Lecture 12: SSMs; Independent Component Analysis; Canonical Correlation Analysis

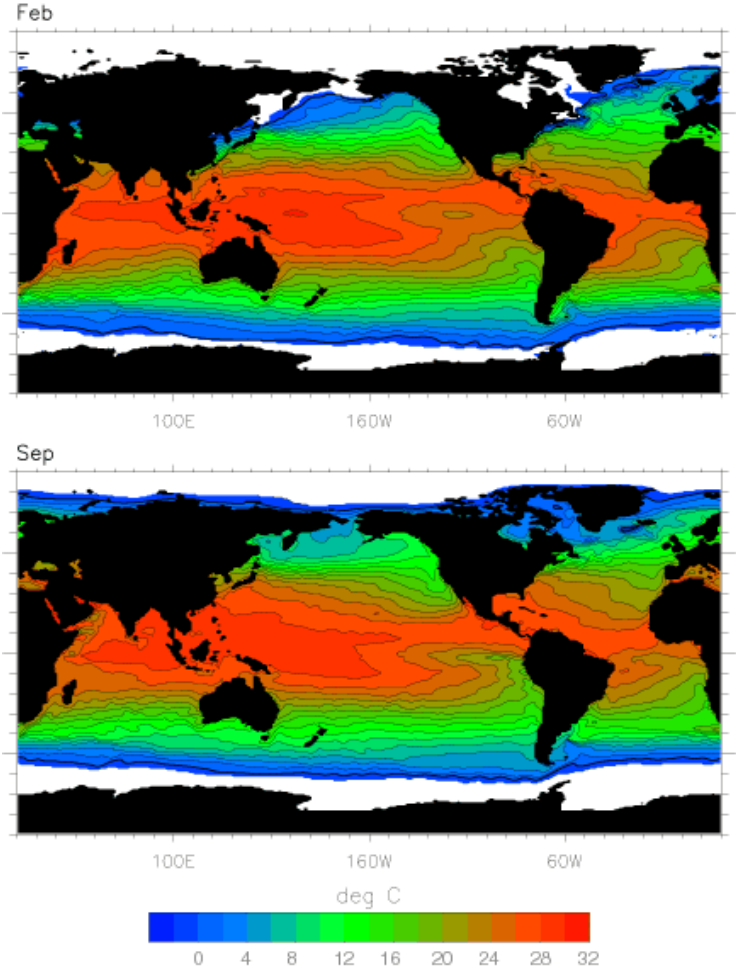
Lester Mackey

May 7, 2014

Beyond linearity in state space modeling

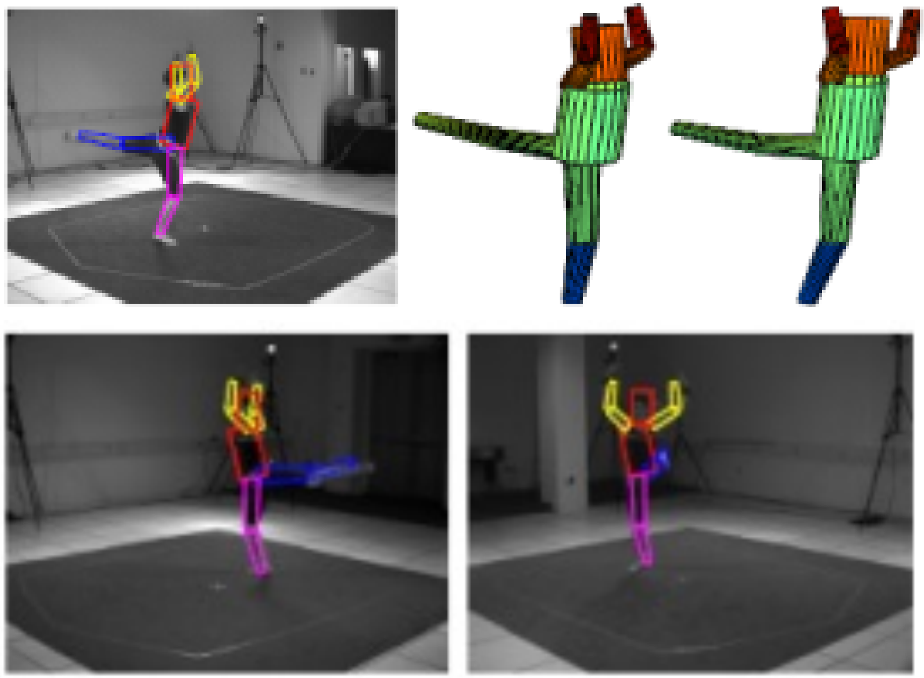
Credit: Alex Simma

Weather forecasting



Dynamics implicitly determined by geophysical simulations

Pose estimation



Observed image is a complex function of the 3D pose, other nearby objects & clutter, lighting conditions, camera calibration, etc.

Approximate nonlinear filters

Credit: Alex Simma

$$q_t(x_t) \propto p(y_t \mid x_t) \cdot \int p(x_t \mid x_{t-1}) q_{t-1}(x_{t-1})\, dx_{t-1}$$

