Orthogonal Projections

Let U be vector subspace of V. Then $\dim(U) + \dim(U^\perp) = \dim(V)$. It is reasonable and important in many problems to express a vector $\mathbf{v} \in V$ as a sum $\mathbf{v} = \mathbf{u} + \mathbf{w}$ where $\mathbf{u} \in U$ and $\mathbf{w} \in U^\perp$. But how do you readily compute $\mathbf{u}, \mathbf{w}$? Are they unique given a choice $\mathbf{v}$? Note that we need only compute $\mathbf{u}$ since then $\mathbf{w}$ is readily computed as $\mathbf{w} = \mathbf{v} - \mathbf{u}$.

Lemma 0.1 *Let $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be an orthonormal basis for V. Then for any $\mathbf{v} \in V$,*

$$\mathbf{v} = \sum_{i=1}^{n} proj_{\mathbf{x}_i}\mathbf{v} = \sum_{i=1}^{n} < \mathbf{x}_i, \mathbf{v} > \mathbf{x}_i$$

Proof: Since $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is a basis for V, then $\mathbf{v} = \sum_{i=1}^{n} a_i\mathbf{x}_i$. Now

$$< \mathbf{v}, \mathbf{x}_j >= \sum_{i=1}^{n} a_i < \mathbf{x}_i, \mathbf{x}_j >= a_j,$$

using the orthonormality of the vectors in the basis. ■

Thus we have a quick way of obtaining the X coordinates of $\mathbf{v}$. Let $V = \mathbf{R}^n$. If M is the matrix whose columns are $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ written with respect to the standard basis, then M is an orthogonal matrix (you can check that $M^TM = I$) and so $M^{-1} = M^T$. Now M represent the change of basis from X basis to standard basis and so $M^{-1} = M^T$ represents the change of basis from standard basis to X basis and we see that for $\mathbf{v}$ written with respect to standard basis, $M^{-1}\mathbf{v} = M^T\mathbf{v}$ yields that $\mathbf{v} = \sum_{i=1}^{n} < \mathbf{x}_i, \mathbf{v} > \mathbf{x}_i$.

In $\mathbf{R}^3$ we could ask for the orthogonal projection of $\mathbf{x}$ onto a plane π (through the origin). You might correctly solve this as projecting $\mathbf{x}$ onto the normal vector $\mathbf{n}$ and then subtracting this from $\mathbf{x}$.

Let $\mathbf{u}_1, \mathbf{u}_2$ be a basis for the plane π. We have various techniques for extending a basis of a subspace to a basis for the entire vector space. We could consider $U = \text{span}(\mathbf{u}_1, \mathbf{u}_2)$ and then find a basis for $U^\perp$. Given that $\dim(U) + \dim(U^\perp) = 3$ in this case, then $\dim(U^\perp) = 1$ and we can find $\mathbf{u}_3$ so that $U^\perp = \text{span}(\mathbf{u}_3)$. Then $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ is a basis for $\mathbf{R}^3$. The following is the general result.

Lemma 0.2 *Let U be a subspace of $\mathbf{R}^n$ and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a basis for U. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}\}$ be a basis for $U^\perp$. Then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \cup \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}\}$ is a basis for $\mathbf{R}^n$.*

Proof: Given that we have n vectors, we need only show that they are linearly independent. Assume

$$\sum_{i=1}^{k} a_i\mathbf{u}_i + \sum_{j=1}^{n-k} b_j\mathbf{v}_j = \mathbf{0}$$

Thus we have two vectors $\mathbf{u} = \sum_{i=1}^{k} a_i\mathbf{u}_i$ in U and $\mathbf{v} = \sum_{j=1}^{n-k} b_j\mathbf{v}_j$ in $U^\perp$ with $\mathbf{u} + \mathbf{v} = \mathbf{0}$. But taking the inner product with $\mathbf{u}$ yields $< \mathbf{0}, \mathbf{u} >=< \mathbf{u} + \mathbf{v}, \mathbf{u} >$ and $< (\mathbf{u} + \mathbf{v}), \mathbf{u} >=< \mathbf{u}, \mathbf{u} > + < \mathbf{u}, \mathbf{v} >=< \mathbf{u}, \mathbf{u} >$ using $\mathbf{v} \in U^\perp$. Also $< \mathbf{0}, \mathbf{u} >= 0$.

