k)} \, \widehat{\mathbf{e}}_1,$$

where $\sigma = \pm 1$.

Lemma (Lemma, Cont.)

Then the matrix

$$\mathsf{H} = \begin{bmatrix} \mathsf{I}_{m-k} & \mathsf{O} \\ \mathsf{O}^{\intercal} & \mathsf{H}_{\mathsf{w}} \end{bmatrix} \in \mathbb{C}^{m\times m}$$

is a Householder reflector in the following sense: there is a vector $\mathbf{z} \in \mathbb{C}^m$, *with* $\|\mathbf{z}\|_{\ell^2(\mathbb{C}^m)} = 1$, *such that*

$$\mathsf{H} = \mathsf{I}_n - 2\mathbf{z}\mathbf{z}^{\mathsf{H}}.$$

Furthermore, if $\mathbf{c} \in \mathbb{C}^{m-k}$ *is any arbitrary vector, then*

$$\mathsf{H}\begin{bmatrix} \mathbf{c} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathsf{I}_{m-k} & \mathsf{O} \\ \mathsf{O}^{\intercal} & \mathsf{H}_{\mathsf{w}} \end{bmatrix}\begin{bmatrix} \mathbf{c} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ r\widehat{\mathbf{e}}_1 \end{bmatrix}, \quad \text{where} \quad r = \sigma e^{\mathrm{i}\alpha_1}\|\mathbf{x}\|_{\ell^2(\mathbb{C}^k)}.$$

Definition (Householder triangularization)

Suppose

$$\mathrm{A} = \mathrm{A}^{(0)} = \left[\mathbf{a}_1^{(0)}, \ldots, \mathbf{a}_n^{(0)}\right] \in \mathbb{C}^{m\times n},$$

with $m \geq n$, is given. The **Householder triangularization process** is a recursive algorithm for converting A into an upper triangular matrix $\mathrm{R} \in \mathbb{C}^{m\times n}$ by applying a sequence of Householder reflections, and is defined as follows: for $s = 1$, suppose that $\mathbf{a}_1^{(0)} \in \mathbb{C}^m$ has as its first element

$$a_{1,1}^{(0)} = \left|a_{1,1}^{(0)}\right| e^{\mathrm{i}\alpha_1}, \quad \alpha_1 \in [0, 2\pi).$$

If

$$\mathbf{a}_1^{(0)} \neq \pm e^{\mathrm{i}\alpha_1} \left\|\mathbf{a}_1^{(0)}\right\|_{\ell^2(\mathbb{C}^m)} \widehat{\mathbf{e}}_1, \tag{8}$$

where $\widehat{\mathbf{e}}_1 \in \mathbb{C}^m$ is the first canonical basis vector, there is a vector $\mathbf{w}_0 \in \mathbb{C}^m$, with $\|\mathbf{w}_0\|_{\ell^2(\mathbb{C}^m)} = 1$, such that

$$\mathrm{H}_{\mathrm{w}_0}\mathbf{a}_1^{(0)} = r_{1,1}\widehat{\mathbf{e}}_1,$$

where $|r_{1,1}| = \left\|\mathbf{a}_1^{(0)}\right\|_{\ell^2(\mathbb{C}^m)}$.

Definition (Cont.)

Set $H_1 = H_{w_0}$. If the condition in (8) fails, set $H_1 = I_m$. In either case, we have

$$H_1 A^{(0)} = \begin{bmatrix} r_{1,1} & \mathbf{f}^T \\ \mathbf{0} & A^{(1)} \end{bmatrix},$$

where $A^{(1)} = \left[\mathbf{a}_1^{(1)}, \ldots, \mathbf{a}_{n-1}^{(1)}\right] \in \mathbb{C}^{(m-1)\times(n-1)}$.
For step $s = k+1$, with $1 \leq k \leq n-1$, suppose that k Householder reflections have been applied resulting in the decomposition

$$H_k \cdots H_1 A = \begin{bmatrix} R^{(k)} & B^{(k)} \\ O^{(k)} & A^{(k)} \end{bmatrix},$$

where $R^{(k)} \in \mathbb{C}^{k\times k}$ is an upper triangular matrix with diagonal elements

$$r_{1,1}, \ldots, r_{k,k} \in \mathbb{C},$$

$B^{(k)} \in \mathbb{C}^{k\times(m-k)}$, $O^{(k)} \in \mathbb{C}^{(m-k)\times k}$ is a zero matrix and

$$A^{(k)} = \left[\mathbf{a}_1^{(k)}, \ldots, \mathbf{a}_{n-k}^{(k)}\right] \in \mathbb{C}^{(m-k)\times(n-k)}.$$

Definition (Cont.)

