Available online at www.sciencedirect.com
SCIENCE DIRECT•
ELSEVIER
Computers and Mathematics with Applications 48 (2004) 177–190
An International Journal
computers & mathematics with applications
www.elsevier.com/locate/camwa

Affine Projections

R. PIZIAK AND P. L. ODELL
Department of Mathmatics, Baylor University
Waco, TX 76798, U.S.A.

(Received and accepted July 2003)

Abstract—This paper presents a "constructive" method for projecting a vector onto an affine subspace of a vector space. It also provides formulas for projecting onto the intersection and "sums" of such subspaces. © 2004 Elsevier Ltd. All rights reserved.

Keywords—Affine subspaces, Projections, Moore-Penrose pseudoinverse.

1. AFFINE SUBSPACES

Most would agree that the central problem of linear algebra is the solving of systems of linear equations. A useful collection of mathematical theory (matrix theory, vector space theory, etc.) has been erected upon this problem which finds application in so many areas (statistics, differential equations, tomography, coding theory, engineering, etc.). A great deal of attention is paid in introductory linear algebra courses to subspaces of a vector space. A (*linear*) *subspace* of a vector space is a nonempty subset distinguished by a closure property. That is, a subspace is a nonempty subset which is closed under the taking of linear combinations. In other words, U is a subspace if U is a nonempty subset of the vector space and if $x, y \in U$ implies $\alpha x + \beta y \in U$ for arbitrary choice of (appropriate) scalars α and β. By induction, this condition is equivalent to being closed under arbitrary finite linear combinations. That is, given $x_1, x_2, \dots, x_k$ in U and any scalars $\alpha_1, \alpha_2, \dots, \alpha_k$, then $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \in U$. In particular then, a subspace is never empty though it could have just the zero vector $\vec{0}$ in it or it could even be the whole space. For real or complex vector spaces, there will be infinitely many vectors in any subspace as soon as there is one that is not zero. Given two subspaces U_1 and U_2, there are two natural constructions that lead to potentially new subspaces. Namely, the *intersection* $U_1 \cap U_2$ is always a subspace as is the *linear sum* $U_1 + U_2 = \{x + y \mid x \in U_1, y \in U_2\}$. These subspaces can be characterized abstractly. The intersection $U_1 \cap U_2$ is the largest subspace contained within both U_1 and U_2 and the sum $U_1 + U_2$ is the smallest subspace that contains both U_1 and U_2. These constructions extend to arbitrary collections of subspaces, but we shall not need them here.

Now consider the set of solutions of a system of linear equations written in matrix form, $Ax = b$. Let $S = \{x \mid Ax = b\}$. If $b \neq \vec{0}$, then S is *not* a subspace. Indeed, S *is* a subspace iff $b = \vec{0}$. However, S does enjoy a restricted kind of closure property. Namely, if x_1 and x_2 are in S, then $\lambda x_1 + (1 - \lambda) x_2 \in S$ as well for any scalar λ. The proof is quick and easy and is based on the

0898-1221/04/$ - see front matter Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
doi:10.1016/j.camwa.2004.07.001

following computation:

$$A(\lambda x_1 + (1-\lambda)x_2) = \lambda Ax_1 + (1-\lambda)Ax_2 = \lambda b + (1-\lambda)b = b.$$

Any subset S that satisfies this kind of closure property is called an *affine subspace*. This closure property is equivalent to being closed under the taking of arbitrary finite affine combinations. That is, $x_1, x_2, \ldots, x_k \in S$ and $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$ implies $\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k \in S$. For fixed x, y, the set of vectors $\lambda x + (1-\lambda)y$ is interpreted as the "line" through x and y. So, one way to look at an affine subspace is as a subset that contains the entire line through any two of its vectors (points). This is why some call affine subspaces "*flats*"? In some sense, then, it would seem more natural to study affine subspaces rather than ordinary ones. Of course, it turns out they are closely related. Given any subset W of a vector space, we can translate it by a fixed vector a. The *translate of W by vector a* is defined by

$$a + W = \{a + w \mid w \in W\} = W + a.$$

The reader may easily verify the first useful fact about translates.

LEMMA 1.1. *The translate of an affine subspace is again an affine subspace.*

Examples of affine subspaces are

(1) the empty set $\emptyset$ which satisfies the closure property vacuously,
(2) singleton sets of vectors (points),
(3) all linear subspaces and, in view of the preceding lemma,
(4) all translates of linear subspaces.

It is easy to identify the linear subspaces among the affine ones since they are the affine subspaces that contain the origin.

LEMMA 1.2. *M is a linear subspace iff M is an affine subspace that contains the origin (i.e., the zero vector).*

Again, we leave the details to the reader but offer the hint that first you show $\vec{0} \in M$ affine implies M is closed under all scalar multiples. Then, note $x + y = 2((1/2)(x+y)) = 2((1/2)x + (1/2)y) = 2((1/2)x + (1-1/2)y)$.

We now come to the significant fact that our list of examples of affine subspaces is complete. That is, affine subspaces are just the translates of linear subspaces.

THEOREM 1.3. *If M is an affine subspace, there is a unique linear subspace U such that $M = a + U$ for some vector a.*

PROOF. To argue uniqueness, suppose $M = a_1 + U_1 = a_2 + U_2$. Then, $U_1 = (a_2 - a_1) + U_2$ so in particular $\vec{0} \in (a_2 - a_1) + U_2$. This says $a_2 - a_1 \in U_2$, but since U_2 is a subspace, $a_1 - a_2 = -(a_2 - a_1) \in U_2$ as well. Thus, $U_2 \supseteq U_2 + (a_2 - a_1) = U_1$. A symmetric argument gives $U_1 \supseteq U_2$, hence, $U_1 = U_2$. For existence, let $U = \{x - y \mid x \in M, y \in M\}$. It is easy to check that U is affine and $\vec{0} \in U$ so the previous lemma applies. ∎

So, to each affine subspace we have associated a unique linear subspace. This allows concepts from linear algebra, like dimension, etc., to be applied to affine subspaces as well. However, this would take us too far afield in this short paper. We note that while the subspace associated with an affine subspace is unique, the translating vector is not. Nonetheless, we have the following simple connection.

