# Principal Component Analysis, Independent Component Analysis

# Singular Value Decomposition (SVD)

Let $\mathbf{X}$ be an $n \times m$ matrix with $n \geq m$ and rank($\mathbf{X}$) = $R$

Then the SVD of $\mathbf{X}$ is $\quad \mathbf{X} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathrm{T}}$

where $\quad \mathbf{U}$ is $n \times m$ and $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_m$

$\mathbf{V}$ is $m \times m$ and $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I}_m$

$\mathbf{S} = \mathrm{diag}\{s_1, \ldots, s_R, 0, \ldots, 0\}$ is $m \times m$

singular values of $\mathbf{X}$ are $s_1 \geq \ldots \geq s_R > 0 \geq \ldots \geq 0$

rank($\mathbf{X}$) = $R$ = # positive singular values of $\mathbf{X}$

SVD is unique (up to sign) for distinct singular values

# Principal Component Analysis (PCA)

$\mathbf{X}$ $(n \times m)$ contains $n$ scores on $m$ variables

PCA model $\mathbf{X} = \mathbf{A}\,\mathbf{B}^{\mathrm{T}} + \mathbf{E}$ $\Leftrightarrow$ $x_{ij} = \sum_{r=1}^{R} a_{ir}\, b_{jr} + e_{ij}$

- variables in $\mathbf{X}$ have mean 0 and variance 1
- $\mathbf{A} = \mathbf{XD}$ $(n \times R)$ contains $R$ components as columns,

  which have mean 0, variance 1 and are uncorrelated
- $\mathbf{B}$ $(m \times R)$ contains loadings of the variables on the components, rank($\mathbf{B}$) = $R$

Objective: Minimize $\left\| \mathbf{X} - \mathbf{A}\,\mathbf{B}^T \right\|^2$

PCA solution $\mathbf{A}\,\mathbf{B}^{\mathrm{T}}$ = truncated SVD of $\mathbf{X}$

$$\mathbf{A}\,\mathbf{B}^{\mathrm{T}} = \mathbf{U}_R\,\mathbf{S}_R\,(\mathbf{V}_R)^{\mathrm{T}}$$

$$\mathbf{A} = n^{1/2}\,\mathbf{U}_R = \mathbf{X}\,\mathbf{V}_R\,(\mathbf{S}_R)^{-1} n^{1/2}$$

$$\mathbf{B}^{\mathrm{T}} = n^{-1/2}\,\mathbf{S}_R\,(\mathbf{V}_R)^{\mathrm{T}}$$

➔ components $\mathbf{A}$ are linear combinations of the data $\mathbf{X}$

➔ components are uncorrelated $\mathbf{A}^{\mathrm{T}}\mathbf{A}/n = (\mathbf{U}_R)^{\mathrm{T}}\mathbf{U}_R = \mathbf{I}_R$

SVD ➔ principal components (columns of **A**) are ordered

➔ explained variances $(s_1)^2/n \geq \ldots \geq (s_R)^2/n$

are the $R$ largest eigenvalues of

$\mathrm{Cov}(\mathbf{X}) = \mathbf{X}^\mathrm{T}\mathbf{X}/n = \mathbf{V}\,\mathbf{S}^2\,\mathbf{V}^\mathrm{T}/n$

## PCA solution can be rotated without loss of fit

$\mathbf{A}\,\mathbf{B}^{\mathrm{T}} = (\mathbf{A}\,\mathbf{Q})(\mathbf{B}\,\mathbf{Q})^{\mathrm{T}}$ for $\mathbf{Q}$ with $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}_R$

$\Rightarrow$ $\left\| \mathbf{X} - \mathbf{A}\,\mathbf{B}^T \right\|^2 = \left\| \mathbf{X} - \mathbf{A}\mathbf{Q}\,\mathbf{Q}^T\mathbf{B}^T \right\|^2$

$\Rightarrow$ uncorrelated components $(\mathbf{A}\,\mathbf{Q})^{\mathrm{T}}(\mathbf{A}\,\mathbf{Q})/n = \mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_R$

- orthogonal rotation **Q** gives a different basis for the component space spanned by the columns of **A**
- components are easier to interpret after a rotation **Q** with simple structure in the loadings matrix (**B Q**)

