Independent Component Analysis (ICA)

Presentation to CDAS
September 10, 2007

LTC Scott T. Nestler
Asst Professor and ORCEN Research Analyst

What is essential is invisible to the eye.”

- The Little Prince, Antoine De Saint-Exupéry, 1943

Shameless Plug

Center for Faculty Development (CFD) talk:

- “Mathematical Perspectives on the Federal Thrift Savings Plan (TSP)”
- Thursday, November 15th at 1350
- Thayer 348

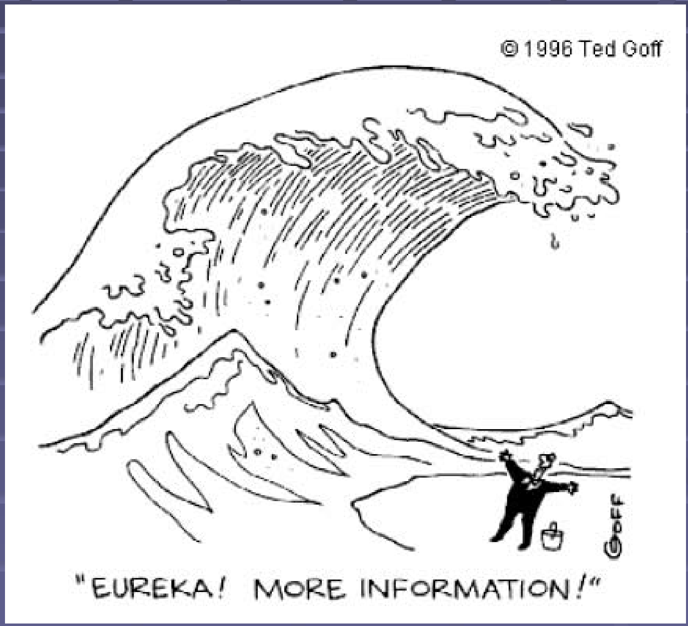
DISCLAIMER: No personal financial advice will be provided at this session.

Outline

- Motivation
- Background and History
- Overview
- Relationship to Other Methods
- Theory and Mechanics of FastICA
- Applications and Examples
- References
- Conclusion

Motivation

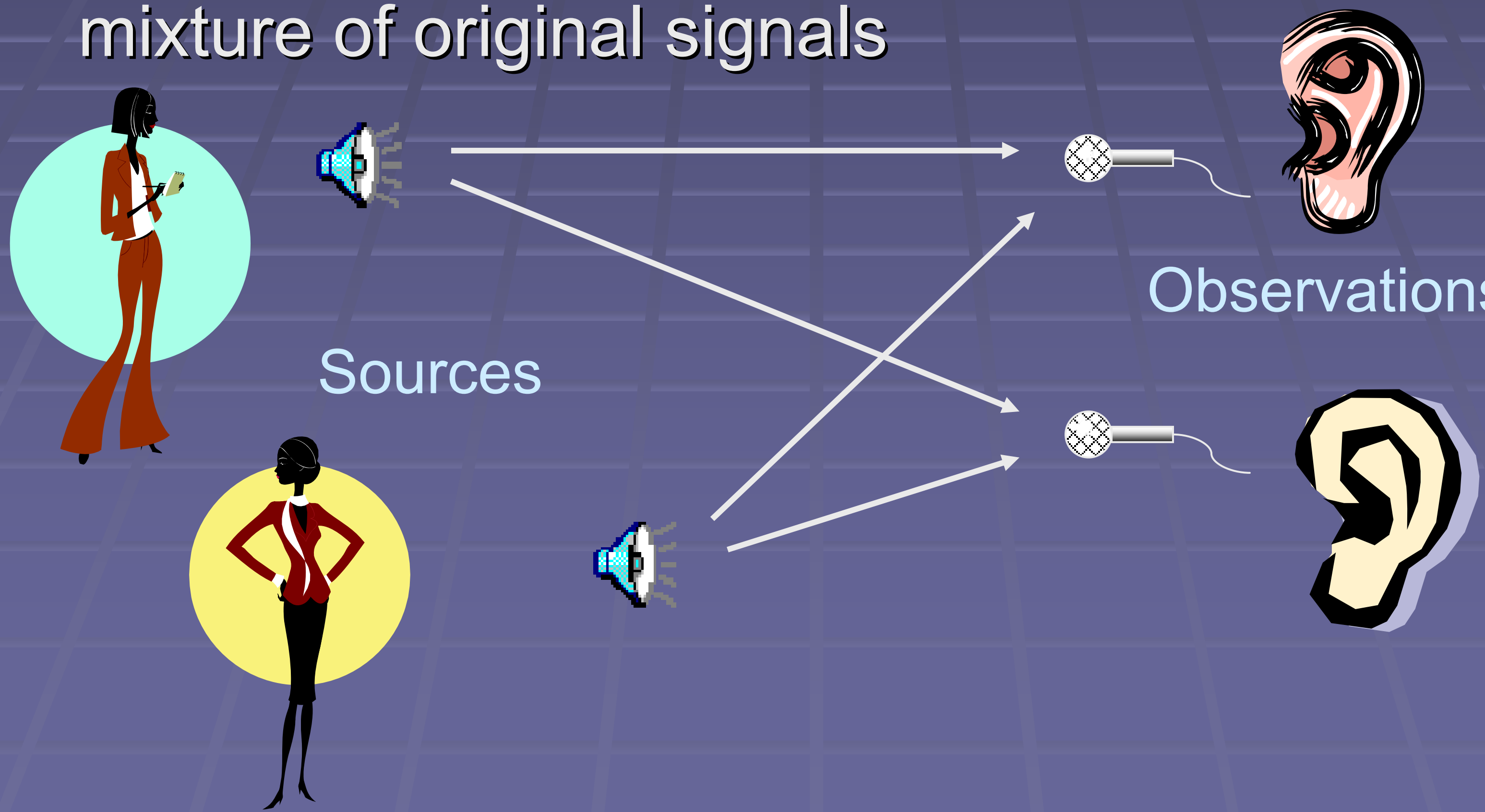
- Do you…
  - Have large amounts of data with relatively small amounts of useful information?
  - Study phenomena that can not be directly observed?
  - Want to discover and exploit hidden relationships?
  - Need a better representation of data without losing too much information?



