Orthogonalization

Def. Let $K = \mathbb{R}$ or $\mathbb{C}$, $V = K^n$, $W \le V$. Then

$$W^\perp := \{\, \mathbf{v} \in V \mid \mathbf{v} \perp \mathbf{w} \quad \forall \mathbf{w} \in W \,\} \le V.$$

Proposition
Every vector $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = \mathbf{w} + \mathbf{u}$, where $\mathbf{w} \in W$ and $\mathbf{u} \in W^\perp$.
Proof. Suppose $\dim W = k$. Let the columns of the matrix $A \in K^{n\times k}$ form a basis of W.
Then $\quad \mathbf{v} \in W^\perp \Leftrightarrow A^*\mathbf{v} = \mathbf{0}$,
so $W^\perp$ is the kernel of the linear map $f : K^n \to K^k, \ \mathbf{v} \mapsto A^*\mathbf{v}$.
Since $\operatorname{rank} f = \operatorname{rank} A^* = k \Rightarrow \dim W^\perp = \dim \operatorname{Ker} f = n - \operatorname{rank} f = n - k$.
On the other hand, $W \cap W^\perp = 0 := \{\, \mathbf{0} \,\}$ because if $\mathbf{w} \in W \cap W^\perp$, then $\langle\, \mathbf{w}, \mathbf{w} \,\rangle = 0 \Rightarrow \mathbf{w} = \mathbf{0}$. So the union of a basis of W and a basis of $W^\perp$ is independent, and has n elements $\Rightarrow$ it is a basis.
So every vector can be written in a form $\mathbf{w} + \mathbf{u}$, $\mathbf{w} \in W$, $\mathbf{u} \in W^\perp$.
This decomposition is unique: if $\mathbf{w}_1 + \mathbf{u}_1 = \mathbf{w}_2 + \mathbf{u}_2$ are two such decompositions then $\mathbf{w}_1 - \mathbf{w}_2 = \mathbf{u}_2 - \mathbf{u}_1 \in W \cap W^\perp = 0$, so $\mathbf{w}_1 = \mathbf{w}_2$ and $\mathbf{u}_1 = \mathbf{u}_2$.

Def. Let $W \le V$, $\mathbf{v} \in V$.
$\mathbf{v}'$ is the **orthogonal projection** of $\mathbf{v}$ on W if $\mathbf{v}' \in W$ and $\mathbf{v} - \mathbf{v}' \in W^\perp$.

Corollary. Every vector has a unique orthogonal projection on W.

Proposition
The closest vector of W to $\mathbf{v} \in V$ is the orthogonal projection of $\mathbf{v}$ to W.
(Closest: $|\mathbf{v} - \mathbf{v}'| = \min\{\, |\mathbf{v} - \mathbf{w}| \mid \mathbf{w} \in W \,\}$.)

Proof. Spse $\mathbf{v} = \mathbf{v}' + \mathbf{v}''$, where $\mathbf{v}' \in W$ and $\mathbf{v}'' \in W^\perp$.
If $\mathbf{w} \in W$, then

$$|\mathbf{v} - \mathbf{w}|^2 = |(\mathbf{v}' - \mathbf{w}) + \mathbf{v}''|^2 = \langle (\mathbf{v}' - \mathbf{w}) + \mathbf{v}'',\ (\mathbf{v}' - \mathbf{w}) + \mathbf{v}'' \,\rangle = |\mathbf{v}' - \mathbf{w}|^2 + |\mathbf{v}''|^2,$$

since $\mathbf{v}' - \mathbf{w} \in W \Rightarrow (\mathbf{v}' - \mathbf{w}) \perp \mathbf{v}''$. But this is $\ge |\mathbf{v}''|^2 = |\mathbf{v} - \mathbf{v}'|^2$.

