1) $\mathcal{R}(M)$ (the column space of M) and $\mathcal{N}(M^T)$ (the left null space of M) are orthogonal complements to each other in $\mathbb{R}^n$, and

2) $\mathcal{R}(M^T)$ (the row space of M) and $\mathcal{N}(M)$ (the right null space of M) are orthogonal complements to each other in $\mathbb{R}^m$.

As a corrolary: *The linear system $M\mathbf{x} = \mathbf{b}$ has solutions if and only if $\mathbf{b} \in \mathcal{R}(M)$ if and only if $\mathbf{b}$ is orthogonal to all solutions of $\mathbf{y}^T M = \mathbf{0}$.*

2.9. **Decomposing linear transformations.** Let $L : U \to V$ be a linear transformation between two inner product spaces $(V, \langle .,. \rangle_V)$, $(U, \langle .,. \rangle_U)$. It makes sense to split the spaces U and V into subspaces which carry information about L, and their orthogonal complements, which are redundant.

For example, only the subspace $\mathcal{R}(L)$ of V is "necessary" to L, so we could decompose $V = \mathcal{R}(L) \oplus \mathcal{R}(L)^\perp$, which, by Theorem 28, is the same as $V = \mathcal{R}(L) \oplus \mathcal{N}(L^*)$.

Also, the subspace $\mathcal{N}(L)$ of U is taken to zero through L, and we may wish to decompose $U = \mathcal{N}(L) \oplus \mathcal{N}(L)^\perp = \mathcal{N}(L) \oplus \mathcal{R}(L^*)$.

Then

$$L : U = \mathcal{R}(L^*) \oplus \mathcal{N}(L) \to V = \mathcal{R}(L) \oplus \mathcal{N}(L^*)$$

Recall that the rank of a matrix equals the one of its transpose. It also equals the rank of its complex conjugate, since $\det(\overline{M}) = \overline{\det M}$. Therefore

$$\operatorname{rank} M = \dim \mathcal{R}(L) = \dim \mathcal{R}(L^*) = \operatorname{rank} M^*$$

Let $B_1 = \{\mathbf{u}_1, \dots \mathbf{u}_r\}$ be an orthonormal basis for $\mathcal{R}(L^*)$ and $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$ an orthonormal basis for $\mathcal{N}(T)$. Then $B_U = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is a basis for U. Similarly, let $B_2 = \{\mathbf{v}_1, \dots \mathbf{v}_r\}$ be an orthonormal basis for $\mathcal{R}(L)$ and $\mathbf{v}_{r+1}, \dots, \mathbf{v}_m$ an orthonormal basis for $\mathcal{N}(L^*)$. Then $B_V = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis for V. It is easily checked that the matrix associated to L in the bases B_U, B_V has the block form

$$\begin{bmatrix} M_0 & 0 \\ 0 & 0 \end{bmatrix}$$

where M_0 is an $k \times k$ invertible matrix associated to the restriction of L:

$$L_0 : \mathcal{R}(L^*) \to \mathcal{R}(L) \ \text{ given by } L_0\mathbf{x} = L\mathbf{x}$$

which is onto, hence also one-to-one: it is invertible!

L_0, and its associated matrix M_0 constitute an invertible "core" of any linear transformation, respectively, matrix.

3. Least squares approximations

The **distance** *between two vectors* $\mathbf{x}, \mathbf{y}$ *in an inner product space is defined as* $\|\mathbf{x} - \mathbf{y}\|$.

The following characterization of orthogonal projections is very useful in approximations. As in Euclidian geometry, the shortest distance between a point and a line (or a plane) is the one measured on a perpendicular line:

Theorem 31. *Let* W *be a subspace of the inner product space* $(V, \langle\, , \,\rangle)$.

For any $\mathbf{x} \in V$ *the point in* W *which is at minimal distance to* $\mathbf{x}$ *is* $P\mathbf{x}$, *the orthogonal projection of* $\mathbf{x}$ *onto* W*:*

$$\|x - P\mathbf{x}\| = \min\{\|\mathbf{x} - \mathbf{w}\| \mid \mathbf{w} \in W\}$$

Proof.