LEMMA 1.4. *Let U be a linear subspace. Then, $a + U = b + U$ iff $b - a \in U$ iff $a \in b + U$.*

Another significant fact is that affine subspaces (in finite dimensions) are precisely the solution sets of systems of linear equations.

THEOREM 1.5. *The affine subspaces of $\mathbb{C}^n$ are exactly the solution sets to systems of linear equations in n variables. That is,* $\{x \in \mathbb{C}^n \mid Ax = b \in \mathbb{C}^m\}$ *is an affine subspace of* $\mathbb{C}^n$. *Moreover, any affine subspace of* $\mathbb{C}^n$ *may be represented in this way.*

PROOF. The reader is referred to [1, pp. 5–6]. ∎

This theorem reinforces our understanding that systems of linear equations can be inconsistent and have no solutions, the solution set is empty, they can have a unique solution, the solution set is a singleton set, or they can have infinitely many solutions, the solution set is a nontrivial affine subspace.

What about the constructions of intersection and of linear sum for linear subspaces? Are there analogous constructions for affine subspaces? Intersections work out more or less naturally but "sums" are a bit tricky.

THEOREM 1.6. *Let* $a + U_1$ *and* $b + U_2$ *be affine subspaces. Then,* $(a + U_1) \cap (b + U_2)$ *is again an affine subspace (possibly the empty one). Moreover, it is the largest affine subspace contained in both* $a + U_1$ *and* $b + U_2$. *If the intersection is not empty, then* $(a + U_1) \cap (b + U_2) = c + (U_1 \cap U_2)$ *where c is any vector in the intersection. We can even tell when the intersection is nonempty. Namely,* $(a + U_1) \cap (b + U_2)$ *is not empty iff* $a - b \in U_1 + U_2$ *and* $(a + U_1) \cap (b + U_2)$ *is a singleton set iff* $a - b \in U_1 + U_2 =$ *the whole space and* $U_1 \cap U_2 = (\vec{0})$. *Finally, the smallest affine subspace containing* $a + U_1$ *and* $b + U_2$ *is* $a + [\mathrm{sp}\{a - b\} + U_1 + U_2] = b + [\mathrm{sp}\{a - b\} + U_1 + U_2]$ *and will be denoted* $(a + U_1) \boxplus (b + U_2)$. *Note that* sp *stands for span.*

PROOF. There are many details to check here but they are more or less routine and safely left to the reader. ∎

It is instructive to think why the "obvious" of sum does not work, namely, $(a+b+U_1+U_2)$. The theorem above opens the door to a discussion of affine and projective geometry. The interested reader is referred to the book by Bennett [2] for an elaboration and discussion of these two kinds of geometry that lie at the very heart of linear algebra. Our focus must be much narrower.

We wish to pursue the analog of orthogonal projections onto linear subspaces by projection onto affine subspaces as well. This is not just an exercise to keep mathematicians busy on weekends, but affine projections actually are used, for example, in the Kaczmarz projection method (see [3]). From now on, perpendicularity plays an important role so inner products are key. To get some nice formulas for projections we will use the Moore-Penrose pseudoinverse. That is the subject of the next section. We end with a summary table.

Table 1.

Projective Geometry	Affine Geometry
Linear subspaces—Closed under the formation of arbitrary linear combinations	Affine subspaces–Closed under the formation of arbitrary affine combinations
Solution sets of homogeneous systems of linear equations	Solution sets of arbitrary systems of linear equations
U a linear subspace	$a + U$ an affine subspace
$U_1 \cap U_2$—The largest linear subspace contained in both U_1 and U_2	$(a + U_1) \cap (b + U_2)$—The largest affine subspace contained in both $a + U_1$ and $b + U_2$
$U_1 + U_2$—The smallest linear subspace containing both U_1 and U_2	$(a + U_1) \boxplus (b + U_2)$—The smallest affine subspace containing both $a + U_1$ and $b + U_2$

2. THE MOORE-PENROSE PSEUDOINVERSE

The matrix theory required for this paper is a very basic understanding of the Moore-Penrose generalized inverse (the pseudoinverse) of a matrix. That is, if A is an m-by-n matrix, then $X = A^+$ (the pseudoinverse) is the unique solution to the four matrix equations

$$AXA = A, \tag{1}$$
$$XAX = X, \tag{2}$$
$$(AX)^* = AX, \tag{3}$$
$$(XA)^* = XA. \tag{4}$$

(Here * denotes conjugate transpose.)

Note that (1) and (3) imply that $(AX)^2 = (AA^+)^2 = AA^+AA^+ = AA^+ = (AA^+)^*$, that is, AA^+ is an Hermitian idempotent matrix. The same is so for A^+A. That is, both are orthogonal projections.

References on the pseudoinverse include Boullion and Odell [4], Ben-Israel and Greville [5], Rao and Mitra [6] and Campbell and Meyer [7]. Recently, two excellent references have appeared, Meyer [3] and Rao and Rao [8]. A handbook compiled by Lutpepohl [9] is also helpful for supplying various identities and formulas. Meyer [3] gives most all the background one needs for this paper. However, Rao and Rao is filled with helpful identities and results for constructive arguments, for formulating mathematical models and especially for performing statistical evaluations.

The following properties of the pseudoinverse are used in our discussion and can be proved by substituting the right side of each equation below into the four defining equations (1)–(4) and relying on the uniqueness of A^+.

$$A^+ = (A^*A)^+A^*, \tag{5}$$
$$A^+ = A^*(AA^*)^+, \tag{6}$$
$$(AA^*)^+ = (A^*)^+A^+, \tag{7}$$
$$(A^*)^+ = (A^+)^*. \tag{8}$$

Also, the following facts are fundamental to our presentation.

