SVD : Singular Value Decomposition

정보컴퓨터공학부 김민환 교수

Contents

Singular Value Decomposition

- To know how to compute the singular value decomposition of a matrix
- To realize the theoretical and practical importance of this method

Introduction

- We have seen that factorizations of matrices into factors with special properties can be very useful
  - $\mathbf{LU}$, $\mathbf{QR}$, $\mathbf{A} = \mathbf{XDX}^{-1}$ or $\mathbf{QDQ}^{T}$, etc

- A factorization is of particular interest if some of the factors are *orthogonal* matrices
  - The reason is that orthogonal transformations preserve norms and angles
  - In particular, they preserve the lengths of the error vectors that are inevitable in numerical calculations

Singular Value Decomposition ?

- One of the most important factorizations that applies to any $m \times n$ matrix $\mathbf{A}$
- Both theoretical and applied interest
- The most useful application is reliable estimation of the rank of a matrix

$$\begin{array}{ccccc} \mathbf{A} & = & \mathbf{U} & \boldsymbol{\Sigma} & \mathbf{V}^T \\ (m\times n) & & (m\times m) & (m\times n) & (n\times n) \end{array}$$

$$\mathbf{A} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^T$$

Orthogonal matrices

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & .. & \mathbf{u}_m \end{bmatrix} \qquad (m \times m)$$

$$\boldsymbol{\Sigma} = \left[\begin{array}{ccc|cc} \sigma_1 & \Lambda & 0 & \Lambda & 0 \\ \mathrm{M} & \mathrm{O} & \mathrm{M} & \Lambda & \mathrm{M} \\ 0 & \Lambda & \sigma_r & \Lambda & 0 \\ \hline \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 0 & \Lambda & 0 & \Lambda & 0 \end{array}\right] = \left[\begin{array}{c|c} \mathbf{D} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \qquad (m \times n)$$

Singular values

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & .. & \mathbf{v}_n \end{bmatrix} \qquad (n \times n)$$

□ Th. 6.5.1 *SVD Theorem*

If $\mathbf{A}$ is an $m \times n$ matrix, then $\mathbf{A}$ has a singular value decomposition

(*Proof*)

$\mathbf{A}^T\mathbf{A}$ is a real symmetric $n \times n$ matrix. By Th. 6.4.1, its eigenvalues are all real and, by Th. 6.4.5, it has an orthogonal diagonalizing matrix $\mathbf{V}$. That is, $\mathbf{V}^T(\mathbf{A}^T\mathbf{A})\mathbf{V} = \mathbf{D}$.

Note that each orthonormal column vector $\mathbf{v}_i$ of $\mathbf{V}$, from Th. 6.3.1, is the eigenvector of $\mathbf{A}^T\mathbf{A}$ and the diagonal element d_{ii} of $\mathbf{D}$ is the eigenvalue corresponding to $\mathbf{v}_i$

Each eigenvalue of $\mathbf{A}^T\mathbf{A}$ must be nonnegative, since

$$\|\mathbf{A}\mathbf{v}_i\|^2 = \mathbf{v}_i^T \mathbf{A}^T \mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i = \lambda_i \|\mathbf{v}_i\|^2 = \lambda_i \geq 0$$

We may assume that the columns of $\mathbf{V}$ have been ordered so that the corresponding eigenvalues satisfy

$$\lambda_1 \geq \lambda_2 \geq \Lambda \geq \lambda_n \geq 0$$

The square root of each eigenvalue is referred to a *singular value* of $\mathbf{A}$

$$\sigma_i = \sqrt{\lambda_i} \quad i = 1, \ldots, n$$

$$\sigma_1 \geq \sigma_2 \geq \Lambda \geq \sigma_n \geq 0$$

Since the $n \times n$ matrix $\mathbf{A}^T\mathbf{A}$ is diagonalizable, by Th. 6.3.2, it has n linearly independent eigenvectors. From the definition of eigenvectors, $(\mathbf{A}^T\mathbf{A})\mathbf{v} = \lambda\mathbf{v}$, the eigenvectors belonging to nonzero eigenvalues are in $\mathrm{Col}(\mathbf{A}^T\mathbf{A})$.

Consequently, the rank of $\mathbf{A}^T\mathbf{A}$ equals the number of nonzero eigenvalues. That is, $\mathrm{Rank}(\mathbf{A}^T\mathbf{A}) = r$.

$$\lambda_1 \geq \lambda_2 \geq \Lambda \geq \lambda_r > 0 \quad \text{and} \quad \lambda_{r+1} = \lambda_{r+2} = \Lambda = \lambda_n = 0$$

(Note that the eigenvectors belonging to zero eigenvalues are in $\mathrm{Null}(\mathbf{A}^T\mathbf{A})$ since $(\mathbf{A}^T\mathbf{A})\mathbf{v} = \mathbf{0}$.)

The same relation holds for the singular values:

$$\sigma_1 \geq \sigma_2 \geq \Lambda \geq \sigma_r > 0 \quad \text{and} \quad \sigma_{r+1} = \sigma_{r+2} = \Lambda = \sigma_n = 0$$

Now let

$$\mathbf{V}_1 = [\,\mathbf{v}_1 \,\Lambda\, \mathbf{v}_r\,], \quad \mathbf{V}_2 = [\,\mathbf{v}_{r+1} \,\Lambda\, \mathbf{v}_n\,]$$

and

$$\mathbf{\Sigma}_1 = \begin{bmatrix} \sigma_1 & & \\ & \mathrm{O} & \\ & & \sigma_r \end{bmatrix} \tag{1}$$

Hence, $\mathbf{\Sigma}_1$ is an $r \times r$ diagonal matrix whose diagonal entries are the nonzero singular values $\sigma_1, \sigma_2, \Lambda, \sigma_r$

The $m \times n$ matrix $\mathbf{\Sigma}$ is then given by

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{O}_{r\times(n-r)} \\ \mathbf{O}_{(m-r)\times r} & \mathbf{O}_{(m-r)\times(n-r)} \end{bmatrix}$$

The column vectors of $\mathbf{V}_2$ are eigenvectors of $\mathbf{A}^T\mathbf{A}$ belonging to $\lambda = 0$. Thus

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_j = \mathbf{0} \quad j = r+1,\ldots,n$$

and, consequently, column vectors of $\mathbf{V}_2$ form an orthonormal basis for $\mathrm{Null}(\mathbf{A}^T\mathbf{A}) = \mathrm{Null}(\mathbf{A})$.

Therefore,

$$\mathbf{A}\mathbf{V}_2 = \mathbf{O}$$

$$\mathbf{A}\mathbf{v}_j \in \mathrm{Null}(\mathbf{A}^T) = \mathrm{Col}(\mathbf{A})^{\perp} \ \ \& \ \ \mathbf{A}\mathbf{v}_j \in \mathrm{Col}(\mathbf{A}) \quad \therefore \mathbf{A}\mathbf{v}_j = \mathbf{0}$$

Since $\mathbf{V}$ is an orthogonal matrix, it follows that

$$\mathbf{I} = \mathbf{V}\mathbf{V}^T = [\ \mathbf{V}_1\ \mathbf{V}_2\ ]\begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = \mathbf{V}_1\mathbf{V}_1^T + \mathbf{V}_2\mathbf{V}_2^T$$

$$\mathbf{A} = \mathbf{A}\mathbf{I} = \mathbf{A}\mathbf{V}_1\mathbf{V}_1^T + \mathbf{A}\mathbf{V}_2\mathbf{V}_2^T = \mathbf{A}\mathbf{V}_1\mathbf{V}_1^T + \mathbf{O}\mathbf{V}_2^T = \mathbf{A}\mathbf{V}_1\mathbf{V}_1^T \qquad (2)$$

If we define

$$\mathbf{u}_j = \frac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j \qquad j = 1, \ldots, r \qquad (3)$$

and

$$\mathbf{U}_1 = [\ \mathbf{u}_1\ \Lambda\ \ \mathbf{u}_r\ ]$$

then it follows that

$$\mathbf{A}\mathbf{V}_1 = \mathbf{U}_1\mathbf{\Sigma}_1 \qquad (4)$$

The column vectors of $\mathbf{U}_1$ form an orthonormal set, since

$$\mathbf{u}_i^T\mathbf{u}_j = \left(\frac{1}{\sigma_i}\mathbf{v}_i^T\mathbf{A}^T\right)\left(\frac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j\right) \quad 1 \le i \le r,\ 1 \le j \le r$$

$$= \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^T(\mathbf{A}^T\mathbf{A}\mathbf{v}_j) = \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^T(\lambda_j\mathbf{v}_j)$$

$$= \frac{\sigma_j}{\sigma_i}\mathbf{v}_i^T\mathbf{v}_j = \delta_{ij} = \begin{cases} 0 & i \ne j \\ 1 & i = j \end{cases}$$

From (3), it follows that each $\mathbf{u}_j$, $1 \le j \le r$, is in $\mathrm{Col}(\mathbf{A})$. The dimension of $\mathrm{Col}(\mathbf{A})$ is r, so $\mathbf{u}_1, \ldots, \mathbf{u}_r$ form an orthonormal basis for $R(\mathbf{A})$

The vector space $R(\mathbf{A})^{\perp} = \mathrm{Null}(\mathbf{A}^T)$ has dimension $m - r$. Let $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ be an orthonormal basis for $\mathrm{Null}(\mathbf{A}^T)$ and set

$$\mathbf{U}_2 = [\,\mathbf{u}_{r+1}\ \mathbf{u}_{r+2}\ \Lambda\ \ \mathbf{u}_m\,]$$

$$\mathbf{U} = [\,\mathbf{U}_1\ \mathbf{U}_2\,]$$

It follows from Th. 5.2.2 that $\mathbf{u}_1, \ldots, \mathbf{u}_m$ form an orthonormal basis for $\mathbf{R}^m$. Hence $\mathbf{U}$ is an orthogonal matrix. Finally, from (4) and (2),

$$\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = [\,\mathbf{U}_1\ \mathbf{U}_2\,]\begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}\begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}$$

$$= \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}_1^T = \mathbf{A}\mathbf{V}_1\mathbf{V}_1^T = \mathbf{A} \qquad \blacksquare$$

Summary

For any $m \times n$ matrix $\mathbf{A}$,

- $\mathbf{A}^T\mathbf{A}$ is the $n \times n$ real symmetric matrix and it is orthogonally diagonalizable from the spectral theorem
- $\mathbf{A}^T\mathbf{A}$ has real nonnegative eigenvalues, say, $\lambda_1,\ldots,\lambda_n$. Let $\mathbf{v}_1,\ldots,\mathbf{v}_n$ be the corresponding eigenvectors such that they form an *orthonormal basis* for $\mathbf{R}^n$

$$\|\mathbf{A}\mathbf{v}_i\|^2 = \mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mathbf{v}_i^T(\lambda_i\mathbf{v}_i) = \lambda_i\mathbf{v}_i^T\mathbf{v}_i = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i \geq 0$$

- Let $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$ without loss of generality

- The *singular values* of $\mathbf{A}$, $\sigma_1, \ldots, \sigma_n$, are defined by

$$\sigma_i = \sqrt{\lambda_i} = \left\| \mathbf{A}\mathbf{v}_i \right\|, \quad \sigma_1 \geq \ldots \geq \sigma_n \geq 0$$

- Let r be the number of positive singular values

$$\sigma_1 \geq \ldots \geq \sigma_r > 0 \quad \text{and} \quad \sigma_{r+1} = \ldots = \sigma_n = 0$$

- Then

$$\mathbf{V} = [\,\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_n\,], \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \quad (m \times n)$$

- For non-zero singular values, $\sigma_1 \geq \ldots \geq \sigma_r > 0$, form

$$\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i \,, \quad i = 1, \ldots, r$$