Proposition
Suppose $\{\, \mathbf{b}_1, \ldots, \mathbf{b}_k \,\}$ is an orthogonal basis in $W \le V = K^n$ $\quad$ (i.e. $\langle\, \mathbf{b}_i, \mathbf{b}_j \,\rangle = 0$ if $i \ne j$)

For any $\mathbf{v} \in V$, the projection of $\mathbf{v}$ to W is

$$\mathbf{v}' = \sum_i \frac{\langle\, \mathbf{b}_i, \mathbf{v} \,\rangle}{|\mathbf{b}_i|^2} \mathbf{b}_i.$$

Proof. $\mathbf{v}' \in \operatorname{span}\{\, \mathbf{b}_1, \ldots, \mathbf{b}_k \,\} = W$. On the other hand

$$\langle\, \mathbf{b}_j, \mathbf{v}' \,\rangle = \sum_i \frac{\langle\, \mathbf{b}_i, \mathbf{v} \,\rangle}{|\mathbf{b}_i|^2} \langle\, \mathbf{b}_j, \mathbf{b}_i \,\rangle = \langle\, \mathbf{b}_j, \mathbf{v} \,\rangle,$$

thus $\langle\, \mathbf{b}_j,\ \mathbf{v} - \mathbf{v}' \,\rangle = 0$ for every j, so $\mathbf{v} - \mathbf{v}' \in W^\perp$.

Exercise. Show that $\{\mathbf{b}_1, \mathbf{b}_2\}$ is an orthog. basis of $W = \operatorname{span}\{\mathbf{b}_1, \mathbf{b}_2\}$, where $\mathbf{b}_1 = (1,2,1)$ and $\mathbf{b}_2 = (2,-1,0)$. What is the orthog. projection of $\mathbf{v} = (1,1,1)$ on W?

Solution. $\langle (1,2,1),(2,-1,0)\rangle = 0$.
$\mathbf{v}' = \frac{4}{6}(1,2,1) + \frac{1}{5}(2,-1,0) = \frac{1}{15}(16,17,10)$.

Theorem (Gram–Schmidt orthogonalization)

Let $\mathbf{b}_1,\dots,\mathbf{b}_k$ be an independent system in $V = K^n$. Then we can find an orthogonal system $\mathbf{c}_1,\dots,\mathbf{c}_k$ such that $\operatorname{span}\{\mathbf{b}_1,\dots,\mathbf{b}_i\} = \operatorname{span}\{\mathbf{c}_1,\dots,\mathbf{c}_i\} \quad \forall i \in \{1,\dots,k\}$,
We can obtain $\mathbf{c}_i$ as
$\mathbf{c}_1 = \mathbf{b}_1$ and

$$\mathbf{c}_i = \mathbf{b}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{c}_j, \mathbf{b}_i\rangle}{|\mathbf{c}_j|^2}\mathbf{c}_j \quad \text{for } i \geq 2.$$

Proof. By induction. Suppose we already constructed $\mathbf{c}_1,\dots,\mathbf{c}_{i-1}$, which satisfy the conditions of the theorem. Then $\mathbf{b}_i' = \sum\limits_{j=1}^{i-1} \frac{\langle \mathbf{c}_j,\mathbf{b}_i\rangle}{|\mathbf{c}_j|^2}\mathbf{c}_j$ is the orthog. proj. of $\mathbf{b}_i$ on $W_{i-1} = \operatorname{span}\{\mathbf{c}_1,\dots,\mathbf{c}_{i-1}\}$, so $\mathbf{c}_i = \mathbf{b}_i - \mathbf{b}_i' \in W^\perp \Rightarrow \mathbf{c}_i \perp \mathbf{c}_j\ \forall i < j$.
Furthermore,
$\operatorname{span}\{\mathbf{c}_1,\dots,\mathbf{c}_i\} = \operatorname{span}\{\mathbf{c}_1,\dots,\mathbf{c}_{i-1},\mathbf{b}_i - \mathbf{b}_i'\} = \operatorname{span}\{\mathbf{c}_1,\dots,\mathbf{c}_{i-1},\mathbf{b}_i\}$, since $\mathbf{b}_i' \in W_{i-1}$, and here $\mathbf{c}_1,\dots,\mathbf{c}_{i-1}$ can be replaced by $\mathbf{b}_1,\dots,\mathbf{b}_{i-1}$ by the induction hypothesis.

Corollary. Every $W \leq V = K^n$ has an orthonormal basis.

Proof. Take an arbitrary basis $\{\mathbf{b}_1,\dots,\mathbf{b}_k\}$ in W. Then the set $\{\mathbf{c}_1,\dots,\mathbf{c}_k\}$ obtained above is also independent, since they span the same k-dimensional subspace.
$\Rightarrow \frac{\mathbf{c}_1}{|\mathbf{c}_1|},\dots,\frac{\mathbf{c}_k}{|\mathbf{c}_k|}$ form an orthonormal basis of W.

