Matrix Decomposition
EVD - SVD

Characteristic Roots and Characteristics Vectors

Any nonzero vector x is said to be a characteristic vector of a matrix A, If there exist a number λ such that $Ax = \lambda x$;

Where A is a square matrix, also then λ is said to be a characteristic root of the matrix A corresponding to the characteristic vector x.

Characteristic root is unique but characteristic vector is not unique.

We calculate characteristics root λ from the characteristic equation $|A - \lambda I| = 0$
For $\lambda = \lambda_i$ the characteristics vector is the solution of x from the following homogeneous system of linear equation $(A - \lambda_i I)x = 0$

Theorem: If A is a real symmetric matrix and λ_i and λ_j are two distinct latent root of A then the corresponding latent vector x_i and x_j are orthogonal.

Multiplicity

Algebraic Multiplicity: The number of repetitions of a certain eigenvalue. If, for a certain matrix, $\lambda=\{3,3,4\}$, then the algebraic multiplicity of 3 would be 2 (as it appears twice) and the algebraic multiplicity of 4 would be 1 (as it appears once). This type of multiplicity is normally represented by the Greek letter α, where $\alpha(\lambda i)$ represents the algebraic multiplicity of λi.

Geometric Multiplicity: the *geometric multiplicity* of an eigenvalue is the number of linearly independent eigenvectors associated with it.

Jordan Decomposition

Camille Jordan (1870)

- Let A be any n×n matrix then there exists a nonsingular matrix P and $J_K(\lambda)$ a k×k matrix form

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}$$

Camille Jordan (1838-1921)

Such that

$$P^{-1}AP = \begin{bmatrix} J_{k_1}(\lambda_1) & 0 & \cdots & 0 \\ 0 & J_{k_2}(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & J_{k_r}(\lambda_r) \end{bmatrix}$$

where $k_1+k_2+ \ldots + k_r = n$. Also λ_i, $i=1,2,\ldots, r$ are the characteristic roots
And k_i are the algebraic multiplicity of λi ,

Jordan Decomposition is used in Differential equation and time series analysis.

Spectral Decomposition

A. L. Cauchy established the Spectral Decomposition in 1829.

CAUCHY, A.L.(1789-1857)

Let A be a $m \times m$ real symmetric matrix. Then there exists an orthogonal matrix P such that

$P^T A P = \Lambda$ or $A = P \Lambda P^T$, where Λ is a diagonal matrix.

Spectral Decomposition and Principal component Analysis

By using spectral decomposition we can write $A = P\Lambda P^T$

In multivariate analysis our data is a matrix. Suppose our data is X matrix. Suppose X is mean centered i.e., $X \to (X - \mu)$
and the variance covariance matrix is ∑. The variance covariance matrix ∑ is real and symmetric.

Using spectral decomposition we can write $\sum = P\Lambda P^T$. Where Λ is a diagonal matrix. $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$

Also

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

tr(∑) = Total variation of Data =tr(Λ)

Spectral Decomposition and Principal component Analysis (Cont.)

The Principal component transformation is the transformation

$$Y=(X-\mu)P$$

Where,

- ❖ $E(Y_i)=0$
- ❖ $V(Y_i)=\lambda_i$
- ❖ $Cov(Y_i, Y_j)=0$ if $i \neq j$
- ❖ $V(Y_1) \geq V(Y_2) \geq . . . \geq V(Y_n)$
- ❖ $\sum_{i=1}^{n} V(Y_i) = tr(\Sigma)$
- ❖ $\prod_{i=1}^{n} V(Y_i) = |\Sigma|$

R code for Spectral Decomposition

```
x<-matrix(c(1,2,3, 2,5,4, 3,4,9),ncol=3,nrow=3)
eigen(x)
```

Application:

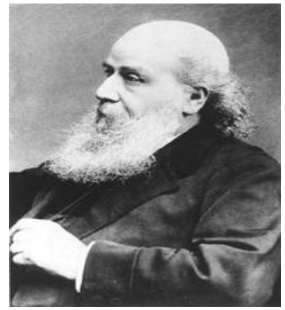
- For Data Reduction.
- Image Processing and Compression.
- K-Selection for K-means clustering
- Multivariate Outliers Detection
- Noise Filtering
- Trend detection in the observations.

Historical background of SVD

There are five mathematicians who were responsible for establishing the existence of the singular value decomposition and developing its theory.

Eugenio Beltrami
(1835-1899)

Camille Jordan
(1838-1921)



James Joseph
Sylvester
(1814-1897)

Erhard Schmidt
(1876-1959)

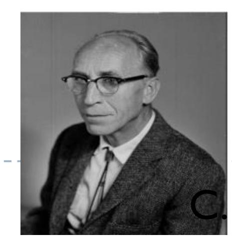
Hermann Weyl
(1885-1955)

The Singular Value Decomposition was originally developed by two mathematician in the mid to late 1800's

1. Eugenio Beltrami , 2.Camille Jordan

Several other mathematicians took part in the final developments of the SVD including James Joseph Sylvester, Erhard Schmidt and Hermann Weyl who studied the SVD into the mid-1900's.

C.Eckart and G.Young prove low rank approximation of SVD (1936).



C.Eckart

What is SVD?

Any real ($m \times n$) matrix X, where ($n \leq m$), can be decomposed,

$$X = U \Lambda V^T$$

- **U is a ($m \times n$) column orthonormal matrix ($U^T U = I$), containing the eigenvectors of the symmetric matrix XX^T.**
- **Λ is a ($n \times n$) diagonal matrix, containing the singular values of matrix X. The number of non zero diagonal elements of Λ corresponds to the rank of X.**
- **V^T is a ($n \times n$) row orthonormal matrix ($V^T V = I$), containing the eigenvectors of the symmetric matrix $X^T X$.**

Singular Value Decomposition

Theorem (Singular Value Decomposition) : Let X be $m \times n$ of rank r, $r \le n \le m$. Then there exist matrices U , V and a diagonal matrix Λ , with positive diagonal elements such that, $X = U \Lambda V^T$

Proof: Since X is $m \times n$ of rank r, $r \le n \le m$. So XX^T and X^TX both of rank r (by using the concept of Grammian matrix) and of dimension $m \times m$ and $n \times n$ respectively. Since XX^T is real symmetric matrix so we can write by spectral decomposition,

$$XX^T = QDQ^T$$

Where Q and D are respectively, the matrices of characteristic vectors and corresponding characteristic roots of XX^T.

