7.7 Singular Value Decomposition

- The SVD is an important general tool in understanding matrix equations. It is not tied to LS
- Can view SVD as a generalization of eigendecomp to non-square matrices

square $\quad Q^{+} A Q = \Lambda = \text{diag}(\lambda_1, \lambda_2 \cdots \lambda_M)$

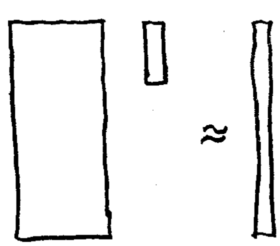
$\qquad Q$ is unitary

non square $\quad U^{+} A V = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \quad \Sigma = \text{diag}(\sigma_1, \cdots \sigma_W)$

$\qquad U, V$ unitary

- SVD in this course is a tool for understanding the pseudoinverse, a general approach to $A\underline{w} = \underline{d}$
- Computationally, using SVD to work directly with A gives more numerical precision than solving $(A^{+}A)^{-1}A^{+}$. The Gram matrix is already factored by A, so square the relative error (eg $10^{-4} \to 10^{-8}$), doubling the significant figures.
- Distinguish

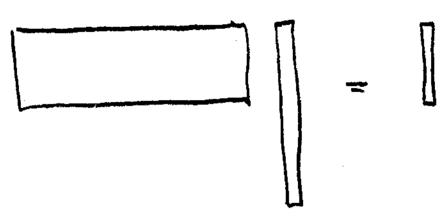
overdetermined



$A \quad \underline{w} \approx \underline{d}$

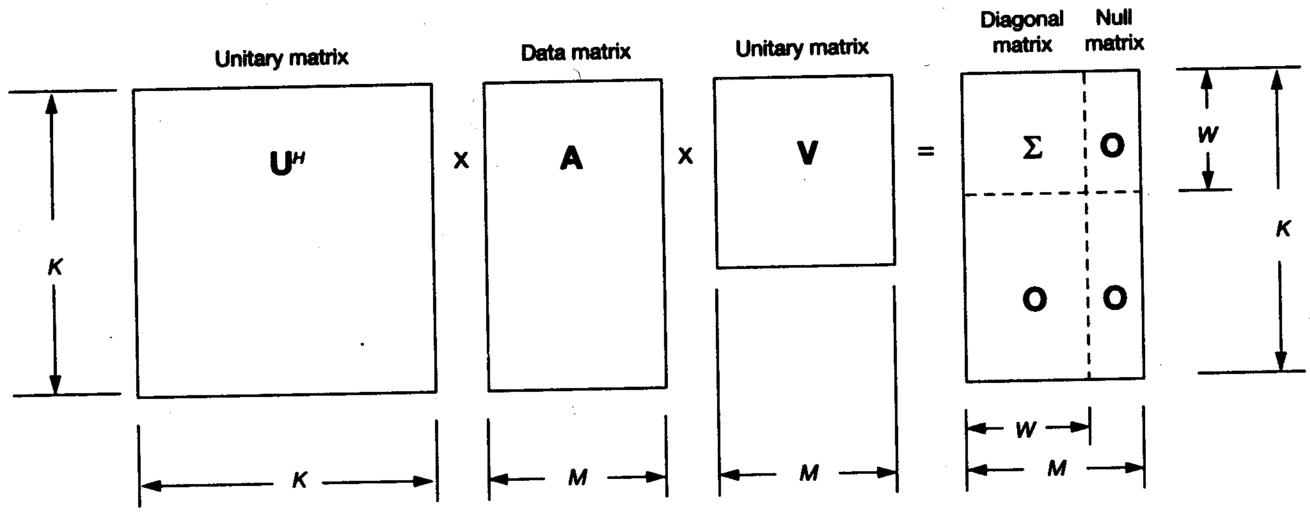
Usually no solution, so min $\|\underline{e}\|^2$

underdetermined



$A \quad \underline{w} = \underline{d}$

Lots of solutions. Which one to pick?