Exercise. Orthogonalize the system $\mathbf{b}_1 = (1,1,1,1)$, $\mathbf{b}_2 = (3,-1,3,-1)$, $\mathbf{b}_3 = (6,2,2,-2)$ in $\mathbb{R}^4$.

Solution. $\mathbf{c}_1 = \mathbf{b}_1 = (1,1,1,1)$.
$\mathbf{c}_2 = \mathbf{b}_2 - \frac{\langle \mathbf{c}_1,\mathbf{b}_2\rangle}{|\mathbf{c}_1|^2}\mathbf{c}_1 = (3,-1,3,-1) - \frac{4}{4}(1,1,1,1) = (2,-2,2,-2)$, but we may substitute this with the parallel $\tilde{\mathbf{c}}_2 = (1,-1,1,-1)$.
$\mathbf{c}_3 = (6,2,2,-2) - \frac{8}{4}(1,1,1,1) - \frac{8}{4}(1,-1,1,-1) = (2,2,-2,-2)$, which can be substitued with the parallel $\tilde{\mathbf{c}}_3 := (1,1,-1,-1)$.
So we got the orthogonal basis $\{(1,1,1,1),(1,-1,1,-1),(1,1,-1,-1)\}$,
or the corresponding orthonormal basis
$\{\frac{1}{2}(1,1,1,1),\frac{1}{2}(1,-1,1,-1),\frac{1}{2}(1,1,-1,-1)\}$ in W.

Exercise. Orthogonalize the vectors $\mathbf{b}_1 = (0,1,i)$ and $\mathbf{b}_2 = (1,-i,1+i)$ in $\mathbb{C}^3$.

Solution. $\mathbf{c}_1 = \mathbf{b}_1 = (0,1,i)$,
$\mathbf{c}_2 = (1,-i,1+i) - \frac{-i+(-i)(1+i)}{2}(0,1,i) = (1,-i,1+i) - \left(0,\ \frac{1-2i}{2},\ \frac{2+i}{2}\right) = \left(1,\ -\frac{1}{2},\ \frac{i}{2}\right) \mid\mid (2,-1,i) =: \tilde{\mathbf{c}}_2$.
So the orthogonal system is $\mathbf{c}_1 = (0,1,i)$ and $\tilde{\mathbf{c}}_2 = (2,-1,i)$.

Best approximate solution of an inconsistent system of equations

Let $A \in \mathbb{R}^{m\times n}$, and suppose that $A\mathbf{x} = \mathbf{b}$ is inconsistent, that is, it has no solution. Then a **best approximate solution** is $\mathbf{x}$ for which $A\mathbf{x} = \mathbf{b}'$, where $\mathbf{b}'$ is the orthog. proj. of $\mathbf{b}$ on the column space of A. (The column space of A is $\{ A\mathbf{x} \,|\, \mathbf{x} \in \mathbb{R}^n \}$, so $\mathbf{b}'$ will be the vector closest to $\mathbf{b}$ for which the system is consistent.)
This is called **the method of least squares**.

Def. For a system of equations $A\mathbf{x} = \mathbf{b}$ the **normal equations** are the ones given by $A^TA\mathbf{x} = A^T\mathbf{b}$.

Theorem. For an inconsistent system of equations $A\mathbf{x} = \mathbf{b}$ a solution of the normal equations is a best approximate solution.

Proof. Let W be the column space of A, that is, $W = \{ A\mathbf{x} \,|\, \mathbf{x} \in \mathbb{R}^n \}$ then $W^\perp = \{ \mathbf{y} \in \mathbb{R}^n \,|\, A^T\mathbf{y} = \mathbf{0} \}$.
Spse $A^TA\mathbf{x} = A^T\mathbf{b}$ for some $\mathbf{x}$. Then $\mathbf{c} := A\mathbf{x} \in W$, and
$A^T(\mathbf{b} - \mathbf{c}) = A^T\mathbf{b} - A^TA\mathbf{x} = \mathbf{0}$, so $\mathbf{b} - \mathbf{c} \in W^\perp$.
So $\mathbf{c}$ is the orthog. proj. of $\mathbf{b}$ on W, i.e. $\mathbf{x}$ is a best approximate solution.