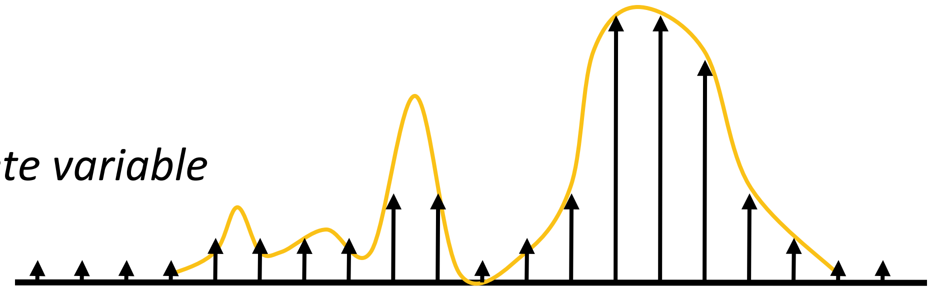
- Typically cannot directly represent these **continuous filtering distributions** or determine a closed form for the prediction integral
- A wide range of approximate nonlinear filters have thus been proposed, including
  - *Histogram filters*
  - *Extended & unscented Kalman filters*
  - *Particle filters*
  - …

Approximate nonlinear filtering methods

Credit: Alex Simma

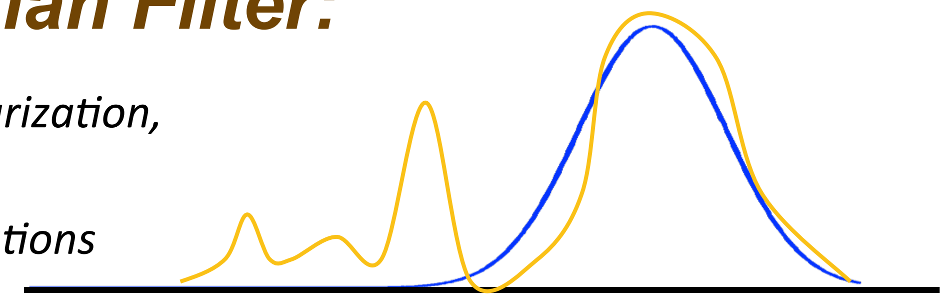
Histogram Filter:

- *Evaluate on fixed discretization grid*
- *Replaces continuous latent variable by discrete variable*
- *Only feasible in low dimensions*
- *Expensive or inaccurate*



