Projections

We have previous seen the case of a projection of a vector onto a single vector (or, equivalently, a one dimensional subspace). We will generalize this idea to projection onto a r-dimensional subspace. For example, when one wants to represent a 3-dimensional object on, say, a computer screen, we must project the object onto a 2-dimensional plane (namely the screen).

For the projection onto a line, L, in $\mathbf{R}^2$ passing through the origin, we have seen that the standard matrix is given by

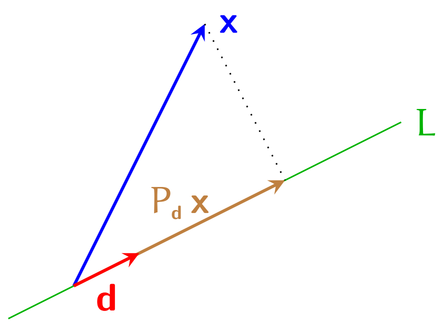
$$P_{\mathbf{d}} = \frac{1}{d_1^2 + d_2^2} \begin{bmatrix} d_1^2 & d_1 d_2 \\ d_1 d_2 & d_2^2 \end{bmatrix}$$

where

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

is any vector parallel to the direction of L.

Projections



Let

$$\mathbf{u} = \frac{1}{\sqrt{d_1^2 + d_2^2}} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

be a *unit* vector in the direction of **d**. Now

$$\mathbf{u}\,\mathbf{u}^T = \frac{1}{d_1^2 + d_2^2} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \begin{bmatrix} d_1 & d_2 \end{bmatrix} = \frac{1}{d_1^2 + d_2^2} \begin{bmatrix} d_1^2 & d_1 d_2 \\ d_1 d_2 & d_2^2 \end{bmatrix}$$

and so

$$P_{\mathbf{d}} = \mathbf{u}\,\mathbf{u}^T.$$

This is a very convenient form for the standard matrix. Note that

$$P_{\mathbf{d}}^T = (\mathbf{u}\,\mathbf{u}^T)^T = \mathbf{u}\,\mathbf{u}^T = P_{\mathbf{d}}$$

and so $P_{\mathbf{d}}$ is a *symmetric* matrix. Since $\mathbf{u}$ is a unit vector

$$\mathbf{u}^T\,\mathbf{u} = 1$$

and so

$$P_{\mathbf{d}}^2 = \mathbf{u}\left(\mathbf{u}^T\,\mathbf{u}\right)\mathbf{u}^T = \mathbf{u}\,\mathbf{u}^T = P_{\mathbf{d}}.$$

A matrix A such that $A^2 = A$ is called *idempotent*. This reflects the fact that if we project a vector that lies in the subspace then it will be unchanged. Furthermore

$$P_{\mathbf{d}} = \mathbf{u}\,\mathbf{u}^T = \frac{1}{\sqrt{d_1^2 + d_2^2}}\begin{bmatrix} d_1\,\mathbf{u} & d_2\,\mathbf{u} \end{bmatrix}$$

and so

$$\text{col}\left(P_{\mathbf{d}}\right) = \text{span}\left(\mathbf{u}\right).$$

Thus the one dimensional subspace which $P_{\mathbf{d}}$ projects onto its column space.

DEFINITION 10.9 (Projection Matrix)

A $n \times n$ matrix P is called a projection matrix if it is symmetric ($P^T = P$) and idempotent ($P^2 = P$).

The following theorem shows that this definition captures the idea of a projection at least in the 1-dimensional case.

THEOREM 10.10

Let P be a rank 1 projection matrix. Then

$$P = \mathbf{u}\,\mathbf{u}^T$$

where $\mathbf{u} \in \text{col}\left(P\right)$ is a unit vector. Furthermore

$$P\,\mathbf{x} = \text{proj}_{\mathbf{u}}\left(\mathbf{x}\right)$$

and so P is the standard matrix for the linear transformation $\text{proj}_{\mathbf{u}}\left(\mathbf{x}\right)$.