This is an immediate consequence of the **Pythagorean theorem**:

$$\text{if } \ \mathbf{y} \perp \mathbf{z} \ \text{ then } \ \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 = \|\mathbf{y} + \mathbf{z}\|^2 \tag{18}$$

(check!) which used for $\mathbf{y} = P\mathbf{x} - \mathbf{w}$ and $\mathbf{z} = \mathbf{x} - P\mathbf{x}$ (orthogonal to W, hence to $\mathbf{y}$) gives

$$\|\mathbf{x} - \mathbf{w}\|^2 = \|\mathbf{w} - P\mathbf{x}\|^2 + \|\mathbf{x} - P\mathbf{x}\|^2, \quad \text{for any } \mathbf{w} \in W$$

implying that $\|\mathbf{x} - \mathbf{w}\| \geq \|\mathbf{x} - P\mathbf{x}\|$ with equality only for $\mathbf{w} = P\mathbf{x}$. □

3.1. **Overdetermined systems: best fit solution.** Let M be an $m \times n$ matrix with entries in $\mathbb{R}$ (one could also work in $\mathbb{C}$). By abuse of notation we will speak of the matrix M both as a matrix and as the linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ which takes $\mathbf{x}$ to $M\mathbf{x}$, denoting by $\mathcal{R}(M)$ the range of the transformation (the column space of M).

The linear system $M\mathbf{x} = \mathbf{b}$ has solutions if and only if $\mathbf{b} \in \mathcal{R}(M)$. If $\mathbf{b} \notin \mathcal{R}(M)$ then the system has no solutions, and it is called *overdetermined*.

In practice overdetermined systems are not uncommon, usual sources being that linear systems are only models for more complicated phenomena, and the collected data is subject to errors and fluctuations. For practical problems it is important to produce a best fit solution: an $\mathbf{x}$ for which the error $M\mathbf{x} - \mathbf{b}$ is as small as possible.

There are many ways of measuring such an error, often this is *the least squares*: find $\mathbf{x}$ which minimizes the square error:

$$S = r_1^2 + \ldots + r_m^2 \quad \text{ where } r_j = (M\mathbf{x})_j - b_j$$

Of course, this is the same as minimizing $\|M\mathbf{x} - \mathbf{b}\|$ where the inner product is the usual dot product on $\mathbb{R}^m$. By Theorem 31 it follows that $M\mathbf{x}$ must equal $P\mathbf{b}$, the orthogonal projection of $\mathbf{b}$ onto the subspace $\mathcal{R}(M)$.

We now need to solve the system $M\overline{\mathbf{x}} = P\mathbf{b}$ (solvable since $P\mathbf{b} \in \mathcal{R}(M)$).

Case I: If M *is one to one*, then there is a unique solution $\overline{\mathbf{x}} = M^{-1}P\mathbf{b}$.

An easier to implement formula can be found as follows. Since $(\mathbf{b} - P\mathbf{b}) \perp \mathcal{R}(M)$ then $(\mathbf{b} - P\mathbf{b}) \in \mathcal{N}(M^*)$ (by Theorem 28 (i)) therfeore $M^*\mathbf{b} = M^*P\mathbf{b}$, so

$$M^*\mathbf{b} = M^*M\overline{\mathbf{x}}$$

which is also called *the normal equation* in statistics.

Remark: If M is one to one, then so is M^*M (since if $M^*M\mathbf{x} = \mathbf{0}$ then $0 = \langle M^*M\mathbf{x}, \mathbf{x}\rangle = \langle M\mathbf{x}, M\mathbf{x}\rangle$ therefore $M\mathbf{x} = \mathbf{0}$).

Since we assumed M is one to one, then M^*M is one to one, therefore it is invertible (being a square matrix), and we can solve

$$\overline{\mathbf{x}} = (M^*M)^{-1}M^*\mathbf{b}$$

Since $M\overline{\mathbf{x}} = P\mathbf{b}$ we also found a formula for the projection

$$P\mathbf{b} = M(M^*M)^{-1}M^*\mathbf{b} \tag{19}$$

Case II: If M is not one to one, then, given one solution $\overline{\mathbf{x}}$ then any vector in the space $\overline{\mathbf{x}} + \mathcal{N}(M)$ is a solution as well. Choosing the vector $\overline{\overline{\mathbf{x}}}$ with the smallest norm in $\overline{\mathbf{x}} + \mathcal{N}(M)$, this gives $\overline{\overline{\mathbf{x}}} = M^+\mathbf{b}$ where M^+ is called the *pseudoinverse* of M. The notion of pseudoinverse will be studied in more detail later on.

