Guided Tour of Machine Learning in Finance

Week 3-1-2-2: Unsupervised Learning

Principal Component Analysis (PCA)

Igor Halperin

NYU Tandon School of Engineering, 2017

PCA approach: eigen-portfolios

Try PCA as a way to **extract factors** directly from the data

Data: history of $N+1$ daily prices of M stocks $S_n^{(i)} \equiv S_{t_n}^{(i)}$ for $i = 1, \ldots, M$ measured at times $t = [t_0, t_1, \ldots, t_N]$

1. Compute **daily returns:**

$$R_{ni} = \frac{S_n^{(i)} - S_{n-1}^{(i)}}{S_{n-1}^{(i)}} \simeq \log \frac{S_n^{(i)}}{S_{n-1}^{(i)}}, \; n = 1, ..., N, \; i = 1, ..., M$$

2. **Standardized returns** (data normalization):

$$X_{ni} = \frac{R_{ni} - \bar{R}_i}{\bar{\sigma}_i}, \quad \bar{R}_i = \frac{1}{N} \sum_{n=1}^{N} R_{ni}, \; \bar{\sigma}_i^2 = \frac{1}{N-1} \sum_{n=1}^{N} \left(R_{ni} - \bar{R}_i \right)^2$$

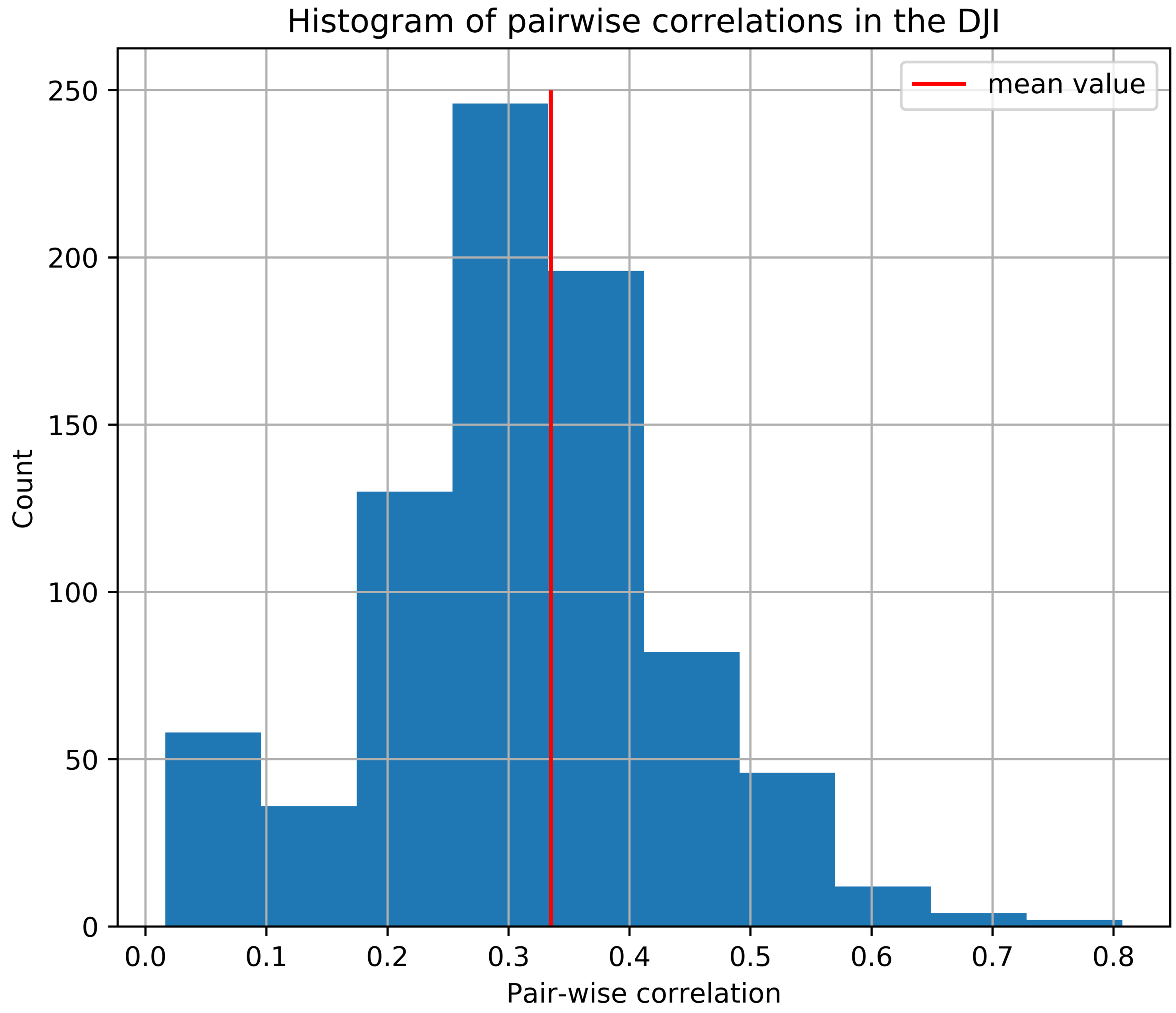
3. The empirical **correlation matrix** is the covariance matrix of standardized returns:

$$C_{ij} = \frac{1}{N-1} \sum_{n=1}^{N} X_{ni} X_{nj} = \frac{1}{N-1} \left(\mathbf{X}^T \mathbf{X} \right)_{ij}$$

This matrix is not diagonal!

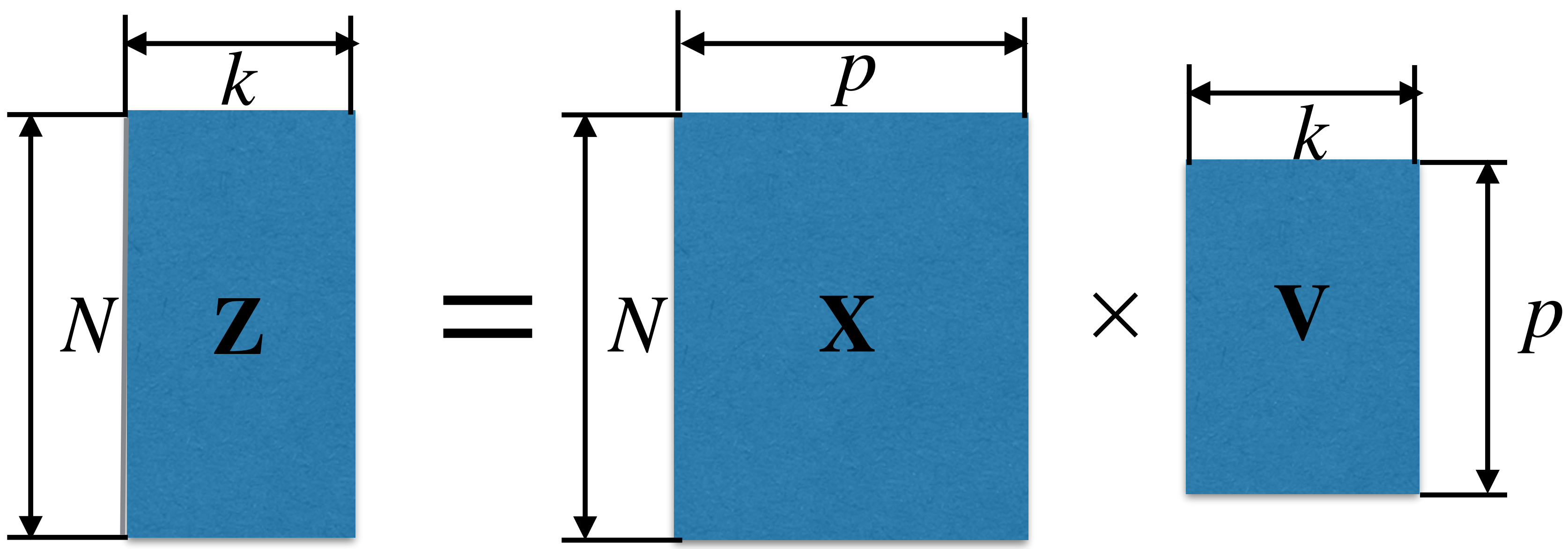
Distribution of pairwise correlations in the DJI

