Lecture

• Monday – Independent Component Analysis (ICA)
• Wednesday – Practicum
• Monday – Putting it all together: Computational Finance
• Wednesday – Introduction to Topological Analysis
Independent Component Analysis (ICA)

• General form

\[ x_i = A s_i \]

where \( A \) is called the mixing matrix and it is assumed to be invertible.

• Center the data

\[ x_M = \frac{1}{n} \sum_{i=1}^{n} x_i \]

\[ x_i^* = A s_i^* \]
Independent Component Analysis (ICA) (continued)

- It follows that
  \[ \frac{1}{n} \sum_{i=1}^{n} s_i^* = 0 \]

- Form outer product:
  \[ \mathbf{x}\mathbf{x}^T \equiv \begin{pmatrix} x_1^2 & x_1x_2 & x_1x_3 & \ldots & x_1x_m \\ x_1x_2 & x_2^2 & x_2x_3 & \ldots & x_2x_m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1x_m & x_2x_m & x_3x_m & \ldots & x_m^2 \end{pmatrix} \]
Independent Component Analysis (ICA) (continued)

- Re-written as:

\[ x^* = (AC^{-1})(Cs^*) \]

\[
\frac{1}{n} \sum_{i=1}^{n} s_i^* (s_i^*)^T = I
\]

- Recall SVD:

\[ A = U\Sigma V^T \]
Independent Component Analysis (ICA) (continued)

\[
\frac{1}{n} \sum_{i=1}^{n} x_i^* (x_i^*)^T = \frac{1}{n} \sum_{i=1}^{n} (A s_i^*) (A s_i^*)^T \\
= \frac{1}{n} \sum_{i=1}^{n} A (s_i^*) (s_i^*)^T A^T \\
= A \left( \frac{1}{n} \sum_{i=1}^{n} s_i^* (s_i^*)^T \right) A^T \\
= (U \Sigma V^T) (U \Sigma V^T)^T \\
= U \Sigma^2 U^T
\]
Independent Component Analysis (ICA) (continued)

• From Theorem 4.7:

\[
\frac{1}{n} \sum_{i=1}^{n} x_i^* (x_i^*)^T = Q D Q^T
\]

\[
\frac{1}{n} \sum_{i=1}^{n} x_i^* (x_i^*)^T = \frac{1}{n} X^* (X^*)^T
\]
Independent Component Analysis (ICA) (continued)

• Written out:

\[
\frac{1}{n} \sum_{i=1}^{n} x_i^*(x_i^*)^T = \frac{1}{n} \begin{pmatrix}
    c_1^* \cdot c_1^* & c_1^* \cdot c_2^* & c_1^* \cdot c_3^* & \cdots & c_1^* \cdot c_m^* \\
    c_2^* \cdot c_1^* & c_2^* \cdot c_2^* & c_2^* \cdot c_3^* & \cdots & c_2^* \cdot c_m^* \\
    \vdots         & \vdots         & \vdots         & \ddots & \vdots         \\
    c_m^* \cdot c_1^* & c_m^* \cdot c_2^* & c_m^* \cdot c_3^* & \cdots & c_m^* \cdot c_m^*
\end{pmatrix}
\]
Independent Component Analysis (ICA)

- Reduced Problem

\[ A = Q D^{1/2} V^T \]

\[ s_i^* = V y_i^* \quad s_i = W x_i \]

where \( W = V D^{-1/2} Q^T \)

\[ V = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \]
Independent Component Analysis (ICA) (continued)

- Kurtosis

\[ \kappa = \frac{1}{n} \sum_{i=1}^{n} (s_i^*)^4 - 3 \]

\[ K = \sum_{j=1}^{m} |\kappa_j|, \]

where \( \kappa_j = \frac{1}{n} \sum_{i=1}^{n} (s_{ij}^*)^4 - 3 \) and \( s_i^* = (s_{i1}^*, s_{i2}^*, \ldots, s_{im}^*) \).
Independent Component Analysis (ICA) (continued)

• Kurtosis (continued)

\[ K(\theta) = |\kappa_1| + |\kappa_2| \]

where \( \kappa_1 = \frac{1}{n} \sum_{i=1}^{n} (y_{i1}^* \cos \theta - y_{i2}^* \sin \theta)^4 - 3 \)

\[ \kappa_2 = \frac{1}{n} \sum_{i=1}^{n} (y_{i1}^* \sin \theta + y_{i2}^* \cos \theta)^4 - 3 \]

and \( y_i^* = (y_{i1}^*, y_{i2}^*)^T \)
Independent Component Analysis (ICA) (continued)

• Summary of ICA

1. Center the data to produce the data set \( \mathbf{X}^* \).

2. Compute the matrix \( \frac{1}{n} \mathbf{X}^*(\mathbf{X}^*)^T \) and then find its SVD which can be written as \( \frac{1}{n} \mathbf{X}^*(\mathbf{X}^*)^T = \mathbf{QDQ}^T \).

3. Compute \( \mathbf{Y}^* = \mathbf{X}^* \mathbf{QD}^{-1/2} \).

4. Pick a contrast function \( K(\mathbf{S}^*) \). Setting \( \mathbf{S}^* = \mathbf{Y}^* \mathbf{V}^T \), determine the \( m \times m \) orthogonal matrices \( \mathbf{V} \) that minimize \( K \).

5. Selecting one of the \( \mathbf{V} \)'s, the unmixing matrix is \( \mathbf{W} = \mathbf{VD}^{-1/2} \mathbf{Q}^T \) and the corresponding source data are \( \mathbf{S} = \mathbf{XW}^T \).
Independent Component Analysis (ICA) (continued)

- Modal Data Analysis

\[
y_i = Ax_i + b
\]

\[
x_i^* = S_x^{-1}(x_i - x_M)
\]

where \(x_M = \frac{1}{n} \sum x_i\) is the mean, and \(S_x\) is a \(m \times m\) diagonal matrix containing the scaling factors for the components of \(x\). Similarly, the normalized \(y\)’s are \(y_i^* = S_y^{-1}(y_i - y_M)\).
Independent Component Analysis (ICA) (continued)

- Modal Data Analysis (continued)

\[ y_i^* = P x_i^* \]

where \( P \) is called the *propagator matrix*.

\[ EP = \sum_{i=1}^{n} \left\| P x_i^* - y_i^* \right\|_2^2 \]

\[ G p_i = (X^*)^T c_i \]

where \( G = (X^*)^T X^* \) and \( c_i \) is the \( i \)th column of \( Y^* \).
Independent Component Analysis (ICA) (continued)

- Modal Data Analysis (continued)

\[ P = (Y^*)^T X^* G^{-1} \]
\[ P = (Y^*)^T (X^+)^T \]

where \( X^+ = ((X^*)^T X^*)^{-1} (X^*)^T \) is known as the Moore-Penrose pseudoinverse of \( X^* \).

\[ y_i = Ax_i + b \]
\[ A = S_y P S_x^{-1}, \]
Independent Component Analysis (ICA) (continued)

• Modal Data Analysis (continued)

\[ y_i = Ax_i + b \]

\[ A = S_y PS_x^{-1}, \]

and \( b = y_M - Ax_M \)
For Next Time

- Read Chapter 4 and 9 (for volatility)
- Start working on Practicums
- Start working on Homework 3 (due on Thursday)