O : Null matrix

- Proof sketch (consult text), overdetermined case
  - Note $\text{rank}(A) = \text{rank}(A^+A)$ since any $\underline{x}$ that makes $A\underline{x} = \underline{0}$ also makes $A^+A\underline{x} = \underline{0}$, and the dimension of this nullspace is $M-W$
  - Therefore, of the M non-neg eigenvals of A^+A

    $$\sigma_1^2, \cdots \sigma_W^2, \sigma_{W+1}^2, \cdots \sigma_M^2 \qquad (\text{formerly } \lambda_1, \cdots \lambda_W, \lambda_{W+1}, \cdots \lambda_M)$$

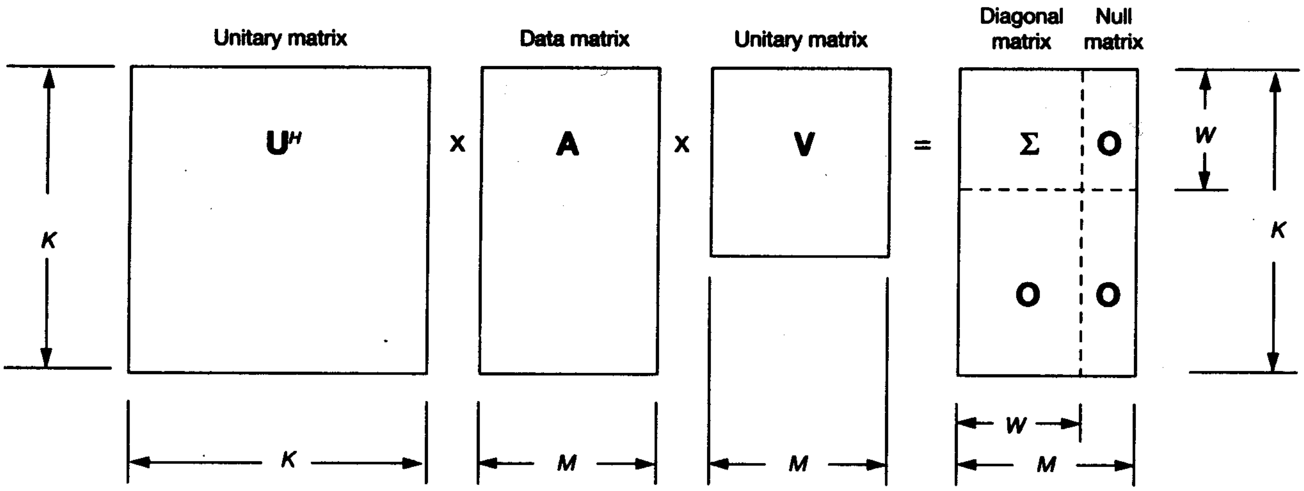
    W are pos, rest are zero
  - For V $(M \times M)$, use the unitary modal matrix of A^+A

    $$V^+A^+AV = \left[\begin{array}{c|c} \Sigma^2 & 0 \\ \hline 0 & 0 \end{array}\right] \qquad \Sigma = \text{diag}(\sigma_1, \sigma_2 \cdots, \sigma_W)$$

    partition $V = [\underbrace{V_1}_{M\times W} \mid \underbrace{V_2}_{M\times(M-W)}]$ the col vecs are orthog

$$\left[\begin{array}{c} V_1^+ \\ \hline V_2^+ \end{array}\right] A^+A \left[V_1 \mid V_2\right] = \left[\begin{array}{c} V_1^+ \\ \hline V_2^+ \end{array}\right] \left[V_1\Sigma^2 \mid 0\right] = \left[\begin{array}{c|c} \Sigma^2 & 0 \\ \hline 0 & 0 \end{array}\right]$$

$$= \left[\begin{array}{c|c} V_1^+A^+AV_1 & V_1^+A^+AV_2 \\ \hline V_2^+A^+AV_1 & V_2^+A^+AV_2 \end{array}\right]$$