Again since X^TX is real symmetric matrix so we can write by spectral decomposition, $X^T X = RMR^T$

https://markheckmann.github.io/notes/svd_image_compression.html

Singular Value Decomposition

Where R is the (orthogonal) matrix of characteristic vectors and M is diagonal matrix of the corresponding characteristic roots.

Since XX^T and X^TX are both of rank *r*, only *r* of their characteristic roots are positive, the remaining being zero. Hence we can write,

$$D = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$$

Also we can write,

$$M = \begin{bmatrix} M_r & 0 \\ 0 & 0 \end{bmatrix}$$

Singular Value Decomposition

We know that the nonzero characteristic roots of XX^T and X^TX are equal so $D_r = M_r$

Partition Q, R conformably with D and M, respectively

i.e., $Q = (Q_r,\ Q_*)$; $R = (R_r,\ R_*)$ such that Q_r is $m \times r$, R_r is $n \times r$ and correspond respectively to the nonzero characteristic roots of XX^T and X^TX. Now take

$$U = Q_r$$

$$V = R_r$$

$$D_r^{1/2} = \Lambda = diag(d_1^{\ 1/2}, d_2^{\ 1/2}, \cdots, d_r^{\ 1/2})$$

Where $d_i,\ \ i = 1,2,\cdots,r$ are the positive characteristic roots of XX^T and hence those of X^TX as well (by using the concept of grammian matrix.)

Singular Value Decomposition

Now define, $S = Q_r D_r^{1/2} R_r^{\ T}$

Now we shall show that *S*=*X* thus completing the proof.

$$
\begin{aligned}
S^T S &= (Q_r D_r^{\ 1/2} R_r^{\ T})^T Q_r D_r^{\ 1/2} R_r^{\ T} \\
&= R_r D_r^{1/2} Q_r^T Q_r D_r^{1/2} R_r^T \\
&= R_r D_r R_r^{\ T} \\
&= R_r M_r R_r^{\ T} \\
&= RMR^T \\
&= X^T X
\end{aligned}
$$

Similarly, $SS^T = XX^T$

From the first relation above we conclude that for an arbitrary orthogonal matrix, say P_1 ,

While from the second we conclude that for an arbitrary orthogonal matrix, say P_2

We must have $S = P_1 X$ $S = XP_2$

Singular Value Decomposition

The preceding, however, implies that for arbitrary orthogonal matrices P_1 , P_2 the matrix X satisfies

$$XX^T = P_1 XX^T P_1^T, \qquad X^T X = P_2^T X^T X P_2$$

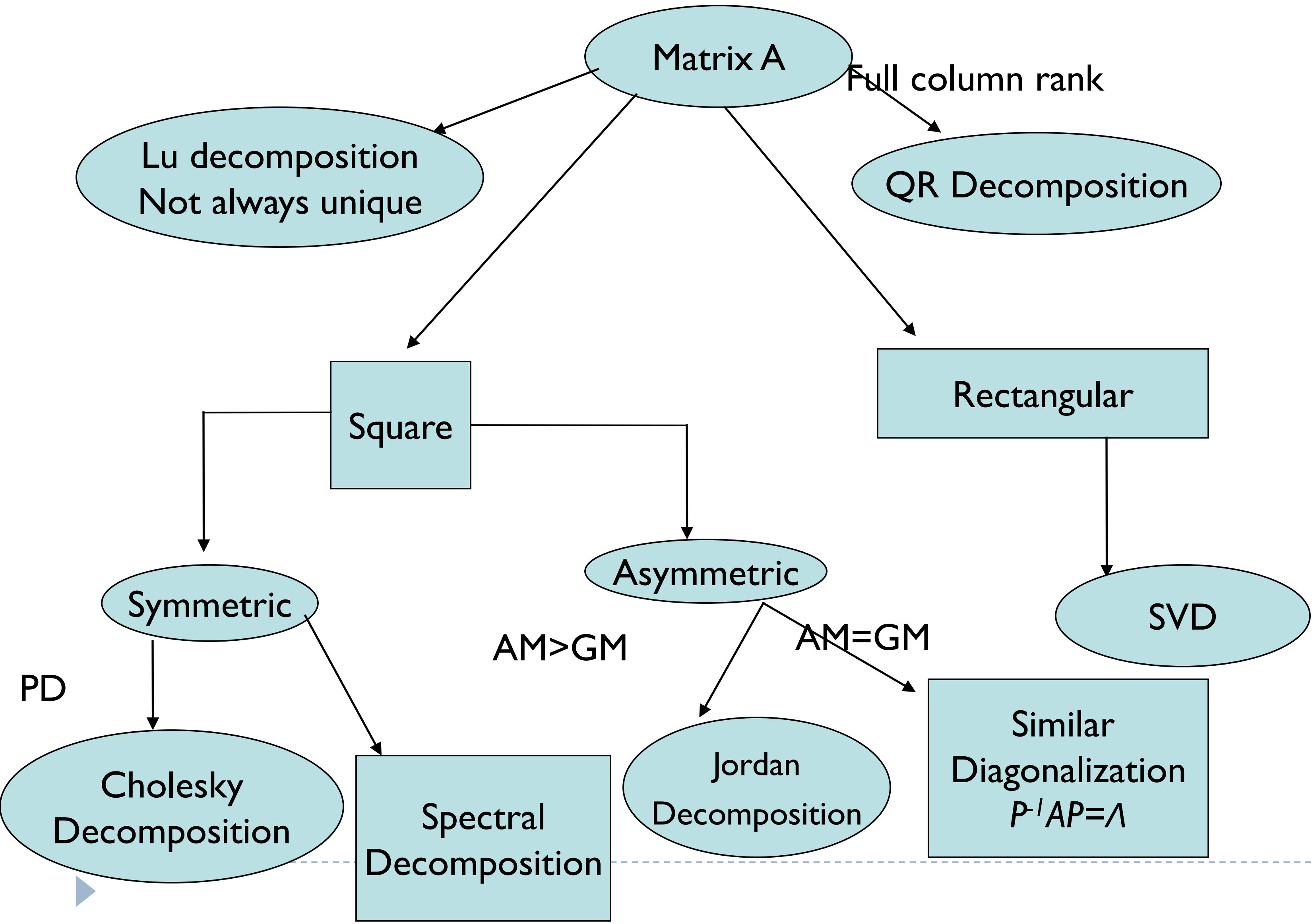
Which in turn implies that, $P_1 = I_m,$ $\qquad P_2 = I_n$

Thus $\qquad X = S = Q_r D_r^{1/2} R_r^T = U \Lambda V^T$

R Code for Singular Value Decomposition

```
x<-matrix(c(1,2,3, 2,5,4, 3,4,9),ncol=3,nrow=3)
sv<-svd(x)
D<-sv$d
U<-sv$u
V<-sv$v
```

Decomposition in Diagram



Properties Of SVD

Rewriting the SVD

$$A = U\Lambda V^T = \sum_{i=1}^{r} u_i \lambda_i v_i^T$$

where

r = rank of A

λ_i = the i-th diagonal element of Λ.

u_i and v_i are the i-th columns of U and V respectively.