# Independent Component Analysis (ICA)

**X** $(n \times m)$ contains $n$ scores on $m$ variables

ICA model $\mathbf{X} = \mathbf{A}\,\mathbf{B}^T + \mathbf{E}$ $\Leftrightarrow$ $x_{ij} = \sum_{r=1}^{R} a_{ir}\, b_{jr} + e_{ij}$

- variables in **X** have mean 0
- **A** = **XD** $(n \times R)$ contains $R$ components as columns,

  which have mean 0, variance 1 and are **independent**
- **B** $(m \times R)$ contains loadings of the variables on the components, rank(**B**) = $R$

Objective: Minimize $\left\| \mathbf{X} - \mathbf{A}\,\mathbf{B}^T \right\|^2$

PCA solution $\mathbf{A}\ \mathbf{B}^{T}$ = truncated SVD of $\mathbf{X}$

ICA solution $(\mathbf{A}\ \mathbf{Q})(\mathbf{B}\ \mathbf{Q})^{T}$ such that the components $(\mathbf{A}\ \mathbf{Q})$ are "approximately independent"

➔ ICA criterion of independent components is used to find rotation $\mathbf{Q}$ for the PCA solution

# ICA model: alternative formulation

$$\mathbf{x} = \mathbf{B}\,\mathbf{a}$$

- $\mathbf{x}$ ($m \times 1$) is random vector of $m$ variables
- $\mathbf{a}$ ($R \times 1$) is random vector of $R$ sources
- $\mathbf{B}$ ($m \times R$) is the mixing matrix

Assumptions:

1. $E(\mathbf{x}) = \mathbf{0}$
2. $E(\mathbf{a}) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{a}) = \mathbf{I}_R$
3. random variables in $\mathbf{a}$ are mutually independent
4. $\mathrm{rank}(\mathbf{B}) = R$

# ICA assumption: non-Gaussian sources

Gaussian sources $\mathbf{a}$ ➔ uncorrelated = independent
➔ PCA = ICA

Gaussian sources $\mathbf{a}$ with $\mathrm{Cov}(\mathbf{a}) = \mathbf{I}_R$

➔ $\mathrm{Cov}(\mathbf{Q}\,\mathbf{a}) = \mathbf{Q}\,\mathrm{Cov}(\mathbf{a})\,\mathbf{Q}^T = \mathbf{Q}\,\mathbf{Q}^T = \mathbf{I}_R$

➔ $\mathbf{Q}\,\mathbf{a}$ is independent for any orthogonal rotation $\mathbf{Q}$

# ICA criteria for independence

Central Limit Theorem :

a linear combination of independent non-Gaussian random variables is "more Gaussian" than the variables themselves

➔ find rotation **Q** such that the sources (**Q a**) are "maximally non-Gaussian"

➔ specify a measure of non-Gaussianity

## Fourth-order Cumulants

$$\mathrm{Cum}(a_1,a_2,a_3,a_4) = \mathrm{E}(a_1 a_2 a_3 a_4) - \mathrm{E}(a_1 a_2)\,\mathrm{E}(a_3 a_4) - \mathrm{E}(a_1 a_3)\,\mathrm{E}(a_2 a_4) - \mathrm{E}(a_1 a_4)\,\mathrm{E}(a_2 a_3)$$

Kurtosis $\quad \mathrm{Kurt}(a) = \mathrm{Cum}(a,a,a,a) = \mathrm{E}(a^4) - 3\,(\mathrm{E}(a^2))^2$

Gaussian **a** ➔ Cum(**a**) $(R \times R \times R \times R)$ is all-zero
Independent **a** ➔ Cum(**a**) is diagonal

ICA criterion: minimize sum-of-squares of off-diagonal elements of Cum(**Q a**)

➔ equivalent to maximizing sum-of-squares of Kurt(**Q a**)

