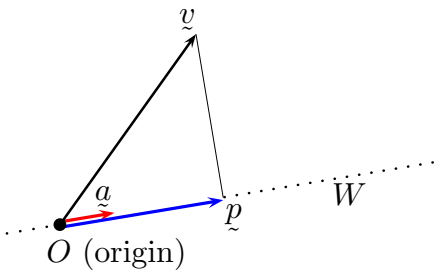
Projection onto a line through the origin

Suppose that $\underset{\sim}{a} \in \mathbb{R}^n$ is a nonzero vector, and let W be the one-dimensional subspace spanned by $\underset{\sim}{a}$. Thus W consists of all scalar multiples of $\underset{\sim}{a}$. Geometrically, it is a straight line through the origin in n-dimensional space. If $\underset{\sim}{v} \in \mathbb{R}^n$ is arbitrary then, as we saw in the first week, the projection of $\underset{\sim}{v}$ onto W is the vector $\underset{\sim}{p} = A\underset{\sim}{x}$, where $\underset{\sim}{x}$ is the unique solution of the linear system $(A^TA)\underset{\sim}{x} = A^T\underset{\sim}{v}$. Here A is any matrix whose columns form a basis for W. In the case we are currently considering, A will have only one column, since $\dim W = 1$; indeed, we may take $A = \underset{\sim}{a}$ (an $n \times 1$ matrix). Now, substituting in to the formula, we find that

$$\underset{\sim}{p} = \underset{\sim}{a}(\underset{\sim}{a}^T\underset{\sim}{a})^{-1}(\underset{\sim}{a}^T\underset{\sim}{v}) = \frac{a \cdot v}{a \cdot a}\underset{\sim}{a},$$

since the product of a row vector by a column vector can be rewritten in terms of the dot product.

Geometrically, $\underset{\sim}{p}$ is that scalar multiple of $\underset{\sim}{a}$ such that $\underset{\sim}{v} - \underset{\sim}{p}$ is perpendicular to W:



It is easy to check that $\underset{\sim}{p}$ depends on the direction of $\underset{\sim}{a}$ but not on its length: indeed, $\|\underset{\sim}{p}\| = \|\underset{\sim}{v}\| \cos\theta$, where θ is the angle between $\underset{\sim}{a}$ and $\underset{\sim}{p}$.

For example, let $\underset{\sim}{a} = (1, -1, 1)^T$, $\underset{\sim}{v} = (0, 1, 2)^T$. Then $\underset{\sim}{a} \cdot \underset{\sim}{v} = 0 - 1 + 2 = 1$ and $\underset{\sim}{a} \cdot \underset{\sim}{a} = 1^2 + (-1)^2 + 1^2 = 3$. So the projection of $\underset{\sim}{v}$ onto $\mathrm{Span}(V)$ is

$$\underset{\sim}{p} = \frac{1}{3}\underset{\sim}{a} = \frac{1}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ -1/3 \\ 1/3 \end{pmatrix}.$$

"Least Squares" approximations

The theory we have been discussing has an important application in the analysis of experimental data, which we now describe.

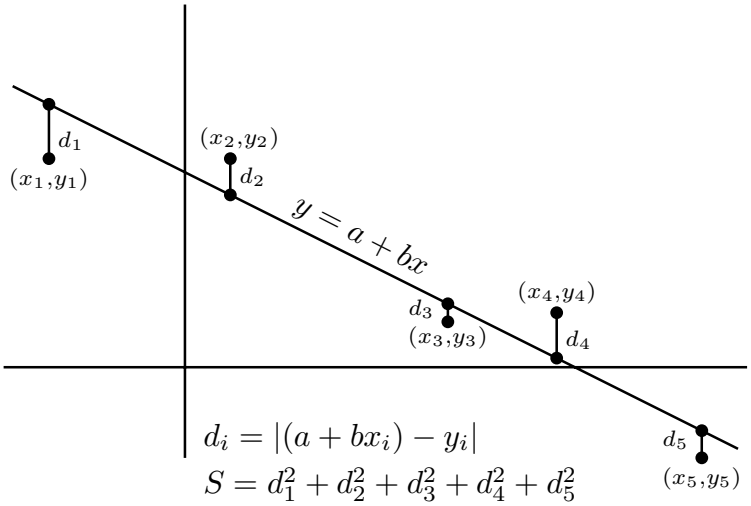
Suppose that we have set of points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\}$ in the (x, y)-plane, and we wish to find the *straight line of best fit*. That is, we want to find a and b so that the line $y = a + bx$ goes as close as is possible to the points (x_1, y_1), (x_2, y_2), $\ldots$, (x_k, y_k).

