LEMMA 1.5. *If S is an affine subspace and $b' \notin S$, then there exists $\hat{x} \in X$ such that $d(b', S) = \|\hat{x} + c - b'\|$. Furthermore, $\hat{x} - (b' - c)$ is orthogonal to x for all $x \in X$. (Note that here we use b' instead of b, to avoid confusion with the system's right-hand side.)*

PROOF. We have $S = \{y : y = x + c, c \neq 0, x \in X\}$, where X is a closed linear subspace of V. Now,

$$\begin{aligned} d(b', S) &= \inf_{y \in S} \|y - b'\| = \inf_{x \in X} \|x + c - b'\| \\ &= \inf_{x \in X} \|x - (b' - c)\| = d(b' - c, X) \\ &= \|\hat{x} - (b' - c)\| = \|\hat{x} + c - b'\|. \end{aligned}$$

The point $\hat{x} \in X$ exists since X is a closed linear subspace. It follows from Theorem 1.3 that $\hat{x} - (b' - c)$ is orthogonal to X. Note that the distance between S and b' is the same as that between X and $b' - c$. ■

From the proof above, we see that $\hat{x} + c$ is the element of S closest to b'. For the case $b' = 0$, we find that $\hat{x} + c$ is orthogonal to X.

Now we return to the problem of finding the "smallest" solution of an underdetermined problem. Assume A has "maximal rank"; that is, m of the column vectors of A are linearly independent. We can write the solutions of the system as $x = x_0 + z$, where x_0 is a particular solution and z is a solution of the homogeneous system $Az = 0$. So the solutions of the system $Ax = b$ form an affine subspace. As a result, if we want to find the solution with the smallest norm (i.e., closest to the origin) we need to find the element of this affine subspace closest to $b' = 0$. From the above, we see that such an element must satisfy two properties. First, it has to be an element of the affine subspace (i.e., a solution to the system $Ax = b$) and second, it has to be orthogonal to the linear subspace X, which is the null space of A (the set of solutions of $Az = 0$). Now consider $x' = A^T(AA^T)^{-1}b$; this vector lies in the affine subspace of the solutions of $Ax = b$, as one can check by multiplying it by A. Furthermore, it is orthogonal to every vector in the space of solutions of $Az = 0$ because $(A^T(AA^T)^{-1}b, z) = ((AA^T)^{-1}b, Az) = 0$. This is enough to make x' the unique solution of our problem.

1.2. Orthonormal Bases

The problem presented in the previous section, of finding an element in a closed linear space that is closest to a vector outside the space, lies in the framework of approximation theory, where we are given a function (or a vector) and try to find an approximation to it as a linear combination of given functions (or vectors). This is done by requiring that the norm of the error (difference between the given function and the approximation) be minimized. In what follows, we shall find coefficients for this optimal linear combination.

DEFINITION. Let S be a linear vector space. A collection of m vectors $\{u_i\}_{i=1}^m$ belonging to S are linearly independent if and only if $\lambda_1 u_1 + \cdots + \lambda_m u_m = 0$ implies $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$.

DEFINITION. Let S be a linear vector space. A collection $\{u_i\}_{i=1}^m$ of vectors belonging to S is called a basis of S if $\{u_i\}$ are linearly independent and any vector in S can be written as a linear combination of them.

Note that the number of elements of a basis can be finite or infinite depending on the space.

THEOREM 1.6. *Let S be an m-dimensional linear inner-product space with m finite. Then any collection of m linearly independent vectors of S is a basis.*

DEFINITION. A set of vectors $\{e_i\}_{i=1}^m$ is orthonormal if the vectors are mutually orthogonal and each has unit length (i.e., $(e_i, e_j) = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise).

The set of all the linear combinations of the vectors $\{u_i\}$ is called the span of $\{u_i\}$ and is written as $\text{Span}\{u_1, u_2, \dots, u_m\}$.