- ICA may be able to help. It is one method (of many) for:
  - extracting useful information from data
  - revealing the driving forces underlying a set of observed phenomenon

Motivation: Cocktail Party Problem

- Multiple (independent) sound sources in room
- Multiple sensors receiving signals which are mixture of original signals



Background and History

- Early 1980s:
  - First appeared in neurophysiological setting in Europe but hidden behind other techniques for awhile
  - Also has origins in signal processing as “blind source separation”
- Mid 1990s: New algorithms and applications appeared in mid-1990s
- Today: Published works found in signal processing, artificial neural networks, statistics, information theory, and application journals and proceedings

What is ICA?

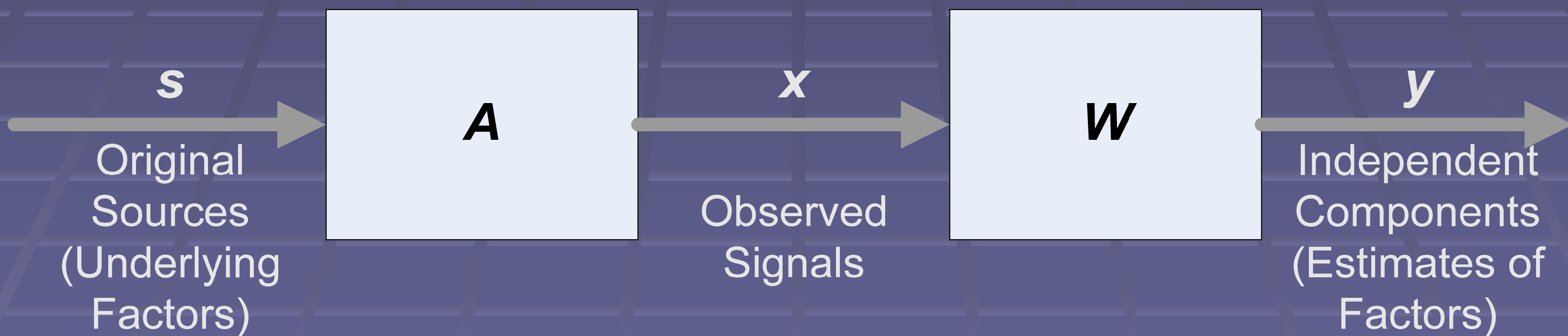
"Independent component analysis (ICA) is a method for finding underlying factors or components from multivariate (multi-dimensional) statistical data. What distinguishes ICA from other methods is that it looks for components that are both *statistically independent*, and *nonGaussian*."

A.Hyvarinen, A.Karhunen, E.Oja

'Independent Component Analysis'