The solution to this problem, of course, depends on how one defines closeness. For this purpose we consider the sum of the squares of the vertical distances from the data points to the line. If L is the line $y = a + bx$, then the point on L with x-coordinate x_i is $(x_i, a + bx_i)$, and this is the point on L that lies vertically above or below the data point (x_i, y_i). The distance between these two points is $|(a + bx_i) - y_i|$, and the quantity

$$S = \big((a + bx_1) - y_1\big)^2 + \big((a + bx_2) - y_2\big)^2 + \cdots + \big((a + bx_k) - y_k\big)^2$$

is the sum of the squares of the vertical distances from the data points to L. Clearly $S \geq 0$, and S can only be zero if $a + bx_i = y_i$ for all i, which would mean that all the

data points lie on L. Our aim is to vary L so as to minimize S; the line for which S is minimal is called the line of best fit.



The observation that relates this problem to vector spaces is that S can be interpreted as the square of the length of a certain vector $\underset{\sim}{z}$ in $\mathbb{R}^k$. Specifically $S = \|\underset{\sim}{z}\|^2$, where

$$\underset{\sim}{z} = \begin{pmatrix} (a+bx_1) - y_1 \\ (a+bx_2) - y_2 \\ \vdots \\ (a+bx_k) - y_k \end{pmatrix}$$
$$= a \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + b \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix},$$

and our task is to find a and b that minimize $\|\underset{\sim}{z}\|$. (Minimizing $\|\underset{\sim}{z}\|$ is of course equivalent to minimizing $\|\underset{\sim}{z}\|^2$.) In other words, we must find a and b such that

$$\underset{\sim}{p} = a \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + b \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$$

is as close as possible to $\underset{\sim}{y} = (y_1, y_2, \ldots, y_k)^T$. If we define W to be the subspace of $\mathbb{R}^k$ spanned by the two column vectors $(1, 1, \ldots, 1)^T$ and $(x_1, x_2, \ldots, x_k)^T$ then $\underset{\sim}{p}$ is the element of W that is closest to $\underset{\sim}{y}$; that is, $\underset{\sim}{p}$ is the projection of $\underset{\sim}{y}$ onto W.† So, by the formulas we have already derived,

$$\begin{pmatrix} a \\ b \end{pmatrix} = (A^T A)^{-1} A^T \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$$

† If the x_i are all equal then $\dim W = 1$, and the formulas below do not apply. However, in this case all the data points lie on a vertical line, and this is the line of best fit.

where A is the matrix whose two columns are $(1,1,\ldots,1)^T$ and $(x_1,x_2,\ldots,x_k)^T$. (For this problem we do not need to calculate the vector $\underset{\sim}{p}$, since the coefficients a and b are what we want.)

As an example, let us find the line of best fit for the four data points $(0,0)$, $(1,1)$, $(3,2)$, $(4,5)$. In the notation used above,

$$A^TA = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 8 \\ 8 & 26 \end{pmatrix},$$

and we have to find a and b satisfying

$$\begin{pmatrix} 4 & 8 \\ 8 & 26 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = A^T \begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 8 \\ 27 \end{pmatrix}.$$

Let us use the row operations technique to solve this system:

$$\left(\begin{array}{cc|c} 4 & 8 & 8 \\ 8 & 26 & 27 \end{array}\right) \xrightarrow{R_1 := \frac{1}{4}R_1} \left(\begin{array}{cc|c} 1 & 2 & 2 \\ 8 & 26 & 27 \end{array}\right) \xrightarrow{R_2 := R_2 - 8R_1} \left(\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 10 & 11 \end{array}\right)$$
$$\xrightarrow{R_2 := 0.1R_2} \left(\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 1 & 1.1 \end{array}\right) \xrightarrow{R_1 := R_1 - 2R_2} \left(\begin{array}{cc|c} 1 & 0 & -0.2 \\ 0 & 1 & 1.1 \end{array}\right).$$

So the equation of the line of best fit is $y = -0.2 + 1.1\,x$.

Best fitting parabolas, cubic curves, etc.