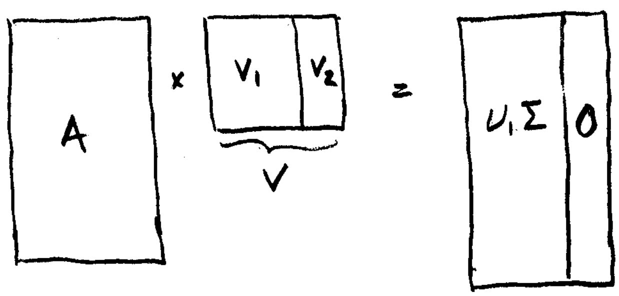
O : Null matrix

– so $V_1^{\dagger} A^{\dagger} A V_1 = \Sigma^2$

or $\underbrace{\Sigma^{-1} V_1^{\dagger} A^{\dagger}}_{U_1^{\dagger}} \; \underbrace{A V_1 \Sigma^{-1}}_{\substack{U_1 \\ K \times W}} = I$

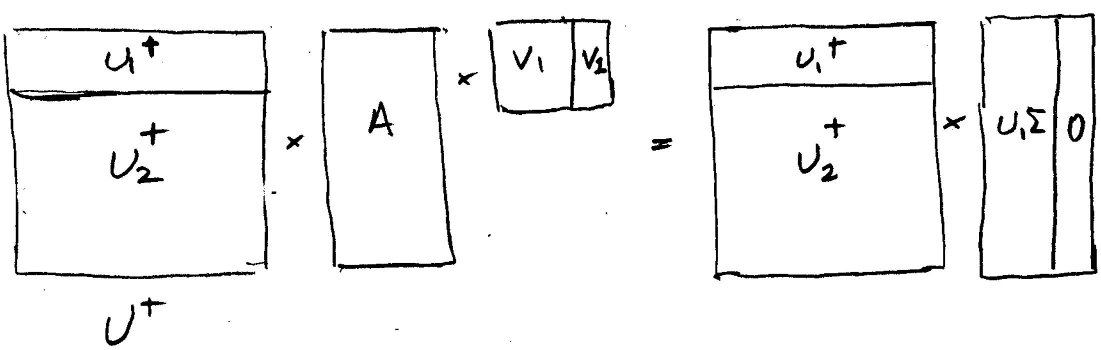
so the W columns of U_1 (K-vectors) are an orthonormal set, and are scaled images of V_1

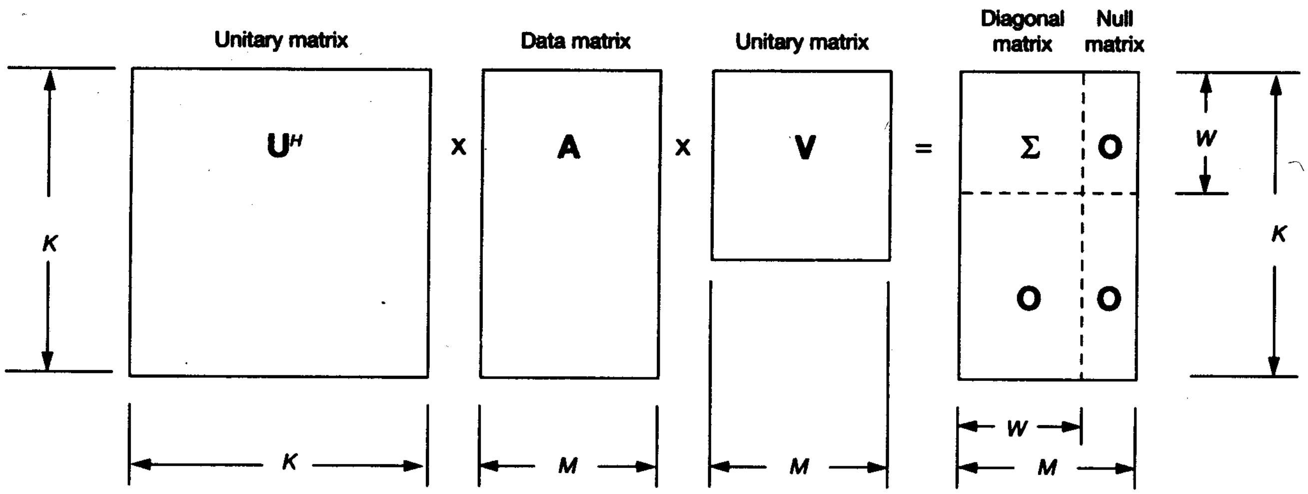
– now



– choose any K–W orthonormal vectors, orthog to U_1 cols (images of V_1) to fill out rest of U