Histogram of pairwise return correlations for stocks in the DJI index:



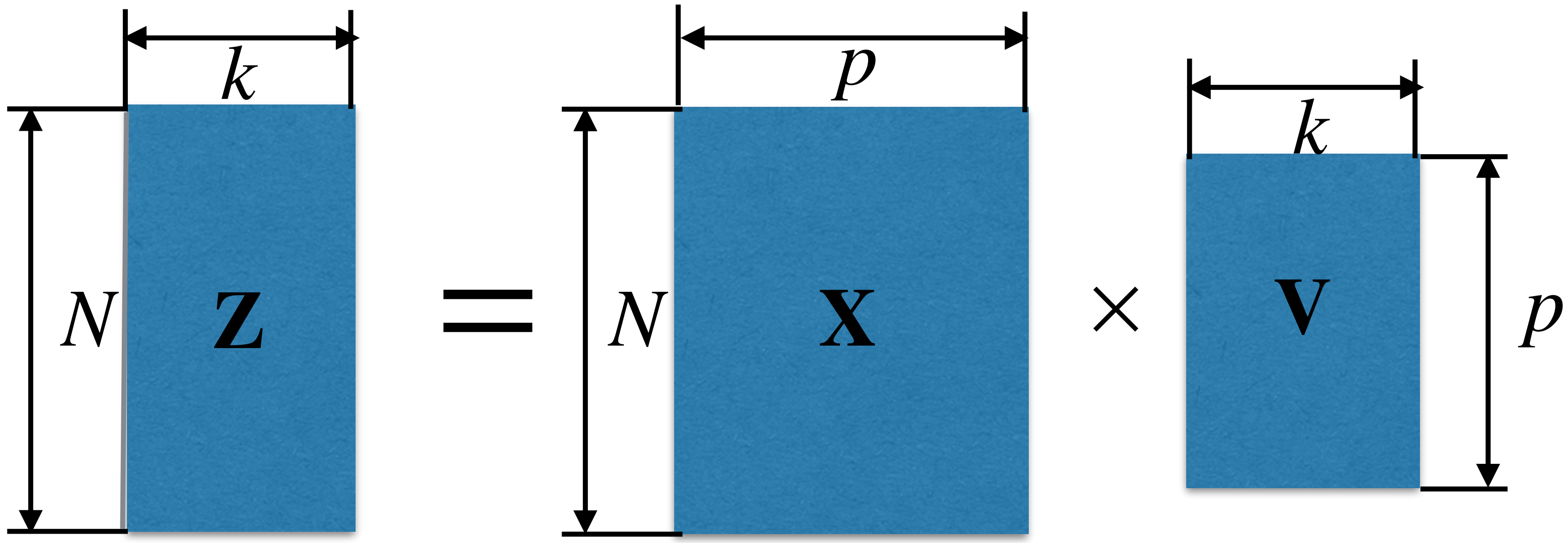
PCA as a coordinate transform

How to make correlation matrix $\mathbf{C}$ diagonal?
Introduce a linear transform (linear encoder) of the data $\mathbf{Z} = \mathbf{X}\mathbf{V}$ parametrized by a $p \times k$ orthogonal matrix $\mathbf{V}$ with $\mathbf{V}\mathbf{V}^T = 1$