3.2. **Another formula for orthogonal projections.** Formula (19) is another useful way of writing projections. Suppose that W is a subspace in $\mathbb{R}^n$ (or $\mathbb{C}^n$) and $\mathbf{x}_1, \dots, \mathbf{x}_r$ is a basis for W (not necessarily orthonormal). The matrix $M = [\mathbf{x}_1, \dots, \mathbf{x}_r]$ has its column space equal to W and has linearly independent columns, therefore a zero null space. Then (19) is the orthogonal projection to W.

4. Orthogonal and unitary matrices, QR factorization

4.1. **Unitary and orthogonal matrices.** Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. The following type of linear transformation are what isomorphisms of inner product spaces should be: linear, invertible, and they preserve the inner product (therefore angles and lengths):

Definition 32. *A linear transformation $U : V \to V$ is called a* **unitary transformation** *if $U^*U = UU^* = I$.*

Remark. U is unitary if and only if U^* is unitary.

In the language of matrices:

Definition 33. *An* **unitary matrix** *is an $n \times n$ matrix with $U^*U = UU^* = I$ and therefore $U^{-1} = U^*$.*

Definition 34. *An unitary matrix with real entries is called an* **orthogonal matrix**.

In other words, Q is an orthogonal matrix means that Q has real elements and $Q^{-1} = Q^T$.

We should immediately state and prove the following properties of unitary matrices, each one can be used as a definition for a unitary matrix:

Theorem 35. *Let U be $n \times n$ a matrix.*

The following statements are equivalent:
(i) U is unitary.
(ii) The columns of U form an orthonormal set of vectors (therefore an orthonormal basis of $\mathbb{C}^n$).
(iii) U preserves angles: $\langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.
(iv) U is an isometry: $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$.

Remark. If U is unitary, also *the rows of U from an orthonormal set.*

Examples. Rotation matrices and reflexion matrices in $\mathbb{R}^n$, and their products, are orthogonal (they preserve the length of vectors).

Remark. An isometry is necessarily one to one, and therefore, it is also onto (in finite dimensional spaces).

Remark. The equivalence between (i) and (iv) is not true in infinite dimensions (unless the isometry is assumed onto).

Proof of Theorem 35.

(i)$\Longleftrightarrow$(ii) is obvious by matrix multiplication (line j of U^* multiplying, place by place, column i of U is exactly the dot product of column j complex conjugated and column i).

(i)$\Longleftrightarrow$(iii) is obvious because $U^*U = I$ is equivalent to $\langle U^*U\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, which is equivalent to (iii).

(iii)$\Rightarrow$(iv) by taking $\mathbf{y} = \mathbf{x}$.

(iv)$\Rightarrow$(iii) follows from the polarization identity. $\square$

4.2. **Rectangular matrices with orthonormal columns. A simple formula for orthogonal projections.**

Let $M = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ be an $n \times k$ matrix whose columns $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are an orthonormal set of vectors. Then necessarily $n \geq k$. If $n = k$ then the matrix M is unitary, but assume here that $n > k$.

Note that $\mathcal{N}(M) = \{\mathbf{0}\}$ since the columns are independent.

Note also that

$$M^*M = I$$

(since line j of M^* multiplying, place by place, column i of M is exactly $\overline{\mathbf{u}_j} \cdot \mathbf{u}_i$ which equals 1 for $i = j$ and 0 otherwise). Then the least squares minimization formula (19) takes the simple form $P = MM^*$:

Theorem 36. *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be an orthonormal set, and $M = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$.*

The orthogonal projection onto $U = Sp(\mathbf{u}_1, \ldots, \mathbf{u}_k) = \mathcal{R}(M)$ (cf. §3.2) is

$$P = MM^*$$

4.3. **QR factorization.** The following decomposition of matrices has countless applications, and extends to infinite dimensions.

If an $m \times k$ matrix $M = [\mathbf{y}_1, \ldots, \mathbf{y}_k]$ has linearly independent columns (hence $m \geq k$ and $\mathcal{N}(M) = \mathbf{0}$) then applying the Gram-Schmidt process on the columns $\mathbf{y}_1, \ldots, \mathbf{y}_k$ amounts to factoring $M = QR$ as described below.

Theorem 37. QR factorization of matrices

Any $m \times k$ matrix $M = [\mathbf{y}_1, \ldots, \mathbf{y}_k]$ with linearly independent columns can be factored as $M = QR$ where Q is an $m \times k$ matrix whose columns form an orthonormal basis for $\mathcal{R}(M)$ (hence $Q^*Q = I$) and R is an $k \times k$ upper triangular matrix with positive entries on its diagonal (hence R is invertible).

In the case of a square matrix M then Q is also square, and it is a unitary matrix.