Proprieties of SVD
Low rank Approximation

Theorem: If $A=U\Lambda V^T$ is the SVD of A and the singular values are sorted as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then for any l < r, the best rank-l approximation to A is

$$\widetilde{A} = \sum_{i=1}^{l} u_i \lambda_i v_i^T \quad ; \quad \left\| A - \widetilde{A} \right\|^2 = \sum_{i=l+1}^{r} \lambda_i^2$$

Low rank approximation technique is very much important for data compression.

Low-rank Approximation

- SVD can be used to compute optimal **low-rank approximations**.
- Approximation of A is Ã of rank ***k*** such that

$$\tilde{A} = \underset{X:rank(X)=k}{Min} \|A - X\|_F$$

Frobenius norm

$$\|A\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}$$

$$||A||^2 = tr\{AA^T\} = tr\{A^TA\}.$$

If $d_1, d_2, \cdots, d_n$ are the characteristics roots of A^TA then $\|A\|^2 = \sum_{i=1}^{n} d_i$

Ã and X are both $m \times n$ matrices.

Low-rank Approximation

- Solution via SVD

$$\tilde{A} = U \,\mathrm{diag}(\lambda_1, \ldots, \lambda_k, \underbrace{0, \ldots, 0}_{\text{set smallest r-k singular values to zero}})V^T$$

set smallest r-k
singular values to zero

$$\underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \bullet & & \\ & \bullet & \\ & & \bullet \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{V^T}$$

K=2

$$\tilde{A} = \sum_{i=1}^{k} \lambda_i u_i v_i^T$$

column notation: sum of rank 1 matrices

Approximation error

- How good (bad) is this approximation?
- It's the best possible, measured by the Frobenius norm of the error:

$$\min_{X:rank(X)=k} \|A - X\|^2_F = \|A - \tilde{A}\|_F^2 = \sum_{i=k+1}^{r} \lambda_i^2$$

- where the λ_i are ordered such that $\lambda_i \geq \lambda_{i+1}$.

Now $\|A - \tilde{A}\|_F^2$

$$\begin{aligned}
&= tr[(U\Lambda V^T - U\Lambda_k V^T)^T (U\Lambda V^T - U\Lambda_k V^T)] \\
&= tr[V(\Lambda - \Lambda_k)^T U^T U(\Lambda - \Lambda_k)V^T] \\
&= tr[V^T V(\Lambda - \Lambda_k)^T (\Lambda - \Lambda_k)] \\
&= tr[(\Lambda - \Lambda_k)^T (\Lambda - \Lambda_k)] \\
&= \sum_{i=k+1}^{min(m,n)} \lambda_i^2
\end{aligned}$$

Row approximation and column approximation

Suppose R_i and c_j represent the i-th row and j-th column of A. The SVD of A and $\tilde{A}$ is

$$A = U\Lambda V^T = \sum_{k=1}^{r} u_k \lambda_k v_k^T \qquad \tilde{A} = U_l \Lambda_l V_l^T = \sum_{k=1}^{l} u_k \lambda_k v_k^T$$

The SVD equation for R_i is

$$R_i = \sum_{k=1}^{r} u_{ik} \lambda_k v_k$$

We can approximate R_i by

$$R_i^l = \sum_{k=1}^{l} u_{ik} \lambda_k v_k \quad ; \quad l<r$$

where i = 1,…,m.

Also the SVD equation for Cj is, where j = 1, 2, …, n

$$C_j = \sum_{k=1}^{r} v_{jk} \lambda_k u_k$$

We can also approximate Cj by

$$C_j^l = \sum_{k=1}^{l} v_{jk} \lambda_k u_k \quad ; l<r$$

Least square solution in inconsistent system

By using SVD we can solve the inconsistent system. This gives the least square solution. $\min_{x} \|Ax - b\|^2$

The least square solution $x_{LS} = A_g b$

where A_g be the MP inverse of A.

The SVD of A_g is

$$A_g = V \Lambda_g U^T$$

This can be written as

$$A_g = \sum_{i=1}^{p} \lambda_i{}^* v_i u_i{}^T$$

Where

$$\lambda_i{}^* = \begin{cases} \lambda_i{}^{-1} & if \quad \lambda_i \neq 0 \\ 0 & if \quad \lambda_i = 0 \end{cases}$$

$X_{m\times n}$, $(\rho(X) = l)$

Seen as a set of vectors

Set of m row vectors X_{io} each row,(written as column) is a linear combination of l $\lambda_i V_i$'s with coefficients in vector equal to i^{th} row of U.

Each row, X_{io} (written as column) can be approximated by the first k $\lambda_i V_i$'s with coefficients vector equal to the first k values of i^{th} row of U.

Set of n column vectors X_{oj} each column is a linear combination of l $\lambda_j U_j$'s with coefficients in vector equal to j^{th} column of V^T.

Each column, X_{oj} can be approximated by the first k $\lambda_j U_j$'s with coefficients vector equal to the first k values of j^{th} column of V^T.

Seen as a matrix of order m×n

A linear combination of l B_i matrices of order m×n each of them has rank 1, with coefficients $\lambda_1 \geq \lambda_2 \ldots\ldots \geq \lambda_l$; $B_i = U_i V_i^T$

$X = \sum_{i=1}^{l} \lambda_i B_i$, can be approximated by $\tilde{X}(\rho(\tilde{X}) = k < l), \tilde{X} = \sum_{i=1}^{k} \lambda_i B_i$ with error $\left\| X - \tilde{X} \right\|^2 = \sum_{i=k+1}^{l} \lambda_i^2$

It is nothing but a projection of X on the k dimensional space generated by k B_i's, projection matrix

$$U_k (U_k^T U_k)^{-1} U_k^T$$

Basic Results of SVD

SVD based PCA

If we reduced variable by using SVD then it performs like PCA.

Suppose X is a mean centered data matrix, Then
X using SVD, $X=U\Lambda V^T$

we can write- $XV = U\Lambda$
Suppose $Y = XV = U\Lambda$
Then the first columns of Y represents the first principal component score and so on.