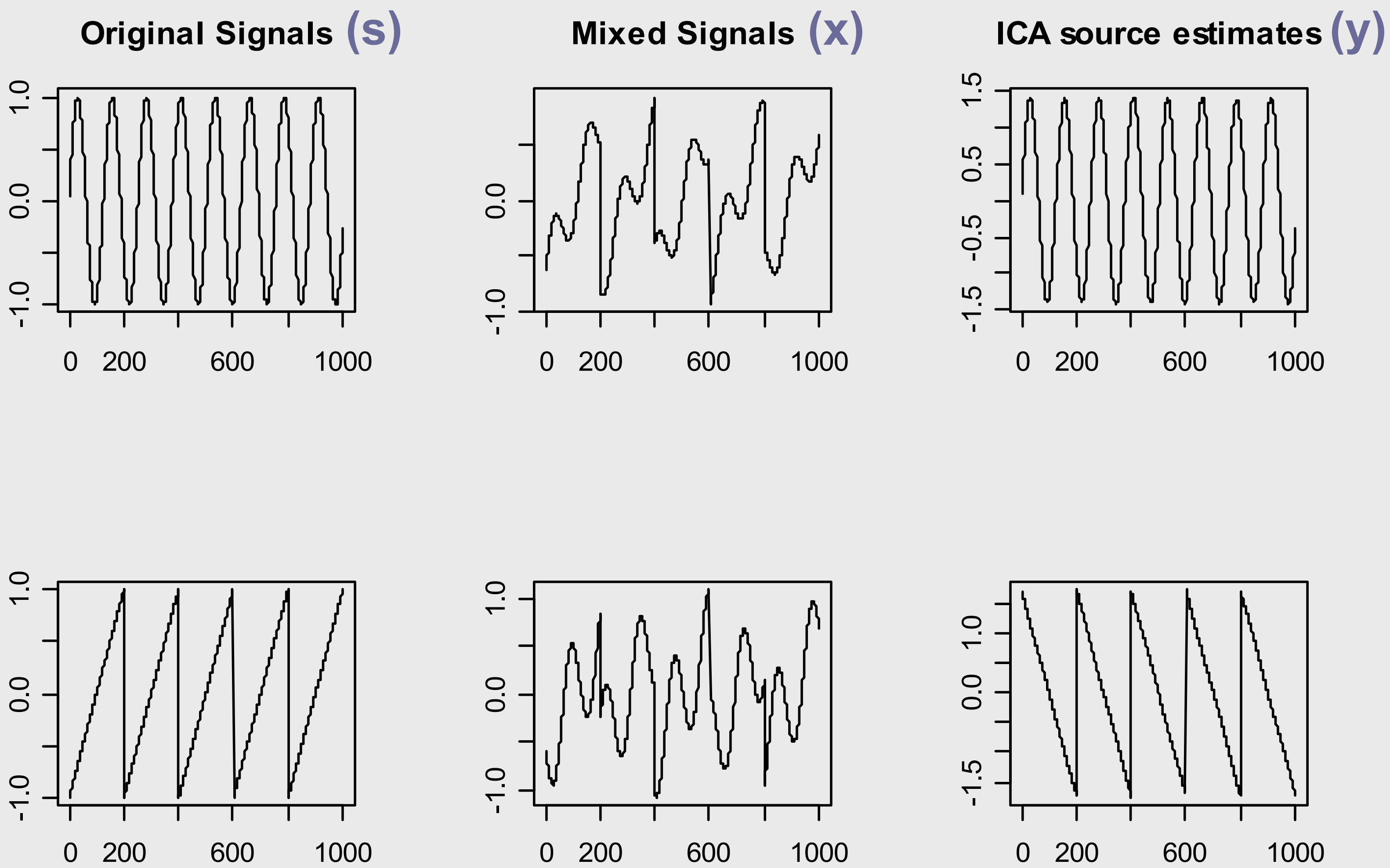
Overview of ICA

- Assumptions:
  - Sources are statistically independent of one another
  - At most 1 source has a Gaussian distribution
  - Signals are stationary



- The original sources **s** are mixed through matrix **A** to form the observed signal **x**.
- The de-mixing matrix **W** transforms the observed signal **x** into the independent components **y**.

Another ICA Example

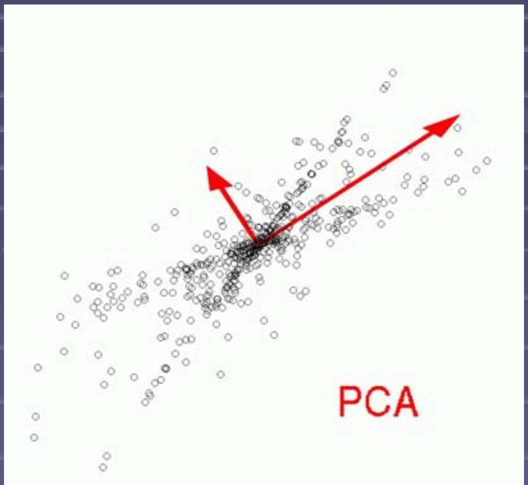


Relationships To Other Methods

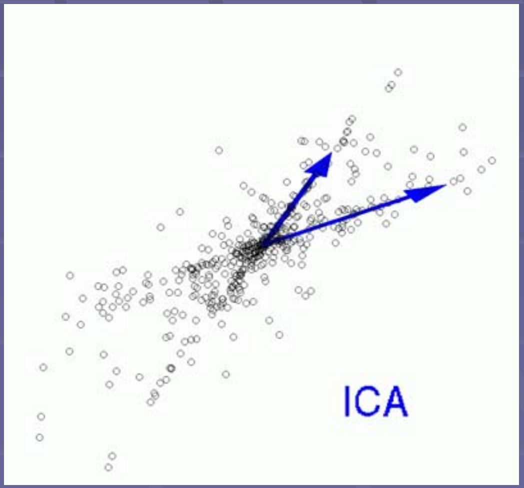
- Principal Component Analysis (PCA)
  - Focus on finding uncorrelated components in Gaussian data
  - Maximizes explained variance
  - Second-order statistics
- Factor Analysis
  - Essentially PCA with extra terms to model noise
- ICA
  - Focus on independent and non-Gaussian components
  - Higher-order statistics
- Projection Pursuit
  - Similar to ICA but seeks one projection at a time (rather than all at once
  - Finds “interesting” directions (not necessarily independent)

ICA versus PCA

- **Principal Component Analysis** (PCA) finds directions of maximal variance in Gaussian data (second-order statistics).



- **Independent Component Analysis** (ICA) finds directions of maximal independence in non-Gaussian data (higher-order statistics).