Exercise. Show that the system of equations given by following the augmented matrix is inconsistent. Give a best approximate solution.

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 4 & 1 & 0 \end{array}\right]$$

Solution.

$$[A|\mathbf{b}] = \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 4 & 1 & 0 \end{array}\right] \mapsto \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & -3 & -1 \\ 0 & -3 & -4 \end{array}\right] \mapsto \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & -3 & -1 \\ 0 & 0 & -3 \end{array}\right]$$

gives a contradiction.
Multiply the augmented matrix with $A^T = \begin{bmatrix} 1 & 2 & 4 \\ 1 & -1 & 1 \end{bmatrix}$ from the left, and solve the system of equations.

$$\left[\begin{array}{cc|c} 21 & 3 & 3 \\ 3 & 3 & 0 \end{array}\right] \mapsto \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & -18 & 3 \end{array}\right] \mapsto \left[\begin{array}{cc|c} 1 & 0 & \frac{1}{6} \\ 0 & 1 & -\frac{1}{6} \end{array}\right] \Rightarrow \mathbf{x} = \begin{bmatrix} \frac{1}{6} \\ -\frac{1}{6} \end{bmatrix}$$

For this $\mathbf{x}$, $A\mathbf{x} = \left(0, \frac{1}{2}, \frac{1}{2}\right)^T$ is indeed the orthog. projection of $(1, 1, 0)$ to the column space of A.

QR decomposition

Def. Spse $A \in \mathbb{R}^{m\times n}$, $r(A) = n$ (so $m \geq n$, and the columns of A are linearly independent).
The **(reduced) QR decomposition** of A is
$A = QR$ if
$Q \in \mathbb{R}^{m\times n}$ is semiorthogonal, that is, $Q^TQ = I$, and
$R \in \mathbb{R}^{n\times n}$ is an upper triangular matrix with positive diagonal elements.

Calculating the QR decomposition, using Gram–Schmidt orthogonalization

Let $\mathbf{b}_1, \dots, \mathbf{b}_m$ be the columns of A,
$\mathbf{c}_1, \dots, \mathbf{c}_m$ the orthog. basis obtained by the G–S method,
$\mathbf{q}_1, \dots, \mathbf{q}_m$, where $\mathbf{q}_i = \frac{\mathbf{c}_i}{|\mathbf{c}_i|}$, the orthonormal basis.

Then $\mathbf{c}_i = \mathbf{b}_i - \sum\limits_{j<i} \alpha_{ij}\mathbf{c}_j$,
$\mathbf{b}_i = \mathbf{c}_i + \sum\limits_{j<i} \alpha_{ij}\mathbf{c}_j = |\mathbf{c}_i|\mathbf{q}_i + \sum \beta_{ij}\mathbf{q}_j$, so
$A = [\mathbf{b}_1|\dots|\mathbf{b}_m] = [\mathbf{q}_1|\dots|\mathbf{q}_m]R$,
where the i'th column of R is $(\beta_{i1}, \dots, \beta_{i,i-1}, |\mathbf{c}_i|, 0, \dots, 0)$
$\Rightarrow R$ is an upper triangular matrix with positive diagonal elements ($|\mathbf{c}_i|$), and the columns of $Q = [\mathbf{q}_1|\dots,\mathbf{q}_m]$ are orthonormal, so $Q^TQ = I_n$.

If we multiply the $\mathbf{c}_i$ by scalars during the orthogonalization, make sure that these are always positiv scalars. In this case the $\mathbf{q}_i$ vectors are uniquely defined.

When we calculated Q, then we can obtain R from it:
$A = QR \Rightarrow Q^TA = Q^TQR = IR = R$, so we get R as Q^TA.