- SVD Based PC is more Numerically Stable.
- If no. of variables is greater than no. of observations then SVD based PCA will give efficient result(Antti Niemistö, Statistical Analysis of Gene Expression Microarray Data,2005)

Application of SVD

- Data Reduction both variables and observations.
- Solving linear least square Problems
- Image Processing and Compression.
- K-Selection for K-means clustering
- Multivariate Outliers Detection
- Noise Filtering
- Trend detection in the observations and the variables.

Origin of biplot

Prof. Ruben Gabriel, "The founder of biplot"
Courtesy of Prof. Purificación Galindo
University of Salamanca, Spain

- Gabriel (1971)
- One of the most important advances in data analysis in recent decades
- Currently…
  - > 50,000 web pages
  - Numerous academic publications
  - Included in most statistical analysis packages
- Still a very new technique to most scientists

What is a biplot?

- “Biplot” = “bi” + “plot”
  - “plot”
    - scatter plot of two rows OR of two columns, or
    - scatter plot summarizing the rows *OR* the columns
  - “bi”
    - BOTH rows AND columns
- 1 biplot >> 2 plots

Practical definition of a biplot

"Any two-way table can be analyzed using a 2D-biplot as soon as it can be sufficiently approximated by a rank-2 matrix." (Gabriel, 1971)

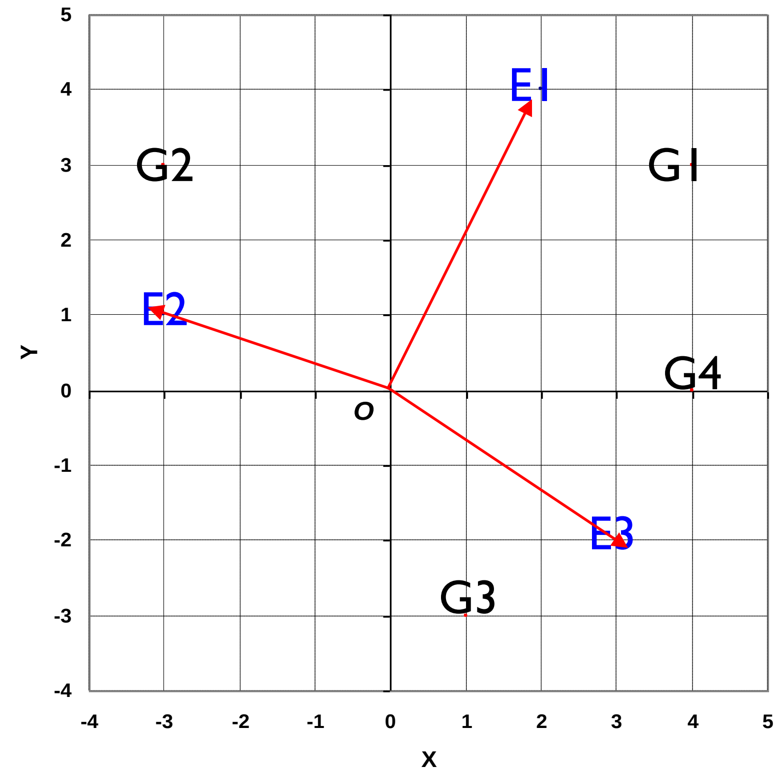
(Now 3D-biplots are also possible…)

Matrix decomposition

P(4, 3) G(3, 2) E(2, 3)

$$
\begin{array}{c|ccc}
 & e1 & e2 & e3 \\
\hline
g1 & 20 & -9 & 6 \\
g2 & 6 & 12 & -15 \\
g3 & -10 & -6 & 9 \\
g4 & 8 & -12 & 12
\end{array}
\longrightarrow
\begin{array}{c|cc}
 & x & y \\
\hline
g1 & 4 & 3 \\
g2 & -3 & 3 \\
g3 & 1 & -3 \\
g4 & 4 & 0
\end{array}
\times
\begin{array}{c|ccc}
 & e1 & e2 & e3 \\
\hline
x & 2 & -3 & 3 \\
y & 4 & 1 & -2
\end{array}
$$

G-by-E table



Singular Value Decomposition (SVD) & Singular Value Partitioning (SVP)

The 'rank' of Y, i.e., the minimum number of PC required to fully represent Y

Matrix characterising the rows

"Singular values"

Matrix characterising the columns

SVD: $$X_{ij} \xrightarrow{SVD} \sum_{k=1}^{r} u_{ik}\boxed{\lambda_k}v_{kj}$$

Common uses value of f

SVP: $$\xrightarrow{SVP} \sum_{k=1}^{r} \left(\underline{u_{ik}\boxed{\lambda_k^{f}}}\right)\left(\underline{\boxed{\lambda_k^{1-f}}v_{kj}}\right)$$

f=1

f=0

f=1/2

Rows scores — Column scores

Plot — Biplot — Plot

Biplot

- The simplest biplot is to show the first two PCs together with the projections of the axes of the original variables
- x-axis represents the scores for the first principal component
- Y-axis the scores for the second principal component.
- The original variables are represented by arrows which graphically indicate the proportion of the original variance explained by the first two principal components.
- The direction of the arrows indicates the relative loadings on the first and second principal components.