Given points (x_1,y_1), (x_2,y_2), …, (x_k,y_k), we may want to find the best fitting polynomial equation of some specified degree. Suppose that the degree is n, so that the equation we want to find has the form $y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$. As in the straight line case, we use the sum of the squares of the vertical distances from the data points to the curve as a measure of how well the curve fits the data. The vertical distance from the point (x_i,y_i) to the curve is the distance between the points (x_i,y_i) and $(x_i, a_0+a_1x_i+a_2x_i^2+\cdots+a_nx_i^n)$, and this is given by $|(a_0+a_1x_i+a_2x_i^2+\cdots+a_nx_i^n)-y_i|$. So the quantity to be minimized is

$$S = \sum_{i=1}^{k} \big((a_0 + a_1x_i + a_2x_i^2 + \cdots + a_nx_i^n) - y_i\big)^2,$$

which is the square of the length of the vector

$$\underset{\sim}{z} = \begin{pmatrix} (a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n) - y_i \\ (a_0 + a_1x_2 + a_2x_2^2 + \cdots + a_nx_2^n) - y_i \\ \vdots \\ (a_0 + a_1x_k + a_2x_k^2 + \cdots + a_nx_k^n) - y_k \end{pmatrix}$$
$$= a_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + a_1 \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} + a_2 \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_k^2 \end{pmatrix} + \cdots + a_n \begin{pmatrix} x_1^n \\ x_2^n \\ \vdots \\ x_k^n \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}.$$

We see immediately that the length of $\underset{\sim}{z}$ is minimal when

$$\underset{\sim}{p} = a_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + a_1 \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} + a_2 \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_k^2 \end{pmatrix} + \cdots + a_n \begin{pmatrix} x_1^n \\ x_2^n \\ \vdots \\ x_k^n \end{pmatrix}$$

is the projection of $(y_1, y_2, \ldots, y_k)^T$ onto the space W spanned by the vectors

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}, \quad \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_k^2 \end{pmatrix}, \quad \ldots, \quad \begin{pmatrix} x_1^n \\ x_2^n \\ \vdots \\ x_k^n \end{pmatrix}.$$

It can be shown that, provided there are at least $n+1$ distinct x_i's, these vectors are linearly independent, and therefore constitute a basis for W. We shall assume that this condition is satisfied.

By the same reasoning that we used in our discussion of the straight line of best fit, we conclude that

$$(A^T A) \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = A^T \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ a_k \end{pmatrix},$$

where A is the matrix whose columns constitute the above basis for W. That is,

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^n \\ 1 & x_2 & x_2^2 & \ldots & x_2^n \\ 1 & x_3 & x_3^2 & \ldots & x_3^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k & x_k^2 & \ldots & x_k^n \end{pmatrix}.$$

To illustrate this, let us find the parabola of best fit through the same four points that we used in our example of the line of best fit: $(0,0)$, $(1,1)$, $(3,2)$, $(4,5)$. This time we have

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{pmatrix} = \begin{pmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{pmatrix}$$

and so we must solve

$$\begin{pmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 1 & 9 & 16 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 8 \\ 27 \\ 99 \end{pmatrix}.$$

To save ourselves some calculation, we can get MAGMA to do this. Of course, we first have to change it into a row vector problem by taking transposes. The equation becomes

$$(\, a \quad b \quad c \,) \begin{pmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{pmatrix} = (\, 8 \quad 27 \quad 99 \,).$$

Now we can use MAGMA's `Solution` function:

```
> R := RealField();
> V := VectorSpace(R,3);
> M := KMatrixSpace(R,3,3);
> v := V![8,27,99];
> ATA := M![4,8,26,8,26,92,26,92,338];
> Solution(ATA,v);
> (3/10 -7/30 1/3)
```

We conclude that the equation of the parabola of best fit is $y = \frac{3}{10} - \frac{7}{30}x + \frac{1}{3}x^2$.

Orthonormal bases

Definition. A set $\{\underset{\sim}{v}_1.\underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_k\}$ is called an *orthogonal* set if $v_i \cdot v_j = 0$ whenever $i \neq j$.

In other words, a set of vectors is said to be orthogonal if distinct elements of the set are perpendicular to one another. For example,

$$\underset{\sim}{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \underset{\sim}{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \underset{\sim}{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

form an orthogonal set of vectors in $\mathbb{R}^3$.

Definition. A set $\{\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_k\}$ is called an *orthonormal* set if it is orthogonal and $v_i \cdot v_i = 1$ for all i.