LEMMA 2.1. *Let $AX = C$ be a matrix equation in the unknown matrix X. Then, a solution exists if and only if the following consistency condition holds:*

$$AA^+C = C.$$

LEMMA 2.2. *If $AA^+C = C$, then $\{x \mid x = A^+C + [I - A^+A]Z, Z \text{ is arbitrary}\}$ is the solution set for $Ax = C$, where x is a vector.*

LEMMA 2.3. *If $AA^+C \neq C$, then $x = A^+C$ is the best approximate solution in the sense that $\|Ax - C\|_2 \geq \|AA^+C - C\|_2$; that is, $x = A^+C$ is the least squares solution of $Ax = C$.*

The proofs of these facts can be found in the references listed above. We pause to note that if y is an m-by-1 (column) vector and A is an m-by-n matrix, then the vector $y_0 = AA^+y$ belongs to the column space of A and is that vector such that $\|y - y_0\|_2 = \|y - AA^+y\|_2$ is minimal. In other words, $\|y - y_0\|_2 = \|y - AA^+y\|_2$ is minimal.

LEMMA 2.4. *$P_{\mathrm{Col}(A)} = AA^+$ is the projection matrix (idempotent and Hermitian) that projects any $y \in \mathbb{C}^n$ onto the column space of A; that is, $P_{\mathrm{Col}(A)}y = AA^+y = y_0 \in \mathrm{Col}(A)$, where $\mathrm{Col}(A)$*

is the column space of A and $\|y-y_0\|_2$ is minimal. Moreover, $y = AA^+y + [I - AA^+]y = y_0 + y_1$, is an orthogonal decomposition of the vector y since the inner product $y_0^* y_1 = \vec{0}$.

We further note that all of the so-called "fundamental" of A can be computed using A^+:

- A^+A is the projection on the column space of A^*,
- $I - AA^+$ is the projection on the orthogonal complement of the column space of A, which is the null space of A^*, and
- $I - A^+A$ is the projection on the orthogonal complement of the column space of A^* which is the null space of A.

Again, we recommend the text by Meyer [3], to those who wish to gain more insight into the pseudoinverse and its uses.

3. THE MATRIX CONNECTION

In order to use matrix theory on subspaces, we need to make a connection between the two. If we have a matrix A, then we can always form the column space $\mathrm{Col}(A)$, that is, the subspace spanned (i.e., generated) by the columns of A and get a perfectly fine linear subspace. We use "sp" as a shorthand for "span". On the other hand, given a subspace, we can always choose a spanning set of vectors, even a basis, and form these into the columns of a matrix so that the column space of this matrix evidently gives the subspace back. For intersections, if $U_1 = \mathrm{Col}(A)$ and $U_2 = \mathrm{Col}(B)$, then $U_1 \cap U_2 = \mathrm{Col}(A) \cap \mathrm{Col}(B) = \{y \mid y = Ax_1 = Bx_2 \text{ for some } x_1 \text{ and } x_2\}$. For the sum, we have

$$U_1 + U_2 = \mathrm{Col}(A) + \mathrm{Col}(B) = \mathrm{Col}\left(\left[A \vdots B\right]\right) = \left\{ y \mid y = \left[A \vdots B\right] \begin{bmatrix} x_1 \\ \cdots \\ x_2 \end{bmatrix} \right\},$$

where $[A \vdots B]$ is matrix A augmented by the matrix B. That is, put matrix B to the right of matrix A making a bigger new matrix. Now if we have an affine subspace M, we associate a matrix A with its unique linear subspace. Then, $M = a + U_1 = a + \mathrm{Col}(A)$. For intersections, $(a+U_1) \cap (b+U_2) = (a+\mathrm{Col}(A)) \cap (b+\mathrm{Col}(B)) = \{y : y = Ax_1 + a = Bx_2 + b \text{ for some } x_1 \text{ and } x_2\}$. For "sums",

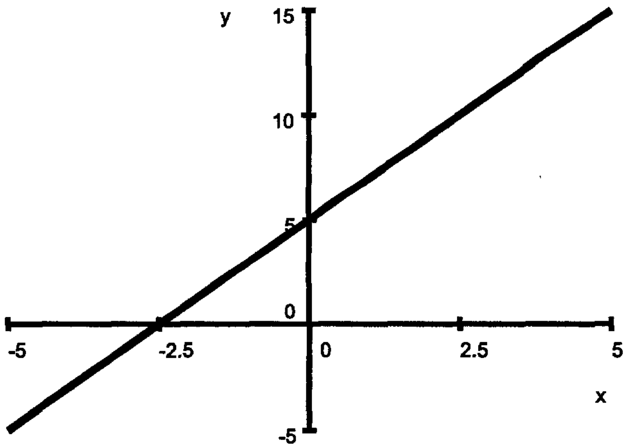
$$\begin{aligned}(a+U_1) \boxplus (b+U_2) &= (a+\mathrm{Col}(A)) \boxplus (b+\mathrm{Col}(B)) = b + [\mathrm{sp}\{a-b\} + U_1 + U_2] \\ &= b + [\mathrm{sp}\{a-b\} + \mathrm{Col}(A) + \mathrm{Col}(B)] = b + \mathrm{Col}\left(\left[A \vdots B \vdots a-b\right]\right) \\ &= \left\{ y \mid y = \left[A \vdots B \vdots (a-b)\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b \right\}.\end{aligned}$$

It is time for some examples. Consider a one dimensional subspace of $\mathbb{R}^2$, that is a line L through the origin say of slope m, $y = mx$. Then, $L_m = \{(x, mx) \mid x \in \mathbb{R}\}$. We associate the rank one matrix $A = \begin{bmatrix} 1 & 0 \\ m & 0 \end{bmatrix}$ to L_m so that $L_m = \left\{ \begin{bmatrix} 1 & 0 \\ m & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \mid x, y \in \mathbb{R} \right\}$. For the affine line $y = mx + y_0$, we have