- Extend the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of $\mathbf{R}^m$. Then $\mathbf{U} = [\, \mathbf{u}_1 \; \mathbf{u}_2 \; \ldots \; \mathbf{u}_m \,]$.

$$\begin{array}{cccc} \mathbf{A} = & \mathbf{U} & \boldsymbol{\Sigma} & \mathbf{V}^T \\ (m \times n) & (m \times m) & (m \times n) & (n \times n) \end{array}$$

(*Ex*.18) Compute the singular value decomposition of $\mathbf{A}$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \mathbf{A}^T\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \begin{matrix} \lambda_1 = 4 \\ \lambda_2 = 0 \end{matrix} \quad \mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\therefore \mathbf{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{u}_1 = \frac{1}{\sigma_1}\mathbf{A}\mathbf{v}_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

A basis for $\mathbf{R}^3$ can be obtained by extending $\{\mathbf{u}_1\}$

$$[\,\mathbf{u}_1\ \mathbf{e}_1\ \mathbf{e}_2\ \mathbf{e}_3\,] = \begin{bmatrix} 1/\sqrt{2} & 1 & 0 & 0 \\ 1/\sqrt{2} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1/\sqrt{2} & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus $B = \{\mathbf{u}_1, \mathbf{e}_1, \mathbf{e}_3\}$ is a basis for $\mathbf{R}^3$. By the Gram-Schmidt process, an orthonormal basis $B' = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ can be obtained

$$\mathbf{p}_1 = (\mathbf{e}_1^T\mathbf{u}_1)\,\mathbf{u}_1 = \mathbf{u}_1/\sqrt{2} = (1/2, 1/2, 0)$$

$$\mathbf{u}_2 = (\mathbf{e}_1 - \mathbf{p}_1)/\left\|\mathbf{e}_1 - \mathbf{p}_1\right\| = (1/\sqrt{2}, -1/\sqrt{2}, 0)$$

$$\mathbf{p}_2 = (\mathbf{e}_3^T\mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{e}_3^T\mathbf{u}_2)\,\mathbf{u}_2 = \mathbf{0}$$

$$\mathbf{u}_3 = (\mathbf{e}_3 - \mathbf{p}_2)/\left\|\mathbf{e}_3 - \mathbf{p}_2\right\| = \mathbf{e}_3 = (0, 0, 1)$$

Therefore, it follows that

$$\mathbf{A} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T$$

$\blacksquare$

Observations

1. The singular values $\sigma_1,\ldots,\sigma_n$ of $\mathbf{A}$ are unique, however, the matrices $\mathbf{U}$ and $\mathbf{V}$ are not unique
2. Since $\mathbf{V}$ diagonalizes $\mathbf{A}^T\mathbf{A}$, that is, $\mathbf{V}^T(\mathbf{A}^T\mathbf{A})\mathbf{V} = \mathbf{D}_n$, it follows that the $\mathbf{v}_j$'s are eigenvectors of $\mathbf{A}^T\mathbf{A}$
3. Since $\mathbf{A}\mathbf{A}^T = (\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T)(\mathbf{V}\,\mathbf{\Sigma}^T\mathbf{U}^T) = \mathbf{U}\,\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T = \mathbf{U}\,\mathbf{D}_m\mathbf{U}^T$, it follows that $\mathbf{U}$ diagonalizes $\mathbf{A}\mathbf{A}^T$ and the $\mathbf{u}_j$'s are eigenvectors of $\mathbf{A}\mathbf{A}^T$

$$\mathbf{\Sigma}\mathbf{\Sigma}^T = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1^T & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma}_1\mathbf{\Sigma}_1^T & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \mathbf{O} \\ & \ddots \ \lambda_r & \\ \mathbf{O} & & \mathbf{O} \end{bmatrix} = \mathbf{D}_m$$

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_j = \mathbf{U}\mathbf{D}_m\mathbf{U}^T\mathbf{u}_j = (\lambda_1\mathbf{u}_1\mathbf{u}_1^T + \cdots + \lambda_r\mathbf{u}_r\mathbf{u}_r^T)\mathbf{u}_j = \lambda_j\mathbf{u}_j$$

4. Comparing the *j*th column of each side of the equation $\mathbf{AV} = \mathbf{U\Sigma}$, we get

$$\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j \qquad j = 1, \ldots, n$$

Similarly, from $\mathbf{A}^T\mathbf{U} = \mathbf{V}\mathbf{\Sigma}^T$, we get

$\mathbf{A}^T = \mathbf{V}\,\mathbf{\Sigma}^T\,\mathbf{U}^T$

$$\mathbf{A}^T\mathbf{u}_j = \sigma_j \mathbf{v}_j \qquad j = 1, \ldots, n$$

$$\mathbf{A}^T\mathbf{u}_j = \mathbf{0} \qquad j = n+1, \ldots, m \qquad (\,m > n\,)$$

The $\mathbf{v}_j$'s are called the *right singular vectors* of $\mathbf{A}$, and the $\mathbf{u}_j$'s are called the *left singular vectors* of $\mathbf{A}$

5. If $\mathbf{A}$ has rank r, then

   (i) $\mathbf{v}_1, \ldots, \mathbf{v}_r$ form an orthonormal basis for $R(\mathbf{A}^T)$

   (ii) $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ form an orthonormal basis for $\mathrm{Null}(\mathbf{A})$

   (iii) $\mathbf{u}_1, \ldots, \mathbf{u}_r$ form an orthonormal basis for $R(\mathbf{A})$

   (iv) $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ form an orthonormal basis for $\mathrm{Null}(\mathbf{A}^T)$

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_j = \mathbf{A}^T(\mathbf{A}\mathbf{v}_j) = \lambda_j\mathbf{v}_j \in R(\mathbf{A}^T), \quad j = 1, \ldots, r$$

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \;\Rightarrow\; \mathbf{A}\mathbf{v}_j = 0\mathbf{u}_j = \mathbf{0} \;\Rightarrow\; \mathbf{v}_j \in \mathrm{Null}(\mathbf{A}), \quad j = r+1, \ldots, n$$

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_j = \mathbf{A}(\mathbf{A}^T\mathbf{u}_j) = \lambda_j\mathbf{u}_j \in R(\mathbf{A}), \quad j = 1, \ldots, r$$

$$\mathbf{A}^T\mathbf{U} = \mathbf{V}\mathbf{\Sigma}^T \;\Rightarrow\; \mathbf{A}^T\mathbf{u}_j = 0\mathbf{v}_j = \mathbf{0} \;\Rightarrow\; \mathbf{u}_j \in \mathrm{Null}(\mathbf{A}^T), \quad j = r+1, \ldots, m$$

Rank($\mathbf{A}$) = r

$$\mathbf{A} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \Lambda + \sigma_n \mathbf{u}_n \mathbf{v}_n^T \quad (m \times n)$$

$R(\mathbf{A})$ $\quad$ $R(\mathbf{A})^{\perp} = \mathrm{Null}(\mathbf{A}^T)$

orthogonal

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \Lambda & \mathbf{u}_r & \mathbf{u}_{r+1} & \Lambda & \mathbf{u}_m \end{bmatrix} \quad (m \times m)$$

$\mathbf{u}_i = (1/\sigma_i)\, \mathbf{A}\mathbf{v}_i$

$\mathbf{A}\mathbf{A}^T$ eigenvectors

pseudo-diagonal

$$\boldsymbol{\Sigma} = \left[\begin{array}{c|c} \mathbf{D}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \quad (m \times n)$$

singular values

orthogonal

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \Lambda & \mathbf{v}_r & \mathbf{v}_{r+1} & \Lambda & \mathbf{v}_n \end{bmatrix} \quad (n \times n)$$

$\mathbf{A}^T\mathbf{A}$ orthonormal eigenvectors

$R(\mathbf{A}^T)$ $\quad$ $R(\mathbf{A}^T)^{\perp} = \mathrm{Null}(\mathbf{A})$

$$\mathbf{A}^T = \mathbf{V}\,\boldsymbol{\Sigma}^T\,\mathbf{U}^T$$

6. The *rank* of the matrix $\mathbf{A}$ is equal to the number of its *nonzero* singular values (where singular values are counted according to multiplicity).

   The reader should be careful not to make a similar assumption about *eigenvalues*

For example, the matrix $\mathbf{M}$ has rank 3 even though all of its eigenvalues are 0

$$\mathbf{M}=\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&0\end{bmatrix}\qquad \det(\mathbf{M}-\lambda\mathbf{I})=\begin{vmatrix}-\lambda&1&0&0\\0&-\lambda&1&0\\0&0&-\lambda&1\\0&0&0&-\lambda\end{vmatrix}=\lambda^4=0$$

Actually, we can obtain 3 nonzero singular values

$$\mathbf{M}^T\mathbf{M}=\begin{bmatrix}0&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},\qquad \begin{aligned}&\det(\mathbf{M}^T\mathbf{M}-\lambda\mathbf{I})=-\lambda(1-\lambda)^3=0\\&\therefore\ \sigma_1=\sigma_2=\sigma_3=1,\ \sigma_4=0\end{aligned}$$

7. In the case that $\mathbf{A}$ has rank $r < n$,

   if we set

$$\mathbf{U}_1 = [\,\mathbf{u}_1 \,\Lambda\, \mathbf{u}_r\,],\ \mathbf{V}_1 = [\,\mathbf{v}_1 \,\Lambda\, \mathbf{v}_r\,],\ \boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_1 & & 0 \\ & \mathrm{O} & \\ 0 & & \sigma_r \end{bmatrix}$$

   then

$$\mathbf{A} = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T \qquad (m \times n) \leftrightarrow (m \times r)\,(r \times r)\,(r \times n)$$

   This factorization is called the *compact form of the singular value decomposition* of $\mathbf{A}$. This form is also useful in many applications

Lemma 6.5.2

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{Q}$ is an $m \times m$ orthogonal matrix, then

$$\|\mathbf{QA}\|_F = \|\mathbf{A}\|_F$$

(*Proof*)

$$\|\mathbf{QA}\|_F^2 = \| [\ \mathbf{Qa}_1\ \mathbf{Qa}_2\ \ldots \mathbf{Qa}_n\ ]\ \|_F^2$$

$$= \sum_{i=1}^{n} \|\mathbf{Qa}_i\|^2 = \sum_{i=1}^{n} \|\mathbf{a}_i\|^2 = \|\mathbf{A}\|_F^2$$

$\|\mathbf{A}\|_F$ Frobenius norm of $\mathbf{A}$

$$\|\mathbf{A}\|_F = (\sigma_1^2 + \sigma_2^2 + \Lambda \ + \sigma_n^2)^{1/2}$$

If $\mathbf{A}_{(m\times n)}$ has singular value decomposition $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, then

$$\|\mathbf{A}\|_F = \|\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\|_F = \|\boldsymbol{\Sigma}\mathbf{V}^T\|_F \qquad \leftarrow \quad \mathit{Lemma}\, 6.5.2$$

$$= \|(\boldsymbol{\Sigma}\mathbf{V}^T)^T\|_F = \|\mathbf{V}\boldsymbol{\Sigma}^T\|_F$$

$$= \|\boldsymbol{\Sigma}^T\|_F$$

$$= (\sigma_1^2 + \sigma_2^2 + \Lambda \ + \sigma_n^2)^{1/2}$$

- Th. 6.5.3

Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ be an $m \times n$ matrix, and let $\mathcal{M}$ denote the set of all $m \times n$ matrices of rank k or less, where $0 < k < \mathrm{Rank}(\mathbf{A})$. If $\mathbf{X}$ is a matrix in $\mathcal{M}$ satisfying

$$\|\mathbf{A}-\mathbf{X}\|_F = \min_{\mathbf{S}\in\mathcal{M}} \|\mathbf{A}-\mathbf{S}\|_F$$

then

$$\|\mathbf{A}-\mathbf{X}\|_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \Lambda + \sigma_n^2)^{1/2}$$