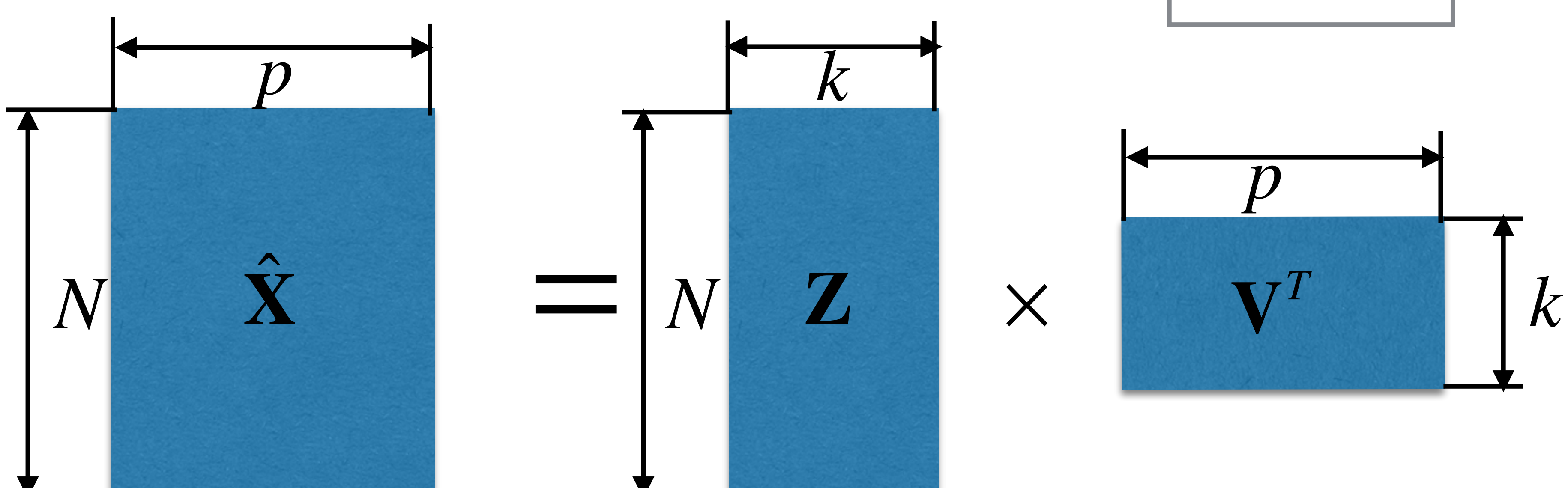


PCA as a coordinate transform

How to make correlation matrix $\mathbf{C}$ diagonal?
Introduce a linear transform (linear encoder) of the data $\mathbf{Z} = \mathbf{XV}$ parametrized by a $p \times k$ orthogonal matrix $\mathbf{V}$ with $\mathbf{VV}^T = 1$



A **decoded signal** (inverse transform) is obtained as $\boxed{\hat{\mathbf{X}} = \mathbf{ZV}^T}$



PCA: Eigenvalue decomposition

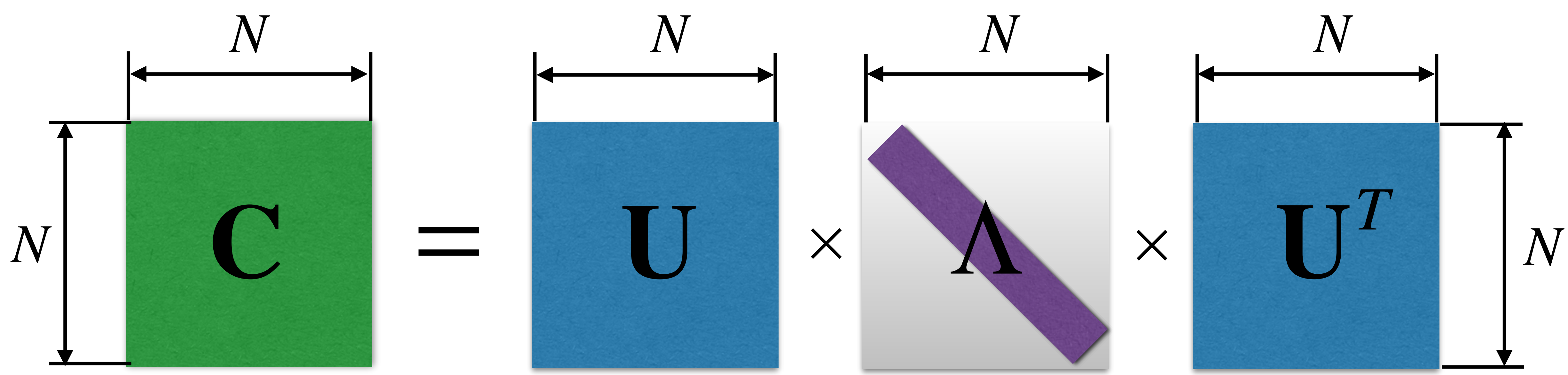
Eigenvalue decomposition of the correlation matrix (note that $\mathbf{C}$ is non-negative definite!)