Exercise Calculate the QR decomposition of the matrix

$$A = \begin{bmatrix} 1 & 3 & 6 \\ 1 & -1 & 2 \\ 1 & 3 & 2 \\ 1 & -1 & -2 \end{bmatrix}$$

Solution. We orthogonalized earlier these column vectors. The orthonormal system we obtained gives

$$Q = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \quad \Rightarrow \quad R = Q^TA = \begin{bmatrix} 2 & 2 & 4 \\ 0 & 4 & 4 \\ 0 & 0 & 4 \end{bmatrix}$$

Application to systems of equations

If $A\mathbf{x} = \mathbf{b}$ is an inconsistent system, and $A = QR$ is a reduced QR decomposition, then

$$\begin{aligned} A^TA\mathbf{x} &= A^T\mathbf{b} \quad \Leftrightarrow \\ R^TQ^TQR\mathbf{x} &= R^TQ^T\mathbf{b} \quad \Leftrightarrow \quad \text{(since } R \text{ is invertible, and } Q^TQ = I) \\ R\mathbf{x} &= Q^T\mathbf{b}, \end{aligned}$$

and this can be solved easily by substitutions, since R is triangular.

Exercise With the previous A, find the best approximate solution of
$A\mathbf{x} = [8 \quad 2 \quad 2 \quad 0]^T$.

Solution.

$$\mathbf{b} = \begin{bmatrix} 8 \\ 2 \\ 2 \\ 0 \end{bmatrix}, \qquad Q^T\mathbf{b} = \begin{bmatrix} 6 \\ 4 \\ 4 \end{bmatrix}, \qquad [R|Q^T\mathbf{b}] = \left[\begin{array}{ccc|c} 2 & 2 & 4 & 6 \\ 0 & 4 & 4 & 4 \\ 0 & 0 & 4 & 4 \end{array}\right]$$

So for the best approximate solution $\mathbf{x}$, we have $x_3 = 1$, then from the second row, $4x_2 + 4 = 4 \Rightarrow x_2 = 0$, and from the first row, $2x_1 + 0 + 4 = 6 \Rightarrow x_1 = 1$, that is, $\mathbf{x} = (1, 0, 1)$.
With this $\mathbf{x}$ we got $A\mathbf{x} = (7, 3, 3, -1)$ to approximate $\mathbf{b} = (8, 2, 2, 0)$.

Orthogonal complement

If we are only given a generating set of $W \leq V = K^n$, and not a basis, we can still use GS's method for finding an orthogonal basis of W, only we have to drop the zero vectors obtained in the process.

Lemma. If $\{ \mathbf{c}_1, \dots, \mathbf{c}_k \}$ is an orthogonal set of nonzero vectors, then it is linearly independent.
Proof. If $\sum \lambda_i \mathbf{c}_i = \mathbf{0}$, then for every j, $0 = \langle \mathbf{c}_j, \mathbf{0} \rangle = \sum \lambda_i \langle \mathbf{c}_j, \mathbf{c}_i \rangle = \lambda_j |\mathbf{c}_j|^2$, but $\mathbf{c}_j \neq \mathbf{0}$, so $\lambda_j = 0$ for every j. □

Now let's see how we get an orthogonal basis from any generating set $\{ \mathbf{v}_1, \dots, \mathbf{v}_m \}$ of $W \leq K^n$.
Spse $\{ \mathbf{c}_1, \dots, \mathbf{c}_k \}$ is already an orthogonal basis of $W_\ell = \operatorname{span}\{ \mathbf{v}_1, \dots, \mathbf{v}_\ell \}$ for some $1 \leq \ell < m$. If the next vector, $\mathbf{v}_{\ell+1}$ is not in W_ℓ, then the orthog. projection $\mathbf{v}'_{\ell+1}$ of $\mathbf{v}_{\ell+1}$ is not $\mathbf{v}_{\ell+1}$, so $\mathbf{c}_{k+1} = \mathbf{v}_{\ell+1} - \mathbf{v}'_{\ell+1} \neq \mathbf{0}$ will do for the next orthogonal element, otherwise the projection is $\mathbf{v}_{\ell+1}$ itself, and we would get $\mathbf{c}_{k+1} = \mathbf{0}$, which we discard.
In the end we get a generating set of nonzero orthogonal vectors, which is a basis of W by the lemma.