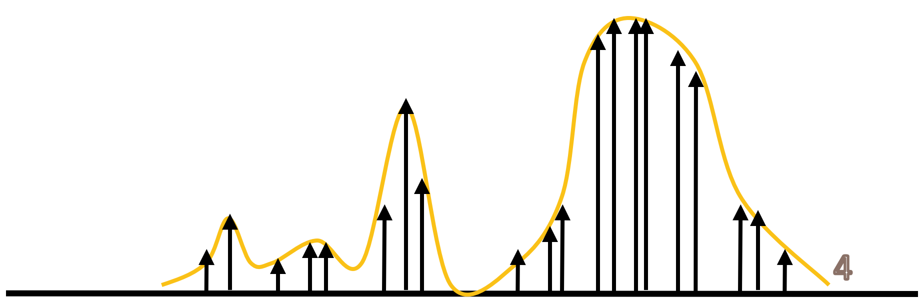
Extended/Unscented Kalman Filter:

- *Approximate posterior as Gaussian via linearization, quadrature, …*
- *Inaccurate for multimodal posterior distributions*



Particle Filter:

- *Approximate filtering distribution with fixed set of particles (delta masses)*
- *Particle locations not fixed; dynamically evaluate states with highest probability*
- *Monte Carlo approximation*



Independent component analysis (ICA)

Courtesy: Rob Tibshirani

$$X = \mathbf{A}S$$

where

- X is a random p-vector representing multivariate input measurements.
- S is a latent source p-vector whose components are independently distributed random variables.
- $\mathbf{A}$ is $p \times p$ mixing matrix.

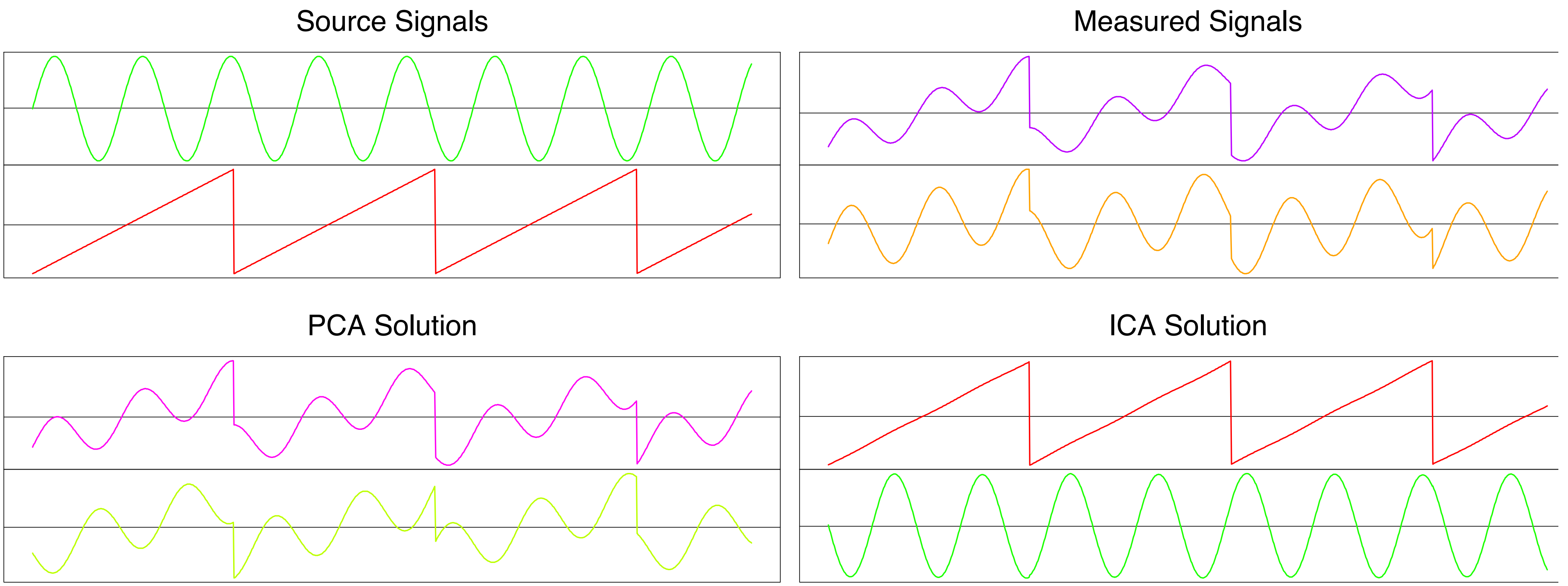
Given realizations $x_1, x_2, \ldots, x_N$ of X, the goals of ICA are to

- Estimate $\mathbf{A}$
- Estimate the source distributions $S_j \sim f_{S_j},\ j = 1, \ldots, p$.