Thus $< \mathbf{u}, \mathbf{u} >= 0$ which yields $\mathbf{u} = \mathbf{0}$ by the axioms for an inner product. Then by linear independence of $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ we deduce that $a_1 = a_2 = \cdots = a_k = 0$. Then $\mathbf{v} = \mathbf{0}$ and so (by the linear independence of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}\}$) we deduce that $b_1 = b_2 = \cdots = b_{n-k} = 0$. We deduce that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \cup \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}\}$ are linearly independent. ■

Return to our original problem of expressing a vector $\mathbf{v} \in V$ as a sum $\mathbf{v} = \mathbf{u} + \mathbf{w}$ where $\mathbf{u} \in U$ and $\mathbf{w} \in U^{\perp}$. We note that $\mathbf{u}, \mathbf{w}$ are unique given a choice for $\mathbf{v}$ since there is a unique way to express any $\mathbf{v} \in V$ in terms of the basis. We form an othornormal basis for U as $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ and an orthonomal basis for $U^{\perp}$ as $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}\}$ yielding an orthonormal basis for V. Now we can compute

$$\mathbf{u} = \sum_{i=1}^{k} < \mathbf{v}, \mathbf{u}_i > \mathbf{u}_i. \tag{1}$$

We can then compute $\mathbf{w} = \mathbf{v} - \mathbf{u}$. We often use this as a check that our computations are working by verifying that $< \mathbf{u}, \mathbf{w} >= 0$.

The equation (1) is very powerful in that we can compute u from v (and hence also compute w) only using an othonormal basis for U and without any explicit reference to $U^{\perp}$.

Least squares application.

When an experiment is run one often creates data of the following type. We compute for each input x_i a value y_i. Assume we have n values x_i, y_i. We do the experiment many times. It is often the case that we want to fit a line to the data believing that y is a linear function of x,

$$y = ax + b$$

Thus we would like to solve for a, b in the matrix equation

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

While this may be possible, we expect, given experimental errors, to be unable to solve this equation. We instead seek a line $y = a'x + b'$ that is close to given data. We seek to find a', b' to minimize the sum over i of the square error from the predicted valuesy_i

$$\min_{a',b'} \sum_{i=1}^{n} (y_i - a'x_i - b')^2$$

which is called the *Least Squares* solution. Putting this another way, if we define $\hat{\mathbf{y}} = (a'x_1 + b', a'x_2 + b', \ldots, a'x_n - b')$, then

$$\min_{a',b'} \sum_{i=1}^{n} (y_i - a'x_i - b')^2 = ||\mathbf{y} - \hat{\mathbf{y}}||^2$$

[*insert picture here*]

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} b \\ a \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

How do we do this? Geometrically it is clear that we choose $\hat{\mathbf{y}}$ to be the orthogonal projection of $\mathbf{y}$ onto colsp(A). Having chosen a vector $\hat{\mathbf{y}}$, we solve the following equation for a, b

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix},$$

to obtain a', b'.

Does $\hat{\mathbf{y}}$ minimize $||\mathbf{y} - \hat{\mathbf{y}}||^2$? Let $\bar{\mathbf{y}} \in \text{colsp}(A)$ be another vector and compute

$$||\mathbf{y} - \bar{\mathbf{y}}||^2 = ||(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})||^2.$$

Using the fact that $\mathbf{y} - \hat{\mathbf{y}} \in \text{colsp}(A)^\perp$ (by our choice of $\hat{\mathbf{y}}$) and $\hat{\mathbf{y}} - \bar{\mathbf{y}} \in \text{colsp}(A)$ (since both $\hat{\mathbf{y}}, \bar{\mathbf{y}} \in \text{colsp}(A)$). Thus the two vectors $\mathbf{y} - \hat{\mathbf{y}}$ and $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ are orthogonal and so by Pythagorean Theorem (the vectors for $\mathbf{y} - \hat{\mathbf{y}}$ and $\hat{\mathbf{y}} - \bar{\mathbf{y}}$ lie in a plane) or by properties of orthogonal vectors and inner products in general we obtain that

$$||\mathbf{y} - \bar{\mathbf{y}}||^2 = ||(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \bar{\mathbf{y}})||^2 = ||\mathbf{y} - \hat{\mathbf{y}}||^2 + ||\hat{\mathbf{y}} - \bar{\mathbf{y}}||^2 \geq ||\mathbf{y} - \hat{\mathbf{y}}||^2.$$

It is now immediate that $\hat{\mathbf{y}}$ is the choice to minimize $||\mathbf{y} - \bar{y}||^2$ over all choices $\bar{\mathbf{y}} \in \text{colsp}(A)$.

This generalizes to a matrix equation $A\mathbf{x} = \mathbf{b}$ where A has more columns, we are unable to solve for $\mathbf{x}$ if $\mathbf{b} \notin \text{colsp}(A)$. We seek a vector $\hat{\mathbf{b}} \in \text{colsp}(A)$ so that $||\mathbf{b} - \hat{\mathbf{b}}|| = \sqrt{\sum_{i=1}^n (b_i - \hat{b}_i)^2}$ is minimized. As in the simple case of two dimensional column space, we choose $\hat{\mathbf{b}} = \text{proj}_{\text{colsp}(A)} \mathbf{b}$.