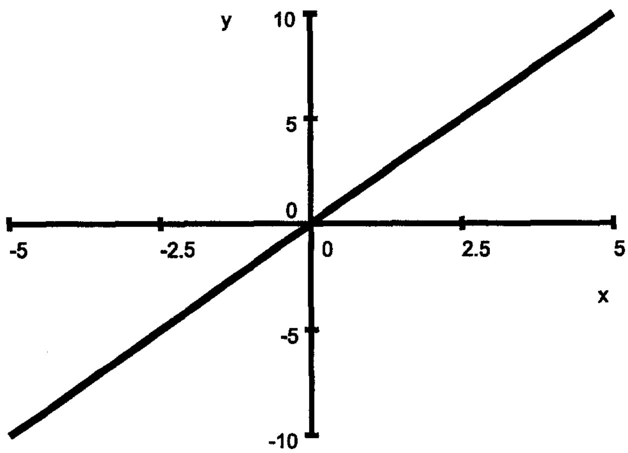
$$L_{m,y_0} = \left\{ \begin{bmatrix} 1 & 0 \\ m & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ y_0 \end{bmatrix} \,\middle|\, x, y \in \mathbb{R} \right\}.$$

As a second illustration, consider the plane (two-dimensional subspace) in $\mathbb{R}^3$ given by $ax + by + cz = 0$, where $c \neq 0$. Here, we associate the rank two matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a}{c} & -\frac{b}{c} & 0 \end{bmatrix}$$



(a) Affine subspace.



(b) Linear subspace.

Figure 1.

so the plane is described by

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a}{c} & -\frac{b}{c} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \middle| \, x, y, z \in \mathbb{R} \right\}.$$

For an affine plane in $\mathbb{R}^3$ given by $ax + by + cz = d$, we have

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a}{c} & -\frac{b}{c} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{d}{c} \end{bmatrix} \middle| \, x, y, z \in \mathbb{R} \right\}.$$

As a final illustration, suppose we wish to represent the affine line $L(t) = (1, 3, 0) + t(1, -1, 2)$ given parametrically in $\mathbb{R}^3$. We associate the rank one matrix

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

so the line is the set of vectors

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \middle| x, y, z \in \mathbb{R} \right\}.$$

We shall refer to these examples later.

4. PROJECTIONS

For orthogonal projections onto subspaces, introductory linear algebra courses usually appeal to orthonormal bases or at least orthogonal bases, to generate formulas. However, we are promoting the pseudoinverse as the way to compute projections. That is, if U is a subspace realized by Col(A), then $P_U = AA^+$. For example, to project on the line of slope m through the origin L_m, recall we associated the matrix $A = \begin{bmatrix} 1 & 0 \\ m & 0 \end{bmatrix}$. One computes $A^+ = \begin{bmatrix} 1/(1+m^2) & m/(1+m^2) \\ 0 & 0 \end{bmatrix}$. (There is an algorithm in Campbell and Meyer [7] that works for small matrices or one can use a computer algebra system like MATHEMATICA to get it.) and so $P_{L_m} = AA^+ = \begin{bmatrix} 1/(1+m^2) & m/(1+m^2) \\ m/(1+m^2) & m^2/(1+m^2) \end{bmatrix}$. Therefore, if you want to project an arbitrary vector (x, y) in $\mathbb{R}^2$ onto the line L_m, you find $\begin{bmatrix} 1/(1+m^2) & m/(1+m^2) \\ m/(1+m^2) & m^2/(1+m^2) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (x+my)/(1+m^2) \\ (mx+m^2y)/(1+m^2) \end{bmatrix}$. Concretely then, if you want to project (4,4) onto the line $y = 3 \times 3$, you take $A = \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix}$ and compute $AA^+ = \begin{bmatrix} 1/10 & 3/10 \\ 3/10 & 9/10 \end{bmatrix}$ so that the projection is $\begin{bmatrix} 1/10 & 3/10 \\ 3/10 & 9/10 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 4.8 \end{bmatrix}$. Similarly, if U is the plane $ax + by + cz = 0$ in $\mathbb{R}^3$, recall we associated the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a}{c} & -\frac{b}{c} & 0 \end{bmatrix}.$$

The formula

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & \beta & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & \beta & 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1+\beta^2}{1+\alpha^2+\beta^2} & \frac{-\beta\alpha}{1+\alpha^2+\beta^2} & \frac{\alpha}{1+\alpha^2+\beta^2} \\ \frac{-\beta\alpha}{1+\alpha^2+\beta^2} & \frac{1+\alpha^2}{1+\alpha^2+\beta^2} & \frac{\beta}{1+\alpha^2+\beta^2} \\ \frac{\alpha}{1+\alpha^2+\beta^2} & \frac{\beta}{1+\alpha^2+\beta^2} & \frac{\alpha^2+\beta^2}{1+\alpha^2+\beta^2} \end{bmatrix}$$

allows us to compute projections onto such planes. Concretely, if you want to project (1,1,3) onto the plane given by $3x + 2y + 6z = 0$, we associate

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{3} & 0 \end{bmatrix}$$

and compute

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{3} & 0 \end{bmatrix}^+ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{40}{49} & \frac{-6}{49} & \frac{-18}{49} \\ \frac{-6}{49} & \frac{45}{49} & \frac{-12}{49} \\ \frac{-18}{49} & \frac{-12}{49} & \frac{13}{49} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{20}{49} \\ \frac{3}{49} \\ \frac{9}{49} \end{bmatrix}.$$

So now the big question is how do you project (orthogonally) onto affine subspaces? A nice discussion and picture is found in [3, p. 436ff]. The idea is to take the affine subspace, translate