$U = [U_1 \,|\, U_2]$

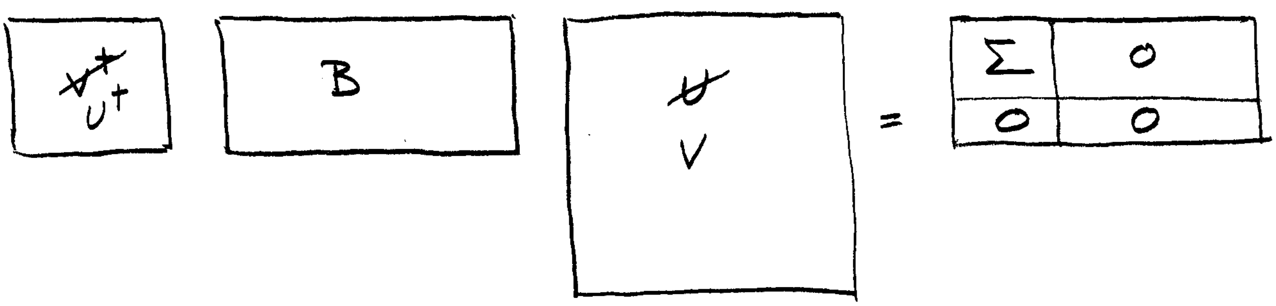




O : Null matrix

• That was for overdetermined. For underdetermined, don't derive again, just transpose.

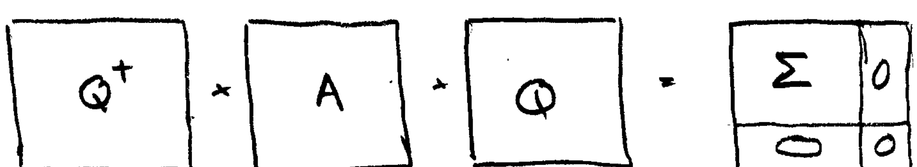
$$V^+ A^+ U = \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array}\right] \qquad \text{define } B = A^+, \quad V \to U, \quad U \to V$$



• In both cases, we have unitary matrices U, V taking A to an embedded diagonal form, a little like a similarity transform. When A is a square matrix,

$$U = V = Q \qquad M \times M$$

and



and Σ^2 is non zero eigen vals of $A^+ A$

- Recall for square matrices, the spectral decomposition of rank-1 matrices

$$A = \sum_{i=1}^{M} \sigma_i \, \underset{\sim}{q}_i \, \underset{\sim}{q}_i^{+}$$

$\underset{\sim}{q}_i$ eigenvecs of A
σ_i eigenvals of A

There is a counterpart for non square matrices.

Terminology: σ_i singular values (eigenvals if square)
V cols – right singular vectors (sometimes right eigen vecs)
U cols – left singular vectors

Recall

$$A \times \underbrace{\begin{bmatrix} V_1 & V_2 \end{bmatrix}}_{V} = \begin{bmatrix} U_1\Sigma & 0 \end{bmatrix}$$

or

$$A = \begin{bmatrix} U_1\Sigma & 0 \end{bmatrix} \times \begin{bmatrix} V_1^{+} \\ V_2^{+} \end{bmatrix}$$

so

$$A = \sum_{i=1}^{W} \sigma_i \, \underline{u}_i \, \underline{v}_i^{+} \qquad A^{+} = \sum_{i=1}^{W} \sigma_i \, \underline{v}_i \, \underline{u}_i^{+}$$

