SVD and the Pseudoinverse

We are now in a position to investigate SVD mechanics in analogy to eigenvalue/eigenvector mechanics. A similar process of finding singular values (eigenvalues) and the corresponding singular vectors (eigenvectors) yields a more general and flexible factorization decomposition of matrix A but the notion of expanding vectors on an eigenbasis remains intact. The terms ‘singular vector’ and ‘eigenvector’ will be used interchangeably.

Singular value decomposition of matrix A can be written as

$$A = UWV^T$$

where

1. U - The columns of U are the eigenvectors of AA^T. U is an m x m matrix containing an orthonormal basis of vectors for both the column space and the left null space of A. For orbit correction, the orbit vector will be expanded in terms of the basis vectors in U. In linear algebra, U contains the left singular vectors of A.

2) W - The 'singular values' or 'principle gains' of A lie on the diagonal of W and are the *square root* of the eigenvalues of both AA^T and A^TA, that is, the eigenvectors in U and the eigenvectors in V share the same eigenvalues!

3) V - The rows of V^T (columns of V) are the eigenvectors of A^TA. V is an n x n matrix containing an orthonormal basis of vectors for both the row space and the null space of A. The column vector of corrector magnets vector will be expanded in terms of the bases in V. V contains the right singular vectors of A.

For our applications, A^TA and AA^T are symmetric, real, positive-definite matrices so that all singular values are real positive numbers. Solving for the two different eigenspaces (U and V) corresponds to finding one diagonalizing transformation in the domain (V, correctors) and another diagonalizing transformation in the range (U, orbit) such that the original matrix becomes diagonal. In the square-matrix eigenvalue/eigenvector problem, only one eigenbasis was required to diagonalize the matrix.

WHY IS SINGULAR VALUE DECOMPOSITION USEFUL IN ACCELERATOR PHYSICS?
Like many problems in mathematical physics, eigenvectors are used to form orthogonal basis sets that can be used to expand the function of interest. Similar to expressing a function in a Fourier series or expanding a wavefunction in quantum mechanics in terms of energy eigenstates, is useful to expand the beam orbit on a set of basis vectors for orbit control.

The SVD factorization of a matrix A generates a set of eigenvectors for *both* the correctors and the orbit. There is a 1:1 correspondence between the i^{th} eigenvector in V, the i^{th} singular value in W and i^{th} eigenvector in U. The 'singular values' (eigenvalues) scale eigenvectors as they are transformed from the corrector eigenspace to the orbit eigenspace (or vice-versa). The process is analogous to square-matrix eigenvector mechanics but can be applied to non-square matrices. What's more, SVD generates the full set of four fundamental subspaces of matrix A.

INPUT/OUTPUT VIEWPOINT

One way to look at SVD is that the v_i are the input vectors of A and u_i are the output vectors of A. The linear mapping y=Ax can be decomposed as

$$y = Ax = UWV^Tx\ .$$

The action of the matrix goes like this:

1. compute coefficients of x along the input directions $v_1 \ldots v_r$.
   V^Tx resolves the input vector x into the orthogonal basis of input vectors v_i.

2. scale the vector coefficients by σ_i on the diagonal of W

3. multiply by U: $y=\Sigma a_i u_i$ is a linear superposition of the singular vectors u_i

The difference from the eigenvalue decomposition for a symmetric matrix A is that the input and output directions are *different*. Since the SVD returns the singular value/eigenvector sets in descending order of the singular values,

v_1 is the most sensitive (highest gain) input direction
u_1 is the most sensitive (highest gain) output direction
$Av_1=\sigma_1 u_1$

SVD gives a clear picture of the gain as a function of input/output directions

Example: Consider a *4 x 4* by matrix A with singular values Σ=diag(12, 10, 0.1, 0.05).
The input components along directions v_1 and v_2 are amplified by about a factor of 10 and come out mostly along the plane spanned by u_1 and u_2 .

The input components along directions v_3 and v_4 are attenuated by ~10.

For some applications, you might say A is effectively *rank 2*!