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

where (assuming that $N > p$):

$\mathbf{\Lambda} = diag(\lambda_1, .., \lambda_p)\ (\lambda_1 \geq \ldots \geq \lambda_p)$ is a diagonal matrix of ordered eigenvalues

$\mathbf{U}$ is a $N \times N$ orthogonal matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$) that stores eigenvectors column-wise



PCA: Eigenvalue decomposition

Eigenvalue decomposition of the correlation matrix (note that $\mathbf{C}$ is non-negative definite!)

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

where (assuming that $N > p$):

$\mathbf{\Lambda} = diag(\lambda_1,..,\lambda_p)\ (\lambda_1 \geq ... \geq \lambda_p)$ is a diagonal matrix of ordered eigenvalues

$\mathbf{U}$ is a $N \times N$ orthogonal matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$) that stores eigenvectors column-wise

Use this to compute the covariance matrix of $\mathbf{Z} = \mathbf{X}\mathbf{V}$

$$Cov[\mathbf{Z}] = \frac{1}{N-1}\mathbf{Z}^T\mathbf{Z} = \frac{1}{N-1}\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{V}^T\mathbf{U}\mathbf{\Lambda}\left(\mathbf{V}^T\mathbf{U}\right)^T$$

PCA: Eigenvalue decomposition

Eigenvalue decomposition of the correlation matrix (note that $\mathbf{C}$ is non-negative definite!)

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

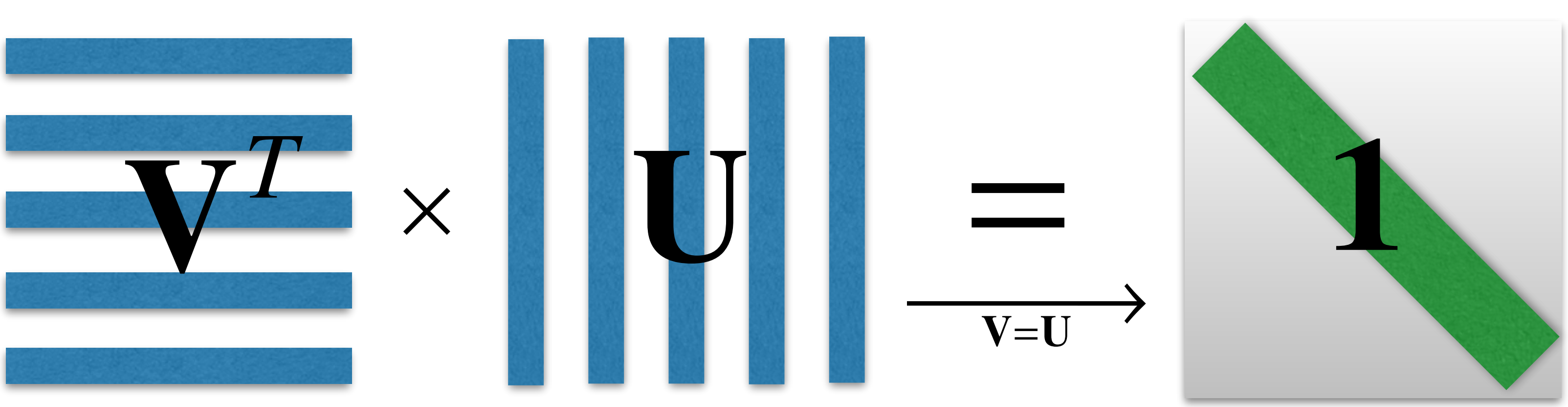
where (assuming that $N > p$):