In particular, if $\mathbf{A}' = \mathbf{U}\boldsymbol{\Sigma}'\mathbf{V}^T$, where $\boldsymbol{\Sigma}' = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \mathrm{O} & & \mathbf{O} \\ & & \sigma_k & \\ \hline & \mathbf{O} & & \mathbf{O} \end{array}\right]$

then

$$\|\mathbf{A}-\mathbf{A}'\|_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \Lambda + \sigma_n^2)^{1/2} = \min_{\mathbf{S}\in\mathcal{M}} \|\mathbf{A}-\mathbf{S}\|_F$$

- An $m \times n$ matrix $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ can be represented by an outer product expansion

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \Lambda + \sigma_n \mathbf{u}_n \mathbf{v}_n^T$$

- If $\mathbf{A}$ is of rank n, then $\mathbf{A}' = \mathbf{U}\mathbf{\Sigma}'\mathbf{V}^T$ will be the matrix of rank $n-1$ that is closest to $\mathbf{A}$ with respect to the Frobenius norm

$$\mathbf{A}' = \mathbf{U}\mathbf{\Sigma}'\mathbf{V}^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \Lambda + \sigma_{n-1} \mathbf{u}_{n-1} \mathbf{v}_{n-1}^T$$

$$= \mathbf{U} \begin{bmatrix} \sigma_1 & & & \\ & \mathrm{O} & & \\ & & \sigma_{n-1} & \\ & & & 0 \\ & & & \end{bmatrix} \mathbf{V}^T$$

- In particular, if $\mathbf{A}$ is a *nonsingular* $n \times n$ matrix, then such a matrix $\mathbf{A}'$ $(= \mathbf{U}\boldsymbol{\Sigma}'\mathbf{V}^T)$ with rank $n-1$ is *singular* and $\left\|\mathbf{A}-\mathbf{A}'\right\|_F = \sigma_n$.

  Thus σ_n may be taken as a measure of how close a square matrix is to being singular

$$\mathbf{A}' = \mathbf{U}\,\boldsymbol{\Sigma}'\,\mathbf{V}^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_{n-1} \mathbf{u}_{n-1} \mathbf{v}_{n-1}^T$$

$$= \mathbf{U} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_{n-1} & \\ & & & 0 \end{bmatrix} \mathbf{V}^T$$

(*Ex*.19) The $n \times n$ upper triangular matrix $\mathbf{A}$ is *nonsingular* since $\det(\mathbf{A}) = 1$. However, it is close to being singular if n is large.

Let us consider the *singular* matrix $\mathbf{B}$. The system $\mathbf{Bx} = \mathbf{0}$ has a nontrivial solution $\mathbf{x} = (2^{n\text{-}2}, 2^{n\text{-}3}, \ldots, 2^0, 1)$.

From Th. 6.5.3,

$$\sigma_n = \min_{\mathbf{X}\ singular} \left\|\mathbf{A} - \mathbf{X}\right\|_F \leq \left\|\mathbf{A} - \mathbf{B}\right\|_F = 1/2^{n-2} \qquad ???\ \ n = 100$$

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & \Lambda & -1 \\ 0 & 1 & \Lambda & -1 \\ : & : & O & : \\ 0 & 0 & \Lambda & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & -1 & \Lambda & -1 \\ 0 & 1 & \Lambda & -1 \\ : & : & O & : \\ -1/2^{n-2} & 0 & \Lambda & 1 \end{bmatrix}$$

Def. Pseudoinverse

Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ be an SVD for an $m \times n$ matrix $\mathbf{A}$

The *pseudoinverse*, or *Moore-Penrose inverse*, of $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}^+$ given by

$$\mathbf{A}^+ = \mathbf{V}\,\boldsymbol{\Sigma}^+\,\mathbf{U}^T$$

where $\boldsymbol{\Sigma}^+$ is the $n \times m$ matrix

$$\boldsymbol{\Sigma}^+ = \left[\begin{array}{c|c} \mathbf{D}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]$$

$$\boldsymbol{\Sigma} = \left[\begin{array}{c|c} \mathbf{D} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \quad (m \times n)$$

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}$$

$$\mathbf{D}^{-1} = \begin{bmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_r \end{bmatrix}$$

$\mathbf{D}$ is, as before, the $r \times r$ diagonal with diagonal entries the positive singular values $\sigma_1 \geq \ldots \geq \sigma_r > 0$ of $\mathbf{A}$

Moore-Penrose Conditions

$\mathbf{A}^+$ is the unique matrix $\mathbf{B}$ that satisfies the conditions

1. $\mathbf{ABA} = \mathbf{A}$
2. $\mathbf{BAB} = \mathbf{B}$
3. $(\mathbf{AB})^T = \mathbf{AB}$
4. $(\mathbf{BA})^T = \mathbf{BA}$

$$\mathbf{AA}^+\mathbf{A} = (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)(\mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^T)(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T) = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{A}$$

$$\mathbf{A}^+\mathbf{AA}^+ = (\mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^T)(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)(\mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^T) = \mathbf{V}\boldsymbol{\Sigma}^+\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+\mathbf{U}^T = \mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^T = \mathbf{A}^+$$

$$\mathbf{AA}^+ = (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)(\mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^T) = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^T = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+\mathbf{U}^T$$

$$= \mathbf{U}\mathbf{I}_{m,r}\mathbf{U}^T = \mathbf{U}\mathbf{I}_{m,r}\mathbf{I}_{m,r}\mathbf{U}^T = \mathbf{U}_{m,r}\mathbf{U}_{m,r}^T = \mathbf{I}_{m,r} = (\mathbf{AA}^+)^T$$

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+ = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{D}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{I}_{m,r}, \quad \mathbf{A}^+\mathbf{A} = \mathbf{I}_{n,r} = (\mathbf{A}^+\mathbf{A})^T$$

Least Squares

Recall that a least square solution for the possibly inconsistent system $\mathbf{Ax} = \mathbf{b}$ is a vector $\mathbf{x}'$ that minimizes the length of the error vector $\boldsymbol{\Delta} = \mathbf{b} - \mathbf{Ax}$,

$$\|\Delta\| = \|\mathbf{b} - \mathbf{Ax}'\| = \min_{\mathbf{x} \in R(\mathbf{A})} \|\mathbf{b} - \mathbf{Ax}\|$$

The vector $\mathbf{x}'$ is not necessary unique. If $\mathbf{A}$ is $m \times n$ with $m < n$, then its nullity ≥ 1. In this case any vector $\mathbf{x}' + \mathbf{z}$ with $\mathbf{z}\ (\neq \mathbf{0}) \in \mathrm{Null}(\mathbf{A})$ will also a least squares solution, because $\mathbf{b} - \mathbf{A}(\mathbf{x}' + \mathbf{z}) = \mathbf{b} - \mathbf{Ax}' - \mathbf{Az} = \mathbf{b} - \mathbf{Ax}'$

If we demand that $\mathbf{x}'$ has also minimum length, then such solution is unique

□ Th. 6.5.4

The least squares problem $\mathbf{Ax} = \mathbf{b}$ has a unique least squares solution $\mathbf{x}'$ of minimal length given by

$$\mathbf{x}' = \mathbf{A}^{+}\mathbf{b}$$

(*Proof*)

Let SVD of the $m \times n$ matrix $\mathbf{A}$ be $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ and let $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be $\mathbf{V}^T\mathbf{x}$. The matrix $\mathbf{U}^T$ is orthogonal, so $\|\mathbf{U}^T\mathbf{z}\| = \|\mathbf{z}\|$ for any m-vector $\mathbf{z}$.

Then because $\boldsymbol{\Sigma}$ has only r nonzero entries located at the upper left $r \times r$ block, the squared error S is

$$S = \left\| \mathbf{b} - \mathbf{A}\mathbf{x} \right\|^2 = \left\| \mathbf{b} - \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T \mathbf{x} \right\|^2 = \left\| \mathbf{U}^T \mathbf{b} - \mathbf{\Sigma}\,\mathbf{V}^T \mathbf{x} \right\|^2 = S_1 + S_2$$

$$S_1 = (\mathbf{u}_1^T \mathbf{b} - \sigma_1 y_1)^2 + \Lambda\ + (\mathbf{u}_r^T \mathbf{b} - \sigma_r y_r)^2$$

$$S_2 = (\mathbf{u}_{r+1}^T \mathbf{b})^2 + \Lambda\ + (\mathbf{u}_n^T \mathbf{b})^2$$

Since S_2 is fixed, S is minimized if S_1 is minimum. In fact, if we could choose $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ such that

$$\mathbf{u}_i^T \mathbf{b} = \sigma_i y_i\,, \quad i = 1, \Lambda\ , r$$

then S_1 would be 0. So, all we need is an $\mathbf{y}$ of the form

$$\mathbf{y} = \left(\frac{\mathbf{u}_1^T \mathbf{b}}{\sigma_1}, \Lambda\ , \frac{\mathbf{u}_r^T \mathbf{b}}{\sigma_r}, *, \Lambda\ , * \right)$$

The $\mathbf{y}$ with minimum magnitude, $\mathbf{y}'$, can be determined by setting the last $n - r$ coordinates equal to 0.

$$\mathbf{y}' = \left(\frac{\mathbf{u}_1^T \mathbf{b}}{\sigma_1}, \Lambda\ , \frac{\mathbf{u}_r^T \mathbf{b}}{\sigma_r}, 0, \Lambda\ , 0 \right) = \mathbf{V}^T \mathbf{x}'$$

Since the matrix $\mathbf{V}^T$ is orthogonal, $\|\mathbf{y}'\| = \|\mathbf{V}^T\mathbf{x}'\| = \|\mathbf{x}'\|$. Hence, $\mathbf{x}'$ is the only least squares solution of minimal length. The solution can be written as

$$\mathbf{x}' = \mathbf{V}\,\mathbf{y}' = \mathbf{V} \left(\frac{\mathbf{u}_1^T \mathbf{b}}{\sigma_1}, \Lambda\ , \frac{\mathbf{u}_r^T \mathbf{b}}{\sigma_r}, 0, \Lambda\ , 0 \right) = \mathbf{V}\,\mathbf{\Sigma}^+\,\mathbf{U}^T\,\mathbf{b} = \mathbf{A}^+\,\mathbf{b}$$

$\blacksquare$

(*Ex*.20) $\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix}$ $\quad \begin{bmatrix} 1 & -1 & : & 1 \\ -1 & 1 & : & -2 \\ -2 & 2 & : & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & : & 1 \\ 0 & 0 & : & -1 \\ 0 & 0 & : & 0 \end{bmatrix}$

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & -1 & -2 \\ -1 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} 6 & -6 \\ -6 & 6 \end{bmatrix}$$

$$\det\left(\begin{bmatrix} 6 & -6 \\ -6 & 6 \end{bmatrix} - \lambda\mathbf{I}\right) = \det\begin{bmatrix} 6-\lambda & -6 \\ -6 & 6-\lambda \end{bmatrix} = (6-\lambda)(6-\lambda) - 6^2 = \lambda^2 - 12\lambda = 0$$

$$\therefore \lambda = 12 \ or \ 0$$

$$\lambda = 12; \begin{bmatrix} -6 & -6 \\ -6 & -6 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \therefore \mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\lambda = 0; \begin{bmatrix} 6 & -6 \\ -6 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \quad \therefore \mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathbf{u}_1 = \frac{1}{\sqrt{12}}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{12}}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