If we wish to determine $\hat{\mathbf{y}} \in \text{colsp}(A)$ we can determine $\hat{\mathbf{y}}$ as the orthogonal projection of $\mathbf{y}$ onto the colsp(A) so that $\mathbf{y} - \hat{\mathbf{y}} \in (\text{colsp}(A))^\perp$. First determine an orthonormal basis for colsp(A).

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \text{proj}_{\mathbf{v}_1} \mathbf{u}_2$$

Using $\mu = \frac{1}{n} \sum_{i=1}^n x_i$ (i.e. the average x value) and $\sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n}(\sum_{i=1}^n x_i)^2$ we have an othonormal basis for colsp(A):

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{n} \\ 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} (x_1 - \mu)/\sigma \\ (x_2 - \mu)/\sigma \\ \vdots \\ (x_n - \mu)/\sigma \end{bmatrix}$$

We compute the orthogonal projection of $\mathbf{y}$ onto colsp(A) as follows

$$\text{proj}_{\mathbf{v}_1} \mathbf{y} = \frac{\sum_{i=1}^n y_i}{\sqrt{n}} \begin{bmatrix} 1/\sqrt{n} \\ 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{bmatrix} = \frac{\sum_{i=1}^n y_i}{n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\text{proj}_{\mathbf{v}_2} \mathbf{y} = \frac{\sum_{i=1}^n y_i(x_i - \mu)}{\sigma} \begin{bmatrix} (x_1 - \mu)/\sigma \\ (x_2 - \mu)/\sigma \\ \vdots \\ (x_n - \mu)/\sigma \end{bmatrix} = \left(-\frac{\mu \sum_{i=1}^n y_i(x_i - \mu)}{\sigma^2} \right) \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \frac{\sum_{i=1}^n y_i(x_i - \mu)}{\sigma^2} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Thus we choose

$$\hat{\mathbf{y}} = \text{proj}_{\mathbf{v}_1}\mathbf{y} + \text{proj}_{\mathbf{v}_2}\mathbf{y} = \left(\frac{\sum_{i=1}^n y_i}{n} - \frac{\mu\sum_{i=1}^n y_i(x_i-\mu)}{\sigma^2}\right)\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} + \left(\frac{\sum_{i=1}^n y_i(x_i-\mu)}{\sigma^2}\right)\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

We can read off the solutions for a', b':

$$a' = \frac{\sum_{i=1}^n y_i(x_i-\mu)}{\sigma^2}, \quad b' = \frac{\sum_{i=1}^n y_i}{n} - \frac{\mu\sum_{i=1}^n y_i(x_i-\mu)}{\sigma^2}$$

These are the formulas you use for fitting lines to data. We would write $\hat{\mathbf{y}} = \text{proj}_{\text{colsp}(A)}\mathbf{y}$.

Fourier Coefficients

One lovely application of orthogonal projections that you are likely to see is to take an arbitrary continuous fuction f and project it orthgonally onto the $2n+1$ dimensional space $U = \text{span}(1, \sin(x), \cos(x), \sin(2x), \cos(2x), \ldots, \sin(nx), \cos(nx))$. We use the orthonormal basis

$$\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \ldots \frac{\sin nx}{\sqrt{\pi}}, \frac{\cos nx}{\sqrt{\pi}}.$$

Our orthogonal projection is then a linear combination of the $2n+1$ functions. The inner product for these functions is

$$< f, g >= \int_0^{2\pi} f(x)g(x)dx$$

and we can verify that the $2n+1$ given functions are orthogonal. To do the orthogonal projection using the inner products, we must remember to normalize the functions.

$$\text{proj}_U f(x) = a + \sum_{i=1}^n b_i \frac{\sin(ix)}{\sqrt{\pi}} + \sum_{i=1}^n c_i \frac{\cos(ix)}{\sqrt{\pi}}$$

where we use the orthonormal basis to assist so that

$$a = \int_0^{2\pi} \frac{1}{\sqrt{2\pi}} f(x)dx, \quad b_i = \int_0^{2\pi} f(x)\frac{\sin(ix)}{\sqrt{\pi}}dx, \quad c_i = \int_0^{2\pi} f(x)\frac{\cos(ix)}{\sqrt{\pi}}dx.$$

The numbers $a, b_1, b_2, \ldots, c_n$ are called the fourier coefficients. Note how we don't need to know about the rest of the vector space to do this orthogonal projection. There are issues with numerical roundoff errors but these ideas seem to work very well in practice.