EIGENVECTORS OF AA^T AND A^TA

Let's look more closely at the basis vectors in U and V. For U the claim was that the columns are the eigenvectors of AA^T. Since AA^T is real symmetric, U is an orthonormal set and $U^T = U^{-1}$ and $U^TU = I$. Substituting the SVD of A into the outer product AA^T

$$AA^T = (UWV^T)(UWV^T)^T = (UWV^T)(VW^TU^T) = UW^2U^T$$

multiplying on the right by U,

$$AA^TU = UW^2$$

Hence, U contains the eigenvectors for AA^T. The eigenvalues are the diagonals on W. Similarly, it can be shown that V contains the eigenvectors of A^TA.

Another way of looking at the same thing:

Let $v_1 \ldots v_r$ be a basis for the row space (correctors)
$u_1 \ldots u_r$ be a basis for the column space (orbits)

By definition

$$Av_i = \sigma u_i$$

If v_i are the eigenvectors of A^TA

$$A^TAv_i = \sigma_i^2 v_i$$

pre-multiplying by A

$$AA^T(Av_i) = AA^T(\sigma_i u_i)$$

so $u_i = Av_i / \sigma_i$ is an eigenvector of AA^T.

MATLAB Demonstration of SVD – Forward multiplication
>>edit SVD_1

SUBSPACES OF A

The SVD factorization of an *m x n* matrix A with rank r is $A = UWV^T$ where W is a quasi-diagonal matrix with singular values on the diagonals

$$W = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

The sparse matrix W also arranges the singular values in descending order

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$$

The columns of U and V are

$U = [u_1\ u_2 \ldots u_m]$ left singular vectors

$V = [v_1\ v_2 \ldots v_n]$ right singular vectors

The fundamental subspaces of A are spanned by

$\text{Range}(A) = \text{span}[u_1\ u_2 \ldots\ u_r]$	basis to expand orbits (column space)
$\text{Range}(A^T) = \text{span}[v_1\ v_2 \ldots\ v_r]$	basis to expand correctors (row space)
$\text{Nullspace}(A) = \text{span}[v_{r+1}\ v_{r+2} \ldots v_n]$	corrector patterns that do not perturb orbit
$\text{Nullspace}(A^T) = \text{span}[u_{r+1}\ u_{r+2} \ldots u_m]$	orbit errors

In it's full glory, the SVD of a well-behaved least-squares problem will look like

$$A = [u_1 u_2 \ldots u_r \quad u_{r+1} \ldots u_{r+m}] \begin{bmatrix} \sigma_1 & & & | & \\ & \ldots & & | & 0 \\ & & \sigma_r & | & \\ - & - & - & | & - \quad - \\ & 0 & & | & 0 \\ & & & | & \end{bmatrix} [v_1 v_2 \ldots v_r \quad v_{r+1} \ldots v_{r+m}]^T$$

MATLAB Demonstration of SVD – Subspaces
>>edit SVD_2

SINGULAR VALUE DECOMPOSITION – FORWARD SOLUTION
Use SVD to decompose the square response matrix: $R = UWV^T$ (m=n).

As a demonstration of SVD matrix mechanics, follow through a simple calculation.

$x = (U\ W\ V^T)\theta$ **=>** $$\boldsymbol{x} = \boldsymbol{U} \cdot \boldsymbol{W} \cdot \begin{bmatrix} - & \boldsymbol{v}_1 & - \\ & \ldots & \\ - & \boldsymbol{v}_n & - \end{bmatrix} \cdot \boldsymbol{\theta}$$

$$x = U \cdot W \cdot \begin{bmatrix} v_1 \cdot \theta \\ | \\ v_n \cdot \theta \end{bmatrix}$$ (project **θ** into **V**-space)

$$\boldsymbol{x} = \boldsymbol{U} \cdot \begin{bmatrix} \boldsymbol{w}_1 & & 0 \\ & \ldots & \\ 0 & & \boldsymbol{w}_n \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{v}_1 \cdot \boldsymbol{\theta} \\ | \\ \boldsymbol{v}_n \cdot \boldsymbol{\theta} \end{bmatrix}$$ (W diagonal)

$$x = U \cdot \begin{bmatrix} w_1 v_1 \cdot \theta \\ | \\ w_n v_n \cdot \theta \end{bmatrix} \qquad \text{(multiply by 'gains' w)}$$

$$x = \begin{bmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} w_1 v_1 \cdot \theta \\ | \\ w_n v_n \cdot \theta \end{bmatrix}$$

$$x = \sum_i (w_i v_i \cdot \theta) u_i \qquad \text{(expansion by eigenvectors)}$$