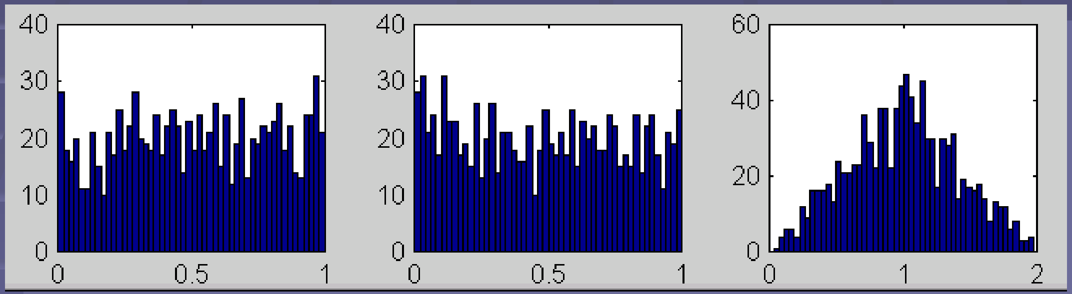
ICA estimation principles

- **Principle 1**: "Nonlinear decorrelation. Find the matrix ***W*** so that for any $i \neq j$, the components y_i and y_j are uncorrelated, *and* the transformed components $g(y_i)$ and $h(y_j)$ are uncorrelated, where g and h are some suitable nonlinear functions."
- **Principle 2**: "Maximum nongaussianity. Find the local maxima of nongaussianity of a linear combination ***y=Wx*** under the constraint that the variance of is constant. Each local maximum gives one independent component."

by A.Hyvarinen, A.Karhunen, E.Oja
'Independent Component Analysis'

Non-Gaussianity is Key to ICA

- The ***central limit theorem*** (***CLT***) tells us that the distribution of a sum of independent random variables tends toward a Gaussian distribution.



RV1 RV2 RV1+RV2

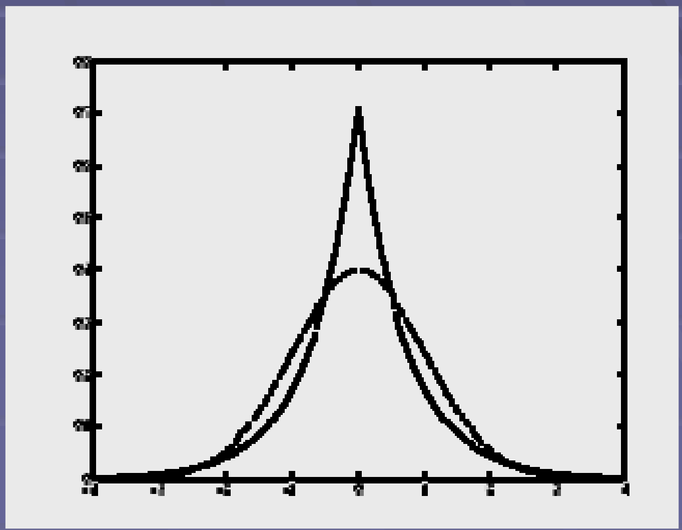
- Similarly, a mixture of two independent signals usually has a distribution that is closer to Gaussian than any of the two original signals

Measures of Non-Gaussianity

- Need a quantitative measure of non-Gaussianity for ICA Estimation.
  - Kurtotis : $kurt(y) = E\{y^4\} - 3(E\{y^2\})^2$
  - Entropy : $H(y) = -\int f(y) \log f(y) dy$
  - Negentropy : $J(y) = H(y_{gauss}) - H(y)$
  - Approximation : $J(y) \approx [E\{G(y)\} - E\{G(v)\}]^2$
    where v is a standard Gaussian RV
    and $G(y) = \frac{1}{a} \log\cosh(a.y)$

Aside on Kurtosis

- Kurtosis can be both positive and negative
- *Sub-Gaussian* RV
  - Negative kurtosis
  - Flat pdf, rather constant near zero, very small for more extreme values
  - Example: Uniform distribution
- *Super-Gaussian* RV
  - Positive kurtosis
  - Spiky pdf, with a heavy tails
  - Example: Laplace distribution



FastICA Algorithm

- Steps
  - Preprocessing
    - Centering
    - Whitening
  - Initialization
  - Iteration for Rotation
- Convergence is cubic (fast)
- No parameters (e.g. learning rate)
- Available for Matlab and R