Extend the set $\{\mathbf{u}_1\}$ to an orthogonal basis of $\mathbf{R}^3$

$$\begin{bmatrix} -1/\sqrt{6} & 1 & 0 & 0 \\ 1/\sqrt{6} & 0 & 1 & 0 \\ 2/\sqrt{6} & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} -1/\sqrt{6} & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{bmatrix} \quad \therefore \mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

By the Gram-Schmidt Orthonalization Process,

$$\mathbf{u}_2' = \mathbf{a}_2 - (\mathbf{a}_2^T\mathbf{u}_1)\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \left(\frac{-1}{\sqrt{6}}\right)\frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} \quad \therefore \mathbf{u}_2 = \frac{1}{\sqrt{30}}\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}$$

$$\mathbf{u}_3' = \mathbf{a}_3 - (\mathbf{a}_3^T\mathbf{u}_1)\mathbf{u}_1 - (\mathbf{a}_3^T\mathbf{u}_2)\mathbf{u}_2 = \frac{2}{5}\begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix} \quad \therefore \mathbf{u}_3 = \frac{1}{\sqrt{5}}\begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = \begin{bmatrix} -1/\sqrt{6} & 5/\sqrt{30} & 0 \\ 1/\sqrt{6} & 1/\sqrt{30} & 2/\sqrt{5} \\ 2/\sqrt{6} & 2/\sqrt{30} & -1/\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T$$

$$\mathbf{A}^+ = \mathbf{V}\,\mathbf{\Sigma}^+\,\mathbf{U}^T = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{12} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{6} & 5/\sqrt{30} & 0 \\ 1/\sqrt{6} & 1/\sqrt{30} & 2/\sqrt{5} \\ 2/\sqrt{6} & 2/\sqrt{30} & -1/\sqrt{5} \end{bmatrix}^T$$

$$= \begin{bmatrix} -1/\sqrt{24} & 0 & 0 \\ 1/\sqrt{24} & 0 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ 5/\sqrt{30} & 1/\sqrt{30} & 2/\sqrt{30} \\ 0 & 2/\sqrt{5} & -1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/12 & -1/12 & -1/6 \\ -1/12 & 1/12 & 1/6 \end{bmatrix}$$

$$\therefore \mathbf{x}' = \mathbf{A}^+\,\mathbf{b} = \begin{bmatrix} 1/12 & -1/12 & -1/6 \\ -1/12 & 1/12 & 1/6 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix} = \frac{11}{12} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \left\| \mathbf{b} - \mathbf{A}\mathbf{x}' \right\| = \frac{\sqrt{54}}{6} \cong 1.22$$

$$\mathbf{u}_1 = \frac{1}{\sqrt{12}}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{12}}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

Another way to get $\mathbf{a}_2$ and $\mathbf{a}_3$ from Null$(\mathbf{A}^T)$

$$\begin{bmatrix} 1 & -1 & -2 \\ -1 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \quad \therefore \mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

By the Gram-Schmidt Orthonalization Process,

$$\mathbf{u}_2' = \mathbf{a}_2 - (\mathbf{a}_2^T\mathbf{u}_1)\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - 0\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \therefore \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{u}_3' = \mathbf{a}_3 - (\mathbf{a}_3^T\mathbf{u}_1)\mathbf{u}_1 - (\mathbf{a}_3^T\mathbf{u}_2)\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \quad \therefore \mathbf{u}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T = \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T$$

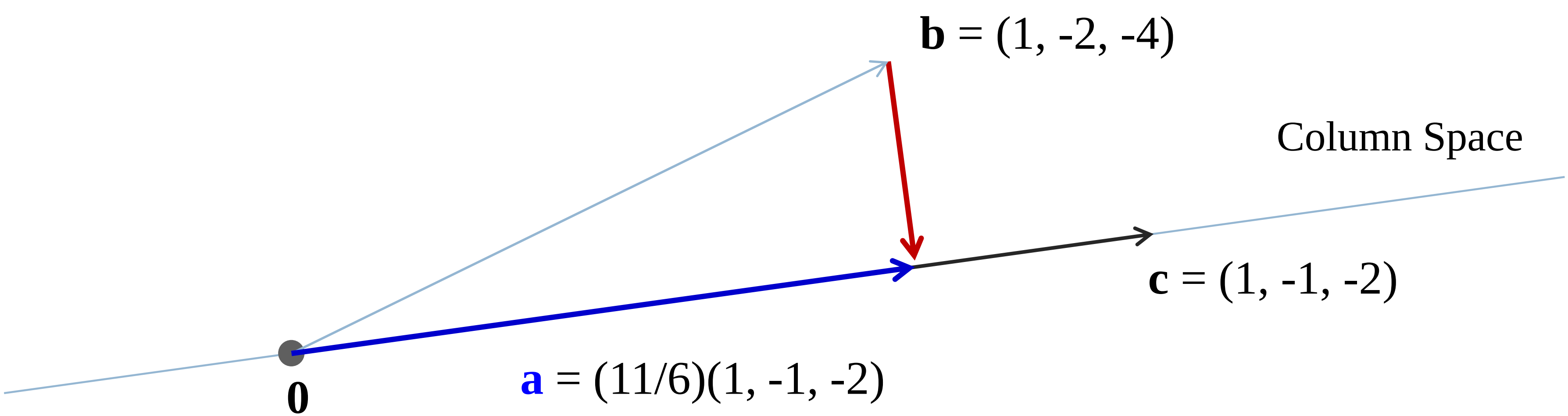
$$\mathbf{A}^+ = \mathbf{V}\,\mathbf{\Sigma}^+\,\mathbf{U}^T = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{12} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \\ 2/\sqrt{6} & 0 & 1/\sqrt{3} \end{bmatrix}^T$$

$$= \begin{bmatrix} -1/\sqrt{24} & 0 & 0 \\ 1/\sqrt{24} & 0 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} = \begin{bmatrix} 1/12 & -1/12 & -1/6 \\ -1/12 & 1/12 & 1/6 \end{bmatrix}$$

$$\therefore \mathbf{x}' = \mathbf{A}^+\mathbf{b} = \begin{bmatrix} 1/12 & -1/12 & -1/6 \\ -1/12 & 1/12 & 1/6 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix} = \frac{11}{12} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \left\| \mathbf{b} - \mathbf{A}\mathbf{x}' \right\| = \frac{\sqrt{54}}{6} \cong 1.22$$

(Another Solution)

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix} \qquad \begin{array}{c} \left[\begin{array}{cccc} 1 & -1 & : & 1 \\ -1 & 1 & : & -2 \\ -2 & 2 & : & -4 \end{array}\right] \\ \begin{array}{l} \mathbf{c} \end{array} \end{array} \sim \left[\begin{array}{cccc} 1 & -1 & : & 1 \\ 0 & 0 & : & -1 \\ 0 & 0 & : & 0 \end{array}\right]$$



$$\mathbf{a} = (11/6)(1, -1, -2)$$

$$= \frac{<\mathbf{b},\mathbf{c}>}{<\mathbf{c},\mathbf{c}>}\mathbf{c}$$

$$\mathbf{A}\mathbf{x}' = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \frac{11}{6} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} = \mathbf{a}$$

$$\begin{bmatrix} 1 & -1 & : & 11/6 \\ -1 & 1 & : & -11/6 \\ -2 & 2 & : & -22/6 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & : & 11/6 \\ 0 & 0 & : & 0 \\ 0 & 0 & : & 0 \end{bmatrix}$$

$$\therefore \mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \left\{ \begin{bmatrix} r + 11/6 \\ r \end{bmatrix}, r \in R \right\}$$

$$\|\mathbf{x}'\|^2 = (r + 11/6)^2 + r^2 \text{ is minimum when } r = -11/12$$

$$\therefore \mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{11}{12} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Applications of SVD

(SVD 1) **Numerical Rank**

The *numerical rank* of an $m \times n$ matrix is the number of singular values that are greater than $\sigma_1 \cdot \max(m,n) \cdot \varepsilon$, where σ_1 is the largest singular value and ε is the *unit round off error* called the *machine epsilon*

(*Ex.*) A 5×5 matrix $\mathbf{A}$ with singular values has numerical rank 3 since $\sigma_1 \max(m,n)\, \varepsilon = 4 \cdot 5 \cdot 5\times10^{-15} = 10^{-13}$

$$\sigma_1 = 4,\ \ \sigma_2 = 1,\ \ \sigma_3 = 10^{-12},\ \ \sigma_4 = 3.1\times10^{-14},\ \ \sigma_5 = 2.6\times10^{-15}$$

(SVD 2) Digital Image Processing

A digital image may be considered as a $m \times n$ matrix $\mathbf{A}$. Generally, $\mathbf{A}$ has many small singular values. Thus it can be approximated by a matrix of much lower rank.

The total storage for $\mathbf{A}_k$ is $k\,(m+n+1)$

$$\mathbf{A}_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \Lambda + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

(SVD 3) Latent Semantic Indexing (Information Retrieval)

If there are m possible key search words and a total of n documents in the collection, then the database can represented by a $m \times n$ matrix $\mathbf{A}$

x	key words \ Documents	doc 1	doc 2	doc 3	doc 4	doc 5	doc 6	doc 7	doc 8
0	determinants	0	6	3	0	1	0	1	1
0	eigenvalues	0	0	0	0	0	5	3	2
0	linear	5	4	4	5	4	0	3	3
0	matrices	6	5	3	3	4	4	3	2
0	numerical	0	0	0	0	3	0	4	3
1	orthogonality	0	0	0	0	4	6	0	2
1	spaces	0	0	5	2	3	3	0	1
0	systems	5	3	3	2	4	2	1	1
0	transformations	0	0	0	5	1	3	1	0
1	vector	0	4	4	3	4	1	0	3

The *database matrix* $\mathbf{Q}$ is formed by scaling each column of $\mathbf{A}$ so that all columns are unit vectors.

$$\mathbf{q}_j = \mathbf{a}_j \,/\, \|\mathbf{a}_j\| \quad j = 1,\ldots,8$$

The search vector $\mathbf{x}$ is also normalized.

