Peeter Joot

May 16, 2008

1

1.1 Application of projection as left pseudoinverse (ie: linear fitting).

Equation **??** provides us a way to find best solutions to general equations of the form:

$$Ax = b$$

Here A is the matrix of a linear transformation, $A : \mathbb{R}^k \rightarrow \mathbb{R}^n$, for some $k < n$. By "best solutions" here, we give this the geometrical meaning, namely, the solution matching the projection of b onto the space.

If b is not completely in the column space $C(A)$ of A, this can have no solution. However, writing

$$b = \text{Proj}_A(b) + b_\perp$$

as the components of b in $C(A)$ and not in $C(A)$ respectively we can at least solve the reduced equation for $\hat{x}$:

$$A\hat{x} = \text{Proj}_A(b) \tag{1}$$

This will be possible even in circumstances that the original equation had no solution. Specifically, the vector b when projected onto the plane can be expressed as some linear combination of the columns of A (a basis for the subspace).

Substuition of our projection result into equation 1 yields:

$$A\hat{x} = \text{Proj}_A(b) = A(A^\text{T}A)^{-1}A^\text{T}b$$

The simplest case here is when A is of full column rank since one can premultiply this complete equation by A^T without any possibility of nulling $A\hat{x}$.

$$\begin{aligned} A^\text{T}A\hat{x} &= A^\text{T}A(A^\text{T}A)^{-1}A^\text{T}b \\ &= A^\text{T}b \end{aligned}$$

Thus our best fit vector is

$$\hat{x} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} b \tag{2}$$

Another way to view this is for any vector x that is not in the null space $N(A)$, then the matrix:

$$A^{+} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} \tag{3}$$

has the action of a left inverse for any full column rank matrix A. Thus when there is a solution to:

$$Ax = b. \tag{4}$$

It can be obtained by pre-multiplication using this "left" inverse.

$$A^{+} A x = x = A^{+} b \tag{5}$$

2 SVD connection.

SVT decomposition is an factoring of $A \in M^{m \times n}$ with orthonormal matrices $U \in M^{m \times m}$

and $V \in M^{n \times n}$ producing the following form:

$$A = U \Sigma V^{\mathrm{T}}$$

Sigma has the form:

$$\Sigma = \begin{bmatrix} D_{r,r} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}$$

where $r = \operatorname{rank}(A)$, and D is a diagonal matrix with the root of the (positive) eigenvalues of $A^{\mathrm{T}} A$.

This provides a generalized spectral decomposition and similarity that applies to both non-square matrices and matrices not otherwise diagonalizable (ie: square matrix with similarity to a Jordon form matrix). Given this decomposition we can write:

$$\Sigma = U^{\mathrm{T}} A V$$

If one were to ask the question of what is the closest that one could get to inverting such a matrix. It's pretty clear that the closest one could get to identity will be with multiplication of a Σ^{+} of the following form:

$$\Sigma^{+} \Sigma = \begin{bmatrix} (D_{r,r})^{-1} & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{bmatrix} \begin{bmatrix} D_{r,r} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} = \begin{bmatrix} I_{r,r} & 0_{r,n-r} \\ 0_{n-r,r} & 0_{n-r,n-r} \end{bmatrix}$$

For a right pseudoinverse we have a similar result:

$$\Sigma\Sigma^+ = \begin{bmatrix} D_{r,r} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} \begin{bmatrix} (D_{r,r})^{-1} & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{bmatrix} = \begin{bmatrix} I_{r,r} & 0_{r,m-r} \\ 0_{m-r,r} & 0_{m-r,m-r} \end{bmatrix}$$

With either of these one can define a corresponding pseudoinverse (left or right) as:

$$A^+ = V\Sigma^+U^{\mathrm{T}} \tag{6}$$

This is a logical definition, but how close is it to the projective left inverse we calculated above in the case where A is not of full column rank?

Multiplication gives:

$$\begin{aligned} A^+A &= V\Sigma^+U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} \\ &= V\Sigma^+\Sigma V^{\mathrm{T}} \\ &= \begin{bmatrix} v_1 & v_2 & \cdots & v_r & v_{r+1} & \cdots & v_n \end{bmatrix} \begin{bmatrix} (D_{r,r})^{-1} & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{bmatrix} \begin{bmatrix} D_{r,r} & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix} \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ \vdots \\ v_r^{\mathrm{T}} \\ {v_{r+1}}^{\mathrm{T}} \\ \vdots \\ v_n^{\mathrm{T}} \end{bmatrix} \end{aligned}$$

Writing $D_{r,r} = [\delta_{ij}\sigma_i]_{ij}$, we have:

$$V\Sigma^+\Sigma V^{\mathrm{T}} = \begin{bmatrix} \frac{v_1}{\sigma_1} & \frac{v_2}{\sigma_2} & \cdots & \frac{v_r}{\sigma_2} & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_1^{\mathrm{T}}\sigma_1 \\ v_2^{\mathrm{T}}\sigma_2 \\ \vdots \\ v_r^{\mathrm{T}}\sigma_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{7}$$

Considering this as the product of block matrices we have a product here of the form

$$\begin{bmatrix} A_{n,r} & 0_{n,n-r} \end{bmatrix} \begin{bmatrix} B_{r,n} \\ 0_{n-r,n} \end{bmatrix} = \begin{bmatrix} A_{n,r}B_{r,n} + 0_{n,n-r}0_{n-r,n} \end{bmatrix} = \begin{bmatrix} A_{n,r}B_{r,n} + 0_{n,n} \end{bmatrix} = \begin{bmatrix} A_{n,r}B_{r,n} \end{bmatrix}$$

Thus we can strip the block zero matrices from equation 7 and write

$$A^+A = V\Sigma^+\Sigma V^{\mathrm{T}} = \begin{bmatrix} \frac{v_1}{\sigma_1} & \frac{v_2}{\sigma_2} & \cdots & \frac{v_r}{\sigma_2} \end{bmatrix} \begin{bmatrix} v_1^{\mathrm{T}}\sigma_1 \\ v_2^{\mathrm{T}}\sigma_2 \\ \vdots \\ v_r^{\mathrm{T}}\sigma_r \end{bmatrix} \tag{8}$$

Eliminating the σ terms we have:

$$A^+A = \begin{bmatrix} \sum_{k=1}^{r} v_k v_k^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix} \begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ \vdots \\ v_r^{\mathrm{T}} \end{bmatrix} \tag{9}$$

We previously calculated a left inverse using the projection matrix associated with a full column rank matrix. For this product to have the properties of a left acting inverse we also expect it to be a projection. Let's disgress slightly before looking at whether equation 9 satisifies this expectation.

2.1 Correlating the SVD derived projection matrix back to A.

We now have to show that this is also the projection matrix associated

with the columns of the original matrix that we have an SVD factorization for

$$A = U\Sigma V^{\mathrm{T}}$$

Once we show this, then we have also demonstrated that the first r (orthonormal) column vectors in the matrix V of this decomposition are a basis for the column space of A itself. Note that we are switching back to the original definition of $V \in M^{n,n}$ here, and not the $V \in M^{n,r}$ of equation **??**.