Suppose that $\mathbf{a}_1^{(k)} \in \mathbb{C}^{m-k}$ has as its first element

$$a_{1,1}^{(k)} = \left| a_{1,1}^{(k)} \right| e^{\mathrm{i}\alpha_1}, \quad \alpha_1 \in [0, 2\pi).$$

If

$$\mathbf{a}_1^{(k)} \neq \pm e^{\mathrm{i}\alpha_1} \left\| \mathbf{a}_1^{(k)} \right\|_{\ell^2(\mathbb{C}^{m-k})} \widehat{\mathbf{e}}_1, \tag{9}$$

where $\widehat{\mathbf{e}}_1 \in \mathbb{C}^{m-k}$ is the first canonical basis vector, there is a vector $\mathbf{w}_k \in \mathbb{C}^{m-k}$, with $\|\mathbf{w}_k\|_{\ell^2(\mathbb{C}^{m-k})} = 1$, such that

$$\mathrm{H}_{\mathrm{w}_k} \mathbf{a}_1^{(k)} = r_{k+1,k+1} \hat{\mathbf{e}}_1,$$

where $|r_{k+1,k+1}| = \left\| \mathbf{a}_1^{(k)} \right\|_{\ell^2(\mathbb{C}^{m-k})}$. Set

$$\mathrm{H}_{k+1} = \begin{bmatrix} \mathrm{I}_k & \mathrm{O} \\ \mathrm{O}^{\mathsf{T}} & \mathrm{H}_{\mathrm{w}_k} \end{bmatrix}.$$

Definition (Cont.)

If the condition in (9) fails, set $\mathrm{H}_{k+1} = \mathrm{I}_m$. Either way,

$$\mathrm{H}_{k+1}\mathrm{H}_k \cdots \mathrm{H}_1\mathrm{A} = \begin{bmatrix} \mathrm{R}^{(k+1)} & \mathrm{B}^{(k+1)} \\ \mathrm{O}^{(k+1)} & \mathrm{A}^{(k+1)} \end{bmatrix},$$

where $\mathrm{R}^{(k+1)} \in \mathbb{C}^{(k+1)\times(k+1)}$ is an upper triangular matrix with diagonal elements

$$r_{1,1}, \ldots, r_{k,k}, r_{k+1,k+1} \in \mathbb{C},$$

$\mathrm{B}^{(k+1)} \in \mathbb{C}^{(k+1)\times(m-k-1)}$, $\mathrm{O}^{(k+1)} \in \mathbb{C}^{(m-k-1)\times(k+1)}$ is a zero matrix and

$$\mathrm{A}^{(k+1)} = \left[\mathbf{a}_1^{(k+1)}, \ldots, \mathbf{a}_{n-k-1}^{(k+1)}\right] \in \mathbb{C}^{(m-k-1)\times(n-k-1)}.$$

Theorem (Householder Triangularization)

Let $A \in \mathbb{C}^{m\times n}$, $m \geq n$ *be given. The Householder triangulation procedure always proceeds to completion. In other words, there exists a sequence of Householder reflectors* $H_1, \ldots, H_n \in \mathbb{C}^{m\times m}$ *such that*

$$H_n \cdots H_1 A = R,$$

where $R \in \mathbb{C}^{m\times n}$ *is upper triangular. Moreover, the product*

$$Q^H = H_n \cdots H_1 \in \mathbb{C}^{m\times m},$$

is a unitary matrix. Consequently, there is a unitary matrix

$$Q = H_1 \cdots H_n$$

and an upper triangular matrix R*, such that* $A = QR$*.*

Proposition (complexity of Householder)

The Householder algorithm requires, approximately, $2mn^2 - \frac{2}{3}n^3$ *operations.*