If we set $\mathbf{y} = \mathbf{Q}^T\mathbf{x}$, then $y_i = \mathbf{q}_i{}^T\mathbf{x} = \cos\theta_i$ where θ_i is the angle between the unit vectors $\mathbf{x}$ and $\mathbf{q}_i$

The documents that best match the search criteria are those corresponding to the entries of $\mathbf{y}$ that are closest to 1. For example, the search result $\mathbf{y}$ for key words (*orthogonality*, *spaces*, *vector*) is given by

$$\mathbf{y} = (\ 0.00,\ 0.23,\ 0.57,\ 0.33,\ 0.64,\ 0.58,\ 0.00,\ 0.54\ )$$

- Let $\mathbf{Q} = \mathbf{P} + \mathbf{E}$, where $\mathbf{P}$ is a perfect database matrix and $\mathbf{E}$ is a matrix representing the errors because of the problems of polysemy (多義性) and synonymy (同義性). Unfortunately, $\mathbf{E}$ is unknown, so we cannot determine $\mathbf{P}$ exactly
- In the method of *latent semantic indexing* (LSI), $\mathbf{Q}$ is approximated by a matrix $\mathbf{Q}_1$ with lower rank. The idea behind the method is that the lower rank matrix may still provide a good approximation to $\mathbf{P}$ and, because of its simpler structure, may actually involve less error; that is, $\|\mathbf{E}_1\| < \|\mathbf{E}\|$

The lower rank approximation can be obtained by truncating the outer product expansion of the SVD of $\mathbf{Q}$. This approach is equivalent to setting

$$\sigma_{r+1} = \sigma_{r+2} = \Lambda = \sigma_n = 0$$

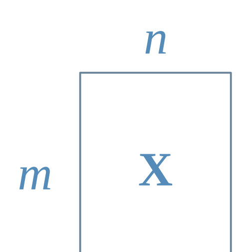
and then setting $\mathbf{Q}_1 = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T$, the compact form of the SVD of the rank r matrix

The basic method using $\mathbf{Q}^T\mathbf{x}$ requires a total of mn scalar multiplications. In contrast, $\mathbf{Q}_1 = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T$ and $\mathbf{Q}_1{}^T\mathbf{x} = \mathbf{V}_1(\mathbf{\Sigma}_1(\mathbf{U}_1{}^T\mathbf{x}))$ requires a total of $r\,(m + n + 1)$ scalar multiplications. For example, if $m = n = 1000$ and $r = 200$, then the search with the lower rank matrix is more than twice as fast ■

(SVD 4) Principal Component Analysis (PCA)

Let assume that a series of n aptitude tests is administered to a group of m individuals and that the *deviations from the mean* for the tests form the columns of an $m \times n$ matrix $\mathbf{X}$

Since the *hypothetical factors* that account for the scores should be uncorrelated, we wish to introduce *mutually orthogonal vectors* $\mathbf{y}_1, \ldots, \mathbf{y}_r$ corresponding to the hypothetical factors. These vectors should span $R(\mathbf{X})$ and hence the number of vectors, r, should be equal to the rank of $\mathbf{X}$



Orthogonal basis of Col($\mathbf{X}$) : $\mathbf{y}_1, \ldots, \mathbf{y}_r$

Without loss of generality, let us number these vectors in decreasing order of *variance*

$$Var(\mathbf{y}_j) = \mathbf{y}_j^T\mathbf{y}_j / (n-1) \quad \text{for } j = 1,\ldots,r$$

The first principal component vector $\mathbf{y}_1$ should account for the most variance. Since $\mathbf{y}_1$ is in $R(\mathbf{X})$, we can represent it as a product $\mathbf{X}\mathbf{x}_1$ for some $\mathbf{x}_1 \in \mathbf{R}^n$.

Then the *covariance matrix* $\mathbf{S}$ is

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

and the variance of $\mathbf{y}_1$ is given by

$$Var(\mathbf{y}_1) = \frac{\mathbf{y}_1^T\mathbf{y}_1}{n-1} = \frac{\mathbf{x}_1^T\mathbf{X}^T\mathbf{X}\mathbf{x}_1}{n-1} = \mathbf{x}_1^T\mathbf{S}\mathbf{x}_1$$

The vector $\mathbf{x}_1$ is chosen to maximize $\mathbf{x}^T\mathbf{S}\mathbf{x}$ over all unit vectors $\mathbf{x}$. This can be accomplished by choosing $\mathbf{x}_1$ to be a unit eigenvector $\mathbf{v}_1$ of $\mathbf{X}^T\mathbf{X}$ belonging to its maximum eigenvalue λ_1 (Why?)

Since $\mathbf{X}^T\mathbf{X}$ is diagonalzable, it has eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$ and their corresponding orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then a nonzero unit vector $\mathbf{x}$ in $\mathbf{R}^n$ is uniquely represented as $\mathbf{x} = c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n$ and

$$\begin{aligned}\rho(\mathbf{x}) &= \mathbf{x}^T(\mathbf{X}^T\mathbf{X})\mathbf{x} = (c_1\mathbf{v}_1^T + \ldots + c_n\mathbf{v}_n^T)\,\mathbf{X}^T\mathbf{X}\,(c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n)\\ &= (c_1\mathbf{v}_1^T + \ldots + c_n\mathbf{v}_n^T)\,(c_1\lambda_1\mathbf{v}_1 + \ldots + c_n\lambda_n\mathbf{v}_n)\\ &= c_1^2\lambda_1 + \ldots + c_n^2\lambda_n\end{aligned}$$

Since $c_1^2 + \ldots + c_n^2 = 1$, $\lambda_n \le \rho(\mathbf{x}) = c_1^2\lambda_1 + \ldots + c_n^2\lambda_n \le \lambda_1$

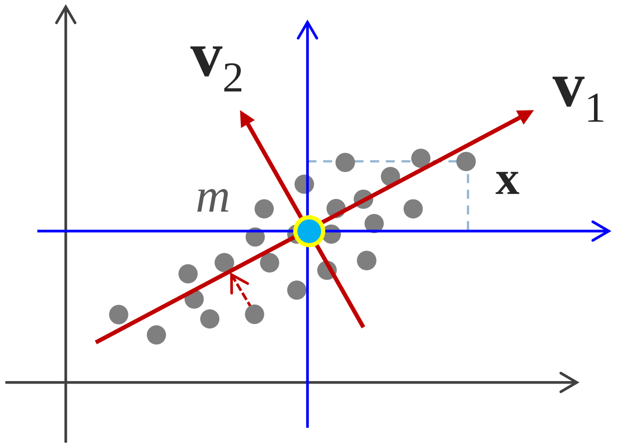
Therefore, the unit vector of maximizing $\rho(\mathbf{x})/(n-1) = \mathbf{x}^T(\mathbf{X}^T\mathbf{X})\mathbf{x}$ is the unit eigenvector $\mathbf{v}_1$ of $\mathbf{X}^T\mathbf{X}$ belonging to its maximum eigenvalue λ_1

On the one hand, if $\mathbf{X} = \mathbf{U\Sigma V}^T$, $\mathbf{XV} = \mathbf{U\Sigma}$. Thus

$$\mathbf{y}_1 = \mathbf{Xv}_1 = \sigma_1\mathbf{u}_1$$

$\mathbf{y}_1$

$(m\times 1)$



PCA Transformation: $\mathbf{p} = \mathbf{V}^T\mathbf{x}$

$$(p_1,p_2) = [\mathbf{v}_1\ \mathbf{v}_2]^T\,(x_1,x_2)$$
$$(x_1,x_2) = p_1\mathbf{v}_1 + p_2\mathbf{v}_2$$
$$\therefore\ (p_1,p_2) = [\mathbf{x}]_B\,,\ \ B = \{\mathbf{v}_1, \mathbf{v}_2\}$$

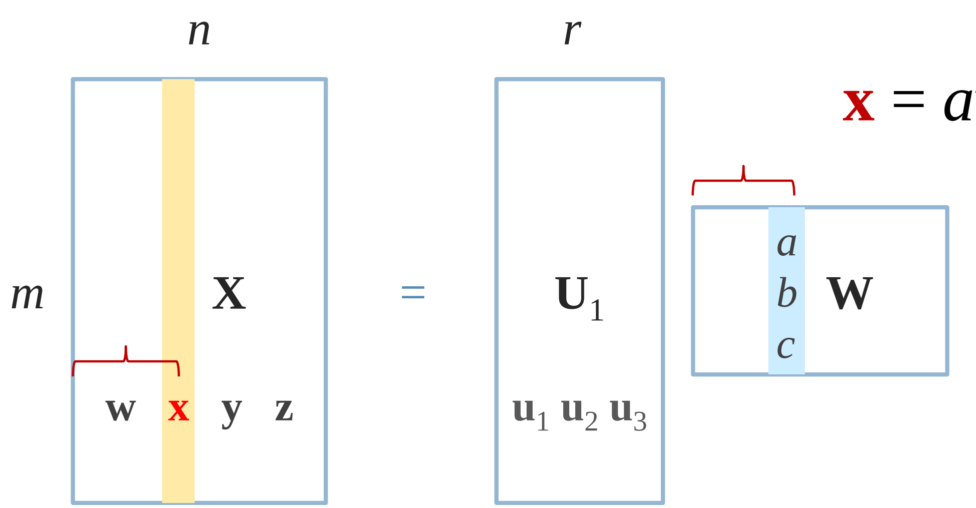
The second principal component vector must be of the form $\mathbf{y}_2 = \mathbf{X}\mathbf{v}_2$, where $\mathbf{v}_2$ is the vector of maximizing $\mathbf{x}^T\mathbf{S}\mathbf{x}$ over all unit vectors that are orthogonal to $\mathbf{v}_1$ and it is just the second right singular vector $\mathbf{v}_2$ of $\mathbf{X}$. Thus

$$\mathbf{y}_2 = \mathbf{X}\mathbf{v}_2 = \sigma_2\mathbf{u}_2$$

In general, the SVD solves the principal component problem. If $\mathbf{X}$ has rank r and SVD $\mathbf{X} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T$ (in compact form), then the principal component vectors are given by

$$\mathbf{y}_1 = \sigma_1\mathbf{u}_1 \text{ , } \mathbf{y}_2 = \sigma_2\mathbf{u}_2 \text{ , } \ldots, \mathbf{y}_r = \sigma_r\mathbf{u}_r$$

If we set $\mathbf{W} = \mathbf{\Sigma}_1 \mathbf{V}_1^{\,T}$, then $\mathbf{X} = \mathbf{U}_1\mathbf{W}$. The columns of $\mathbf{U}_1$ correspond to the hypothetical intelligence factors. The entries in each column measure how well the individuals exhibit that particular intelligent activity. The matrix $\mathbf{W}$ measures to what extent each test depends on the hypothetical factors ■



$$\mathbf{x} = a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3$$

(*Ex*.21) Correlation, Covariance, PCA

We want to see how closely two sets of mathematical scores are correlated. To allow for any differences in difficulty, each score is adjusted to have a mean of 0

	Midterm	Final	Assignments
S1	198	200	196
S2	160	165	165
S3	158	158	133
S4	150	165	91
S5	175	182	151
S6	134	135	101
S7	152	136	80
Average	161	163	131

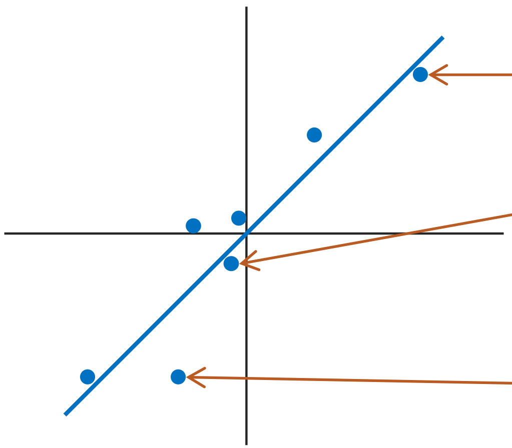
$$\mathbf{X} = \begin{bmatrix} 37 & 37 & 65 \\ -1 & 2 & 34 \\ -3 & -5 & 2 \\ -11 & 2 & -40 \\ 14 & 19 & 20 \\ -27 & -28 & -30 \\ -9 & -27 & -51 \end{bmatrix}$$

$$\begin{array}{ccc} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{array}$$

To see correlation between two scores, the cosine of the angle between them may be computed. A cosine value near to 1 indicates that they are highly correlated, For example,

$$\cos\theta = \mathbf{x}_1^T\mathbf{x}_2 / \|\mathbf{x}_1\|\|\mathbf{x}_2\| \approx 0.92$$

Correlation between actual pairs can be represented as a line $y = mx$. The slope can be determined by the *least squares* method



$$\mathbf{X} = \begin{bmatrix} 37 & 37 & 65 \\ -1 & 2 & 34 \\ -3 & -5 & 2 \\ -11 & 2 & -40 \\ 14 & 19 & 20 \\ -27 & -28 & -30 \\ -9 & -27 & -51 \end{bmatrix}$$

If the column vectors of $\mathbf{X}$ are normalized to have unit length, then the cosine values can be simply computed. Let the normalized matrix be $\mathbf{U}$.