As in the case of square matrices, it is a sum of outer products of eigenvecs and their images

- Cols of U_1 span the range space of A $\underline{y}: \underline{y} = A\underline{x}$

Cols of V_2 span the null space of A $\underline{x}: \ A\underline{x} = \underline{0}$

- There is another common formulation of SVD

$$\underset{\substack{U^+ \\ M\times K}}{\begin{bmatrix} U_1^+ \\ \hline U_2^+ \end{bmatrix}} \times \underset{K\times M}{\begin{bmatrix} A \end{bmatrix}} \times \underset{\substack{V \\ M\times M}}{\begin{bmatrix} V_1 & \vert & V_2 \end{bmatrix}} = \underset{\substack{M\times M \\ \text{diagonal}}}{\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}}$$

It's neater in some ways, since singular value matrix is square, diagonal. Unfortunately, U is not square, though still unitary:

$$\begin{bmatrix} U_1^+ \\ \hline U_2^+ \end{bmatrix} \times \begin{bmatrix} U_1 & \vert & U_2 \end{bmatrix} = \begin{bmatrix} I \end{bmatrix}$$

To obtain this, don't build up U_2 to the full $K-W$ vectors — stop at $M-W$.

It diagonalizes both the Gram matrix A^+A and outer product AA^+:

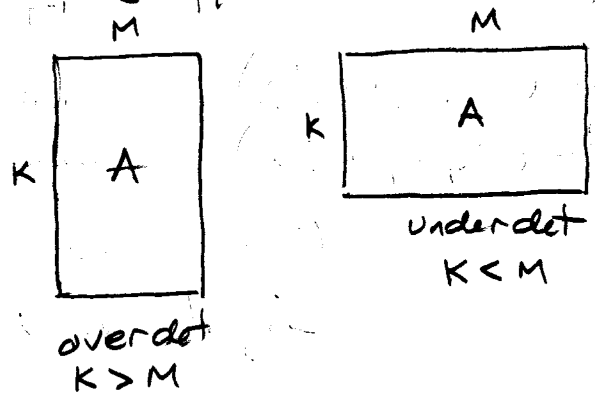
$$A = U \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array}\right] V^+$$

so $$A^+A = V \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array}\right] U^+ U \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array}\right] V^+ = V \underset{M\times M}{\left[\begin{array}{c|c} \Sigma^2 & 0 \\ \hline 0 & 0 \end{array}\right]} V^+$$

and $$AA^+ = U \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array}\right] V^+ V \left[\begin{array}{c|c} \Sigma & 0 \\ \hline 0 & 0 \end{array}\right] U^+ = \underset{K\times M}{U} \underset{M\times M}{\left[\begin{array}{c|c} \Sigma^2 & 0 \\ \hline 0 & 0 \end{array}\right]} \underset{M\times K}{U^+}$$

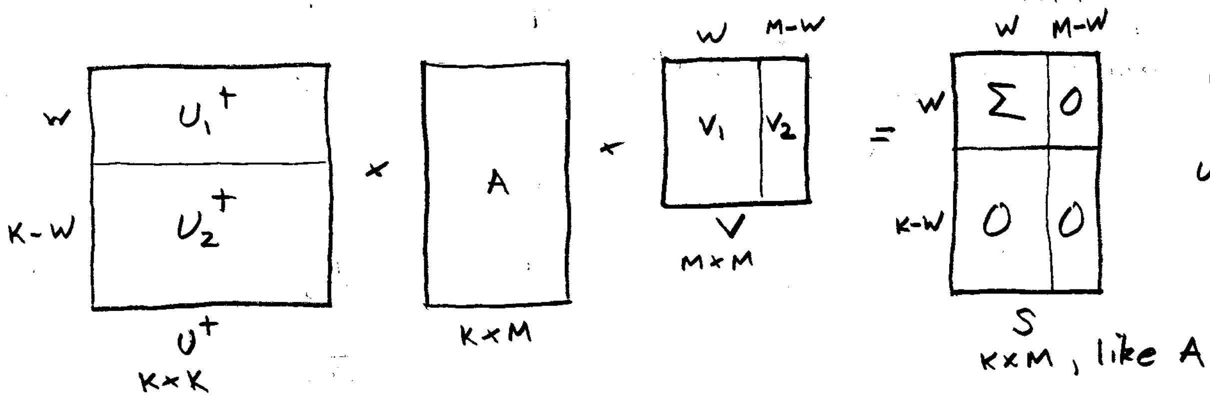
- What the SVD tells us (Haykin formulation)