The final equation expresses the orbit as a linear combination of orbit eigenvectors u_i. The coefficients of the expansion are projections of the corrector vector θ into the corrector eigenvectors v_i weighted by the singular values w_i.

RESPONSE MATRIX EXAMPLE: INPUT IS AN EIGENVECTOR v_i

Suppose the corrector column vector θ is equal to corrector eigenvector v_i. In this case,

$$x = \begin{bmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} w_1 & & 0 \\ & \dots & \\ 0 & & w_n \end{bmatrix} \cdot \begin{bmatrix} - & v_1 & - \\ & \dots & \\ - & v_n & - \end{bmatrix} \cdot v_i$$

$$x = \begin{bmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} w_1 & & 0 \\ & \dots & \\ 0 & & w_n \end{bmatrix} \cdot \begin{bmatrix} 0 \\ v_i \cdot v_i \\ 0 \end{bmatrix}$$

$$x = \begin{bmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} 0 \\ w_i \\ 0 \end{bmatrix} \qquad (v_i \cdot v_i = 1)$$

$$x = w_i u_i$$

In other words, corrector eigenvector v_i produces orbit eigenvector u_i scaled by the singular value quantity w_i.

An important point can be made here. Remember that we called the v-vectors the input vector basis and the u-vectors the output vector basis. If we 'zero out' small singular values in the w-matrix then projections of the input vector (θ) onto the corresponding v-vectors will not pass through. In this sense, we are creating a 'low pass filter' on the transformation from θ-space to x-space. This will be even more important in the next section (pseudoinverse) where we have the opportunity to low-pass the (noisy) measured orbit.

MATLAB Demonstration of SVD – Vector expansion on a SVD eigenbasis

>>edit SVD_3

THE PSEUDOINVERSE
If a matrix A has the singular value decomposition

$$A=UWV^T$$

then the pseudo-inverse or Moore-Penrose inverse of A is

$$A^+=V^TW^{-1}U$$

If A is 'tall' (m>n) and has full rank

$$A^+=(A^TA)^{-1}A^T \qquad \text{(it gives the least-squares solution } x_{lsq}=A^+b)$$

If A is 'short' (m<n) and has full rank

$$A^+=A^T(AA^T)^{-1} \qquad \text{(it gives the least-norm solution } x_{l\text{-}n}=A^+b)$$

In general, $x_{pinv}=A^+b$ is the minimum-norm, least-squares solution.

MATLAB Demonstration of SVD – Pseudoinverse
>>edit SVD_4

SINGULAR VALUE DECOMPOSITION – BACKWARD SOLUTION (INVERSE)
Again the response matrix R is decomposed using SVD: $R^{-1} = VW^{-1}U^T$
Where W^{-1} has the inverse elements of W along the diagonal. If an element of W is zero, the inverse is set to zero.

We now repeat the matrix mechanics outlined above for the inverse problem:

$\theta = (V\, W^{-1}\, U^T)x$ =>

$$\boldsymbol{\theta} = \boldsymbol{V}\cdot\boldsymbol{W}^{-1}\cdot\begin{bmatrix} - & \boldsymbol{u}_1 & - \\ & \ldots & \\ - & \boldsymbol{u}_n & - \end{bmatrix}\cdot\boldsymbol{x}$$

$$\boldsymbol{\theta} = \boldsymbol{V}\cdot\boldsymbol{W}^{-1}\cdot\begin{bmatrix} \boldsymbol{u}_1\cdot\boldsymbol{x} \\ | \\ \boldsymbol{u}_n\cdot\boldsymbol{x} \end{bmatrix} \qquad \text{(project } \mathbf{x} \text{ into } \mathbf{U}\text{-space)}$$