The cocktail party problem

Courtesy: Rob Tibshirani

In a room there are p independent sources of sound, and p microphones placed around the room hear different mixtures.



Here each of the $x_{ij} = x_j(t_i)$ and recovered sources are a time-series sampled uniformly at times t_i.

Independent vs. uncorrelated

Courtesy: Rob Tibshirani

WoLOG can assume that $\mathrm{E}(S) = 0$ and $\mathrm{Cov}(S) = \mathbf{I}$, and hence $\mathrm{Cov}(X) = \mathrm{Cov}(\mathbf{A}S) = \mathbf{A}\mathbf{A}^T$.

Suppose $X = \mathbf{A}S$ with S unit variance, uncorrelated

Let $\mathbf{R}$ be any orthogonal $p \times p$ matrix. Then

$$X = \mathbf{A}S = \mathbf{A}\mathbf{R}^T\mathbf{R}S = \mathbf{A}^* S^*$$

and $\mathrm{Cov}(S^*) = I$

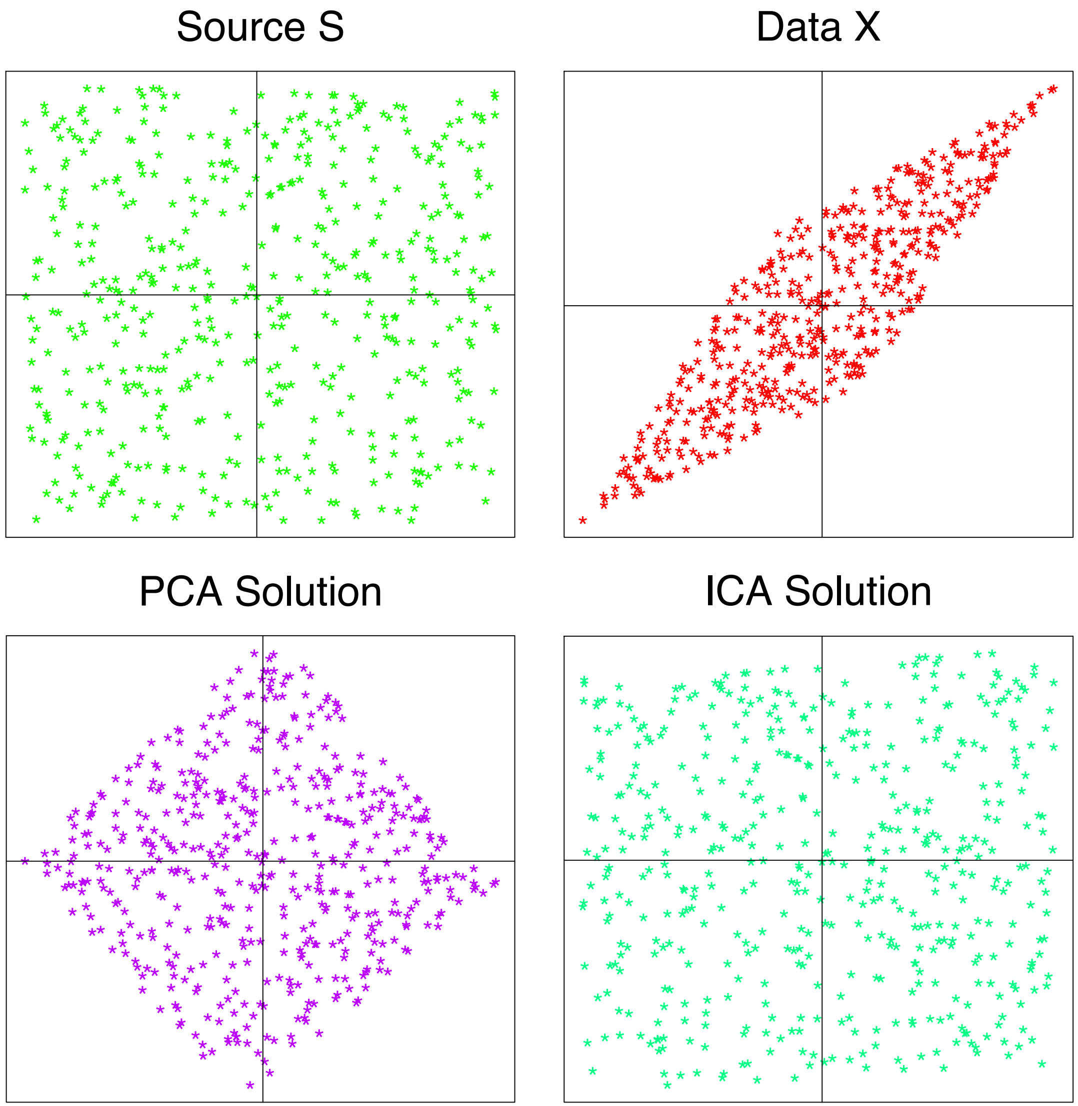
It is not enough to find uncorrelated variables, as they are not unique under rotations.