Data matrix A



W, rank $\le \min(K,M)$: full rank if $W = \min(K,M)$, rank deficient if $W < \min(K,M)$

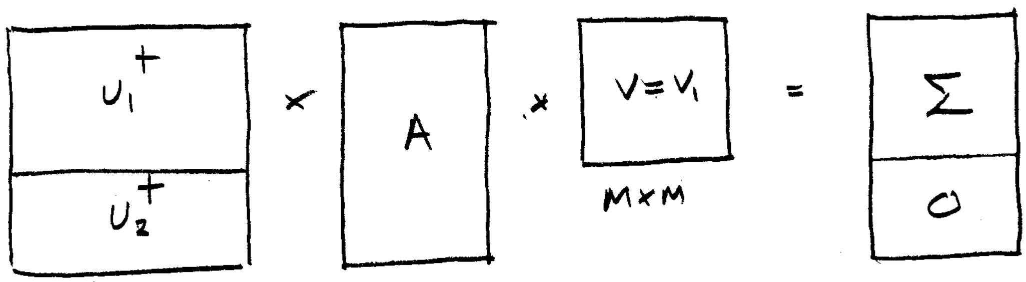
SVD for overdetermined:



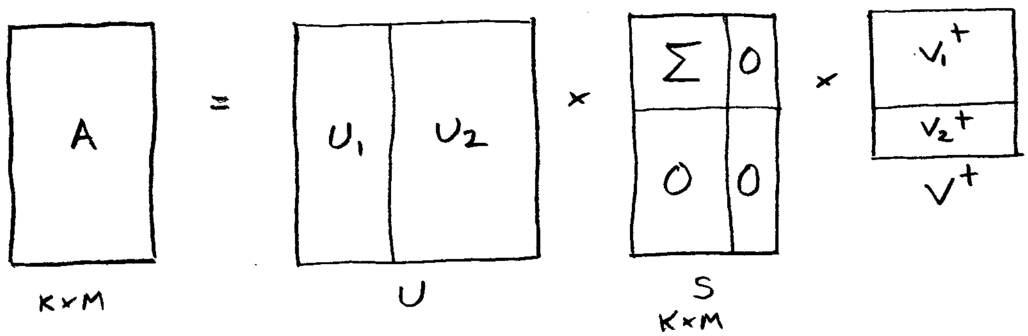
$$U^+ A V = S$$

$$U_1^+ A V_1 = \Sigma$$

and if full rank $W = M$:



and backwards (rank deficient)



$$A = U S V^+$$

$$= U_1 \Sigma V_1^+$$

SVD for underdetermined, rank deficient

$$\begin{array}{r} W \\ K-W \end{array}\underbrace{\begin{bmatrix} U_1^+ \\ \hline U_2^+ \end{bmatrix}}_{\substack{U^+ \\ K\times K}} \times \underbrace{\begin{bmatrix} A \end{bmatrix}}_{K\times M} \times \begin{array}{cc} W & M-W \\ \end{array}\begin{bmatrix} V_1 & V_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}}_{S,\ K\times M \text{ like } A}$$

$$U^+ A V = S$$
$$U_1^+ A V_1 = \Sigma$$

and for full (row) rank

$$W{=}K\ \underbrace{\begin{bmatrix} U = U_1 \end{bmatrix}}_{K\times K} \times \begin{bmatrix} A \end{bmatrix} \times \begin{array}{cc} W{=}K & M-W \end{array}\begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} \Sigma & 0 \end{bmatrix}$$

and backwards (rank deficient)