The set $\{\underset{\sim}{i}, \underset{\sim}{j}, \underset{\sim}{k}\}$ is also an orthonormal set in $\mathbb{R}^3$. The following three vectors form an orthogonal set that is not orthonormal:

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Whenever we have an orthogonal set of vectors that are all nonzero, we can produce an orthonormal set of vectors by a process known as *normalization*, which consists of replacing each $\underset{\sim}{v}$ in the set by $(1/\|\underset{\sim}{v}\|)\underset{\sim}{v}$. In other words, divide each vector by its own length: this produces a scalar multiple of the vector having length 1. In the example above, the orthonormal set obtained in this way is

$$\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The following theorem states one of the key facts about orthogonal sets.

Theorem. *If $\{\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_k\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$ then it is also a linearly independent set.*

Proof. Suppose that λ_1, λ_2, ..., λ_k are scalars such that $\lambda_1 \underset{\sim}{v}_1 + \lambda_2 \underset{\sim}{v}_2 + \cdots + \lambda_k \underset{\sim}{v}_k = \underset{\sim}{0}$. Then, for any i,

$$\begin{aligned} 0 = \underset{\sim}{0} \cdot \underset{\sim}{v}_i &= (\lambda_1 \underset{\sim}{v}_1 + \lambda_2 \underset{\sim}{v}_2 + \cdots + \lambda_k \underset{\sim}{v}_k) \cdot \underset{\sim}{v}_i \\ &= \lambda_1 (\underset{\sim}{v}_1 \cdot \underset{\sim}{v}_i) + \lambda_2 (\underset{\sim}{v}_2 \cdot \underset{\sim}{v}_i) + \cdots + \lambda_k (\underset{\sim}{v}_k \cdot \underset{\sim}{v}_i) \\ &= \lambda_i (\underset{\sim}{v}_i \cdot \underset{\sim}{v}_i), \end{aligned}$$

all the other terms being zero since $\underset{\sim}{v}_j \cdot \underset{\sim}{v}_i = 0$ when $j \neq i$. Now since we have assumed that $\underset{\sim}{v}_i \neq \underset{\sim}{0}$ it follows that $\underset{\sim}{v}_i \cdot \underset{\sim}{v}_i = \|\underset{\sim}{v}_i\|^2 \neq 0$, and so the above equation shows that $\lambda_i = 0$. Furthermore, i was arbitrary, and so we have shown that the only solution of $\sum_{i=1}^{n} \lambda_i \underset{\sim}{v}_i = \underset{\sim}{0}$ is given by $\lambda_i = 0$ for all i. That is, $\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_k$ are linearly independent, as claimed. □

By this theorem, an orthogonal set of nonzero vectors necessarily constitutes a basis for the subspace it spans. Normalization will then yield an orthonormal basis for this subspace.

Projections using orthogonal bases

Orthogonal bases are important because many formulas become much simpler to work with when expressed in terms of orthogonal bases. The formula for the projection onto a subspace is a case in point.

Let A be an $n \times k$ matrix. The columns of A form an orthonormal set of vectors if and only if $A^T A = I$, the identity matrix. This fact is quite straightforward to prove, as follows. Let $\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_k$ be the columns of A. Then $\underset{\sim}{v}_1^T, \underset{\sim}{v}_2^T, \ldots, \underset{\sim}{v}_k^T$ are the rows of A^T, and, by the way matrix multiplication is defined, the (i, j) entry of $A^T A$ is $\underset{\sim}{v}_i^T \underset{\sim}{v}_j$, the product of the i-th row of A^T and the j-th column of A. But the product of a row vector by a column vector can alternatively be expressed as the dot product of two column vectors; so in fact we can say that the (i, j) entry of $A^T A$ is the dot product of the i-th and j-th columns of A. The columns of A comprise an orthonormal set if and only if $\underset{\sim}{v}_i \cdot \underset{\sim}{v}_j$ is zero for $i \neq j$ and 1 when $i = j$. Since $\underset{\sim}{v}_i \cdot \underset{\sim}{v}_j$ is the (i, j)-entry of $A^T A$, this condition is equivalent to the main-diagonal entries of $A^T A$ being 1 and the other entries being 0, and of course this is the same as saying that $A^T A = I$. So the columns of A form an orthonormal set if and only if $A^T A = I$, as claimed.