If we set $\mathbf{C} = \mathbf{U}^T\mathbf{U}$, then the (i, j) entry of $\mathbf{C}$ represents the correlation between the ith and jth sets of scores. The matrix $\mathbf{C}$ is called a *correlation matrix*

$$\mathbf{U} = \begin{bmatrix} 0.74 & 0.65 & 0.62 \\ -0.02 & 0.03 & 0.33 \\ -0.06 & -0.09 & 0.02 \\ -0.22 & 0.03 & -0.38 \\ 0.28 & 0.33 & 0.19 \\ -0.54 & -0.49 & -0.29 \\ -0.18 & -0.47 & -0.49 \end{bmatrix}$$

$$\mathbf{C} = \mathbf{U}^T\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0.92 & 0.83 \\ 0.92 & 1 & 0.83 \\ 0.83 & 0.83 & 1 \end{bmatrix}$$

Another statistically important quantity that is closely related to the correlation matrix is the *covariance matrix*. Given k data sets each containing n values of a variable, we can form vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of *deviations from the mean for each set*. The covariance matrix $\mathbf{S}$ is defined by

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X} = \frac{1}{n-1}\left[\mathbf{x}_1 \cdots \mathbf{x}_k\right]^T\left[\mathbf{x}_1 \cdots \mathbf{x}_k\right]_{(n \times k)} \quad ; k \text{ variables}$$

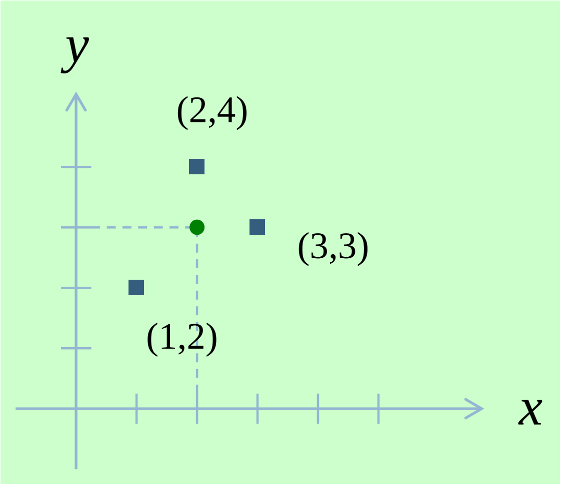
$$= \begin{bmatrix} \sigma_1^2 & \gamma_{12} & \cdots & \gamma_{1k} \\ \gamma_{21} & \sigma_2^2 & \cdots & \gamma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k1} & \gamma_{k2} & \cdots & \sigma_k^2 \end{bmatrix} \qquad \begin{aligned} \sigma_i^2 &= \operatorname{var}(\mathbf{x}_i) = \mathbf{x}_i^T\mathbf{x}_i/(n-1) \\ \gamma_{ij} &= \operatorname{cov}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j/(n-1) \end{aligned}$$

For example, the covariance matrix for the three sets of mathematical scores is

$$\mathbf{S} = \frac{1}{6}\mathbf{X}^T\mathbf{X} = \frac{1}{6}\begin{bmatrix} 2506 & 2625 & 4354 \\ 2625 & 3276 & 4980 \\ 4354 & 4980 & 10886 \end{bmatrix}$$

Let us consider *another example* for easily comparing the correlation matrix, the covariance matrix, and the PCA result

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 3 \end{bmatrix}$$



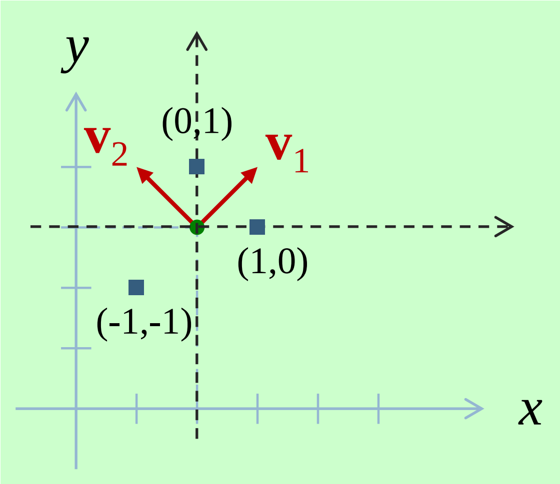
$$\mathbf{X} = \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{C} = \mathbf{U}^T \mathbf{U} = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \quad \mathbf{S} = \frac{1}{2} \mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$$

$$\det(\mathbf{X}^T\mathbf{X}-\lambda\mathbf{I})=\begin{vmatrix}2-\lambda & 1\\ 1 & 2-\lambda\end{vmatrix}=(\lambda-3)(\lambda-1)=0 \quad \therefore \lambda_1=3,\ \lambda_2=1$$

$$\lambda_1=3; \quad \begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix}\sim\begin{bmatrix}-1 & 1\\ 0 & 0\end{bmatrix} \quad \therefore \mathbf{v}_1=\begin{bmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}$$

$$\lambda_2=1; \quad \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\sim\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix} \quad \therefore \mathbf{v}_2=\begin{bmatrix}-1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}$$



$$\mathbf{u}_1=\frac{1}{\sigma_1}\mathbf{X}\mathbf{v}_1=\frac{1}{\sqrt{3}}\begin{bmatrix}-1 & -1\\ 0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}=\begin{bmatrix}-2/\sqrt{6}\\ 1/\sqrt{6}\\ 1/\sqrt{6}\end{bmatrix}$$

$$\mathbf{u}_2=\frac{1}{\sigma_2}\mathbf{X}\mathbf{v}_2=\begin{bmatrix}-1 & -1\\ 0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}-1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}=\begin{bmatrix}0\\ 1/\sqrt{2}\\ -1/\sqrt{2}\end{bmatrix}$$

Principal Component Vectors :

$$\mathbf{y}_1 = \sigma_1 \mathbf{u}_1 = \sqrt{3}\,(-2/\sqrt{6}, 1/\sqrt{6}, 1/\sqrt{6}) = (-\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$$

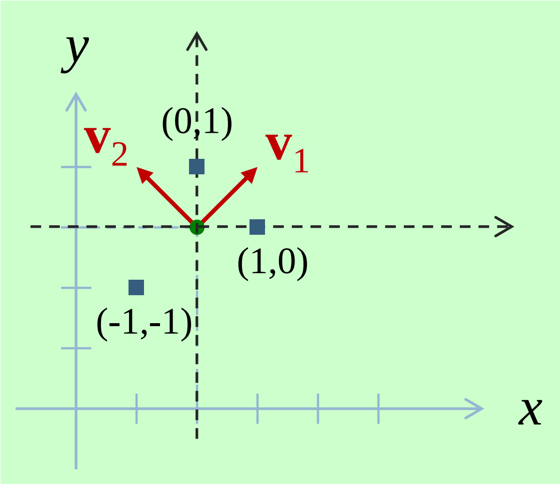
$$\mathbf{y}_2 = \sigma_2 \mathbf{u}_2 = (0, 1/\sqrt{2}, -1/\sqrt{2})$$

PCA Transformation :

$$\left[\mathbf{y}_1\ \mathbf{y}_2\right]^T = \begin{bmatrix} -\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix}$$

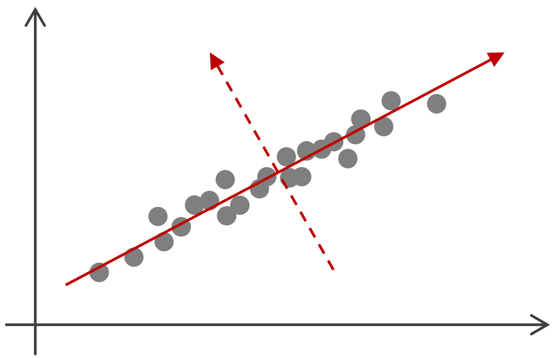
$$= \mathbf{V}^T \left[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3\right] = \mathbf{V}^T \mathbf{X}^T$$

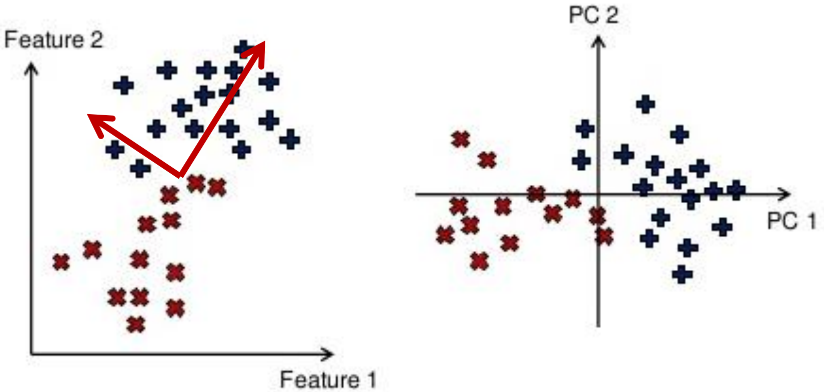


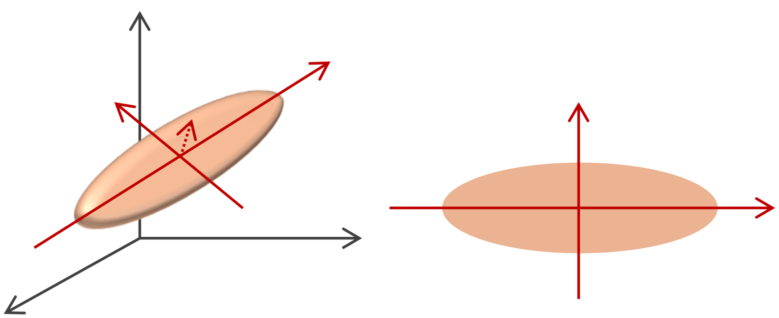
$$\mathbf{X}\mathbf{V} = \mathbf{U}\boldsymbol{\Sigma} = [\mathbf{y}_1\ \mathbf{y}_2] = \mathbf{Y} \quad \therefore\ \mathbf{Y}^T = \mathbf{V}^T \mathbf{X}^T$$

Applications of PCA

- Dimension Reduction, Image Compression, Pattern Classification, etc







Quadratic Forms

To know the definition of quadratic form and its applications

1. Eigenvalues and Eigenvectors
2. Diagonalization
3. Hermitian Matrices
4. Singular Value Decomposition
5. Quadratic Forms

Introduction

- Quadratic Expression

$$ax^2 + by^2 + cxy + dx + ey + f \quad \text{or}$$

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + kz + l$$

- We are interested in only the quadratic (or principal) terms. The sum forms of those are called *quadratic forms*, which can be written as matrix product $\mathbf{x}^T\mathbf{A}\mathbf{x}$

$$3x^2 + 7y^2 - 2xy = [\,x \quad y\,]\begin{bmatrix} 3 & -1 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \text{ or } [\,x \quad y\,]\begin{bmatrix} 3 & -2 \\ 0 & 7 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

- We prefer the first decomposition over the second, because the first square matrix is *symmetric* and *orthogonally diagonalizable*

$$ax^2 + by^2 + cxy = [\,x \quad y\,]\begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz = [\,x \quad y \quad z\,]\begin{bmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

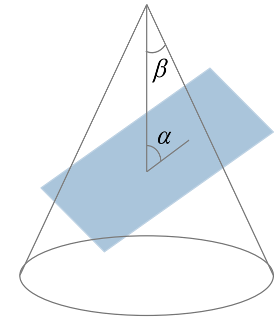
Def. Quadratic Form

A quadratic form (in a n variables) is a function $q : \mathbf{R}^n \to \mathbf{R}$ of the form

$$q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

for some symmetric $n \times n$ matrix $\mathbf{A}$ and any n-vector $\mathbf{x}$. We say that $\mathbf{A}$ is the matrix associated with q