Projections

Proof: Since P has rank 1, we have

$$P = \mathbf{u}\,\mathbf{v}^T = \begin{bmatrix} v_1\,\mathbf{u} & v_2\,\mathbf{u} & \cdots & v_n\,\mathbf{u} \end{bmatrix}$$

where we can assume that, without loss of generality, $\mathbf{u}$ is a unit vector (why?). We must show $\mathbf{v} = \mathbf{u}$. Since P is idempotent, we have

$$O = P^2 - P = \mathbf{u}\,\left(\mathbf{v}^T\,\mathbf{u}\right)\mathbf{v}^T - \mathbf{u}\,\mathbf{v}^T$$

Now $\mathbf{v}^T\,\mathbf{u}$ is a scalar (it is a dot product). Therefore we have

$$\mathbf{u}\,\left(\mathbf{v}^T\,\mathbf{u}\right)\mathbf{v}^T = \left(\mathbf{v}^T\,\mathbf{u}\right)\mathbf{u}\,\mathbf{v}^T.$$

Thus

$$O = \left(\mathbf{v}^T\,\mathbf{u} - 1\right)\mathbf{u}\,\mathbf{v}^T$$

and so

$$\mathbf{v}^T\,\mathbf{u} = 1.$$

Furthermore P is symmetric and so

$$P^T = \mathbf{v}\,\mathbf{u}^T = \mathbf{u}\,\mathbf{v}^T = P.$$

Projections

Now $\mathbf{u}$ is a unit vector, $\mathbf{u}^T\,\mathbf{u} = 1$ and so

$$\begin{array}{lllllllll} \mathbf{v} & = & \mathbf{v}\left(\mathbf{u}^T\mathbf{u}\right) & = & \left(\mathbf{v}\,\mathbf{u}^T\right)\mathbf{u} & = & \mathbf{u}\,\mathbf{v}^T\,\mathbf{u} & & \text{since } P \text{ is symmetric} \\ & & & = & \mathbf{u}\left(\mathbf{v}^T\,\mathbf{u}\right) & = & \mathbf{u} & & \text{since } P \text{ is idempotent} \end{array}$$

as required. Clearly $\mathbf{u} \in \operatorname{col}(P)$. Finally, since $\mathbf{u}$ is a unit vector, we have

$$P\,\mathbf{x} = \mathbf{u}\,\mathbf{u}^T\,\mathbf{x} = (\mathbf{u}\cdot\mathbf{x})\,\mathbf{u} = \operatorname{proj}_{\mathbf{u}}(\mathbf{x})\,. \qquad \square$$

This result generalizes our earlier observations in $\mathbf{R}^2$ to $\mathbf{R}^n$. The projection of a vector $\mathbf{x} \in \mathbf{R}^n$ onto a direction given by the unit vector $\mathbf{u} \in \mathbf{R}^n$ is given by $\mathbf{u}\,\mathbf{u}^T\,\mathbf{x}$. Note that there are two choices for the unit vector ($\pm\,\mathbf{u}$) but that the projection formula is independent of that choice.

PROJECTIONS

Example

Find the projection of $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ onto the direction $\mathbf{d} = \begin{bmatrix} 4 \\ 0 \\ -3 \end{bmatrix}$.

Solution: A unit vector in the direction of $\mathbf{d}$ is $\mathbf{u} = \frac{1}{5}\begin{bmatrix} 4 \\ 0 \\ -3 \end{bmatrix}$. Therefore

$$\mathrm{P} = \mathbf{u}\,\mathbf{u}^{\mathrm{T}} = \frac{1}{25}\begin{bmatrix} 4 \\ 0 \\ -3 \end{bmatrix}\begin{bmatrix} 4 & 0 & -3 \end{bmatrix} = \frac{1}{25}\begin{bmatrix} 16 & 0 & -12 \\ 0 & 0 & 0 \\ -12 & 0 & 9 \end{bmatrix}.$$

Thus

$$\mathrm{proj}_{\mathbf{d}}(\mathbf{x}) = \mathrm{P}\,\mathbf{x} = \frac{1}{25}\begin{bmatrix} 16 & 0 & -12 \\ 0 & 0 & 0 \\ -12 & 0 & 9 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \frac{1}{25}\begin{bmatrix} -8 \\ 0 \\ 6 \end{bmatrix}.$$