Suppose we are given a set of vectors $\{e_i\}_{i=1}^m$ that are an orthonormal basis for a subspace S of a real vector space. If b is an element outside the space, we want to find the element $\hat{b} \in S$, where $\hat{b} = \sum_{i=1}^m c_i e_i$ such that $\|b - \sum_{i=1}^m c_i e_i\|$ is minimized. Specifically, we have

$$\begin{aligned}
\left\| b - \sum_{i=1}^m c_i e_i \right\|^2 &= \left(b - \sum_{i=1}^m c_i e_i \,,\, b - \sum_{j=1}^m c_j e_j \right) \\
&= (b, b) - 2\sum_{i=1}^m c_i (b, e_i) + \left(\sum_{i=1}^m c_i e_i \,,\, \sum_{j=1}^m c_j e_j \right) \\
&= (b, b) - 2\sum_{i=1}^m c_i (b, e_i) + \sum_{i,j=1}^m c_i c_j (e_i, e_j) \\
&= (b, b) - 2\sum_{i=1}^m c_i (b, e_i) + \sum_{i=1}^m c_i^2 \\
&= \|b\|^2 - \sum_{i=1}^m (b, e_i)^2 + \sum_{i=1}^m (c_i - (b, e_i))^2,
\end{aligned}$$

where we have used the orthonormality of the e_i to simplify the expression. As is readily seen, the norm of the error is a minimum when

$c_i = (b, e_i), i = 1 \dots m$, so that $\hat{b}$ is the projection of b onto S. It is easy to check that $b - \hat{b}$ is orthogonal to any element in S. Also, we see that the following inequality, called Bessel's inequality, holds:

$$\sum_{i=1}^{m} (b, e_i)^2 \leq \|b\|^2.$$

When the basis is not orthonormal, steps similar to the above yield

$$\begin{aligned}
\left\| b - \sum_{i=1}^{m} c_i g_i \right\|^2 &= \left(b - \sum_{i=1}^{m} c_i g_i, b - \sum_{j=1}^{m} c_j g_j \right) \\
&= (b, b) - 2\sum_{i=1}^{m} c_i (b, g_i) + \left(\sum_{i=1}^{m} c_i g_i, \sum_{j=1}^{m} c_j g_j \right) \\
&= (b, b) - 2\sum_{i=1}^{m} c_i (b, g_i) + \sum_{i,j=1}^{m} c_i c_j (g_i, g_j).
\end{aligned}$$

If we differentiate the last expression with respect to c_i and set the derivatives equal to zero, we get

$$Gc = r,$$

where G is the matrix with entries $g_{ij} = (g_i, g_j)$, $c = (c_1, \dots, c_m)^T$, and $r = ((g_1, b), \dots, (g_m, b))^T$. This system can be ill-conditioned so that its numerical solution presents a problem. The question that arises is how to find, given a set of vectors, a new set that is orthonormal. This is done through the Gram-Schmidt process, which we now describe.

Let $\{u_i\}_{i=1}^m$ be a basis of a linear subspace. The following algorithm will give an orthonormal set of vectors $e_1, e_2, \dots, e_m$ such that $\text{Span}\{e_1, e_2, \dots, e_m\} = \text{Span}\{u_1, u_2, \dots, u_m\}$.

1. Normalize u_1 (i.e., let $e_1 = u_1/\|u_1\|$).
2. We want a vector e_2 that is orthonormal to e_1. In other words we look for a vector e_2 satisfying $(e_2, e_1) = 0$ and $\|e_2\| = 1$. Take $e_2 = u_2 - (u_2, e_1)e_1$ and then normalize.
3. In general, e_j is found recursively by taking

$$e_j = u_j - \sum_{i=1}^{j-1} (u_j, e_i) e_i$$

and normalizing.

The Gram-Schmidt process can be implemented numerically very efficiently. The solution of the recursion above is equivalent to finding