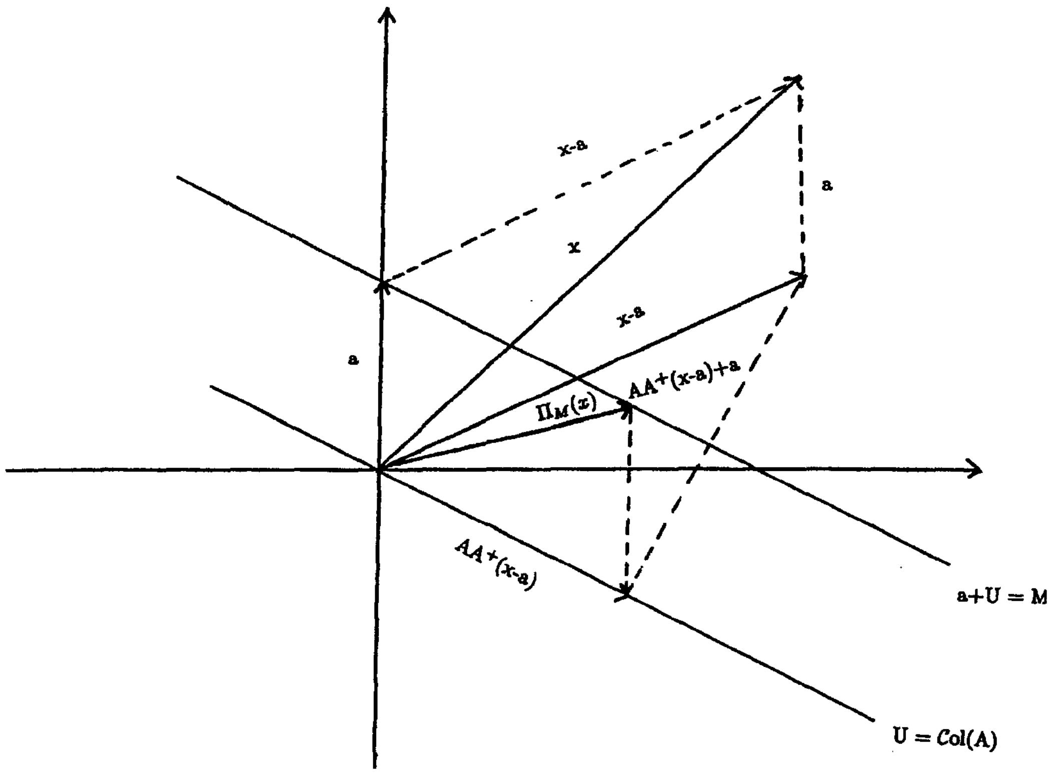


Figure 2.

it back to the origin and get its uniquely associated linear subspace, project there where we know how to do a projection, and translate back.

So our official definition of an affine projection is this. If x is a vector in $\mathbb{R}^n$, and M is an affine subspace of $\mathbb{R}^n$, the *affine projection of x onto M* where $M = a + U$ is

$$\Pi_M(x) = a + P_U(x - a) = a + P_U(x) - P_U(a) = P_U(x) + P_{U^\perp}(a).$$

This simple algebraic equivalence suggests another geometric interpretation using the orthogonal complement of U as the next figure shows.

So for example, to project onto the affine line L_{m,y_0} in $\mathbb{R}^2$, we find $\Pi_{L_{m,y_0}}(x,y) =$

$$\begin{bmatrix} \frac{1}{1+m^2} & \frac{m}{1+m^2} \\ \frac{m}{1+m^2} & \frac{m^2}{1+m^2} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 0 \\ y_0 \end{bmatrix} \right) + \begin{bmatrix} 0 \\ y_0 \end{bmatrix} = \begin{bmatrix} \frac{x + m(y - y_0)}{1+m^2} \\ \frac{mx + m^2(y - y_0)}{1+m^2} + y_0 \end{bmatrix}.$$

To project onto the affine plane $ax + by + cz = d$, we calculate $\Pi(x,y,z) =$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a}{c} & -\frac{b}{c} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{a}{c} & -\frac{b}{c} & 0 \end{bmatrix}^+ \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \frac{d}{c} \end{bmatrix} \right) + \begin{bmatrix} 0 \\ 0 \\ \frac{d}{c} \end{bmatrix}.$$

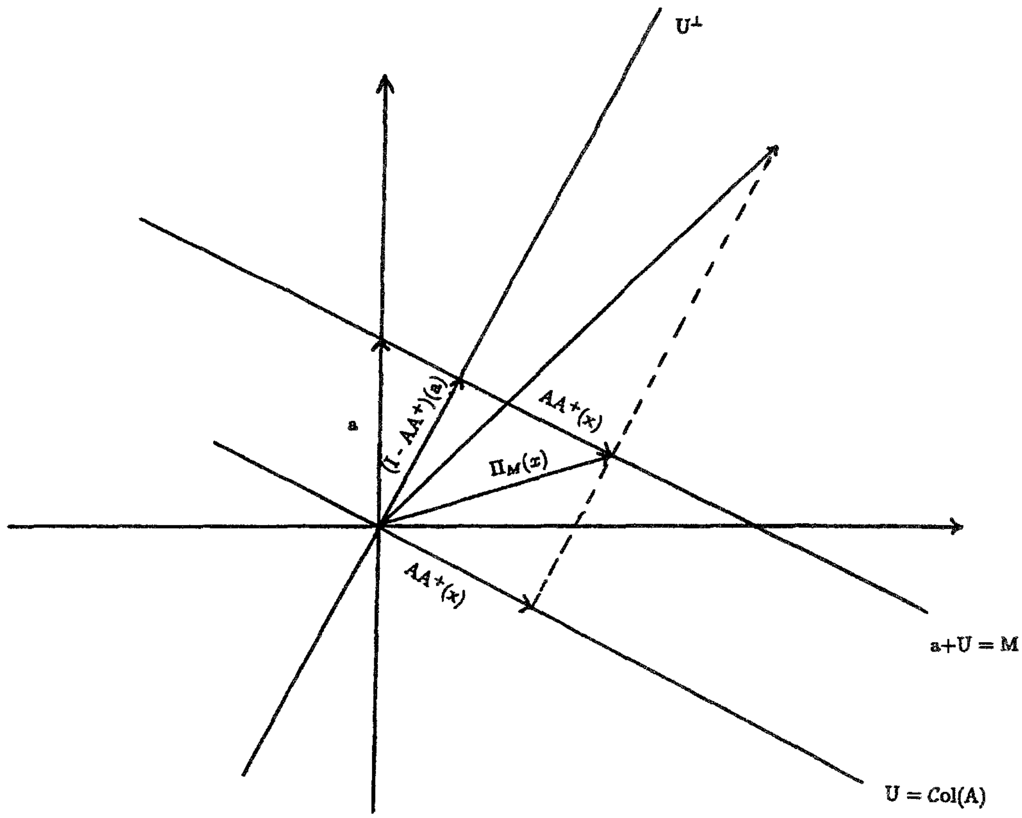


Figure 3.