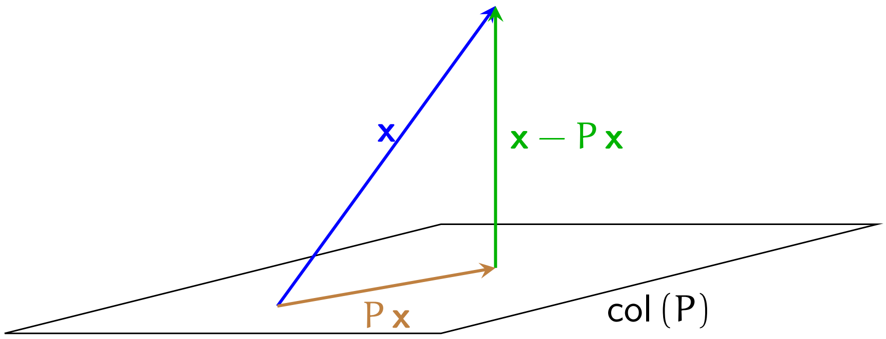
PROJECTIONS

What is the effect of projection matrices of higher rank? We know that $\mathrm{P}\,\mathbf{x}$ is a linear combination of the columns of P and so P might act as a *projection onto its column space*. At this stage we have two options to address this question. First we could consider more examples. For example, we could examine the projection of a vector onto a 2-dimensional subspace in $\mathbf{R}^3$ (that is, onto a plane passing through the origin). The second alternative is to try to generalize the properties we have found in the rank 1 case.

Exercise: Do Problems 6-12 on pages 378-379 (pages 364-365 in the second edition) in Poole. This computes the projection of a vector in $\mathbf{R}^3$ onto a plane passing through the origin.

PROJECTIONS

The idea of projection is to split a vector $\mathbf{x}$ into a component which lies in a given subspace ($P\,\mathbf{x}$) and a component which is perpendicular to that subspace ($\mathbf{x} - P\,\mathbf{x}$).



The two crucial properties for a projection are

- Any vector $\mathbf{v} \in \operatorname{col}(P)$ remains unchanged; that is $P\,\mathbf{v} = \mathbf{v}$.
- For any $\mathbf{x}$, the "orthogonal" component $\mathbf{x} - P\,\mathbf{x}$ is orthogonal to *every* vector in $\operatorname{col}(P)$.

PROJECTIONS

The question that we must address is whether we have captured these two geometric properties in our (algebraic) definition of a projection matrix. As a first step, we can check that these two properties hold in the 1-dimensional case by using the explicit form of $P = \mathbf{u}\,\mathbf{u}^T$.

We see that if $\mathbf{v} \in \operatorname{col}(P)$ then $\mathbf{v} = c\,\mathbf{u}$ for some scalar c and so

$$P\,\mathbf{v} = \mathbf{u}\,\mathbf{u}^T c\,\mathbf{u} = c\,\mathbf{u} = \mathbf{v}$$

since $\mathbf{u}$ is a unit vector. Thus, in the rank 1 case, P leaves unchanged anything in its column space. In addition, for any $\mathbf{x}$, we have

$$\mathbf{u}^T(\mathbf{x} - P\,\mathbf{x}) = \mathbf{u}^T\mathbf{x} - \left(\mathbf{u}^T\mathbf{u}\right)\mathbf{u}^T\mathbf{x} = 0$$

since $\mathbf{u}$ is a unit vector. Therefore $\mathbf{x} - P\,\mathbf{x}$ is orthogonal to $\mathbf{u}$ and so is orthogonal to any vector in $\operatorname{col}(P)$.

Projections

Let us consider the general case. Suppose P is a projection matrix. Does the associated linear transformation satisfy the two requirements for a projection? For any $\mathbf{x}$, we have

$$P(\mathbf{x} - P\mathbf{x}) = P\mathbf{x} - P^2\mathbf{x} = \mathbf{0}$$

since $P^2 = P$ and so $\mathbf{x} - P\mathbf{x} \in \text{null}(P)$. Thus we want the null space of P to be "orthogonal" to the column space of P. Let $\mathbf{y} \in \text{null}(P)$. Since $P^T = P$, the rows of P are the same as the columns of P and so

$$P = \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \vdots \\ \mathbf{p}_n^T \end{bmatrix}$$

where $\mathbf{p}_i$ are the columns of P. Thus

Projections

$$P\mathbf{y} = \begin{bmatrix} \mathbf{p}_1^T\mathbf{y} \\ \mathbf{p}_2^T\mathbf{y} \\ \vdots \\ \mathbf{p}_n^T\mathbf{y} \end{bmatrix} = \mathbf{0}.$$

Thus $\mathbf{y}$ is orthogonal to any vector in $\text{col}(P) = \text{span}(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n)$.