$$\underbrace{\begin{bmatrix} A \end{bmatrix}}_{K\times M} = \underbrace{\begin{bmatrix} U_1 & U_2 \end{bmatrix}}_{\substack{U \\ K\times K}} \times \underbrace{\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}}_{S,\ K\times M} \times \underbrace{\begin{bmatrix} V_1^+ \\ \hline V_2^+ \end{bmatrix}}_{M\times M}$$

$$A = U S V^+$$
$$= U_1 \Sigma V_1^+$$

For both over and under determined we have $A = U_1 \Sigma V_1^+$, where

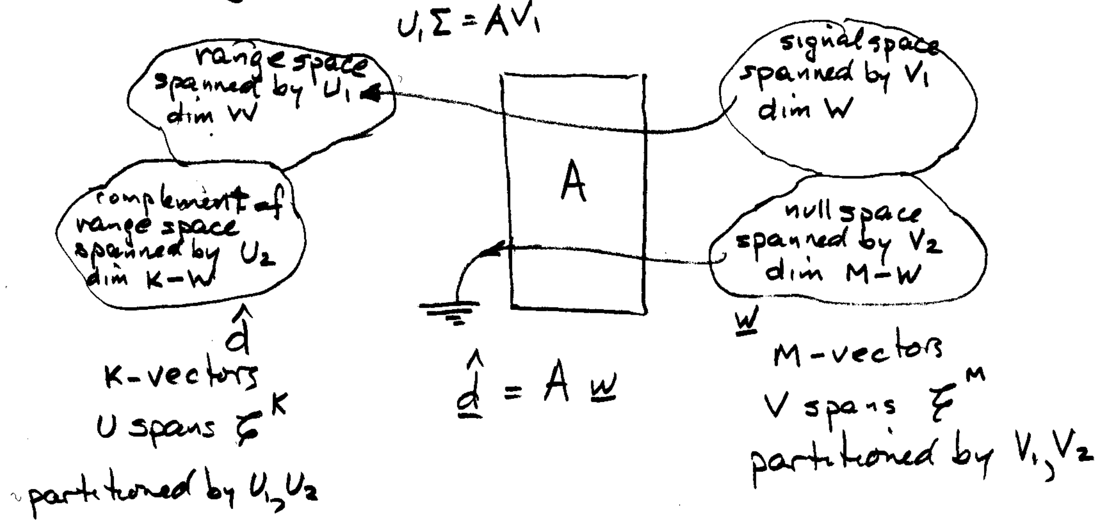
overdet: the cols of V_1 are the W eigenvecs of A^+A with nonzero eigenvals

underdet: the cols of U_1 are the W eigenvecs of AA^+ with non zero eigenvals

And from $A = U_1 \Sigma V_1^{+}$

we have $A = \sum_{i=1}^{W} \sigma_i \underline{u}_i \underline{v}_i^{+}$ $\qquad A^{+} = \sum_{i=1}^{W} \sigma_i \underline{v}_i \underline{u}_i^{+}$

- This gives us a fascinating interpretation:



For any input vector $\underline{w}$

$$A\underline{w} = \sum_{i=1}^{W} \sigma_i \underline{u}_i \underbrace{\underline{v}_i^{+}\underline{w}}_{\substack{\text{component} \\ \omega_i}} = \sum_{i=1}^{W} \sigma_i \omega_i \underline{u}_i = \hat{\underline{d}}$$

i.e. calculate the components of $\underline{w}$ with respect to the signal space basis vectors $\underline{v}_i$, $i = 1,.. W$ (cols of V_1). Each gets scaled by its σ_i to become the coefficient of the image $\underline{u}_i$ of $\underline{v}_i$.

Together, $\hat{\underline{d}} = U_1 \Sigma V_1^{+} \underline{w} = A\underline{w}$

Any part of $\underline{w}$ lying in the null space has a zero singular value and does not appear at the output

$$\underline{w}_{null} = V_2 (V_2^{+} \underline{w}), \quad A\underline{w}_{null} = \underline{0}$$