

# Scientific Computing for Biologists

## Singular Value Decomposition and Biplots

Instructor: Paul M. Magwene

02 October 2012

# Overview of Lecture

- Singular Value Decomposition
  - Algebra of SVD
  - Geometry of SVD
  - Relationship to Eigendecomposition
  - Applications of SVD
- Biplots
  - Simultaneous representation of rows and columns of a matrix

# Hands-on Session

- SVD and Biplots in R
- Applications of SVD in R
  - 'Seriation' using SVD
  - Matrix approximation and image compression using SVD

# Matrix Decomposition

A matrix decomposition is a factoring of a matrix into simpler parts.

Some familiar factorizations:

$$12 = 3 \times 4 = 2 \times 6$$

$$x^2 - 4 = (x - 2)(x + 2)$$

Matrix decomposition is the same idea!

# Eigendecomposition

You've already been introduced to one way to decompose a square matrix,  $A$ :

$$A = VDV^{-1}$$

where:

- $V$  is a matrix of eigenvectors (in columns)
- $D$  is a diagonal matrix with eigenvalues along diagonal.

when  $A$  is real-valued and symmetric then  $V$  is orthogonal.

# Singular Value Decomposition

$$A = U S V^T$$

$n \times p$

$n \times n$   
left  
singular  
vectors

$n \times p$   
singular  
values

$p \times p$   
right  
singular  
vectors

- When written like this  $U$  and  $V$  are orthonormal.
- Sometimes SVD is written as:

$$\underset{(n \times p)}{A} = \underset{(n \times p)}{U} \underset{(p \times p)}{S} \underset{(p \times p)}{V^T}$$

## Facts about SVD

- Singular Value Decomposition is often referred to as giving the “basic structure” of a matrix
- The rank of  $A$  is equivalent to the number of non-zero singular values in  $A = USV^T$

$$\text{rank}(A) \leq \min(n, p)$$

- The Euclidean norm ( $L_2$ ) norm of a matrix is the relative amount it stretches a vector:

$$|A|_E = \frac{|Ax|}{|x|}$$

The  $L_2$  norm of  $A$  is given by  $S_{11}$ .

# Geometric Interpretation of SVD

Any matrix,  $A_{n \times p}$ , represents a linear transformation from  $\mathbb{R}^p \mapsto \mathbb{R}^n$ .

SVD can be thought of decomposing the transformation specified by  $A$  into a simple form:

$$A = (\text{rotation})(\text{scaling})(\text{rotation})$$

- $U$  and  $V$  are orthonormal (orthogonal) matrices  $\leadsto$  Orthonormal matrices represent rigid rotations (or rotation plus reflection).
- Diagonal matrices represent “stretching”



# SVD Example

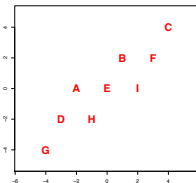
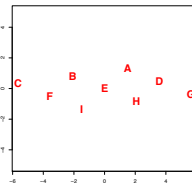
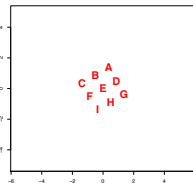
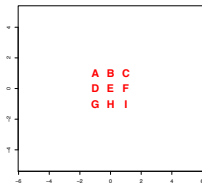
$$A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} = U S V^T$$

where

$$U = \begin{bmatrix} -0.75 & -0.66 \\ -0.66 & 0.75 \end{bmatrix}, S = \begin{bmatrix} 4.13 & 0 \\ 0 & 0.97 \end{bmatrix}, V^T = \begin{bmatrix} -0.86 & -0.50 \\ 0.50 & -0.86 \end{bmatrix}$$

## Geometry

$$x \xrightarrow{\text{rotation (+ reflection)}} V^T x \xrightarrow{\text{scaling}} S V^T x \xrightarrow{\text{rotation (+ reflection)}} U S V^T x$$



# Relationship of SVD to Eigendecomposition

$$A = USV^T \quad (1)$$

$$A^T A = (VSU^T)(USV^T) \quad (2)$$

$$= VSU^T USV^T \quad (3)$$

$$= VSSV^T \quad (4)$$

Equation 4 follows from the fact that  $U$  is orthonormal ( $U^T U = I$ )

If we let  $D = SS$ , we can rewrite equation 4 as:

$$A^T A = VDV \quad \text{Eigendecomposition!} \quad (5)$$

- The singular values  $S_{ii}$  are  $\sqrt{D_{ii}}$  where  $d_{ii}$  are the eigenvalues of  $A^T A$ .
- The columns of  $V$  are the eigenvectors of  $A^T A$ .

## Using SVD to do PCA

Let  $X$  be a mean-centered  $n \times p$  data matrix. The covariance matrix is given by:

$$C = \frac{1}{n-1} X^T X$$

By SVD we can write  $X = USV^T$ , therefore:

$$\begin{aligned} C &= \frac{1}{n-1} VSU^T USV^T \\ &= \frac{1}{n-1} VSSV^T \end{aligned}$$

- The PC vectors are given by the columns of  $V$  (rows of  $V^T$ )
- The PC scores are given by  $UD$ , where  $D = SS$

# Another Way of Thinking about SVD

Observations in  
measurement space

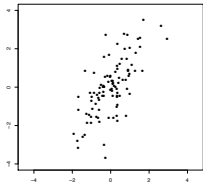
=

Observations  
in PC Space

← rotation

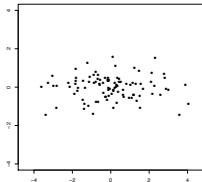
Inverse of  
Eigenvectors

$X$



=

$US$



$V^T$

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

where  $\theta \approx \frac{\pi}{3}$

# Applications of SVD

- Pseudoinverse of an arbitrary matrix
- Matrix approximation
- Biplots
- ... and many more ...