The fact that P is idempotent guarantees that $\mathbf{x} - P\mathbf{x} \in \text{null}(P)$. The fact that P is symmetric guarantees that anything in the null space of P is orthogonal to any vector in the column space of P. In particular, if $\mathbf{y}$ is in *both* $\text{null}(P)$ and $\text{col}(P)$ then $\mathbf{y}^T\mathbf{y} = 0$ and so $\mathbf{y} = \mathbf{0}$. Finally, for $\mathbf{v} \in \text{col}(P)$ we have $\mathbf{v} - P\mathbf{v} \in \text{col}(P)$ (why?) and $\mathbf{v} - P\mathbf{v} \in \text{null}(P)$. Thus $\mathbf{v} - P\mathbf{v} = \mathbf{0}$ and so

$$P\mathbf{v} = \mathbf{v}$$

as required for a projection. Thus the linear transformation associated with a projection matrix is a projection in the geometric sense.

Projections

Theorem 10.11

Let P be a $n \times n$ projection matrix of rank r. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a basis for $\operatorname{col}(P)$ *and* $\{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}$ *be a basis for* $\operatorname{null}(P)$. *Then* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *is a basis for* $\mathbf{R}^n$. *Furthermore, if* $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ *then*

$$P\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_r\mathbf{v}_r.$$

Geometrically, we are splitting a vector $\mathbf{v}$ into a component that lies in the subspace, the *parallel component*

$$\operatorname{proj}_{\operatorname{col}(P)}(\mathbf{v}) = \mathbf{v}_{\parallel} = P\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_r\mathbf{v}_r,$$

and a component that is orthogonal to the subspace, the *orthogonal (or perpendicular) component*

$$\operatorname{perp}_{\operatorname{col}(P)}(\mathbf{v}) = \mathbf{v}_{\perp} = \mathbf{v} - P\mathbf{v} = c_{r+1}\mathbf{v}_{r+1} + c_{r+2}\mathbf{v}_{r+2} + \cdots + c_n\mathbf{v}_n.$$

Projections

We have yet to find an efficient way (or for that matter any way) to compute P. We need a symmetric and idempotent matrix with a prescribed column space.

Note that, for $i = 1, 2, \ldots, r$,

$$\mathbf{v}_i \cdot \mathbf{v} = \mathbf{v}_i^T\mathbf{v} = c_1\mathbf{v}_i^T\mathbf{v}_1 + c_2\mathbf{v}_i^T\mathbf{v}_2 + \cdots + c_r\mathbf{v}_i^T\mathbf{v}_r$$

since $\mathbf{v}_i$ is orthogonal to everything in $\operatorname{null}(P)$. This gives a system of r equations for the r unknowns $c_1, c_2, \ldots, c_r$

$$\begin{bmatrix} \mathbf{v}_1^T\mathbf{v}_1 & \mathbf{v}_1^T\mathbf{v}_2 & \cdots & \mathbf{v}_1^T\mathbf{v}_r \\ \mathbf{v}_2^T\mathbf{v}_1 & \mathbf{v}_2^T\mathbf{v}_2 & \cdots & \mathbf{v}_2^T\mathbf{v}_r \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_r^T\mathbf{v}_1 & \mathbf{v}_r^T\mathbf{v}_2 & \cdots & \mathbf{v}_r^T\mathbf{v}_r \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T\mathbf{v} \\ \mathbf{v}_2^T\mathbf{v} \\ \vdots \\ \mathbf{v}_r^T\mathbf{v} \end{bmatrix}.$$

Projections

This system can be rewritten

$$\begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_r \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \mathbf{v}.$$

Since $\{\mathbf{v}_1, \ \mathbf{v}_2, \ \ldots, \ \mathbf{v}_r\}$ are linearly independent (they form a basis), the rank of the coefficient matrix is r and so this system has a unique solution for all $\mathbf{v}$. Let

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix};$$

a matrix whose column space is given by span $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r)$, then this system takes the form

$$A^T A \mathbf{c} = A^T \mathbf{v}.$$

Projections

Example

Find the projection of $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ onto the subspace spanned by

$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

Solution: Now

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 5 & 0 \end{bmatrix}.$$

The system is

$$\left[A^T A \mid A^T \mathbf{v}\right] = \left[\begin{array}{cc|c} 35 & 5 & 3 \\ 5 & 5 & 1 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & \frac{1}{15} \\ 0 & 1 & \frac{2}{15} \end{array}\right]$$

Projections

and so the projection is

$$P\,\mathbf{v} = \tfrac{1}{15}\,\mathbf{v}_1 + \tfrac{2}{15}\,\mathbf{v}_2 = \tfrac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix}.$$

This is a cumbersome way to compute the projection (particularly if one wants to project many vectors onto the same subspace). With the rank 1 case, we found a very nice form for P by choosing a special basis for the subspace; namely a unit vector to span the 1-dimensional subspace. Can we do a similar thing with the higher rank cases?