We can also use this method for finding an orthogonal basis of $W^\perp$:
If $\{ \mathbf{v}_1, \dots, \mathbf{v}_m \}$ is a generating set of W, complete it to a generating set of V (to make sure that it spans K^n, we may simply add $\mathbf{e}_1, \dots, \mathbf{e}_n$). Then apply GS's method to this larger system of vectors $\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{e}_1, \dots, \mathbf{e}_n$ (or whatever we chose, taking the generating set of W first). The nonzero orthogonal vectors $\mathbf{c}_1, \dots \mathbf{c}_k$, that we got when we just finished processing $\mathbf{v}_1, \dots, \mathbf{v}_m$ will give a basis of W, the nonzero orthog. vectors obtained after this will give a basis of $W^\perp$. We may stop when the number of nonzero orthog. vectors reached n because then they already generate K^n.

Exercise. Consider the vectors $\mathbf{v}_1 = (1, 0, -1, 1)$, $\mathbf{v}_2 = (1, 0, 0, 2)$, $\mathbf{v}_3 = (0, 0, 1, 1)$ in $\mathbb{R}^4$. Give an orthogonal basis of $W = \operatorname{span}\{ \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \}$ and of $W^\perp$. (See Problem 24 in the excercise sheet.)

Solution. Complete the set to $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$, that is,

$$(1, 0, -1, 1), (1, 0, 0, 2), (0, 0, 1, 1), (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1).$$

We get $\mathbf{c}_1 = (1, 0, -1, 1)$, $\mathbf{c}_2 = (0, 0, 1, 1)$, and we discard $\mathbf{c}_3 = \mathbf{0}$, so an orthogonal basis of W is $\{(1, 0, -1, 1), (0, 0, 1, 1) \}$. continuing with the $\mathbf{e}_i$ vectors we get $\mathbf{c}_3 = (2, 0, 1, -1)$ $\mathbf{c}_4 = (0, 1, 0, 0)$, and this completes the basis, so an orthogonal basis of $W^\perp$ is $\{(2, 0, 1, -1), (0, 1, 0, 0) \}$.

The reduced and the full QR decomposition

Def. $A = \hat{Q}\hat{R}$ is a full QR decomposition of $A \in \mathbb{R}^{m \times n}$ if $Q \in R^{m \times m}$ is orthogonal, and $R \in \mathbb{R}^{m \times n}$ is upper triangular with positive elements in the main diagonal.

A reduced decompostion $A = QR$ can be turned to a full decomposition by completing the columns of Q to an orthonormal basis of $\mathbb{R}^m$ (using GS's method as above), and adding extra $\mathbf{0}$ rows to R: $\hat{Q} = [Q|Q_1]$, and $\hat{R} = \left[\frac{R}{0}\right]$. Conversely, every full decomposition can be reduced by taking only the first n columns of $\hat{Q}$ as Q, and the first n rows of $\hat{R}$ (that is, deleting the zero rows).

Example From an earlier exercise the reduced and full QR decomposition is

$$A = \begin{bmatrix} 1 & 3 & 6 \\ 1 & -1 & 2 \\ 1 & 3 & 2 \\ 1 & -1 & -2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 2 & 4 \\ 0 & 4 & 4 \\ 0 & 0 & 4 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 2 & 4 \\ 0 & 4 & 4 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

Though the reduced QR decomposition is uniqe, the full decomposition is clearly not: there are several ways to complete and orthonormal sysem to an orthonormal basis (for example, we can take the negative of any of the new vectors).

Householder reflections

We have seen earlier that any vector $\mathbf{a}$ can be mapped to a vector $\mathbf{b}$ of equal length by a reflection across a hyperplane whose normal vector is the vector $\mathbf{a} - \mathbf{b}$.

This gives a way to construct a full QR decomposition of the matrix A so that we change each colomn from left to right to a column of an upper triangular matrix by multiplication from the left by the orthogonal matrices of appropriate reflections.

Let $\mathbf{a}_1$ be the first column of A, and Q the matrix of the reflection that maps $\mathbf{a}_1$ to $(|\mathbf{a}_1|, 0, \ldots, 0)^T$. If in $Q_i \cdots Q_1 A$ the first i columns are already "above" the diagonal then let $\tilde{Q}_{i+1}$ be the matrix of the reflection mapping the first column of the lower right $(m-i) \times (n-i)$ submatrix of $Q_1 \cdots Q_i A$, than choose $Q_{i+1} = \begin{bmatrix} I & 0 \\ 0 & \tilde{Q}_{i+1} \end{bmatrix}$, which is also orthogonal. Then we get: $Q_n \ldots Q_1 A = R$ upper triangular, so $A = Q_1^{-1} \cdots Q_n^{-1} R = (Q_1^T \cdots Q_n^T) R = Q_1 \cdots Q_n R$, since the reflections are symmetric (self adjoint). Hhere the first term is orthogonal because it is a product of orthogonal matrices.