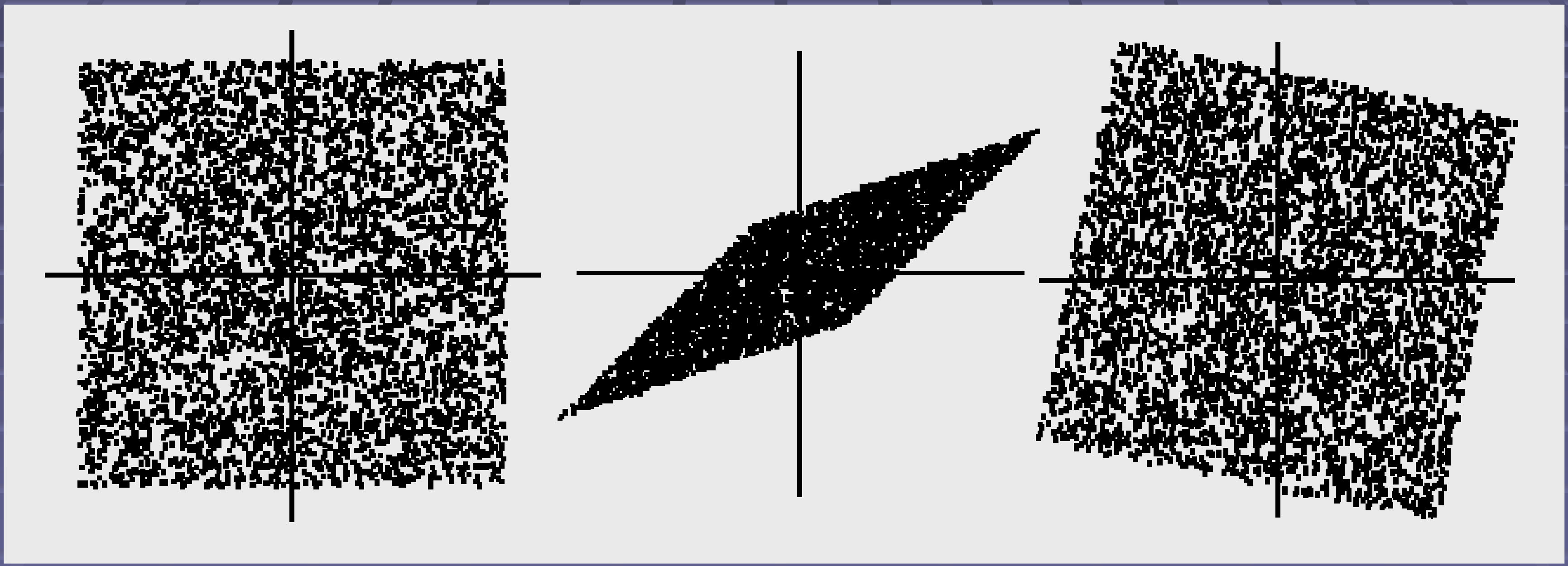
FastICA - Preprocessing

- Centering:
  - Make the observed signals (**x**) zero-mean variables

- Whitening
  - Linearly transform **x** so that its components are uncorrelated and have unit variance.
  - Saying that a variable is “white” means that its covariance matrix equals __**the identity matrix**__.
  - Uncorrelated variables are only partly independent; but this helps to simplify the problem.
  - Whitening can be computed by eigenvalue decomposition (EVD) of the covariance matrix

$$E\{xx^T\} = EDE^T$$

Preprocessing - Whitening



Original (Underlying) Data

Observed Signals

After Whitening

Computing the Rotation

This is based on maximizing of an objective function G(.) of the non-Gaussianity measure.

$$Obj(\mathbf{W}) = \sum_{t=1}^{T} G(\mathbf{W}^T \mathbf{x}_t) - \Lambda(\mathbf{W}^T \mathbf{W} - \mathbf{I})$$

$$\frac{\partial Obj}{\partial \mathbf{W}} = \mathbf{X} g(\mathbf{W}^T \mathbf{X})^T - \Lambda \mathbf{W} = \mathbf{0}$$

where g(.) is derivative of G(.),
W is the rotation transform sought
Λ is Lagrange multiplier to enforce that W is an orthogonal transform i.e. a rotation

Solve by fixed point iterations

The effect of **Λ** is an orthogonal de-correlation

<u>Fixed Point Algorithm</u>

Input: **X**

Random init of **W**

Iterate until convergence:

$$Y = \mathbf{W}^T \mathbf{X}$$
$$\mathbf{W} = \mathbf{X} g(\mathbf{S})^T$$
$$\mathbf{W} = \mathbf{W} \sqrt{(\mathbf{W}^T \mathbf{W})^{-1}}$$

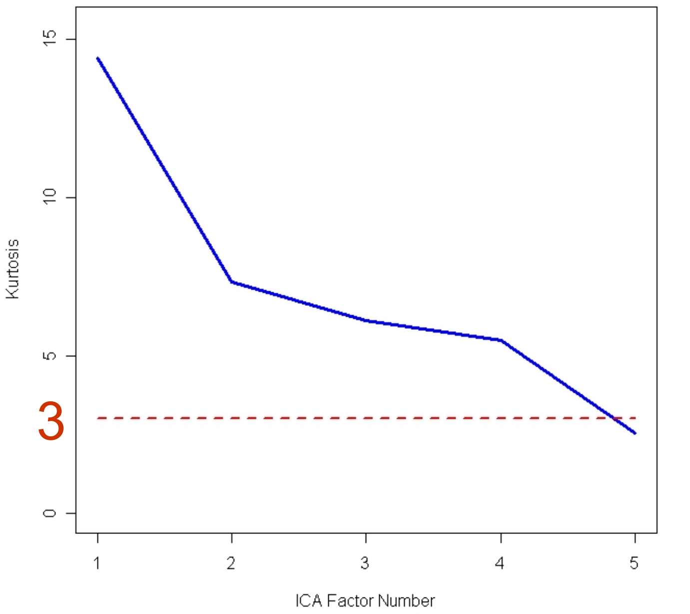
Output: **W, Y**

Component Order

- PCA which has well-defined and intuitive explanation of the ordering of its components (i.e. eigenvalues of covariance matrix)
- ICA, however, deserves further investigation on this particular problem since a particular kind of ordering is not readily at hand.
- Can follow a heuristic scheme called: *testing-and-acceptance* (TNA)

How Many ICs to Keep?