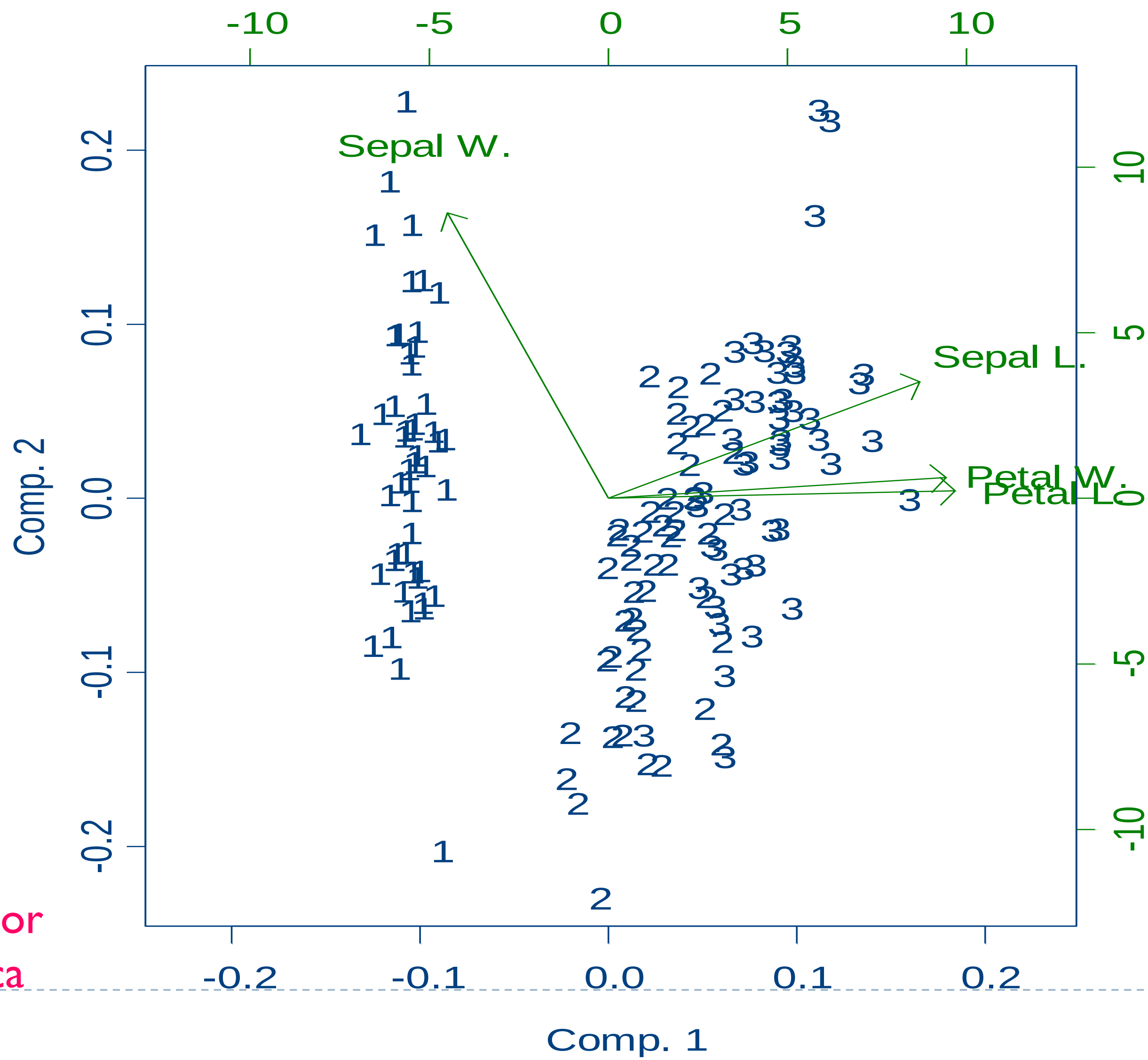
- Biplot analysis can help to understand the multivariate data

i) Graphically

ii) Effectively

iii) Conveniently.