Concretely then to project (x, y, z) onto the plane given by $3x + 2y + 6z = 6$, we compute

$$\Pi(x,y,z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{3} & 0 \end{bmatrix}^{+} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{40}{49}x - \frac{6}{49}y - \frac{18}{49}(z-1) \\ \frac{-6}{49}x + \frac{45}{49}y - \frac{12}{49}(z-1) \\ \frac{-18}{49}x - \frac{12}{49}y + \frac{13}{49}(z-1) + 1 \end{bmatrix}.$$

So to compute the projection of (1,1,3) onto this plane compute that

$$\Pi(1,1,3) = \begin{bmatrix} -\frac{2}{49} \\ \frac{15}{49} \\ \frac{45}{49} \end{bmatrix}.$$

As one final illustration, let us suppose we want to project onto the affine line $L(t) = (1,3,0) + t(1,-1,2)$ in $\mathbb{R}^3$. We compute

$$\Pi(x,y,z) = \begin{bmatrix} \frac{1}{6} & -\frac{1}{6} & \frac{2}{6} \\ -\frac{1}{6} & \frac{1}{6} & -\frac{2}{6} \\ \frac{2}{6} & -\frac{2}{6} & \frac{4}{6} \end{bmatrix} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \right) + \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{6(x-1)} - \frac{1}{6(y-3)} + \frac{2}{6z} + 1 \\ -\frac{1}{6(x-1)} + \frac{1}{6(y-3)} - \frac{2}{6z} + 3 \\ \frac{2}{6(x-1)} - \frac{2}{6(y-3)} + \frac{4}{6z} \end{bmatrix}$$

so for example,

$$\Pi(1,1,5) = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}.$$

Now several questions arise about properties of affine projections and how they mirror properties we know about orthogonal projections. These are given in the following table.

Table 2.

Ordinary Projections	Affine Projections
P_U projects onto linear subspace U	Π_M projects onto affine subspace $M = a + U$
U is exactly the set of fixed points of P_U	M is exactly the set of fixed points of Π_M
$P_{U^\perp}(x) = (I - P_U)(x)$	$\Pi_{a+U^\perp}(x) = (I - \Pi_{a+U})(x)$
P_U is idempotent—$P_U(P_U(x)) = P_U(x)$	Π_M is idempotent—$\Pi_M(\Pi_M(x)) = \Pi_M(x)$

In addition, if $M = a + U$, we have

$$x - \Pi_M(x) = (x - a) - P_U(x - a) = P_{U^\perp}(x - a), \tag{9}$$

$$\|\Pi_M(y) - \Pi_M(z)\| = \|P_U(y) - P_U(z)\| \tag{10}$$

and

$$\|\Pi_M(x)\|^2 = \|P_U(x)\|^2 + \|P_{U^\perp}(a)\|^2. \tag{11}$$

As long as we are talking about norms, since ordinary projections solve a minimization (least squares) problem, so do affine projections. After all, translation by a fixed vector does not change distances. More precisely, let y be an arbitrary vector and suppose we seek the closest vector in $M = a + U$ to y where $U = \mathrm{Col}(A)$. Then,

$$\begin{aligned} \|(Ax + a) - y\| &= \|Ax - (y - a)\| = \left\|Ax - AA^+(y - a)\right\| \\ &\geq \left\|AA^+(y - a) - (y - a)\right\| = \left\|AA^+(y - a) + a - y\right\| = \|\Pi_M(y) - y\|. \end{aligned}$$

Finally, we produce formulas for projecting onto the intersection and "sum" affine subspaces. These formulas extend the work done in [10]. We work on intersections first. Of course, it only makes sense to project when the intersection is not empty. We have seen that $(a+U_1)\cap(b+U_2) = (a+\mathrm{Col}(A))\cap(b+\mathrm{Col}(B)) \neq \emptyset$ iff $b-a \in U_1+U_2 = \mathrm{Col}(A)+\mathrm{Col}(B) = \mathrm{Col}([A\vdots B])$. Thus,

$$b-a = Ax_1 + Bx_2 = \begin{bmatrix} A\vdots B \end{bmatrix} \begin{bmatrix} x_1 \\ \cdots \\ x_2 \end{bmatrix}.$$

Using (6),

$$\begin{bmatrix} A\vdots B \end{bmatrix}^+ = \begin{bmatrix} A^* \\ \cdots \\ B^* \end{bmatrix} \left(\begin{bmatrix} A\vdots B \end{bmatrix} \begin{bmatrix} A^* \\ \cdots \\ B^* \end{bmatrix} \right)^+ = \begin{bmatrix} A^*(AA^*+BB^*)^+ \\ \cdots \\ B^*(AA^*+BB^*)^+ \end{bmatrix}.$$

For notational convenience, let $D = AA^*+BB^* = D^*$. Thus, the projection onto $\mathrm{Col}(A)+\mathrm{Col}(B)$ is

$$\begin{bmatrix} A\vdots B \end{bmatrix} \begin{bmatrix} A\vdots B \end{bmatrix}^+ = \begin{bmatrix} A\vdots B \end{bmatrix} \begin{bmatrix} A^*D^+ \\ \cdots \\ B^*D^+ \end{bmatrix} = AA^*D^+ + BB^*D^+ = DD^+.$$

Thus, we see $(a+\mathrm{Col}(A))\cap(b+\mathrm{Col}(B)) \neq \emptyset$ iff $b-a = DD^+(b-a)$. But $b-a = DD^+(b-a)$ iff $b-a = AA^*D^+(b-a) + BB^*D^+(b-a)$ iff $AA*D^+(b-a)+a = -BB*D^+(b-a)+b$. Also since $\mathrm{Col}(A)$ and $\mathrm{Col}(B)$ are contained in $\mathrm{Col}(A)+\mathrm{Col}(B)$, $DD^+A = A$ and $DD^+B = B$. But $DD^+A = A$ implies $(AA^*+BB^*)D^+A = A$ so $BB^*D^+A = A - AA^*D^+A$ whence $BB^*D^+AA^* = AA^* - AA^*D^+AA^*$. But this matrix is self-adjoint so

$$BB^*D^+AA^* = (BB^*D^+AA^*)^* = AA^*D^+BB^*.$$

We now have background for the next theorem.