“Scree” Plot



R^2 Values from Regression

# ICs Kept	F Fund	C Fund	S Fund	I Fund
5	1	1	1	1
4	.9980	.9994	.9998	.9949
3	.0611	.8983	.9652	.9869
2	.0611	.8962	.9629	.1276

- Dropping to 3 or fewer ICs significantly reduces fit
- First 4 ICs have excess kurtosis

→ Keep all 4 Independent Components (ICs)

Application Domains

- Image de-noising
- Medical signal processing (fMRI, ECG)
- Feature extraction & face recognition
- Time series analysis (financial, economic)
- Telecommunications (CDMA signals)
- Topic extraction
- Scientific Data Mining

Functional Brain Imaging

- Functional magnetic resonance imaging (fMRI) data are noisy and complex.
- ICA identifies concurrent hemodynamic processes.
- Does not require *a priori* knowledge of time courses or spatial distributions.

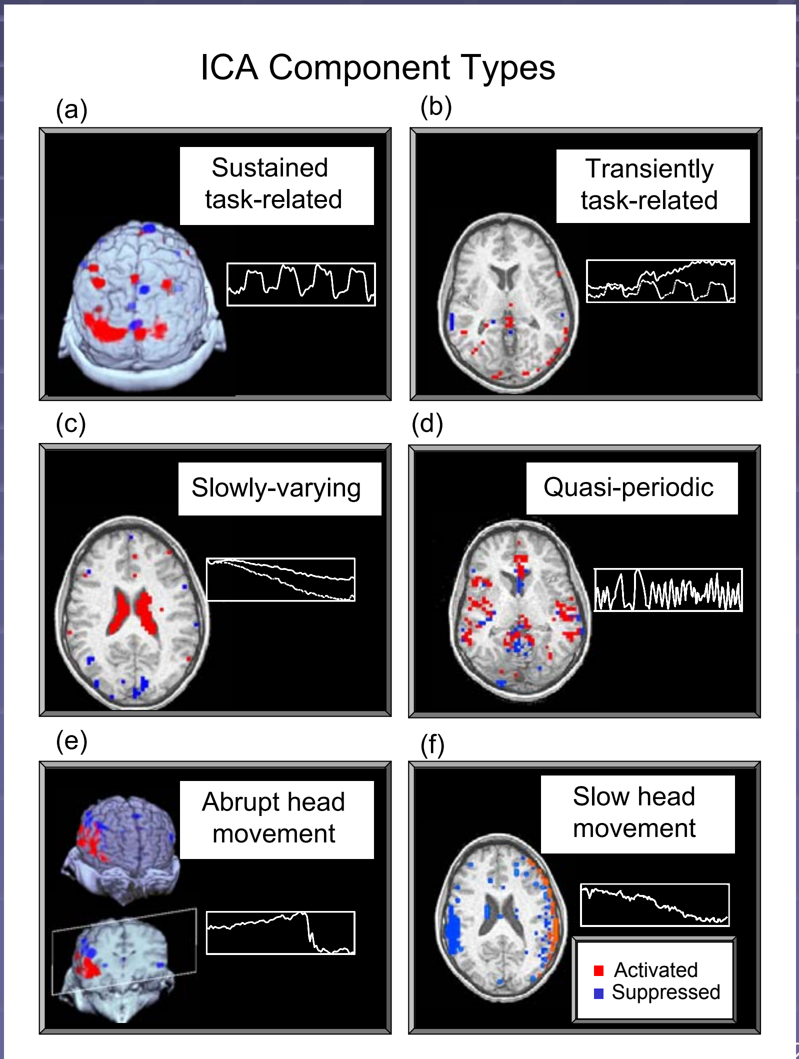
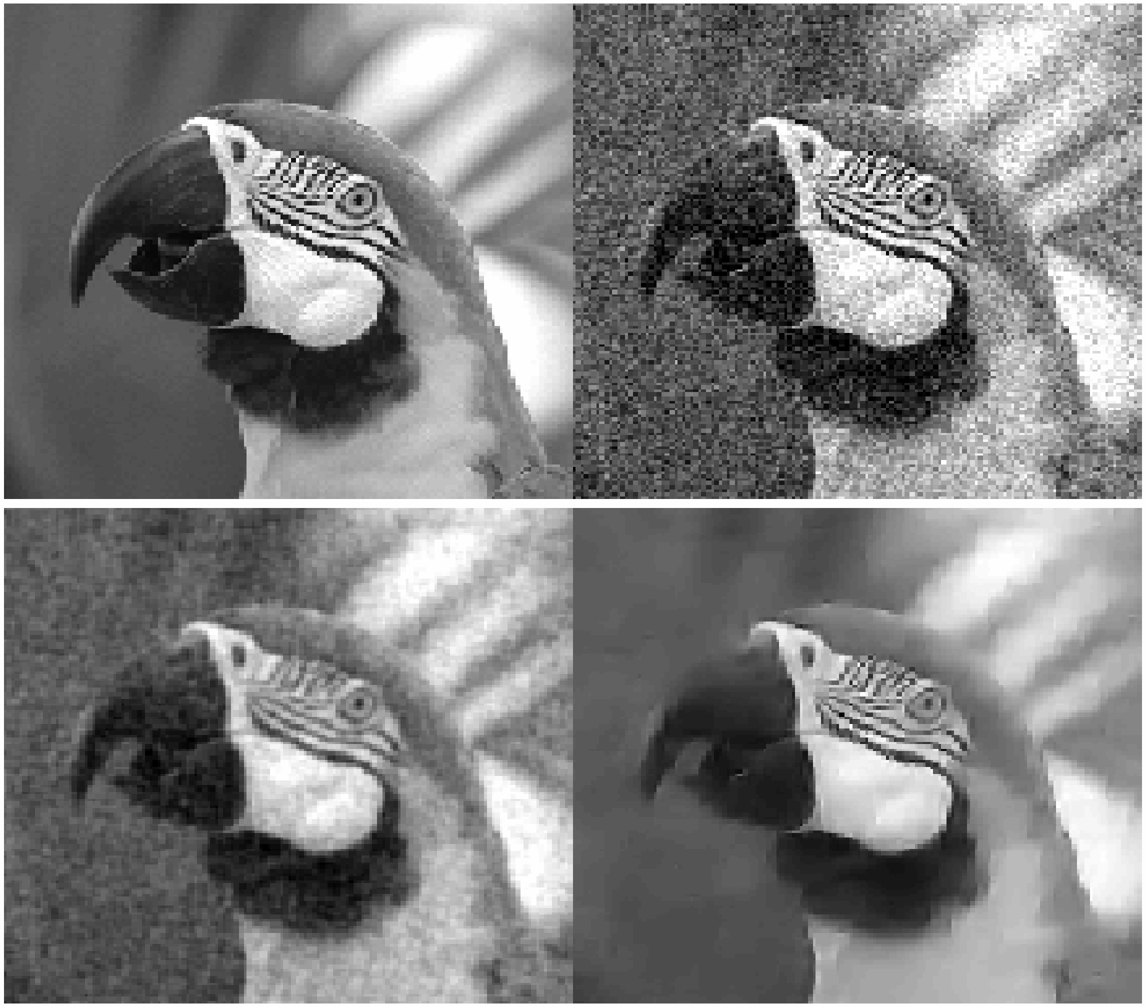