Example. Determine the full QR decomposition of the matrix A below by Householder reflections.

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & -3 \\ -2 & 5 & -7 \end{bmatrix}$$

Solution. $\begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix} \mapsto \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$ by the reflection across the hyperplane with normalvector $(2, -2, 2)$, or equivalently $\mathbf{a} = (1, -1, 1)$. Its matrix is $Q_1 = I - \frac{2}{3}\mathbf{a}\mathbf{a}^T$, so

$$Q_1 A = \begin{bmatrix} 1/3 & 2/3 & -2/3 \\ 2/3 & 1/3 & 2/3 \\ -2/3 & 2/3 & 1/3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & -3 \\ -2 & 5 & -7 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 3 \\ 0 & 4 & -5 \\ 0 & 3 & -5 \end{bmatrix}$$

Next we need to map $(4,3)$ to $(5,0)$, so $\tilde{Q}_2 = I - \frac{2}{10}\begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix} = \begin{bmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{bmatrix}$, and

$$Q_2 Q_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4/5 & 3/5 \\ 0 & 3/5 & -4/5 \end{bmatrix} \begin{bmatrix} 3 & -2 & 3 \\ 0 & 4 & -5 \\ 0 & 3 & -5 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 3 \\ 0 & 5 & -7 \\ 0 & 0 & 1 \end{bmatrix} = R, \quad \text{and}$$

$$A = Q_1 Q_2 R = \frac{1}{15}\begin{bmatrix} 5 & 2 & 14 \\ 10 & 10 & -5 \\ -10 & 11 & 2 \end{bmatrix} \begin{bmatrix} 3 & -2 & 3 \\ 0 & 5 & -7 \\ 0 & 0 & 1 \end{bmatrix}$$

Givens rotations

What is the matrix of a rotation of $\mathbb{R}^2$ by degree α? Using complex numbers: $x+iy \mapsto (x+iy)(\cos\alpha + i\sin\alpha) = (x\cos\alpha - y\sin\alpha) + i(x\sin\alpha + y\cos\alpha)$. So if A is the matrix of this transformation in the standard basis $\{1, i\}$, then

$$A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\alpha - y\sin\alpha \\ x\sin\alpha + y\cos\alpha \end{bmatrix} \Rightarrow A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

For any $\begin{bmatrix} a \\ b \end{bmatrix}$ there exists a rotation $\begin{bmatrix} a \\ b \end{bmatrix} \mapsto \begin{bmatrix} r \\ 0 \end{bmatrix}$, where $r = \sqrt{a^2+b^2}$. This will be the rotation about the origin by angle $-\alpha$, where α is the angle of $a+bi$. Its matrix is

$$\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} a/r & b/r \\ -b/r & a/r \end{bmatrix}. \quad \text{Indeed, } \begin{bmatrix} a/r & b/r \\ -b/r & a/r \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}.$$

Def. A **Givens rotation** is a rotation in the plane spanned by two coordinate axes. It leaves the other axes fixed. Its matrix is

$$\begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & \cos\alpha & & & & -\sin\alpha & & & \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & \sin\alpha & & & & \cos\alpha & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix}$$

Proposition. For $\forall$ $\mathbf{v} = (\ldots, a, \ldots, b \ldots)$ vector $\exists$ a Givens rotation, which maps it to $(\ldots, r \ldots, 0 \ldots)$, where $r = \sqrt{a^2+b^2}$, and in every vecor only the i'th and j'th component can change.

Proof. The matrix

$$\begin{bmatrix} 1 &&&&&&&&&& \\ & \ddots &&&&&&&&& \\ && 1 &&&&&&&& \\ &&& a/r &&&& b/r &&& \\ &&&& 1 &&&&&& \\ &&&&& \ddots &&&&& \\ &&&&&& 1 &&&& \\ &&& -b/r &&&& a/r &&& \\ &&&&&&&& 1 && \\ &&&&&&&&& \ddots & \\ &&&&&&&&&& 1 \end{bmatrix} \quad \text{will do.}$$

We can do Givens rotations to replace all but the first element of the first column by 0. Then we continue with the second column. The later rotations do not change the elements of the previous columns, which we already made zero. In the end, if the last element of the diagonal turns out to be negative, we have to do a reflecion in the end.

