Singular Value Decomposition

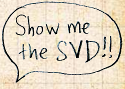
Matrixology (Linear Algebra)—Episode 25/25
MATH 124, Spring, 2015

Prof. Peter Dodds

Dept. of Mathematics & Statistics | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont

Show me the SVD!!

These slides are brought to you by:

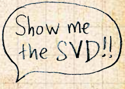


Outline

The Fundamental Theorem of Linear Algebra

Hubs and Authorities

Approximating matrices with SVD



Fundamental Theorem of Linear Algebra

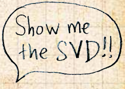
- Applies to any $m \times n$ matrix A.
- Symmetry of A and A^T.

Where $\vec{x}$ lives:

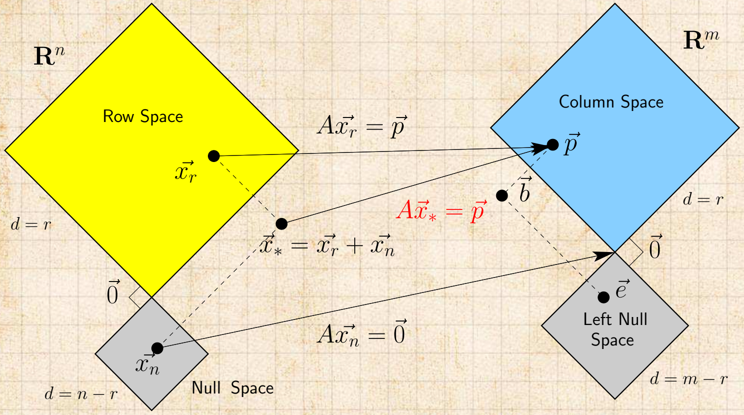
- Row space $C(A^T) \subset R^n$.
- (Right) Nullspace $N(A) \subset R^n$.
- dim $C(A^T)$ + dim $N(A)$ = $r + (n - r) = n$
- Orthogonality: $C(A^T) \bigotimes N(A) = R^n$

Where $\vec{b}$ lives:

- Column space $C(A) \subset R^m$.
- Left Nullspace $N(A^T) \subset R^m$.
- dim $C(A)$ + dim $N(A^T)$ = $r + (m - r) = m$
- Orthogonality: $C(A) \bigotimes N(A^T) = R^m$



Best solution $\vec{x}_*$ when $\vec{b} = \vec{p} + \vec{e}$:



Show me the SVD!!

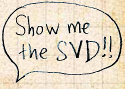
Fundamental Theorem of Linear Algebra

Now we see:

- Each of the four fundamental subspaces has a 'best' orthonormal basis
- The $\hat{v}_i$ span R^n
- We find the $\hat{v}_i$ as eigenvectors of $A^T A$.
- The $\hat{u}_i$ span R^m
- We find the $\hat{u}_i$ as eigenvectors of AA^T.

Happy bases

- $\{\hat{v}_1, \ldots, \hat{v}_r\}$ span Row space
- $\{\hat{v}_{r+1}, \ldots, \hat{v}_n\}$ span Null space
- $\{\hat{u}_1, \ldots, \hat{u}_r\}$ span Column space
- $\{\hat{u}_{r+1}, \ldots, \hat{u}_m\}$ span Left Null space



Fundamental Theorem of Linear Algebra

How $A\vec{x}$ works:

- $$\boxed{A\hat{v}_i = \sigma_i \hat{u}_i} \text{ for } i = 1, \ldots, r.$$

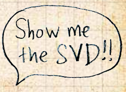
  and

  $$\boxed{A\hat{v}_i = \hat{0}} \text{ for } i = r+1, \ldots, n.$$

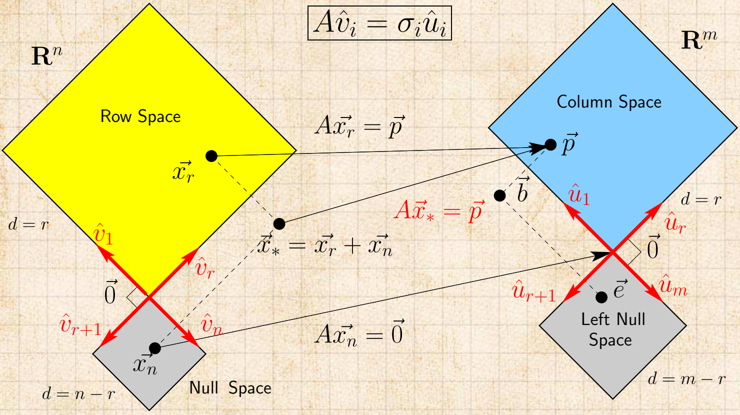
- Matrix version:

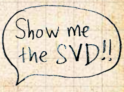
  $$\boxed{A = U\Sigma V^T}$$

- A sends each $\hat{v}_i \in C(A^T)$ to its partner $\hat{u}_i \in C(A)$ with a positive stretch/shrink factor $\sigma_i > 0$.
- A is diagonal with respect to these bases.
- When viewed in the right way, every A is a diagonal matrix Σ.



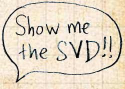
The complete big picture:





Hubs and Authorities

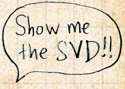
- Idea: allow nodes in a knowledge network to have two attributes:
  1. Authority: how much knowledge, information, etc., held by a node on a topic.
  2. Hubness (or Hubosity or Hubbishness or Hubtasticness): how well a node 'knows' where to find information on a given topic.
- Original work due to the legendary Jon Kleinberg.
- Best hubs point to best authorities.
- Recursive: Hubs authoritatively link to hubs, authorities hubbishly link to other authorities.
- More: look for dense links between sets of 'good' hubs pointing to sets of 'good' authorities.
- Known as the HITS algorithm (Hyperlink-Induced Topics Search).



Hubs and Authorities

- Give each node two scores:
  1. x_i = authority score for node i
  2. y_i = hubtasticness score for node i
- We connect the scores of neighboring nodes.
- I: a good authority is linked to by good hubs.
- Means x_i should increase as $\sum_{j=1}^{N} a_{ji} y_j$ increases.
- Note: indices are ji meaning j has a directed link to i.
- II: good hubs point to good authorities.
- Means y_i should increase as $\sum_{j=1}^{N} a_{ij} x_j$ increases.
- Linearity assumption:

$$\vec{x} \propto A^T \vec{y} \text{ and } \vec{y} \propto A \vec{x}$$



Hubs and Authorities

- So let's say we have

$$\vec{x} = c_1 A^T \vec{y} \text{ and } \vec{y} = c_2 A \vec{x}$$

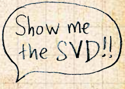
where c_1 and c_2 must be positive.

- Above equations combine to give

$$\vec{x} = c_1 A^T c_2 A \vec{x} = \lambda A^T A \vec{x}.$$

where $\lambda = c_1 c_2 > 0$.

- It's all good: we have the heart of singular value decomposition before us...



We can do this:

- $A^T A$ is symmetric.
- $A^T A$ is semi-positive definite so its eigenvalues are all ≥ 0.
- $A^T A$'s eigenvalues are the square of A's singular values.
- $A^T A$'s eigenvectors form a joyful orthogonal basis.
- The splendid Perron-Frobenius theorem tells us that only the dominant eigenvalue's eigenvector can be chosen to have non-negative entries.
- So: linear assumption leads to a solvable system.
- What would be very good: find networks where we have independent measures of node 'importance' and see how importance is actually distributed.

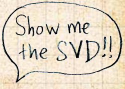


Image approximation (80x60)

Idea: use SVD to approximate images

- Interpret elements of matrix A as color values of an image.
- Truncate series SVD representation of A:

$$A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i \hat{u}_i \hat{v}_i^T$$

- Use fact that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$.
- Rank $r \leq$ min(m, n).
- Rank $r \leq$ # of pixels on shortest side.
- For color: approximate 3 matrices (RGB).