# Procedure for finding an ICA solution

1. compute unrotated PCA solution $\mathbf{A} = n^{1/2}\,\mathbf{U}_R$

   $\mathbf{B}^{\mathrm{T}} = n^{-1/2}\,\mathbf{S}_R\,(\mathbf{V}_R)^{\mathrm{T}}$

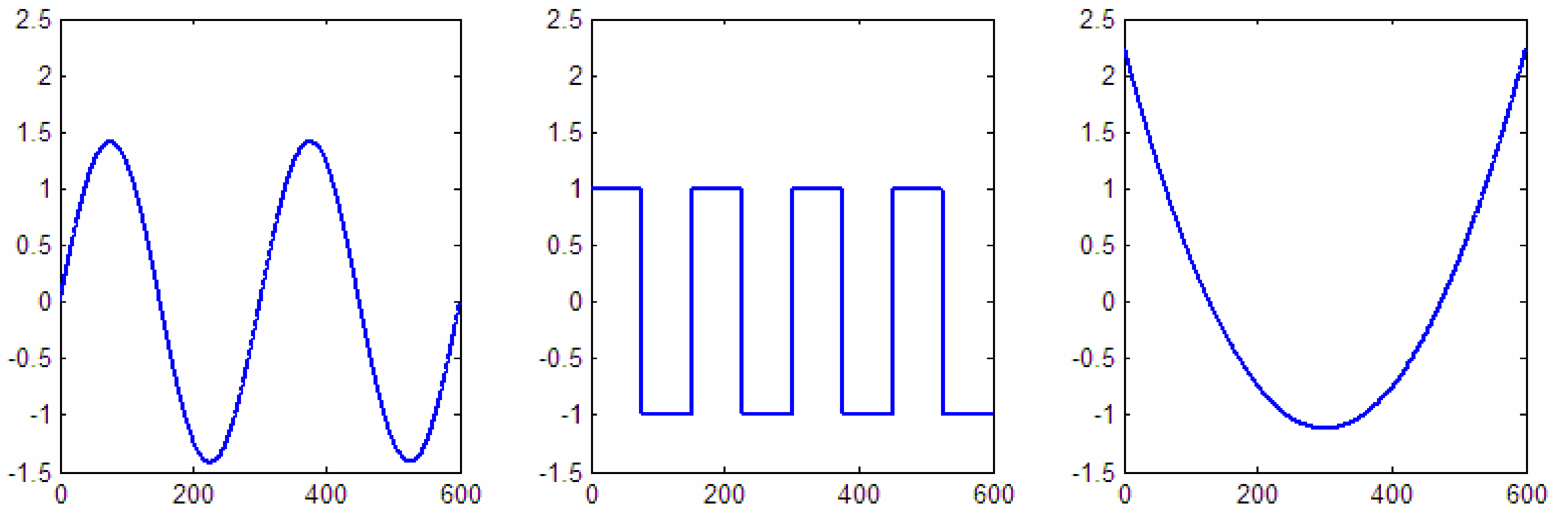
2. compute Kurt(**a**) from samples in **A**

3. find **Q** such that Kurt(**Q** **a**) is maximal

| | | |
|---|---|---|
| algorithms: | MD | (Comon, 1994) |
| | JADE | (Cardoso, 1993) |
| | Fixed-point | (Hyvärinen, 1999) |