THEOREM 4.1. *$(a+\mathrm{Col}(A))\cap(b+\mathrm{Col}(B))$ is not empty iff $b-a \in \mathrm{Col}(A)+\mathrm{Col}(B)$ iff $b-a = DD^+(b-a)$, where $D = AA^*+BB^*$. Moreover, if the intersection is not empty, $(a+\mathrm{Col}(A))\cap(b+\mathrm{Col}(B)) = c+\mathrm{Col}(C)$, where $c = AA^*D^+(b-a)+a$ and $C = [BB^*D^+A\vdots AA^*D^+B]$.*

PROOF. The first part has already been argued above, so let us look at the "moreover". Let us say $y \in (a+\mathrm{Col}(A))\cap(b+\mathrm{Col}(B))$. Then, $y = Ax_1+a = Bx_2+b$ for suitable x_1 and x_2. But then $Ax_1+a = Bx_2+b$ so $b-a = Ax_1-Bx_2$. In matrix notation,

$$\begin{bmatrix} A\vdots -B \end{bmatrix} \begin{bmatrix} x_1 \\ \cdots \\ x_2 \end{bmatrix} = b-a.$$

By Lemma 2.2, the solution to this system take the form

$$\begin{aligned} \begin{bmatrix} x_1 \\ \cdots \\ x_2 \end{bmatrix} &= \begin{bmatrix} A\vdots -B \end{bmatrix}^+ (b-a) + \left[I - \begin{bmatrix} A\vdots -B \end{bmatrix}^+ \begin{bmatrix} A\vdots -B \end{bmatrix} \right] z \\ &= \begin{bmatrix} A^*D^+(b-a) \\ \cdots \\ -B^*D+(b-a) \end{bmatrix} + \begin{bmatrix} I-A^*D^+A & A^*D^+B \\ B^*D^+A & I-B^*D^+B \end{bmatrix} \begin{bmatrix} z_1 \\ \cdots \\ z_2 \end{bmatrix}, \end{aligned}$$

where z is arbitrary. Then, $y = Ax_1+a = [AA^*D^+(b-a)+a] + [A - AA^*D^+A\vdots AA^*D^+B]z_1 = c+[BB^*D^+A\vdots AA^*D^+B]z_1$. This puts y in $c+\mathrm{Col}(C)$.

Conversely, suppose $y \in c+\mathrm{Col}(C)$. Then, $y = [BB^*D^+A\vdots AA^*D^+]z$ for some z in $\mathrm{Col}(C)$. But then, $y = ((AA^*D^+)(b-a)+a) + [A - AA^*D^+A\vdots AA^*D^+B]z = A[A^*D^+(b-a)] + A[I -$

$A^*D^+A\vdots A^*D^+B]z+a = Aw_1+a$ which is in $a+\mathrm{Col}(A)$. Similarly, $y = (-BB^*D+(b-a)+b)+[BB^*D+A\vdots B-BB^*D+B]z = B[-B^*D+(b-a)+B[B^*D+A\vdots I-B^*D+B]z+b = Bw_2+b$ which is in $b+\mathrm{Col}(B)$. Thus, y is in the intersection which completes the proof. ∎

The reader may now verify the following useful reduction.

$$\left[BB^*D^+A\vdots AA^*D^+B\right]\left[BB^*D^+A\vdots AA^*D^+B\right]^+ = [BB^*D^+AA^*]\,[BB^*D^+AA^*]^+. \qquad (12)$$

We are now in a position to exhibit a formula for the projection onto the intersection of two affine subspaces. As a corollary, we also get a formula for the orthogonal projection onto the intersection of two linear subspaces.

COROLLARY 4.2.

(1) $\Pi_{(a+\mathrm{Col}(A))\cap(b+\mathrm{Col}(B))}(x) = [BB^*D^+AA^*][BB^*D^+AA^*]^+(x-c)+c$,

(2) $P_{\mathrm{Col}(A)\cap\mathrm{Col}(B)}(x) = [BB^*D^+AA^*][BB^*D^+AA^*]^+(x)$.

We illustrate with an example. Consider the plane $3x-6y-2z=15$ in $\mathbb{R}^3$ to which we associate the matrix

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

and vector

$$a = \begin{bmatrix} 0 \\ 0 \\ -\frac{15}{2} \end{bmatrix}.$$

Also consider the plane $2x+y-2z=5$ to which we associate the matrix

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

and vector

$$b = \begin{bmatrix} 0 \\ 0 \\ \frac{5}{2} \end{bmatrix}.$$

We shall compute the affine projection onto the intersection of these two affine planes which will turn out to be an affine line. First,

$$AA^* = \begin{bmatrix} 1 & 0 & \frac{3}{2} \\ 0 & 1 & -3 \\ \frac{3}{2} & -3 & \frac{45}{4} \end{bmatrix}, \quad BB^* = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \\ 1 & \frac{1}{2} & \frac{5}{4} \end{bmatrix}, \quad D^+ = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{5} \\ -\frac{1}{4} & \frac{3}{4} & \frac{1}{5} \\ -\frac{1}{5} & \frac{1}{5} & 25 \end{bmatrix}.$$

Now $DD^+ = I$ ensuring that these planes do intersect without appealing to other considerations like dimension. Next,

$$c = \begin{bmatrix} \frac{1}{5} \\ -\frac{7}{5} \\ -3 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} \frac{49}{100} & \frac{7}{100} & \frac{21}{40} \\ \frac{7}{100} & \frac{1}{100} & \frac{3}{40} \\ \frac{21}{40} & \frac{3}{40} & \frac{9}{16} \end{bmatrix}$$

Table 3.