Hence methods based on second order moments, like principal components and Gaussian factor analysis, cannot recover $\mathbf{A}$.

ICA uses independence, and non-Gaussianity of S, to recover $\mathbf{A}$ — e.g. higher order moments.

Independent vs. uncorrelated

Courtesy: Rob Tibshirani



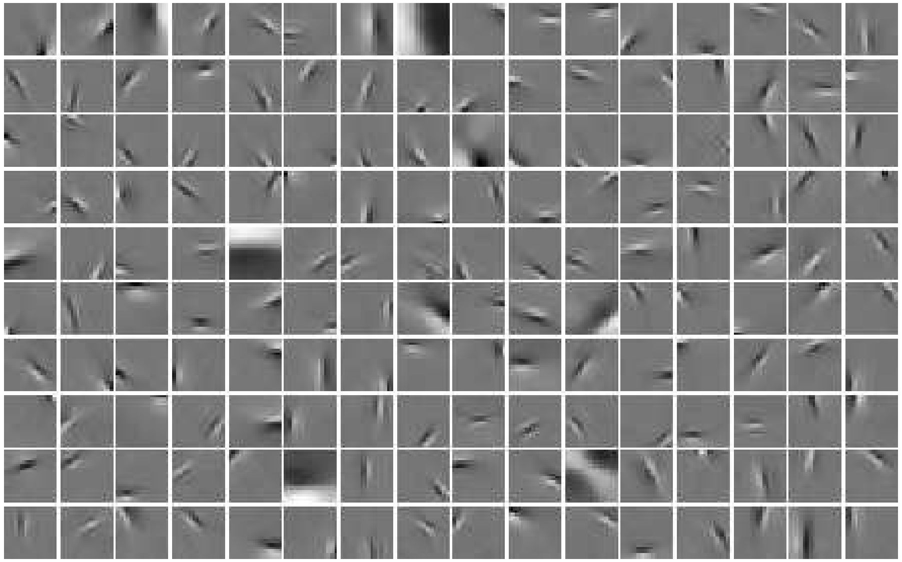
Principal components are uncorrelated linear combinations of X, chosen to successively maximize variance.

Independent components are also uncorrelated linear combinations of X, chosen to be as independent as possible.