4. ICA solution is (**A** **Q**, **B** **Q**)

# PCA versus ICA example



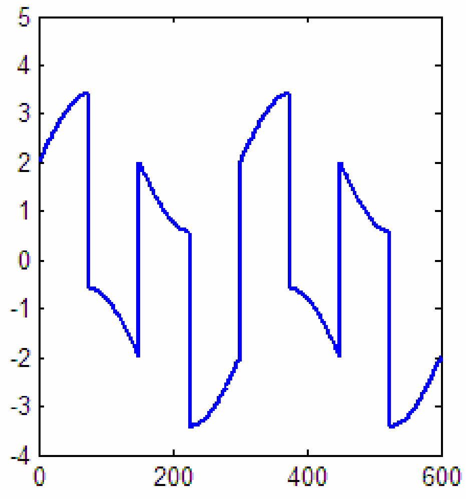

sources: sinus block quadratic

$R = 3$ sources $n = 600$ samples

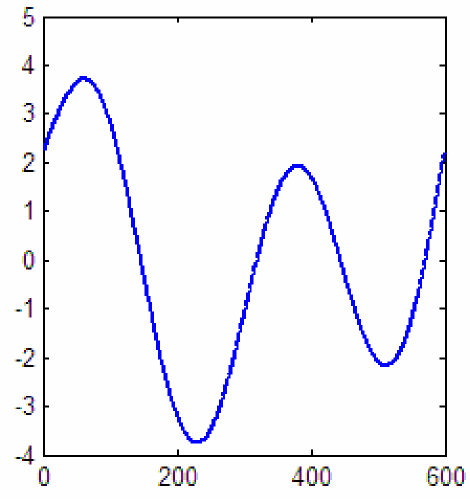
$m = 3$ variables

$$\text{mixing matrix} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

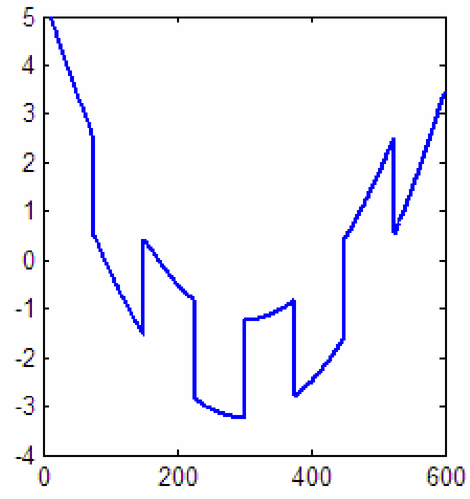
observed variables:




1

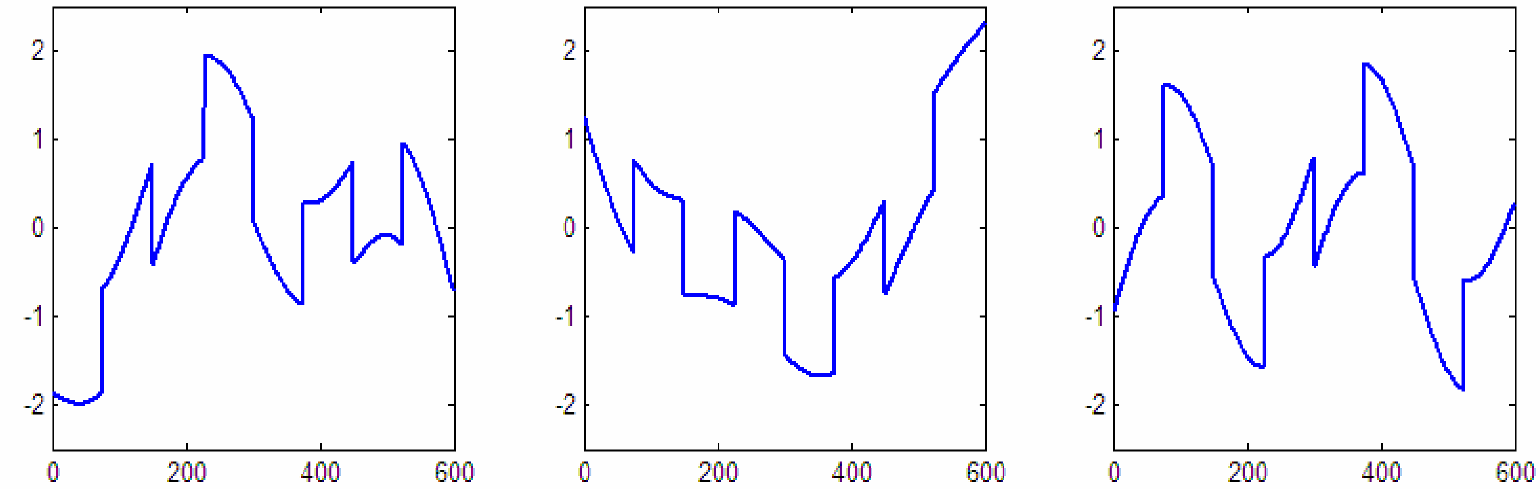



2

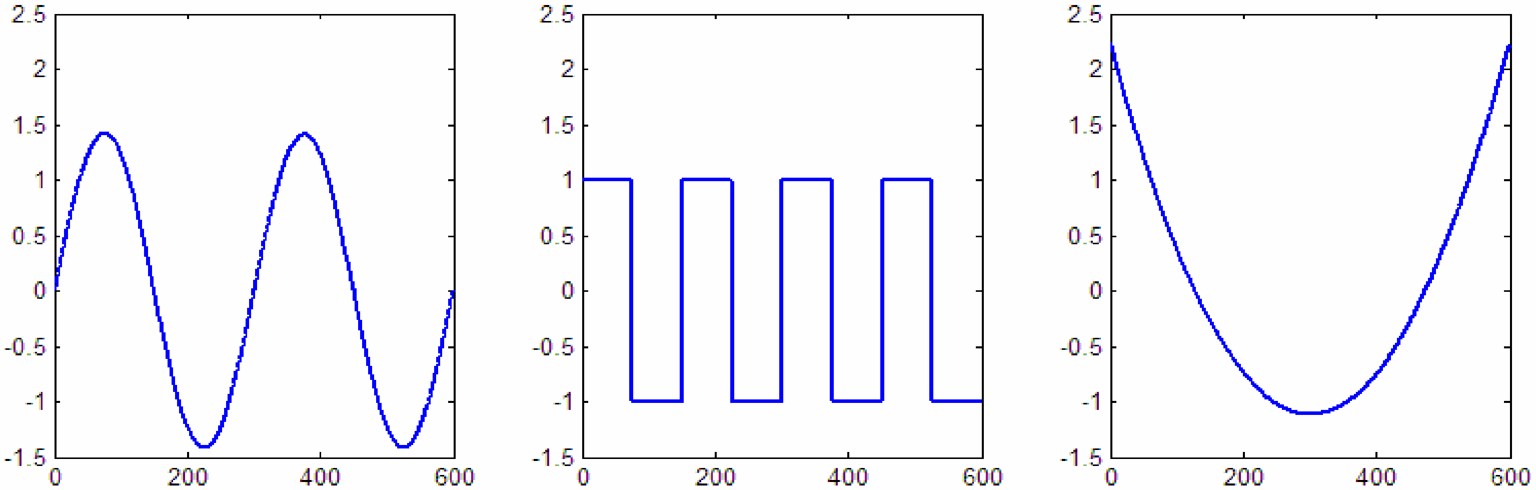



3

# PCA solution:



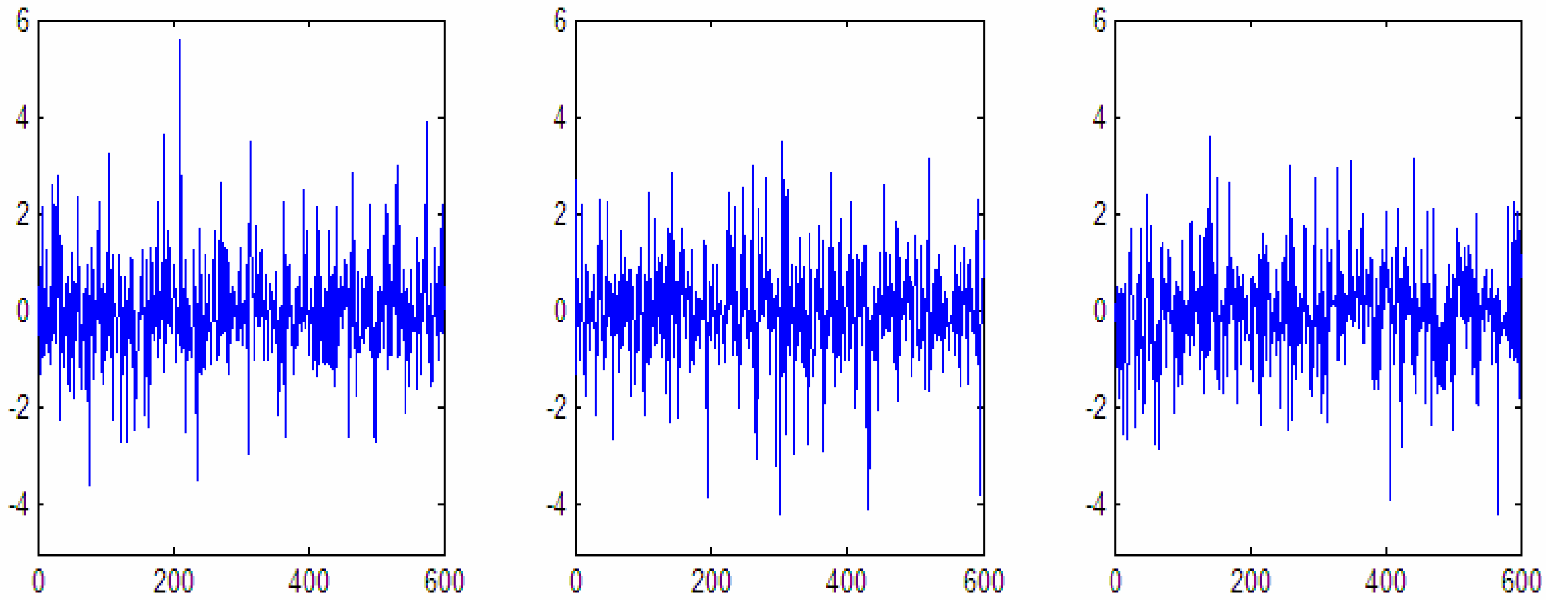

# ICA solution:



$$\text{mixing matrix} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\text{ICA mixing matrix estimate} = \begin{bmatrix} 1.0006 & 2.0016 & -0.0039 \\ 2.0017 & -0.0010 & 0.9966 \\ 0.0075 & 0.9954 & 2.0064 \end{bmatrix}$$

# ICA example with random sources




sources: Laplace distribution $f(x) = \frac{\sqrt{2}}{2} \exp\left(-\sqrt{2}\,|x|\right)$

$R = 3$ sources $\quad n = 600$ samples

$m = 3$ variables $\quad$ mixing matrix $= \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

estimate using kurtosis $= \begin{bmatrix} 0.81 & 2.00 & 0.17 \\ 1.96 & -0.06 & 1.08 \\ -0.17 & 0.83 & 2.06 \end{bmatrix}$

estimate using fastICA $= \begin{bmatrix} 0.89 & 1.97 & -0.07 \\ 1.94 & -0.07 & 1.11 \\ -0.16 & 0.96 & 2.00 \end{bmatrix}$

# Additional remarks about ICA

- ICA solution is unique for non-Gaussian sources (up to signs and permutations)