Projections

The coefficent matrix is

$$\begin{bmatrix}\mathbf{v}_1^T\,\mathbf{v}_1 & \mathbf{v}_1^T\,\mathbf{v}_2 & \cdots & \mathbf{v}_1^T\,\mathbf{v}_r\\ \mathbf{v}_2^T\,\mathbf{v}_1 & \mathbf{v}_2^T\,\mathbf{v}_2 & \cdots & \mathbf{v}_2^T\,\mathbf{v}_r\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{v}_r^T\,\mathbf{v}_1 & \mathbf{v}_r^T\,\mathbf{v}_2 & \cdots & \mathbf{v}_r^T\,\mathbf{v}_r\end{bmatrix}.$$

If we choose an orthogonal basis for the column space then the coefficient matrix would become a *diagonal* matrix. In this case, the solution of the system is immediate and is given by

$$c_j = \frac{\mathbf{v}_j^T\,\mathbf{v}}{\mathbf{v}_j^T\,\mathbf{v}_j}.$$

Note that

$$c_j\,\mathbf{v}_j = \frac{\mathbf{v}_j^T\,\mathbf{v}}{\mathbf{v}_j^T\,\mathbf{v}_j}\,\mathbf{v}_j = \frac{\mathbf{v}_j^T\,\mathbf{v}\,\mathbf{v}_j}{\mathbf{v}_j^T\,\mathbf{v}_j} = \frac{\mathbf{v}_j\,\mathbf{v}_j^T}{\mathbf{v}_j^T\,\mathbf{v}_j}\,\mathbf{v}$$

since $\mathbf{v}_j^T\,\mathbf{v}$ is a scalar and so

Projections

$$P\,\mathbf{v} = \left(\frac{\mathbf{v}_1\,\mathbf{v}_1^T}{\mathbf{v}_1^T\,\mathbf{v}_1} + \frac{\mathbf{v}_2\,\mathbf{v}_2^T}{\mathbf{v}_2^T\,\mathbf{v}_2} + \cdots + \frac{\mathbf{v}_r\,\mathbf{v}_r^T}{\mathbf{v}_r^T\,\mathbf{v}_r}\right)\mathbf{v}.$$

Furthermore if we chose an orthonormal basis then the coefficient matrix will be an identity matrix. In this case the solution simplifies to

$$P\,\mathbf{v} = \left(\mathbf{v}_1\,\mathbf{v}_1^T + \mathbf{v}_2\,\mathbf{v}_2^T + \cdots + \mathbf{v}_r\,\mathbf{v}_r^T\right)\mathbf{v}$$

and we see

$$P = \mathbf{v}_1\,\mathbf{v}_1^T + \mathbf{v}_2\,\mathbf{v}_2^T + \cdots + \mathbf{v}_r\,\mathbf{v}_r^T.$$

Gram Schmidt Process

Reference: §5.2-5.3 in Poole.

We have seen that if $\{\mathbf{q}_1,\ \mathbf{q}_2,\ \ldots,\ \mathbf{q}_k\}$ is an orthonormal basis for a subspace W then

$$\mathrm{proj}_W(\mathbf{x}) = \left(\mathbf{q}_1\,\mathbf{q}_1^T + \mathbf{q}_2\,\mathbf{q}_2^T + \cdots + \mathbf{q}_k\,\mathbf{q}_k^T\right)\mathbf{x} = Q\,Q^T\mathbf{x}$$

where

$$Q = \begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k\end{bmatrix}.$$

Thus we can decompose the projection onto W

$$\mathrm{proj}_W(\mathbf{x}) = \mathrm{proj}_{\mathbf{q}_1}(\mathbf{x}) + \mathrm{proj}_{\mathbf{q}_2}(\mathbf{x}) + \cdots + \mathrm{proj}_{\mathbf{q}_k}(\mathbf{x})$$

to the sum of projections onto each of the vectors in the basis. Thus, if we *choose* an orthonormal basis, then the projection onto W can be computed by computing the projection onto each of the 1-dimensional subspaces given by $\mathrm{span}(\mathbf{q}_i)$.