Biplot of Iris Data



1= Setosa
2= Versicolor
3= Virginica

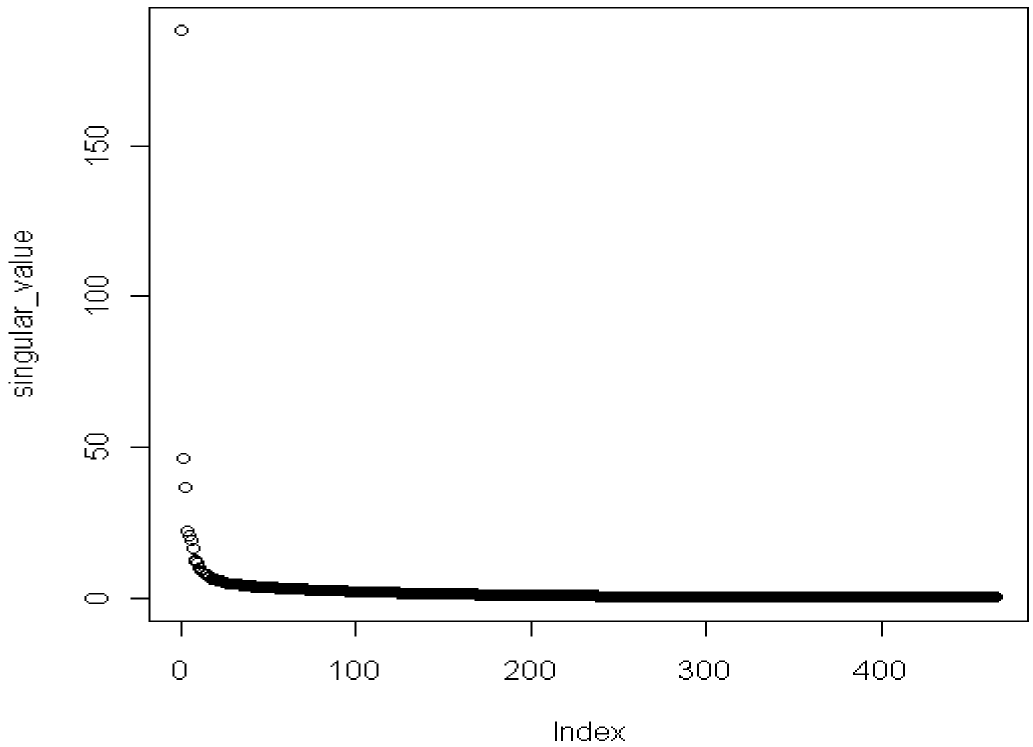
Image Compression Example

Pansy Flower image, collected from

http://www.ats.ucla.edu/stat/r/code/pansy.jpg

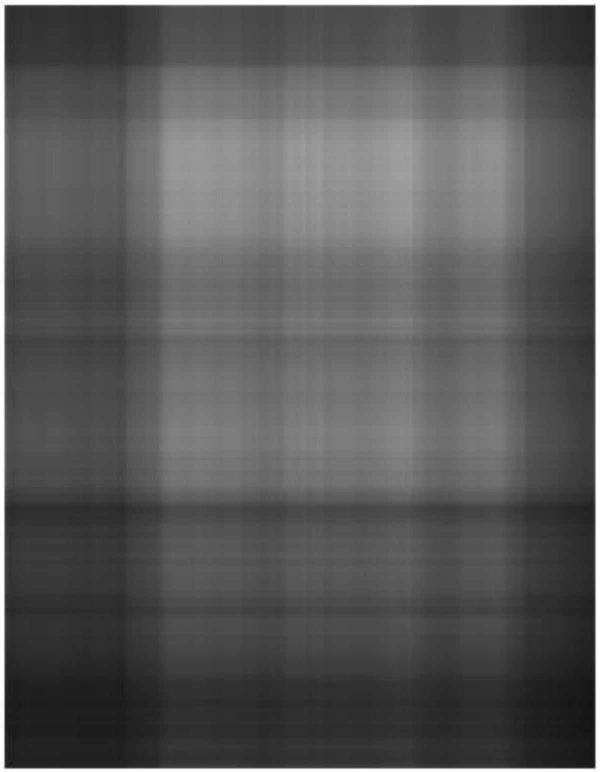
This image is 600×465 pixels

Singular values of flowers image

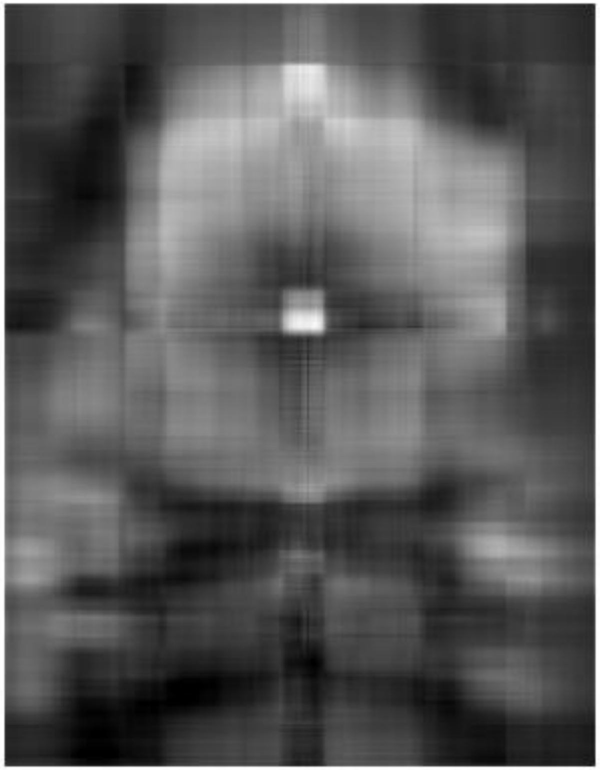


Plot of the singular values

Low rank Approximation to flowers image



Rank-1 approximation



Rank- 5 approximation

Low rank Approximation to flowers image

Rank-20 approximation

Rank-30 approximation

Low rank Approximation to flowers image

Rank-50 approximation

Rank-80 approximation

Low rank Approximation to flowers image

Rank-100 approximation

Rank-120 approximation

Low rank Approximation to flowers image

Rank-150 approximation

True Image

Outlier Detection Using SVD

Nishith and Nasser (2007,MSc. Thesis) propose a graphical method of outliers detection using SVD.

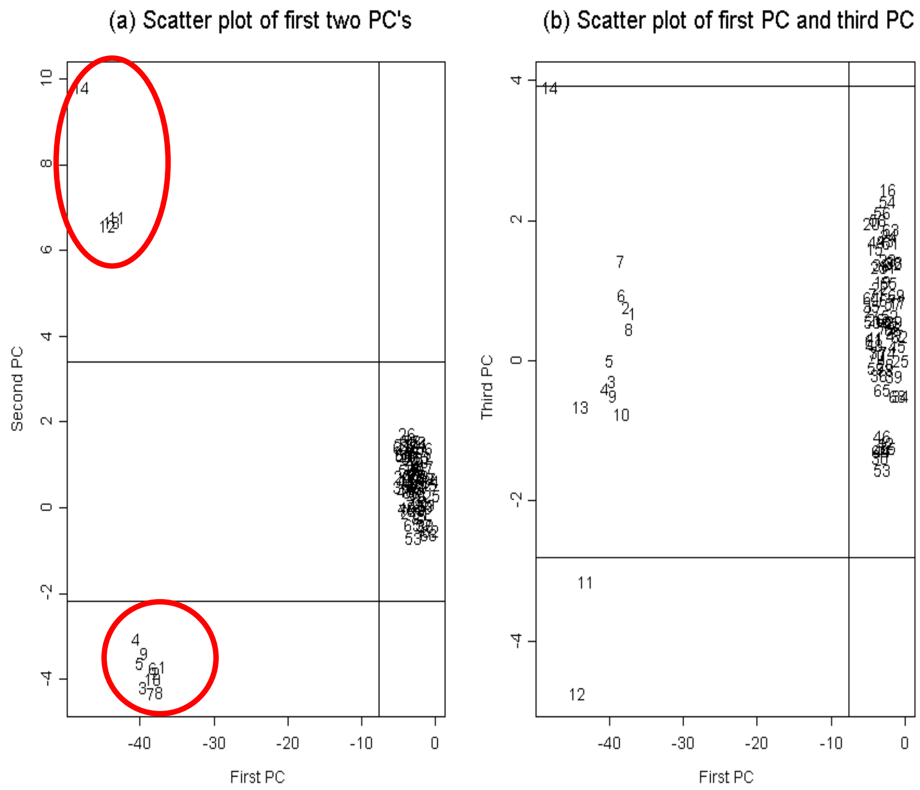
It is suitable for both general multivariate data and regression data. For this we construct the scatter plots of first two PC's, and first PC and third PC. We also make a box in the scatter plot whose range lies

median(1*stPC*) ± 3 × *mad*(1*stPC*) in the *X*-axis and *median*(2*ndPC*/3*rdPC*) ± 3 × *mad*(2*ndPC*/3*rdPC*) in the *Y*-axis.

Where *mad* = median absolute deviation.

The points that are outside the box can be considered as extreme outliers. The points outside one side of the box is termed as outliers. Along with the box we may construct another smaller box bounded by 2.5/2 MAD line

Outlier Detection Using SVD (Cont.)