$\mathbf{\Lambda} = diag(\lambda_1, .., \lambda_p)\ (\lambda_1 \geq \ldots \geq \lambda_p)$ is a diagonal matrix of ordered eigenvalues

$\mathbf{U}$ is a $N \times N$ orthogonal matrix $(\mathbf{U}\mathbf{U}^T = \mathbf{I})$ that stores eigenvectors column-wise

Use this to compute the covariance matrix of $\mathbf{Z} = \mathbf{X}\mathbf{V}$

$$Cov[\mathbf{Z}] = \frac{1}{N-1}\mathbf{Z}^T\mathbf{Z} = \frac{1}{N-1}\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{V}^T\mathbf{U}\mathbf{\Lambda}\left(\mathbf{V}^T\mathbf{U}\right)^T \xrightarrow[\mathbf{V}=\mathbf{U}]{} \mathbf{\Lambda}$$



PCA as a coordinate transform

Eigenvalue decomposition of the correlation matrix (note that $\mathbf{C}$ is non-negative definite!)

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

where (assuming that $N > p$):

$\mathbf{\Lambda} = diag(\lambda_1,..,\lambda_p)\ (\lambda_1 \geq \ldots \geq \lambda_p)$ is a diagonal matrix of ordered eigenvalues

$\mathbf{U}$ is a $N \times N$ orthogonal matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$) that stores eigenvectors column-wise

Use this to compute the covariance matrix of $\mathbf{Z} = \mathbf{X}\mathbf{V}$

$$Cov[\mathbf{Z}] = \frac{1}{N-1}\mathbf{Z}^T\mathbf{Z} = \frac{1}{N-1}\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{V}^T\mathbf{U}\mathbf{\Lambda}\left(\mathbf{V}^T\mathbf{U}\right)^T \xrightarrow[\mathbf{V}=\mathbf{U}]{} \mathbf{\Lambda}$$