In the course of the above discussion, we showed that if A is a matrix with k columns $\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_k$, then

$$A^T A = \begin{pmatrix} \underset{\sim}{v}_1 \cdot \underset{\sim}{v}_1 & \underset{\sim}{v}_1 \cdot \underset{\sim}{v}_2 & \cdots & \underset{\sim}{v}_1 \cdot \underset{\sim}{v}_k \\ \underset{\sim}{v}_2 \cdot \underset{\sim}{v}_1 & \underset{\sim}{v}_2 \cdot \underset{\sim}{v}_2 & \cdots & \underset{\sim}{v}_2 \cdot \underset{\sim}{v}_k \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{v}_k \cdot \underset{\sim}{v}_1 & \underset{\sim}{v}_k \cdot \underset{\sim}{v}_2 & \cdots & \underset{\sim}{v}_k \cdot \underset{\sim}{v}_k \end{pmatrix}. \tag{1}$$

This result is true for all matrices A, without the assumption that the columns are orthonormal. When the columns are orthonormal, the right hand side of Eq. (1) is just the identity matrix.

Definition. The matrix on the right hand side of Eq. (1) above is called the *Gram matrix* of the set of vectors $\{\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_k\}$.

We turn now to consideration of projections onto a subspace for which we have an orthogonal basis. First, let us suppose that $\{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_k\}$ is an orthonormal set of

vectors in $\mathbb{R}^n$, and let $W = \mathrm{Span}(\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_k)$. If $\underset{\sim}{v} \in \mathbb{R}^n$ then the projection of $\underset{\sim}{v}$ onto W is given by

$$\underset{\sim}{p} = A(A^T A)^{-1} A^T \underset{\sim}{v} = A A^T \underset{\sim}{v} \tag{2}$$

(since $A^T A = I$). Note that although we know that $A^T A = I$, we have no simple formula for AA^T. To evaluate the right hand side of Eq. (2) we bracket it as $A(A^T \underset{\sim}{v})$, and observe that

$$A^T \underset{\sim}{v} = \begin{pmatrix} \underset{\sim}{a}_1^T \\ \underset{\sim}{a}_2^T \\ \vdots \\ \underset{\sim}{a}_k^T \end{pmatrix} \underset{\sim}{v} = \begin{pmatrix} \underset{\sim}{a}_1^T \underset{\sim}{v} \\ \underset{\sim}{a}_2^T \underset{\sim}{v} \\ \vdots \\ \underset{\sim}{a}_k^T \underset{\sim}{v} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{a}_1 \cdot \underset{\sim}{v} \\ \underset{\sim}{a}_1 \cdot \underset{\sim}{v} \\ \vdots \\ \underset{\sim}{a}_1 \cdot \underset{\sim}{v} \end{pmatrix}.$$

So we find that

$$\begin{aligned} \underset{\sim}{p} = A \begin{pmatrix} \underset{\sim}{a}_1 \cdot \underset{\sim}{v} \\ \underset{\sim}{a}_2 \cdot \underset{\sim}{v} \\ \vdots \\ \underset{\sim}{a}_k \cdot \underset{\sim}{v} \end{pmatrix} &= \begin{pmatrix} \underset{\sim}{a}_1 & \underset{\sim}{a}_2 & \ldots & \underset{\sim}{a}_k \end{pmatrix} \begin{pmatrix} \underset{\sim}{a}_1 \cdot \underset{\sim}{v} \\ \underset{\sim}{a}_2 \cdot \underset{\sim}{v} \\ \vdots \\ \underset{\sim}{a}_k \cdot \underset{\sim}{v} \end{pmatrix} \\ &= (\underset{\sim}{a}_1 \cdot \underset{\sim}{v}) \underset{\sim}{a}_1 + (\underset{\sim}{a}_2 \cdot \underset{\sim}{v}) \underset{\sim}{a}_2 + \cdots + (\underset{\sim}{a}_k \cdot \underset{\sim}{v}) \underset{\sim}{a}_k. \end{aligned}$$

Comparing this with the formula that we obtained earlier for the projection of $\underset{\sim}{v}$ onto a 1-dimensional subspace, we see that in fact $\underset{\sim}{p}$ is the sum of the projections of $\underset{\sim}{v}$ on to the one-dimensional subspaces spanned by the $\underset{\sim}{a}_i$, for each i from 1 to k. (The projection of $\underset{\sim}{v}$ onto the one-dimensional space $\mathrm{Span}(\underset{\sim}{a}_i)$ is $\frac{\underset{\sim}{a}_i \cdot \underset{\sim}{v}}{\underset{\sim}{a}_i \cdot \underset{\sim}{a}_i} \underset{\sim}{a}_i = (\underset{\sim}{a}_i \cdot v) \underset{\sim}{a}_i$, since $\underset{\sim}{a}_i \cdot \underset{\sim}{a}_i = 1$.)