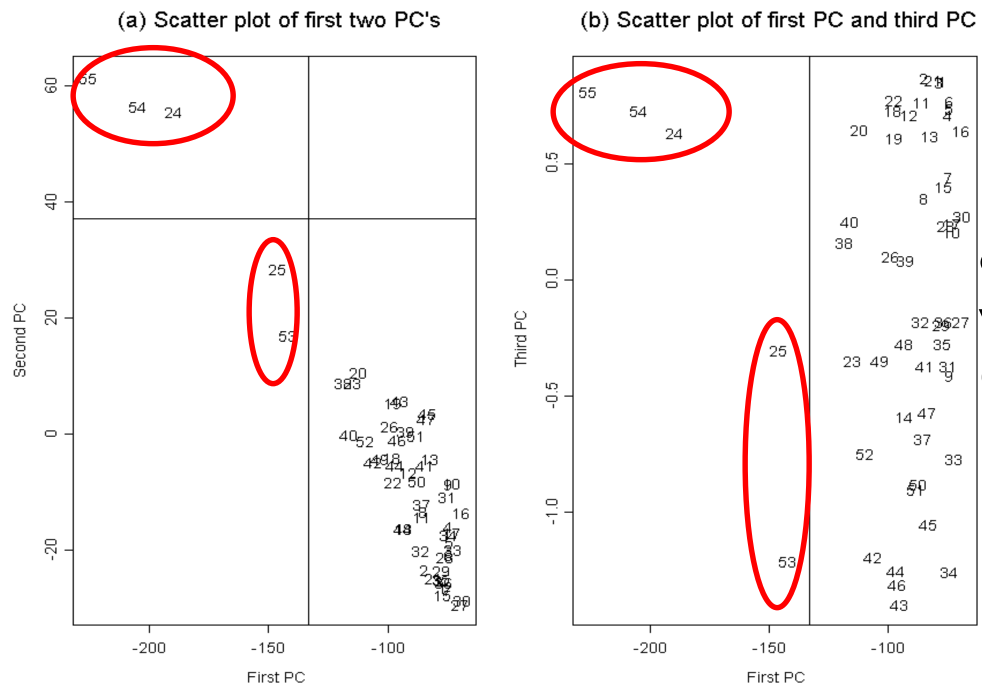


HAWKINS-BRADU-KASS (1984) DATA

Data set containing 75 observations with 14 influential observations. Among them there are ten high leverage outliers (cases 1-10) and for high leverage points (cases 11-14) -Imon (2005).

Scatter plot of Hawkins, Bradu and kass data (a) scatter plot of first two PC's and (b) scatter plot of first and third PC.

Outlier Detection Using SVD (Cont.)



Scatter plot of modified Brown data (a) scatter plot of first two PC's and (b) scatter plot of first and third PC.

MODIFIED BROWN DATA

Data set given by Brown (1980).

Ryan (1997) pointed out that the original data on the 53 patients which contains 1 outlier (observation number 24).

Imon and Hadi(2005) modified this data set by putting two more outliers as cases 54 and 55.
Also they showed that observations 24, 54 and 55 are outliers by using generalized standardized Pearson residual (GSPR)

Cluster Detection Using SVD

Singular Value Decomposition is also used for cluster detection (Nishith, Nasser and Suboron, 2011).

The methods for clustering data using first three PC's are given below,

median (1st PC) $\pm$ *k* $\times$ *mad (1st PC)* in the X-axis and *median (2nd PC/3rd PC)* $\pm$ *k* $\times$ *mad (2nd PC/3rd PC)* in the Y-axis.

Where *mad* = median absolute deviation. The value of *k* = 1, 2, 3.

Consequences of SVD

Generally many missing values may present in the data. It may also contain unusual observations. Both types of problem can not handle Classical singular value decomposition.

Robust singular value decomposition can solve both types of problems.

Robust singular value decomposition can be obtained by alternating L1 regression approach (Douglas M. Hawkins, Li Liu, and S. Stanley Young, (2001)).

The Alternating L1 Regression Algorithm for Robust Singular Value Decomposition.

There is no obvious choice of the initial values of u_1

Initialize the leading left singular vector u_1

Fit the L1 regression coefficient c_j by minimizing $\sum_{i=1}^{n}|x_{ij} - c_j u_{i1}|$; $j=1,2,\ldots,p$

Calculate right singular vector $\mathbf{v}_1 = c/\|c\|$, where $\|.\|$ refers to Euclidean norm.

Again fit the L1 regression coefficient d_i by minimizing $\sum_{j=1}^{p}|x_{ij} - d_i v_{j1}|$; $i=1,2,\ldots,n$

Calculate the resulting estimate of the left eigenvector $u_i = d/\|d\|$

Iterate this process untill it converge.

For the second and subsequent of the SVD, we replaced X by a deflated matrix obtained by subtracting the most recently found them in the SVD $X \leftarrow X-\lambda_k u_k v_k^T$

References

- Brown B.W., Jr. (1980). Prediction analysis for binary data. in *Biostatistics Casebook*, R.G. Miller, Jr., B. Efron, B.W. Brown, Jr., L.E. Moses (Eds.), New York: Wiley.
- Dhrymes, Phoebus J. (1984), *Mathematics for Econometrics,* 2nd ed. Springer Verlag, New York.
- Hawkins D. M., Bradu D. and Kass G.V.(1984),Location of several outliers in multiple regression data using elemental sets. *Technometrics*, 20, 197-208.
- Imon A. H. M. R. (2005). Identifying multiple influential observations in linear Regression. *Journal of Applied Statistics* 32, 73 – 90.
- Kumar, N. , Nasser, M., and Sarker, S.C., 2011. "A New Singular Value Decomposition Based Robust Graphical Clustering Technique and Its Application in Climatic Data" *Journal of Geography and Geology*, *Canadian Center of Science and Education* ,Vol-3, No. 1, 227-238.
- Ryan T.P. (1997). *Modern Regression Methods*, Wiley, New York.
- **Stewart, G.W. (1998). *Matrix Algorithms*, Vol 1. Basic Decompositions, Siam, Philadelphia.**
- Matrix Decomposition. http://fedc.wiwi.hu-berlin.de/xplore/ebooks/html/csa/node36.html

Power Method

Power Method

- Assume that the eigenvalues for an n x n matrix A can be ordered such that
  $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_{n-2}| \geq |\lambda_{n-1}| > |\lambda_n|$
- Then λ_1 is the dominant eigenvalue and $|\lambda_1|$ is the spectral radius of A, denoted $\rho(A)$
- The i^{th} eigenvector will be denoted using *superscripts* as $\mathbf{x}^i$, *subscripts* being reserved for the *components* of $\mathbf{x}$

Power Methods: The Direct Method

- Assume an n x n matrix A has n linearly independent eigenvectors $\mathbf{e}^1, \mathbf{e}^2, \ldots, \mathbf{e}^n$ ordered by decreasing eigenvalues
  $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_{n-2}| \geq |\lambda_{n-1}| > |\lambda_n|$
- Given any vector $\mathbf{y}^0 \neq \mathbf{0}$, there exist constants c_i, i = 1,…,n, such that
  $\mathbf{y}^0 = c_1\mathbf{e}^1 + c_2\mathbf{e}^2 + \ldots + c_n\mathbf{e}^n$

The Direct Method (continued)

- If $\mathbf{y}^0$ is *not* orthogonal to $\mathbf{e}^1$, i.e., $(\mathbf{y}^0)^T\mathbf{e}^1 \neq 0$,
- $\mathbf{y}^1 = A\mathbf{y}^0 = A(c_1\mathbf{e}^1 + c_2\mathbf{e}^2 + \ldots + c_n\mathbf{e}^n)$
- $= Ac_1\mathbf{e}^1 + Ac_2\mathbf{e}^2 + \ldots + Ac_n\mathbf{e}^n$
- $= c_1A\mathbf{e}^1 + c_2A\mathbf{e}^2 + \ldots + c_nA\mathbf{e}^n$
- Can you simplify the previous line?