- PCA step is also known as "prewhitening the data" transformation to $\mathbf{z} = n^{1/2}(\mathbf{S}_R)^{-1}(\mathbf{V}_R)^{\mathrm{T}}\,\mathbf{x} = \mathbf{Q}\,\mathbf{a}$
- PCA versus ICA ≈ second-order versus higher-order
- Other ICA criteria than kurtosis have been formulated
- ICA without initial PCA step is also possible
- ICA applications: signal processing (Blind Source Separation), neuro-imaging (fMRI, EEG), pattern recognition

# PICA = ICA with Gaussian errors

PICA model $\mathbf{x} = \mathbf{B}\,\mathbf{a} + \mathbf{e}$

$\mathbf{e}$ is a Gaussian vector with $E(\mathbf{e}) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{e}) = \sigma^2\,\mathbf{I}_m$

$\mathbf{e}$ and $\mathbf{a}$ are independent

$\mathrm{Cov}(\mathbf{x}) = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \sigma^2\,\mathbf{I}_m$ $\mathrm{Cov}(\mathbf{x}) \approx \mathbf{X}^{\mathrm{T}}\mathbf{X}/n = \mathbf{V}\,\mathbf{S}^2\,\mathbf{V}^{\mathrm{T}}/n$

PICA estimates are: $\mathbf{B} = \mathbf{V}_R\,((\mathbf{S}_R)^2/n - \sigma^2\,\mathbf{I}_R)^{1/2}\,\mathbf{Q}$

$$\mathbf{A} = \mathbf{U}_R\,\mathbf{S}_R\,((\mathbf{S}_R)^2/n - \sigma^2\,\mathbf{I}_R)^{-1/2}\,\mathbf{Q}$$