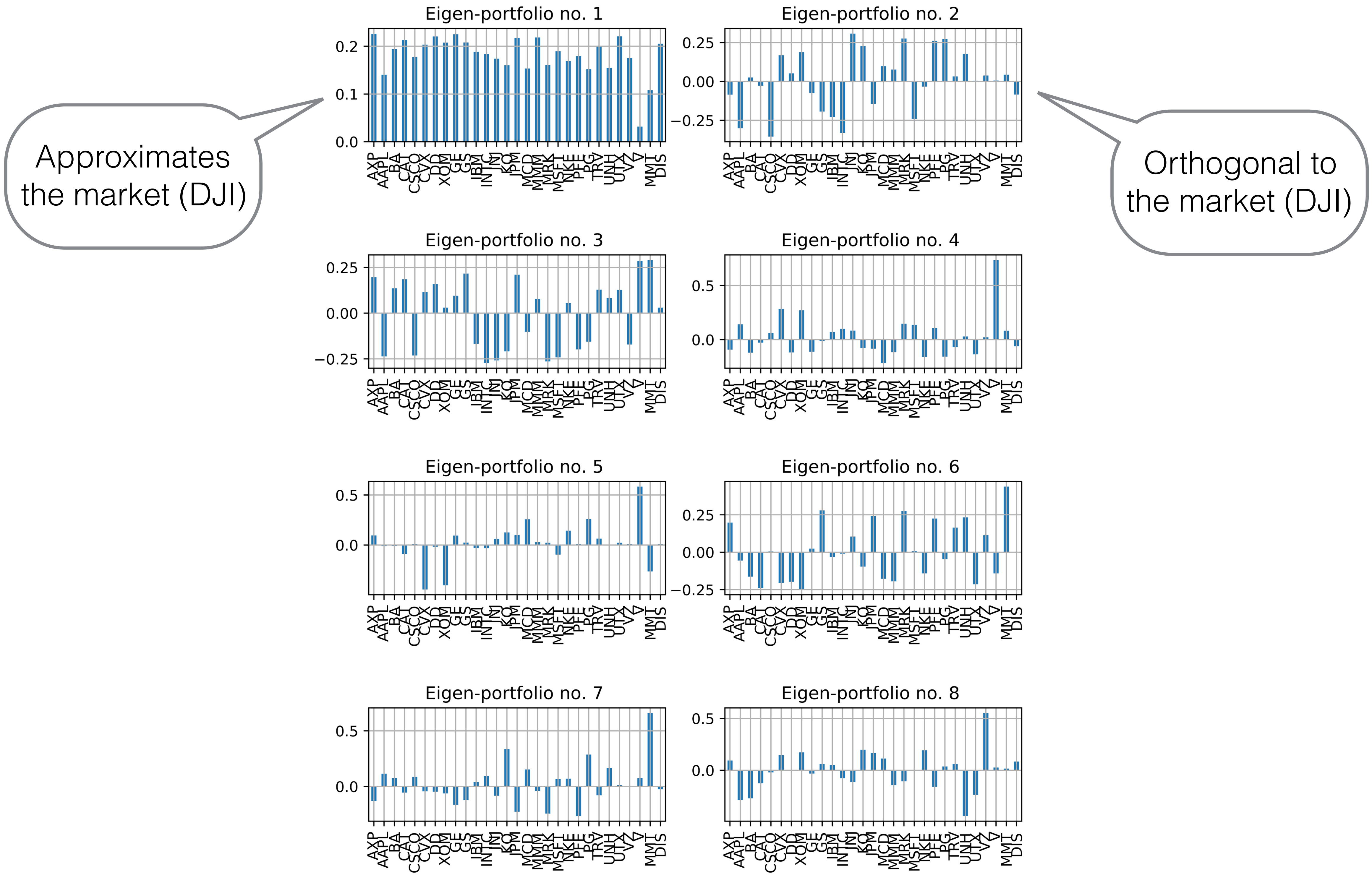
Therefore, when $\mathbf{V} = \mathbf{U}$ (and $k = p$), encoding $\mathbf{Z} = \mathbf{X}\mathbf{V}$ **preserves the total variation of data**

$$TotVar[\mathbf{X}] = \frac{1}{N-1}Tr\left[\mathbf{X}^T\mathbf{X}\right] = \mathrm{Tr}[\mathbf{C}] = \mathrm{Tr}\left[\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\right] = \mathrm{Tr}\left[\mathbf{\Lambda}\mathbf{U}^T\mathbf{U}\right] = \mathrm{Tr}[\mathbf{\Lambda}]$$

$$TotVar[\mathbf{Z}] = \frac{1}{N-1}Tr\left[\mathbf{Z}^T\mathbf{Z}\right] = \mathrm{Tr}\left[\mathbf{V}^T\mathbf{U}\mathbf{\Lambda}\left(\mathbf{V}^T\mathbf{U}\right)^T\right] \xrightarrow[\mathbf{V}=\mathbf{U}]{} \mathrm{Tr}[\mathbf{\Lambda}]$$

PCA for stocks in DJI: eigenvectors

Eigenvectors = weights of stocks in “eigen-portfolios”



Control question

Select all correct answers

1. The PCA implements a Linear Encoder that converts correlated inputs $\mathbf{X}$ into uncorrelated features $\mathbf{Z}$ by a linear transform $\mathbf{Z} = \mathbf{X}\mathbf{V}$
2. If the orthogonal matrix $\mathbf{V}$ has dimension $p \times p$, i.e. it keeps all eigenvectors of correlation matrix of $\mathbf{X}$, then $\mathbf{Z}$ preserves the total variation of $\mathbf{X}$.
3. When the data is noisy or non-stationary, $\mathbf{Z}$ can have a higher variance than $\mathbf{X}$.
4. The PCA is a probabilistic method that enables simulating from data.

Correct answer: 1, 2