The Direct Method (continued)

- If $\mathbf{y}^0$ is *not* orthogonal to $\mathbf{e}^1$, i.e., $(\mathbf{y}^0)^T\mathbf{e}^1 \neq 0$,
- $\mathbf{y}^1 = A\mathbf{y}^0 = A(c_1\mathbf{e}^1 + c_2\mathbf{e}^2 + \ldots + c_n\mathbf{e}^n)$
- $\quad = Ac_1\mathbf{e}^1 + Ac_2\mathbf{e}^2 + \ldots + Ac_n\mathbf{e}^n$
- $\quad = c_1A\mathbf{e}^1 + c_2A\mathbf{e}^2 + \ldots + c_nA\mathbf{e}^n$
- $\mathbf{y}^1 = c_1\lambda_1\mathbf{e}^1 + c_2\lambda_2\mathbf{e}^2 + \ldots + c_n\lambda_n\mathbf{e}^n$
- What is $\mathbf{y}^2 = A\mathbf{y}^1$?

The Direct Method (continued)

$$\mathbf{y}^{\mathbf{2}} = c_1\lambda_1^2\mathbf{e}^1 + c_2\lambda_2^2\mathbf{e}^2 + \ldots + c_n\lambda_n^2\mathbf{e}^n = \sum_{k=1}^{n} c_k\lambda_k^2\mathbf{e}^{\mathbf{k}}$$

in general,

$$\mathbf{y}^i = c_1\lambda_1^i\mathbf{e}^1 + c_2\lambda_2^i\mathbf{e}^2 + \ldots + c_n\lambda_n^i\mathbf{e}^n = \sum_{k=1}^{n} c_k\lambda_k^i\mathbf{e}^{\mathbf{k}}$$

or

$$\mathbf{y}^i = \lambda_1^i\left(c_1\mathbf{e}^1 + \sum_{k=2}^{n} c_k\left(\frac{\lambda_k}{\lambda_1}\right)^i\mathbf{e}^{\mathbf{k}} \right)$$

The Direct Method (continued)

$$\mathbf{y}^{i} = \lambda_1^{i}\left(c_1\mathbf{e}^{1} + \sum_{k=2}^{n} c_k \left(\frac{\lambda_k}{\lambda_1}\right)^{i} \mathbf{e}^{\mathbf{k}} \right)$$

So what?

Recall that $\left|\frac{\lambda_k}{\lambda_1}\right| < 1$, so $\left(\frac{\lambda_k}{\lambda_1}\right)^{i} \rightarrow 0$ as $i \rightarrow \infty$

The Direct Method (continued)

Since

$$\mathbf{y}^i = \lambda_1^i \left(c_1 \mathbf{e}^1 + \sum_{k=2}^{n} c_k \left(\frac{\lambda_k}{\lambda_1} \right)^i \mathbf{e}^{\mathbf{k}} \right)$$

then

$$\mathbf{y}^i \approx \lambda_1^i c_1 \mathbf{e}^1 \text{ as i} \rightarrow \infty$$

The Direct Method (continued)

- Note: any nonzero multiple of an eigenvector is also an eigenvector
- Why?
- Suppose e is an eigenvector of A, i.e., $A\mathbf{e}=\lambda\mathbf{e}$ and $c \neq 0$ is a scalar such that $\mathbf{x} = c\mathbf{e}$
- $A\mathbf{x} = A(c\mathbf{e}) = c\,(A\mathbf{e}) = c\,(\lambda\mathbf{e}) = \lambda\,(c\mathbf{e}) = \lambda\mathbf{x}$

The Direct Method (continued)

Since $\mathbf{y}^i \approx \lambda_1^i c_1 \mathbf{e}^1$ and $\mathbf{e}^1$ is an eigenvector

$\mathbf{y}^i$ will become arbitrarily close to an eigenvector

and we can prevent $\mathbf{y}^i$ from growing inordinately

by normalizing

$$\mathbf{y}^i = \frac{\mathbf{A}\mathbf{y}^{i-1}}{\left\|\mathbf{A}\mathbf{y}^{i-1}\right\|}$$

Direct Method (continued)

- Given an eigenvector $\mathbf{e}$ for the matrix $\mathbf{A}$
- We have $\mathbf{Ae} = \lambda\mathbf{e}$ and $\mathbf{e} \neq \mathbf{0}$, so $\mathbf{e}^T\mathbf{e} \neq 0$ (a scalar)
- Thus, $\mathbf{e}^T\mathbf{Ae} = \mathbf{e}^T\lambda\mathbf{e} = \lambda\mathbf{e}^T\mathbf{e} \neq 0$
- So $\lambda = (\mathbf{e}^T\mathbf{Ae}) / (\mathbf{e}^T\mathbf{e})$

Direct Method (completed)

$$\mu_i = \frac{\left(\mathbf{y}^i\right)^T \mathbf{A}\mathbf{y}^i}{\left(\mathbf{y}^i\right)^T \mathbf{y}^i}$$ approximates the dominant eigenvalue

$\mathbf{r}^{\mathrm{i}} \equiv \left(\mu_i \mathbf{I} - \mathbf{A}\right)\mathbf{y}^i$ is the residual error vector

and

$\mathbf{r}^{\mathrm{i}} = \mathbf{0}$ when μ_i is an eigenvalue with eigenvector $\mathbf{y}^i$

Direct Method Algorithm

1. Choose $\varepsilon > 0$, $m > 0$, and $\mathbf{y} \neq \mathbf{0}$.

   Set $i = 0$ and compute $\mathbf{x} = \mathbf{Ay}$ and .

2. Do

$$y = \frac{x}{\|x\|}$$

$$x = Ay$$

$$\mu = \frac{y^T x}{y^T y}$$

$$r = \mu y - x$$

$$i = i + 1$$

3. While $(\|r\| > \varepsilon)$ and $(i < m)$