Image De-noising

Original image

Noisy image

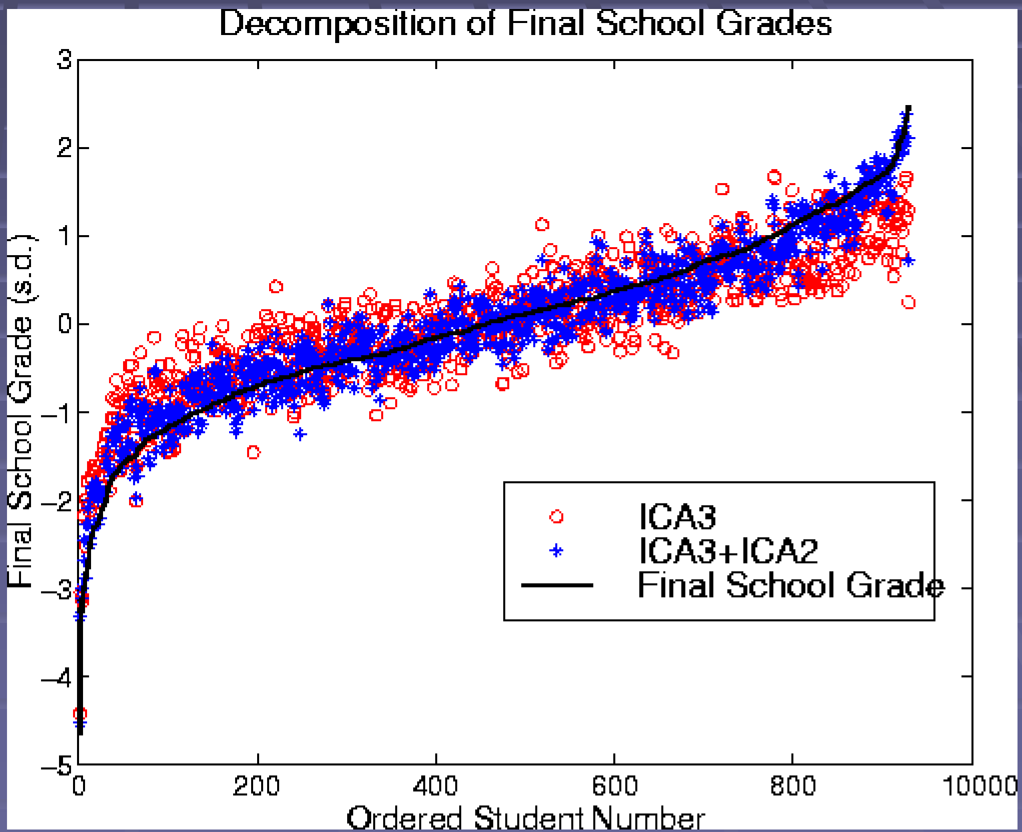


Wiener filtering

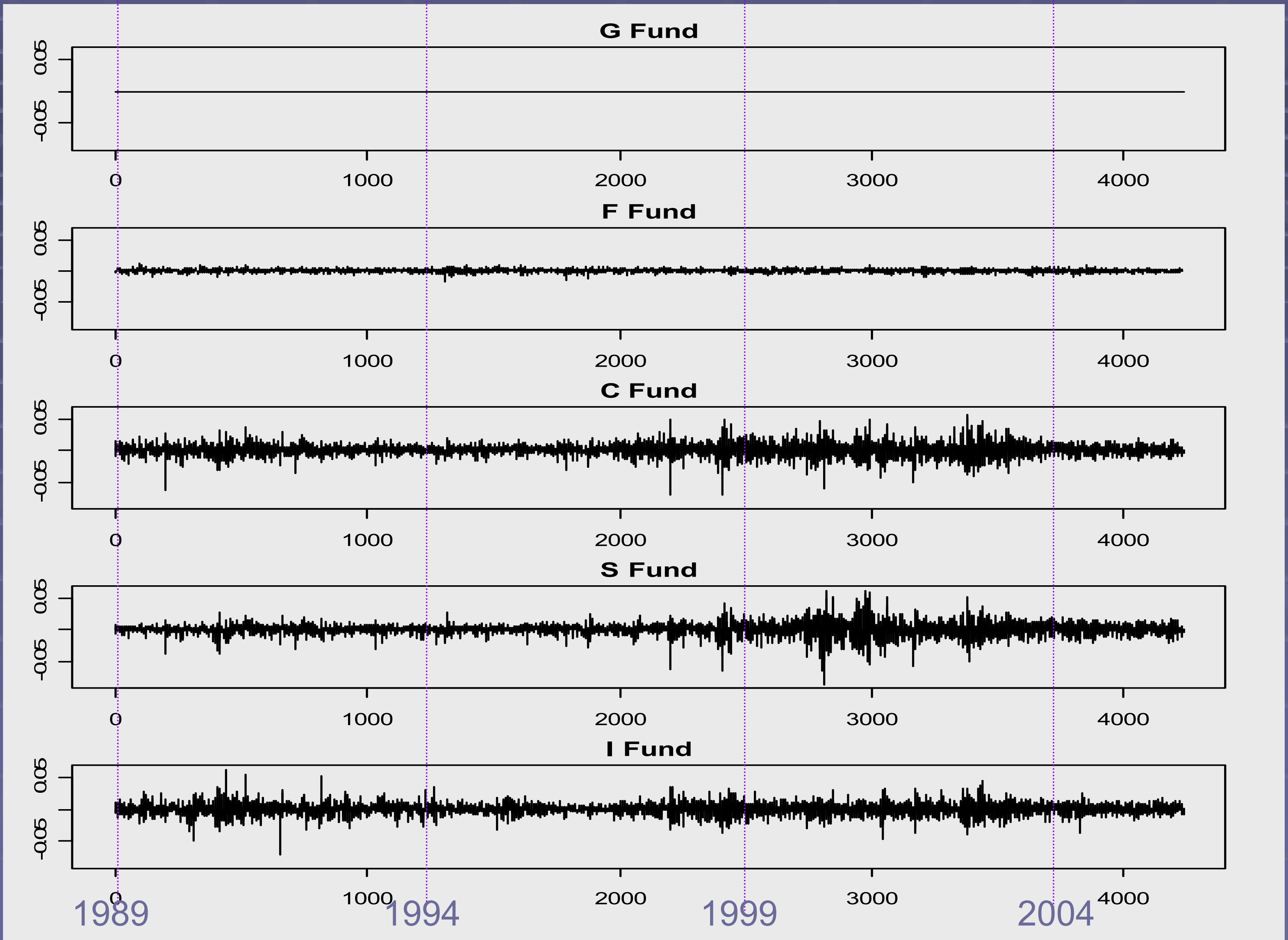
ICA filtering

Data Mining

- ICA was applied to Armed Forces Vocation Aptitude Battery (ASVAB) test scores and Navy Fire Control School grades.
- Two ICA components contributed to final school grade.
- ICA may suggest more efficient and balanced selection criteria.



Financial Time Series



Recommended Reading

Books

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). Independent Component Analysis. London, John Wiley and Sons

Stone, J. (2004). Independent Component Analysis: A Tutorial Introduction. Cambridge, MA, MIT Press.

Roberts, S., and Everson, R., editors, (2001). Independent component analysis: principles and practice. Cambridge UK, Cambridge University Press.

Articles

Back, A. and A Weigend (1997). "A First Application of Independent Component Analysis to Extracting Structure from Stock Returns." International Journal of Neural Systems 8(4): 473-484.

Cardoso, J.-F. (1998). "Blind Signal Separation: Statistical Principals." Proceedings of the IEEE 86(10): 2009-2025.

Hyvarinen, A. (1999). Survey on Independent Component Analysis. Neural Computing Surveys 2: 94-128.

Hyvarinen, A and Oja, E. (2000). "Independent component analysis: algorithms and applications." Neural Networks 13(4-5): 411-430.

Concluding Remarks on ICA

- A flexible tool which searches linear transformations of the observed data into components that are statistically maximally independent

- Recent extensions include:
  - More signals than mixtures
  - Allowing for Gaussian noise term

- Algorithms readily available

- Many possible applications prospects

Every problem becomes very childish when once it is explained to you."

- Sherlock Holmes (The Dancing Men, A.C. Doyle, 1905)

QUESTIONS?