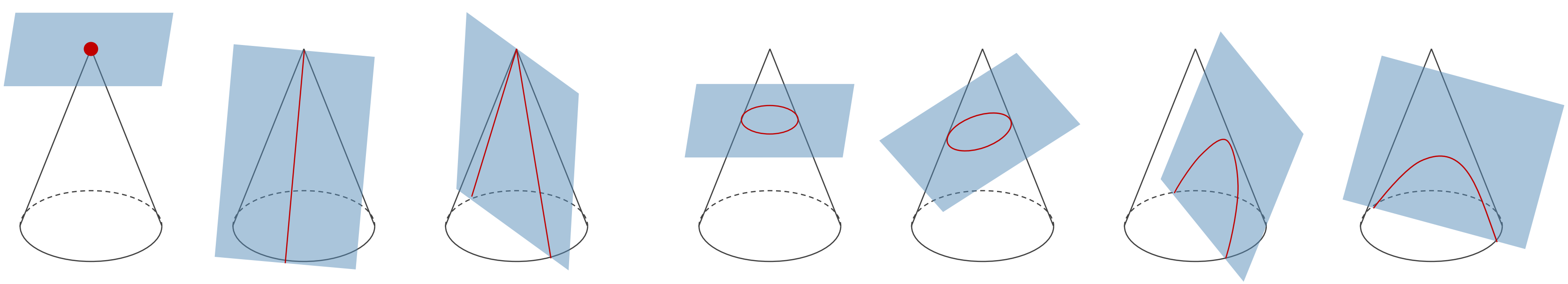
Conic Sections



A graph of a quadratic equation $ax^2 + bxy + cy^2 + dx + ey + f = 0$ is called a conic section

(1) Degenerate : a single point, a line, a pair of lines

(2) Non-degenerate : a circle, an ellipse, a parabola, a hyperbola



$\alpha = 90$ $\quad \beta < \alpha < 90$ $\quad \alpha = \beta$ $\quad 0 \leq \alpha < \beta$

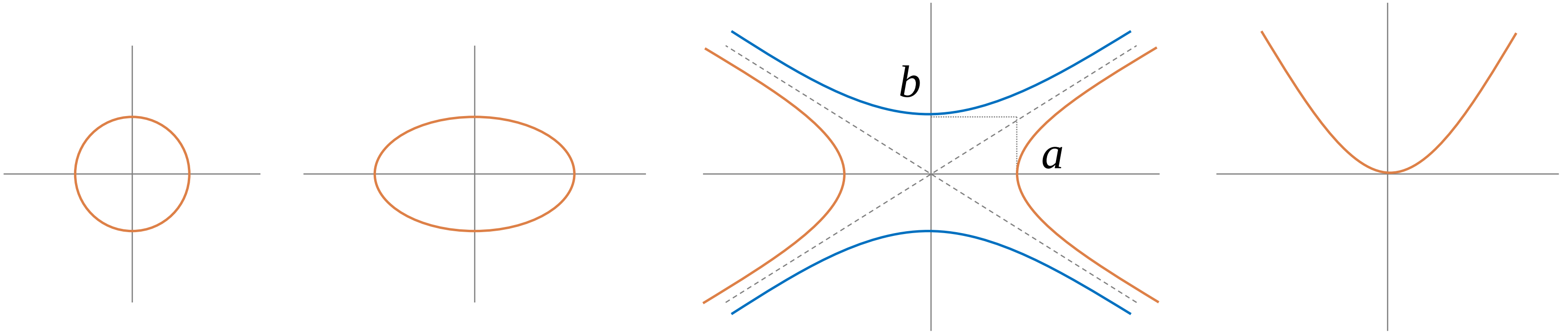
Conic sections in standard position

- They can be easily sketched on x-y plane

circle $x^2 + y^2 = r^2$ ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$

hyperbola $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$ or $-\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$

parabola $x^2 = ay$ or $y^2 = ax$



- What about the conics that are not in standard position?

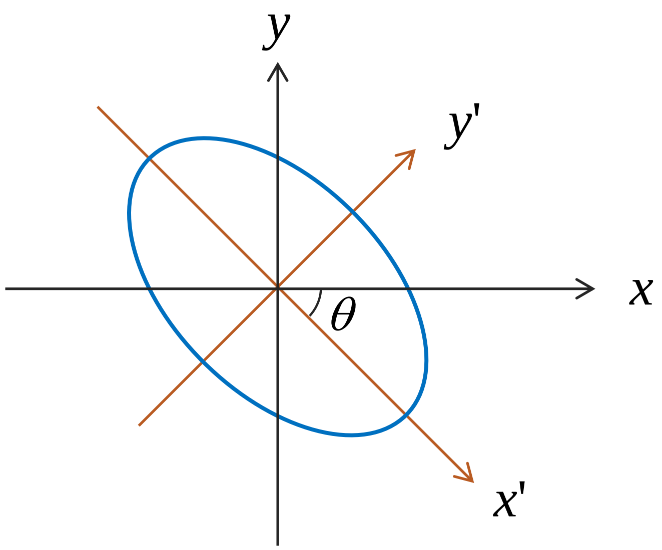
$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

Case 1. The conic section is translated horizontally when the x^2 and x terms have nonzero coefficients

Case 2. The conic section is translated vertically when the y^2 and y terms both have nonzero coefficients

Case 3. The conic section is rotated from the standard position by an angle θ that is not a multiple of $90°$ when the coefficient of the xy term is nonzero

- There is little problem if the center or vertex of the conic section has been translated.
- However, if the conic section has been rotated from the standard position, it is necessary to change coordinates so that the equation in terms of the new x'-y' coordinates involves no $x'y'$ term



Let $\mathbf{x} = (x,y)$ and $\mathbf{x}' = (x',y')$. Since the new coordinates differ from the old coordinates by a rotation, we have

$$\mathbf{x} = \mathbf{Q}\mathbf{x}' \quad \text{or} \quad \mathbf{x}' = \mathbf{Q}^T\mathbf{x}$$

where

$$\mathbf{Q} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \mathbf{Q}^T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

If $0 < \theta < \pi$, then the matrix $\mathbf{Q}$ corresponds to a rotation of radians in the clockwise direction and $\mathbf{Q}^T$ in the counterclockwise direction. With this change of variables, the equation

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

$$\mathbf{x}^T\mathbf{A}\mathbf{x} + [d \quad e]\,\mathbf{x} + f = 0$$

$$\Rightarrow (\mathbf{Q}\mathbf{x}')^T\mathbf{A}(\mathbf{Q}\mathbf{x}') + [d \quad e](\mathbf{Q}\mathbf{x}') + f = 0$$

becomes

$$(\mathbf{x}')^T \mathbf{Q}^T \mathbf{A} \mathbf{Q}\, \mathbf{x}' + [\, d' \;\; e' \,]\, \mathbf{x}' + f = 0$$

This equation will involve no $x'y'$ term if and only if $\mathbf{Q}^T\mathbf{A}\mathbf{Q}$ is diagonal. Since $\mathbf{A}$ is symmetric, it is possible to find a pair of *orthonormal eigenvectors* $\mathbf{q}_1 = (x_1, -y_1)$ and $\mathbf{q}_2 = (y_1, x_1)$. Thus, if we set $\cos\theta = x_1$ and $\sin\theta = y_1$, then

$$\mathbf{Q} = [\, \mathbf{q}_1 \;\; \mathbf{q}_2 \,] = \begin{bmatrix} x_1 & y_1 \\ -y_1 & x_1 \end{bmatrix}$$

diagonalizes $\mathbf{A}$ and the above equation simplifies to

$$\lambda_1 (x')^2 + \lambda_2 (y')^2 + d'x' + e'y' + f = 0$$

(*Ex*.22) Consider the conic section

$$3x^2 + 2xy + 3y^2 - 8 = 0$$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = [\, x \ y \,] \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 8$$

The matrix $\mathbf{A}$ has eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 4$, with corresponding unit eigenvectors $\mathbf{q}_1 = (1/\sqrt{2}, -1/\sqrt{2})$ and $\mathbf{q}_2 = (1/\sqrt{2}, 1/\sqrt{2})$. Let $\mathbf{Q} = [\ \mathbf{q}_1\ \mathbf{q}_2\ ]$ and set $\mathbf{x} = \mathbf{Q}\mathbf{x}'$.

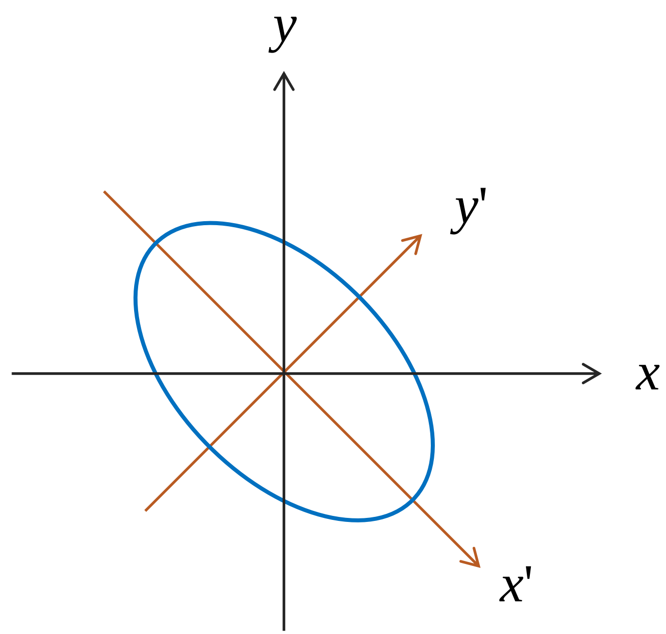
$$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \cos 45^\circ & \sin 45^\circ \\ -\sin 45^\circ & \cos 45^\circ \end{bmatrix}$$

Thus the conic equation becomes

$$(\mathbf{x}')^T \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \mathbf{x}' + [0 \;\; 0]\mathbf{Q}\,\mathbf{x}' - 8 = 0 \quad \leftarrow \; \mathbf{Q}^T\mathbf{A}\mathbf{Q} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

$$\Rightarrow 2(x')^2 + 4(y')^2 = 8 \text{ or } \frac{(x')^2}{4} + \frac{(y')^2}{2} = 1$$

In the new coordinate system, the direction of x'-axis is determined by the point $x' = 1$, $y' = 0$. To translate this to the xy-coordinate system, we multiply $\mathbf{Q}\,[1\ 0]^T = \mathbf{q}_1$. Similarly, the direction of y'-axis is $\mathbf{Q}\,\mathbf{e}_2 = \mathbf{q}_2$. The eigenvectors that form the columns of $\mathbf{Q}$ tell us the direction of the new coordinate axes



~~Homework~~: Draw the following two conics

$$3x^2 + 2xy + 3y^2 + 8\sqrt{2}\,y - 4 = 0 \text{ and}$$

$$-x^2 - y^2 + 6xy - 1 = 0$$

$$3x^2+2xy+3y^2+8\sqrt{2}y-4=[\,x\ \ y\,]\begin{bmatrix}3&1\\1&3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}+\begin{bmatrix}0&8\sqrt{2}\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}-4=0$$