If the matrix M has real entries, then Q and R have real entries, and if $k = n$ then Q is an orthogonal matrix.

Remark 38. *A similar factorization can be written $A = Q_1 R_1$ with Q_1 an $m \times m$ unitary matrix and R_1 an $m \times k$ rectangular matrix whose first k rows are the upper triangular matrix R and the last $m - k$ rows are zero:*

$$A = Q_1 R_1 = [Q \ Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = QR$$

Proof of Theorem 37.

Every step of the Gram-Schmidt process (8), (9), ... (11) is completed by a special normalization: after obtaining an orthonormal set $\mathbf{u}_1, \ldots, \mathbf{u}_k$ let

$$\mathbf{q}_j = \gamma_j \, \mathbf{u}_j \quad \text{for all } j = 1, \ldots k, \quad \text{where } \gamma_j \in \mathbb{C}, \ |\gamma_j| = 1 \tag{20}$$

where the numbers γ_j will be suitably determined to obtain a special orthonormal set $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

First, replace $\mathbf{u}_1, \ldots, \mathbf{u}_k$ in (8), (9),...,(11) by the orthonormal set $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

Then invert: write $\mathbf{y}_j$'s in terms of $\mathbf{q}_j$'s. Since $\mathbf{q}_1 \in Sp(\mathbf{y}_1)$, $\mathbf{q}_2 \in Sp(\mathbf{y}_1, \mathbf{y}_2)$,..., $\mathbf{q}_j \in Sp(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j)$, ... then $\mathbf{y}_1 \in Sp(\mathbf{q}_1)$, $\mathbf{y}_2 \in Sp(\mathbf{q}_1, \mathbf{q}_2)$,..., $\mathbf{y}_j \in Sp(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_j)$, ... and therefore there are scalars c_{ij} so that

$$\mathbf{y}_j = c_{1j}\mathbf{q}_1 + c_{2j}\mathbf{q}_2 + \ldots + c_{jj}\mathbf{q}_j \quad \text{for each } j = 1, \ldots, k \tag{21}$$

and since $\mathbf{q}_1, \ldots, \mathbf{q}_k$ are orthonormal, then

$$c_{ij} = \langle \mathbf{q}_i, \mathbf{y}_j \rangle \tag{22}$$

Relations (21), (22) can be written in matrix form as

$$M = QR$$

with

$$Q = [\mathbf{q}_1, \ldots, \mathbf{q}_k], \quad R = \begin{bmatrix} \langle \mathbf{q}_1, \mathbf{y}_1 \rangle & \langle \mathbf{q}_1, \mathbf{y}_2 \rangle & \ldots & \langle \mathbf{q}_1, \mathbf{y}_k \rangle \\ 0 & \langle \mathbf{q}_2, \mathbf{y}_2 \rangle & \ldots & \langle \mathbf{q}_2, \mathbf{y}_k \rangle \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \langle \mathbf{q}_k, \mathbf{y}_k \rangle \end{bmatrix}$$

We have the freedom of choosing the constants γ_j of modulus 1, and they can be chosen so that all diagonal elements of R are positive (since $\langle \mathbf{q}_j, \mathbf{y}_j \rangle = \overline{\gamma_j} \langle \mathbf{u}_j, \mathbf{y}_j \rangle$ choose $\gamma_j = \langle \mathbf{u}_j, \mathbf{y}_j \rangle / |\langle \mathbf{u}_j, \mathbf{y}_j \rangle|$). $\square$

For numerical calculations the Gram-Schmidt process described above accumulates round-off errors. For large m and k other more efficient, numerically stable, algorithms exist, and should be used.

Applications of the QR factorization to solving linear systems.

1. Suppose M is an invertible square matrix. To solve $M\mathbf{x} = \mathbf{b}$, factoring $M = QR$, the system is $QR\mathbf{x} = \mathbf{b}$, or $R\mathbf{x} = Q^*\mathbf{b}$ which can be easily solved since R is triangular.

2. Suppose M is an $m \times k$ rectangular matrix, of full rank k. Since $m > k$ the linear system $M\mathbf{x} = \mathbf{b}$ may be overdetermined. Using the QR factorization in Remark 38, the system is $Q_1R_1\mathbf{x} = \mathbf{b}$, or $R_1\mathbf{x} = Q_1^*\mathbf{b}$, which is easy to see if it has solutions: the last $m-k$ rows of $Q_1^*\mathbf{b}$ must be zero. If this is the case, the system can be easily solved since R is upper triangular.