$$\boldsymbol{\theta} = \boldsymbol{V}\cdot\begin{bmatrix} \boldsymbol{w}_1^{-1} & & 0 \\ & \ldots & \\ 0 & & \boldsymbol{w}_n^{-1} \end{bmatrix}\cdot\begin{bmatrix} \boldsymbol{u}_1\cdot\boldsymbol{x} \\ | \\ \boldsymbol{u}_n\cdot\boldsymbol{x} \end{bmatrix} \qquad (\mathbf{W} \text{ diagonal})$$

$$\boldsymbol{\theta} = \boldsymbol{V} \cdot \begin{bmatrix} \boldsymbol{w}_1^{-1}\boldsymbol{u}_1 \cdot \boldsymbol{x} \\ | \\ \boldsymbol{w}_n^{-1}\boldsymbol{u}_n \cdot \boldsymbol{x} \end{bmatrix} \quad \text{(multiply by inverse gains } w_i^{-1}\text{)}$$

$$\boldsymbol{\theta} = \begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \dots & \boldsymbol{v}_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{w}_1^{-1}\boldsymbol{u}_1 \cdot \boldsymbol{x} \\ | \\ \boldsymbol{w}_n^{-1}\boldsymbol{u}_n \cdot \boldsymbol{x} \end{bmatrix}$$

$$\boldsymbol{\theta} = \sum_{\boldsymbol{i}} (\boldsymbol{w}_{\boldsymbol{i}}^{-1}\boldsymbol{u}_{\boldsymbol{i}} \cdot \boldsymbol{x})\boldsymbol{v}_{\boldsymbol{i}} \quad \text{(expansion by eigenvectors)}$$

The final equation expresses the corrector set θ as a linear combination of corrector eigenvectors v_i. The coefficients of the expansion are projections of the orbit vector x into the orbit eigenvectors u_i weighted by the inverse singular values w_i^{-1}. The process is analogous to the forward problem outlined above.

ORBIT CORRECTION EXAMPLE (m=n)

As a simple example, suppose the orbit vector x is equal to orbit eigenvector u_i. In this case,

$$\boldsymbol{\theta} = \begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \dots & \boldsymbol{v}_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{w}_1^{-1} & & 0 \\ & \dots & \\ 0 & & \boldsymbol{w}_n^{-1} \end{bmatrix} \cdot \begin{bmatrix} - & \boldsymbol{u}_1 & - \\ & \dots & \\ - & \boldsymbol{u}_n & - \end{bmatrix} \cdot \boldsymbol{u}_{\boldsymbol{i}}$$

$$\boldsymbol{\theta} = \begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \dots & \boldsymbol{v}_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{w}_1^{-1} & & 0 \\ & \dots & \\ 0 & & \boldsymbol{w}_n^{-1} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \boldsymbol{u}_{\boldsymbol{i}} \cdot \boldsymbol{u}_{\boldsymbol{i}} \\ 0 \end{bmatrix}$$

$$\boldsymbol{\theta} = \begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \dots & \boldsymbol{v}_m \\ | & & | \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \boldsymbol{w}_{\boldsymbol{i}}^{-1} \\ 0 \end{bmatrix} \quad (\boldsymbol{u}_{\boldsymbol{i}} \cdot \boldsymbol{u}_{\boldsymbol{i}} = 1)$$

$$\boldsymbol{\theta} = \boldsymbol{w}_{\boldsymbol{i}}^{-1}\boldsymbol{v}_{\boldsymbol{i}}$$

In other words, orbit eigenvector u_i produces corrector eigenvector v_i scaled by the singular value quantity w_i^{-1}.

Again we point out that by zeroing-out small singular values (large w^{-1}) we can low-pass filter the input vector, x. This gives the opportunity to filter out noise in the measurement.

MATLAB Demonstration of SVD – Backsubstitution

```
>>edit SVD_5
```