## Pseudoinverse via SVD

The pseudoinverse of a matrix is a generalization of the concept of a matrix inverse. Only square matrices have a matrix inverse; the pseudoinverse applies to an arbitrary  $n \times p$  matrix.

Given an  $n \times p$  matrix  $A$  find matrix  $A^+$  such that:

$$\begin{aligned}AA^+A &= A \\A^+AA^+ &= A^+ \\(AA^+)^T &= AA^+ \\(A^+A)^T &= A^+A\end{aligned}$$

Moore-Penrose Inverse via SVD:

$$\begin{aligned}\text{if } A &= USV^T \\A^+ &= VS^+U^T\end{aligned}$$

where  $S^+$  has the reciprocal of non-zero elements of  $S$ .

# SVD for Matrix Approximation

If  $A = USV^T$  then the optimal (least-squares)  $k$ -dimensional approximation of  $A$  (where  $k < \text{rank}(A)$ ) is given by:

$$\tilde{A} = US^*V^T$$

where:

$$\begin{aligned} S_{ii}^* &= S_{ii} \text{ for } i \leq k \\ S_{ii}^* &= 0 \text{ for } i > k \end{aligned}$$

# Biplots

- Technique for simultaneously displaying row and column data (observations and variables)
- Invented by K. Gabriel (see also papers by Gower)

Given a data matrix,  $X$ , we can use SVD to approximate  $X$  as so:

$$\tilde{X}_k = US^*V^T$$

(k-dimensional approximation to  $X$ )

We can rewrite  $\tilde{X}_k$  as a product of two matrices:

$$X = GH^T$$

where

$$G = U(S^*)^\alpha \quad \text{and} \quad H^T = (S^*)^{1-\alpha}V^T$$



## Biplots, cont.

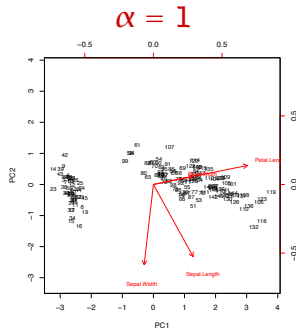
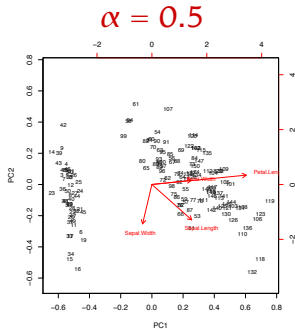
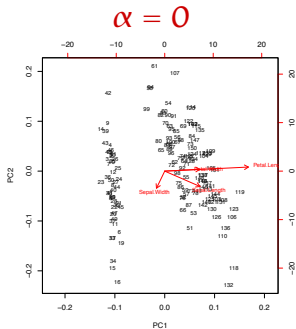
$$\begin{aligned}G &= \mathbf{U}(\mathbf{S}^*)^\alpha \text{ (row effects)} \\ \mathbf{H}^T &= (\mathbf{S}^*)^{1-\alpha} \mathbf{V}^T \text{ (columns effects)}\end{aligned}$$

Different choices of  $\alpha$  emphasize different relationships in the data.

- $\alpha = 0$ , column-metric preserving biplot; optimally approximates variance-covariance structure. Cosine of angles between vectors approximate correlations; distances between points approximate Mahalanobis distance ( “correlation biplot”)
- $\alpha = 1$ , row-metric preserving biplot; optimally approximates Euclidean distances among observations. Coordinates of observations correspond to PC scores; coordinates of variables correspond to eigenvector coefficients (“distance biplot”). PCs are ‘sphered’.
- $\alpha = 0.5$ , optimally approximates observational values (“symmetric biplot”)

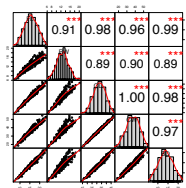
# Biplots, Example

- Observations drawn as points in space of PCs
- Variables drawn as vectors in PC space



PCA Biplots of Iris data set.

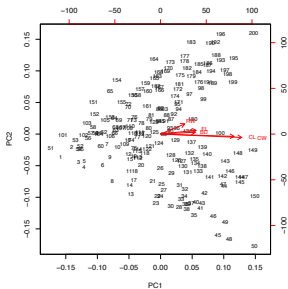
# Biplots, Example II



Crabs Dataset  
from MASS library

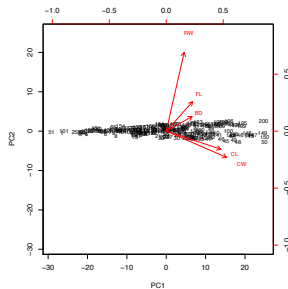
Column preserving

$$\alpha = 0$$



Row preserving

$$\alpha = 1$$



Biplots for dataset of five morphological measurements on crabs.