<table>
<tr><th colspan="2">Intersection</th></tr>
<tr><th>Ordinary</th><th>Affine</th></tr>
<tr><td>$y \in \mathrm{Col}(A) \cap \mathrm{Col}(B) \Leftrightarrow$
(1) $y = \left[A - AA^*D^+A \vdots AA^*D^+B\right] z$
for some z or
(2) $y = \left[A - AA^*D^+A \vdots B - BB^*D^+B\right] z$
for some z or
(3) $y = \left[BB^*D^+A \vdots B - BB^*D^+B\right] z$
for some z or
(4) $y = \left[BB^*D^+A \vdots AA^*D^+B\right] z$
for some z

$P_{\mathrm{Col}(A)}(x) = AA^+(x)$

$P_{\mathrm{Col}(A)\cap\mathrm{Col}(B)}(x)$
$= \left[BB^*D^+A \vdots AA^*D^+B\right]\left[BB^*D^+A \vdots AA^*D^+B\right]^+(x)$
$= [BB*D^+AA*]\,[BB*D^+AA*]^+\,(x)$</td><td>$y \in (a + \mathrm{Col}(A)) \cap (b + \mathrm{Col}(B)) \Leftrightarrow$
(1) $y = \left[A - AA^*D^+A \vdots AA^*D^+B\right] z$
$+AA^*D^+(b-a) + a$ for some z or
(2) $y = \left[A - AA^*D^+A \vdots B - BB^*D^+B\right] z$
$+AA^*D^+(b-a) + a$ for some z or
(3) $y = \left[BB^*D^+A \vdots B - BB^*D^+B\right] Z$
$+BB^*D^+(a-b) + b$ for some z or
(4) $y = \left[BB^*D^+A \vdots AA^*D^+B\right] Z + BB^*D^+(a-b) + b$
for some z

$\Pi_{a+\mathrm{Col}(A)}(x)$
$= AA^+(x-a) + a = \Pi_{(a+\mathrm{Col}(A))\cap(b+\mathrm{Col}(B))}(x)$
$= \left[BB^*D^+A \vdots AA^*D^+B\right]\left[BB^*D^+A \vdots AA^*D^+B\right]^+$
$(x - (AA^*D^+(b-a) + a) + (A^*D^+(b-a) + a)$
$=[BB^*D^+AA^*][BB^*D^+AA^*]^+(x-(BB^*D^+(a-b)+b)$
$+[BB^*D^+(a-b) + b]$</td></tr>
<tr><th colspan="2">Sum</th></tr>
<tr><td>$y \in \mathrm{Col}(A) + \mathrm{Col}(B) = \mathrm{Col}\left(\left[A \vdots B\right]\right) \Longleftrightarrow$
(1) $y = \left[A \vdots B\right] z$ for some z or
(2) $y = \left[A \vdots B\right]\left[A \vdots B\right]^+ y$

$P_{\mathrm{Col}(A)+\mathrm{Col}(B)}(x)$
$= \left[A \vdots B\right]\left[A \vdots B\right]^+ (x) = DD^+(x)$
where $D = AA* + BB*$</td><td>$y \in (a + \mathrm{Col}(A)) \boxplus (b + \mathrm{Col}(B))$
(1) $y = \left[A \vdots B \vdots (a-b)\right] z + b$ for some z or
(2) $y = \left[A \vdots B \vdots (b-a)\right] z + a$

$\Pi_{(a+\mathrm{Col}(A))\boxplus(b+\mathrm{Col}(B))}(x)$
$= \left[A \vdots B \vdots (a-b)\right]\left[A \vdots B \vdots (a-b)\right]^+ [x-b] + b$
$= [D + (a-b)(a-b)^*]\,[D + (a-b)(a-b)^*]^+\,[x-b] + b$
where $D = AA* + BB*$</td></tr>
</table>

and so

$$\Pi(x,y,z) = \begin{bmatrix} \frac{196}{425}\left(x - \frac{1}{5}\right) + \frac{28}{425}\left(y + \frac{7}{5}\right) + \frac{42}{85}(z+3) + \frac{1}{5} \\ \frac{28}{425}\left(x - \frac{1}{5}\right) + \frac{4}{425}\left(y + \frac{7}{5}\right) + \frac{6}{85}(z+3) - \frac{7}{5} \\ \frac{42}{85}\left(x - \frac{1}{5}\right) + \frac{6}{85}\left(y + \frac{7}{5}\right) + \frac{9}{17}(z+3) - 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} \\ -\frac{7}{5} \\ -3 \end{bmatrix} + \begin{bmatrix} 14t \\ 2t \\ 15t \end{bmatrix},$$

where $t = (14/215)(x - 1/5) + (2/415)(y + 7/5) + (15/425)(z + 3)$.

Since projecting onto the "sum" of affine is straightforward, we finish with summarizing our findings in Table 3.

REFERENCES

1. T.R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, (1970).
2. M.K. Bennett, *Affine and Projective Geometry*, Wiley-Interscience, (1995).
3. C.D. Meyer, Jr., *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics, (2000).
4. T.L. Boullion and P.L. Odell, *Generalized Inverse Matrices*, Wiley-Interscience, (1971).
5. A. Ben-Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications*, John Wiley & Sons, (1974).
6. C.R. Rao and S.K. Mitra, *Generalized Inverses of Matrices and its Applications*, John Wiley & Sons, (1971).
7. S.K. Campbell and C.D. Meyer, Jr., *Generalized Inverses of Linear Transformations*, Dover, (1979).
8. C.R. Rao and M.B. Rao, *Matrix Algebra and its Applications to Statistics and Economics*, World Scientific, (1998).
9. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, (1996).
10. R. Piziak, P.L. Odell and R. Hahn, Constructing projections on sums and intersections, *Computers Math. Applic.* **37** (1), 67–74, (1999).
11. S.K. Berberian, *Introduction to Hilbert Space*, Oxford University, New York, (1961).