Exercise. Calculate the QR decomposition of the matrix A below by using Givens rotations.

$$A = \begin{bmatrix} 4 & 5 & 5 \\ 3 & -15 & -5 \\ 12 & 40 & 5 \end{bmatrix}$$

Solution First we map the first column vector $(4, 3, 12)$ to $(5, 0, 12)$ applying a rotation in the first two coordinates, this way making the element a_{21} zero, while only changing elements in the first two rows.

$$Q_1 A = \begin{bmatrix} 4/5 & 3/5 & 0 \\ -3/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 5 & 5 \\ 3 & -15 & -5 \\ 12 & 40 & 5 \end{bmatrix} = \begin{bmatrix} 5 & -5 & 1 \\ 0 & -15 & -7 \\ 12 & 40 & 5 \end{bmatrix}$$

Then we use rotation in the first and third coordinate (thus leaving the zero we already obtained intact) to map $(5, 0, 12)$ to $(13, 0, 0)$.

$$Q_2 Q_1 A = \begin{bmatrix} 5/13 & 0 & 12/13 \\ 0 & 1 & 0 \\ -12/13 & 0 & 5/13 \end{bmatrix} \begin{bmatrix} 5 & -5 & 1 \\ 0 & -15 & -7 \\ 12 & 40 & 5 \end{bmatrix} = \begin{bmatrix} 13 & 35 & 5 \\ 0 & -15 & -7 \\ 0 & 20 & 1 \end{bmatrix}$$

Finally we use rotation in the second and third coordinate mapping $(0, -15, 20)$ to $(0, 25, 0)$.

$$Q_3 Q_2 Q_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3/5 & 4/5 \\ 0 & -4/5 & -3/5 \end{bmatrix} \begin{bmatrix} 13 & 35 & 5 \\ 0 & -15 & -7 \\ 0 & 20 & 1 \end{bmatrix} = \begin{bmatrix} 13 & 35 & 5 \\ 0 & 25 & 5 \\ 0 & 0 & 5 \end{bmatrix} = R$$

Then

$$Q = Q_3 Q_2 Q_1 = \frac{1}{13} \begin{bmatrix} 4 & 3 & 12 \\ -3 & -12 & 4 \\ 12 & -4 & -3 \end{bmatrix}$$

and

$$A = Q^T R = \frac{1}{3} \begin{bmatrix} 4 & -3 & 12 \\ 3 & -12 & -4 \\ 12 & 4 & -3 \end{bmatrix} \begin{bmatrix} 13 & 35 & 5 \\ 0 & 25 & 5 \\ 0 & 0 & 5 \end{bmatrix}$$

Remark. A QR decomposition, in a broader sense, exists even if the columns of A are not independent. In that case the Q of the reduced decomposition will be an $m \times r$ matrix where $r = \operatorname{rank} A$, and R will be an $r \times n$ upper triangular matrix which can also have zeros in its diagonal. The other methods for the full decomposition can also be applied. In such a case even the reduced decomposition will not necessarily be unique.

Applications

The determinant, rank and image can be easily determined from a QR form, and it is more stable numerically than the Gaussian method. To approximate the eigenvalues of a matrix (provided they are all real, for instance when A is symmetric), one can use the following algorithm:

$$\begin{aligned} A_1 :=& A \\ A_1 =& Q_1 R_1 \text{ a (full) QR decomposition} \\ &\vdots \\ A_i =& Q_i R_i \text{ a QR decomposition} \\ A_{i+1} :=& R_i Q_i \\ A_{i+1} =& Q_{i+1} R_{i+1} \text{ a QR decomposition} \\ &\vdots \end{aligned}$$

Note that here $A_{i+1} = Q_i^{-1} A_i Q_i$, so all A_i are similar, and thus have the same eigenvalues. Under certain conditions A_i converges to an upper triangular matrix with the eigenvalues in its diagonal.