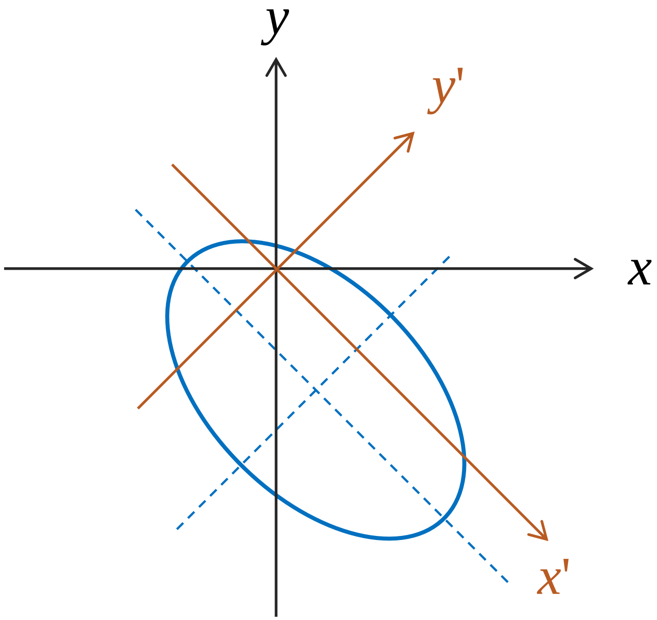
$$\mathbf{Q}^T\mathbf{A}\mathbf{Q}=\begin{bmatrix}1/\sqrt{2}&-1/\sqrt{2}\\1/\sqrt{2}&1/\sqrt{2}\end{bmatrix}\begin{bmatrix}3&1\\1&3\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&1/\sqrt{2}\\-1/\sqrt{2}&1/\sqrt{2}\end{bmatrix}=\begin{bmatrix}2&0\\0&4\end{bmatrix}=\mathbf{D}$$

$\mathbf{q}_1$ $\mathbf{q}_2$

$$2x'^2+4y'^2-8x'+8y'-4=0 \quad \leftarrow \quad \mathbf{x}=\mathbf{Q}\mathbf{x}'$$

$$2(x'-2)^2+4(y'+1)^2-16=0$$

$$\frac{(x'-2)^2}{(2\sqrt{2})^2}+\frac{(y'+1)^2}{2^2}=1$$



Summary

A quadratic equation in the variables x and y can be written in the form

$$\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x} + f = 0 \qquad (1)$$

where $\mathbf{x} = (x,y)$, $\mathbf{A}$ is a 2×2 symmetirc matrix, $\mathbf{B}$ is a 1×2 matrix, and f is a scalar

If $\mathbf{A}$ is nonsingular, then, by rotating and translating the axes, it is possible to rewrite the equation in the form

$$\lambda_1(x')^2 + \lambda_2(y')^2 + f' = 0 \qquad (2)$$

where λ_1 and λ_2 are the eigenvalues of $\mathbf{A}$

If the equation (2) represents a real nondegerate conic, it will be either an ellipse or a hyperbola, depending on whether λ_1 and λ_2 agree in sign or differ in sign.

If $\mathbf{A}$ is singular and exactly one of its eigenvalues is zero, the quadratic equation can be reduced to either

$$\lambda_1(x')^2 + e'y' + f' = 0 \quad \text{or} \quad \lambda_2(y')^2 + d'x' + f' = 0$$

These equations will represent parabolas, provided that e' and d' are nonzero

There is no reason to limit ourselves to two variables. Indeed, a quadratic equation in n variables is also one of the form in (1)

□ Th. 6.6.1 *Principal Axes Theorem*

If $\mathbf{A}$ is a real symmetric $n \times n$ matrix, then there is a change of variables $\mathbf{u} = \mathbf{Q}^T\mathbf{x}$ such that $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{u}^T\mathbf{D}\mathbf{u}$, where $\mathbf{D}$ is a diagonal matrix

(*Proof*)

If $\mathbf{A}$ is a real symmetric matrix, there is an orthogonal matrix $\mathbf{Q}$ that diagonalizes $\mathbf{A}$; that is, $\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{D}$. If we set $\mathbf{u} = \mathbf{Q}^T\mathbf{x}$, then $\mathbf{x} = \mathbf{Q}\mathbf{u}$ and

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{u}^T\mathbf{Q}^T\mathbf{A}\mathbf{Q}\mathbf{u} = \mathbf{u}^T\mathbf{D}\mathbf{u}$$

An Application to Calculus

Definition. Let $F(\mathbf{x})$ be a real-valued vector function on $\mathbf{R}^n$. A point $\mathbf{x}_0$ in $\mathbf{R}^n$ is said to be a *stationary point* of F if all the first partial derivatives of F at $\mathbf{x}_0$ exist and are zero

If $F(\mathbf{x})$ has either a local maximum or a local minimum at a point $\mathbf{x}_0$ and the first partials of F exist at $\mathbf{x}_0$, they will all be zero. Thus, if $F(\mathbf{x})$ has first partial everywhere, its local maxima and minima will occur at stationary points

Consider the quadratic form

$$f(x,y) = ax^2 + 2bxy + cy^2 = \mathbf{x}^T\mathbf{A}\mathbf{x} = [\,x\ \ y\,]\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

The first partials of f are $f_x = 2ax + 2by$ and $f_y = 2bx + 2cy$. Setting these equal to zero, we see that $(0,0)$ is a stationary point. Moreover, if the matrix $\mathbf{A}$ is *nonsingular*, this will be the only critical point.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = 2\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = 2\mathbf{A}\mathbf{x} = \mathbf{0}$$

Thus, if $\mathbf{A}$ is nonsingular, f will have either a global minimum, a global maximum, or a saddle point at $(0,0)$

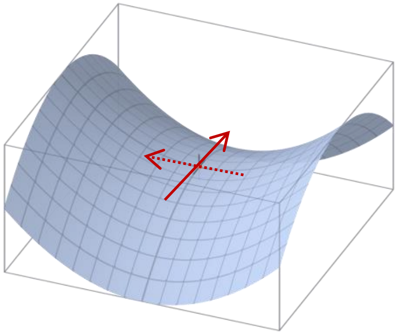
Let us write f in the form

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} \quad \text{where } \det(\mathbf{A}) \neq 0 \text{ and } \mathbf{x} = (x,y)$$

Since $f(\mathbf{0}) = 0$, it follows that

1. f will have a global minimum at $\mathbf{0}$ if and only if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$
2. f will have a global maximum at $\mathbf{0}$ if and only if $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$
3. $\mathbf{0}$ is a saddle point if $\mathbf{x}^T\mathbf{A}\mathbf{x}$ changes sign

$$f(x) = x^2 - y^2 = [\,x \;\; y\,]\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$



Def. Definite for Quadratic Forms

A quadratic form $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ is said to be *definite* if it takes on *only one sign* as $\mathbf{x}$ varies over all nonzero vectors in $\mathbf{R}^n$.

This form is *positive definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all nonzero $\mathbf{x}$ in $\mathbf{R}^n$ and *negative definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$ for all nonzero $\mathbf{x}$.

A quadratic form is said to be *indefinite* if it takes on values that differ in sign.

If $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ and assumes the value 0 for some $\mathbf{x} \neq \mathbf{0}$, then $f(\mathbf{x})$ is said to be *positive semidefinite*

If $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0$ and assumes the value 0 for some $\mathbf{x} \neq \mathbf{0}$, then $f(\mathbf{x})$ is said to be *negative semidefinite*

Def. Definite for Matrices

A *real symmetric* matrix $\mathbf{A}$ is said to be

1. *positive definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all nonzero $\mathbf{x}$ in $\mathbf{R}^n$
2. *negative definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$ for all nonzero $\mathbf{x}$ in $\mathbf{R}^n$
3. *positive semidefinite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all nonzero $\mathbf{x}$ in $\mathbf{R}^n$
4. *negative semidefinite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0$ for all nonzero $\mathbf{x}$ in $\mathbf{R}^n$
5. *indefinite* if $\mathbf{x}^T\mathbf{A}\mathbf{x}$ takes on values that differ in sign

Th. 6.6.2

Let $\mathbf{A}$ be a real symmetric $n \times n$ matrix. Then $\mathbf{A}$ is positive definite if and only if all its eigenvalues are positive

(*Proof*)

If $\mathbf{A}$ is positive definite and λ is an eigenvalue of $\mathbf{A}$, then, for any eigenvector $\mathbf{v}$ belonging to λ,

$$\mathbf{v}^T\mathbf{A}\mathbf{v} = \lambda\mathbf{v}^T\mathbf{v} = \lambda\|\mathbf{v}\|^2 > 0$$

Hence

$$\lambda = \frac{\mathbf{v}^T\mathbf{A}\mathbf{v}}{\|\mathbf{v}\|^2} > 0$$

Conversely, suppose that all the eigenvalues of $\mathbf{A}$ are positive. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an *orthonormal* set of eigenvectors of $\mathbf{A}$. If $\mathbf{x}$ is any nonzero vector in $\mathbf{R}^n$, then $\mathbf{x}$ can be written in the form

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \Lambda + a_n\mathbf{v}_n$$

where $a_i = \mathbf{x}^T\mathbf{v}_i$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} a_i^2 = \|\mathbf{x}\|^2 > 0$

It follows that

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T(a_1\lambda_1\mathbf{v}_1 + \Lambda + a_n\lambda_n\mathbf{v}_n) = a_1\lambda_1\mathbf{x}^T\mathbf{v}_1 + \Lambda + a_n\lambda_n\mathbf{x}^T\mathbf{v}_n$$

$$= \sum_{i=1}^{n} a_i^2\lambda_i \geq (\min \lambda_i)\|\mathbf{x}\|^2 > 0$$

and hence $\mathbf{A}$ is positive definite. $\blacksquare$

(*Ex*.23) Determine whether the stationary point $(0,0)$ of the quadratic form

$$f(x,y) = 2x^2 - 4xy + 5y^2$$

is a global minimum or a saddle point

(*Solution*)

The matrix $\mathbf{A}$ of the quadratic form has $\lambda_1 = 6$ and $\lambda_2 = 1$. Since both eigenvalues are positive, $\mathbf{A}$ is positive definite and hence the stationary point $(0,0)$ is a global minimum

Suppose that a function $F(x,y)$ with a stationary point (x_0,y_0) has continuous third partials in a neighborhood of (x_0,y_0). Let the Hessian of F at (x_0,y_0) be

$$\mathbf{H} = \begin{bmatrix} F_{xx}(x_0,y_0) & F_{xy}(x_0,y_0) \\ F_{xy}(x_0,y_0) & F_{yy}(x_0,y_0) \end{bmatrix}$$

and let λ_1 and λ_2 be the eigenvalues of $\mathbf{H}$.

If $\mathbf{H}$ is nonsingular, then λ_1 and λ_2 are nonzero and we can classify the stationary points as follows:

1. F has a minimum at (x_0,y_0) if $\lambda_1 > 0$ and $\lambda_2 > 0$
2. F has a maximum at (x_0,y_0) if $\lambda_1 < 0$ and $\lambda_2 < 0$
3. F has a saddle point at (x_0,y_0) if λ_1 and λ_2 differ in sign

(*Ex*.24) Classify the stationary points of the function

$$F(x,y) = (1/3)x^3 + xy^2 - 4xy + 1$$

(*Solution*)

From $F_x = x^2 + y^2 - 4y = 0$ and $F_y = 2x(y-2) = 0,$ we can determine the four stationary points of F, (0,0), (0,4), (2,2), and (-2,2). The second partials of F are

$$F_{xx} = 2x, \quad F_{xy} = 2y - 4, \quad F_{yy} = 2x$$

For each stationary point (x_0, y_0), we determine the eigenvalues of the matrix

$$\begin{bmatrix} 2x_0 & 2y_0 - 4 \\ 2y_0 - 4 & 2x_0 \end{bmatrix}$$ called *Hessian* of F

Stationary Point	λ_1	λ_2	Description
(0,0)	4	-4	Saddle Point
(0,4)	4	-4	Saddle Point
(2,2)	4	4	Local Minimum
(-2,2)	-4	-4	Local Maximum

- This method of classifying stationary points can be generalized to functions of more than two variables. Refer to Example 6 in our text for the function

$$F(x, y, z) = x^2 + xz - 3\cos y + z^2$$