ICA representation of natural images

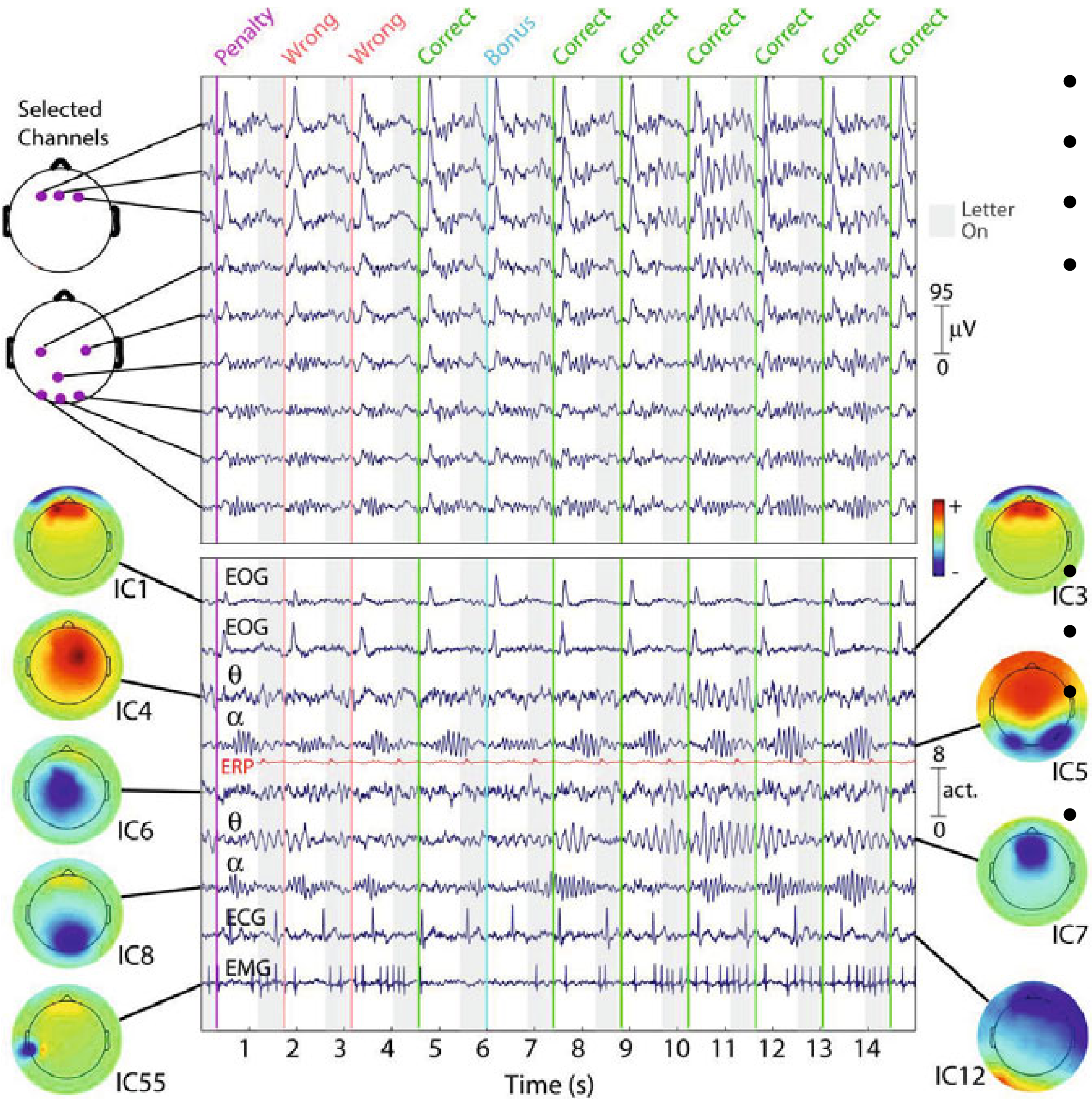
Courtesy: Rob Tibshirani

Pixel blocks are treated as vectors, and then the collection of such vectors for an image forms an image database. ICA can lead to a sparse coding for the image, using a natural basis.



see http://www.cis.hut.fi/projects/ica/imageica/ (Patrik Hoyer and Aapo Hyvärinen, Helsinki University of Technology)

ICA and electroencephalographic (EEG) data

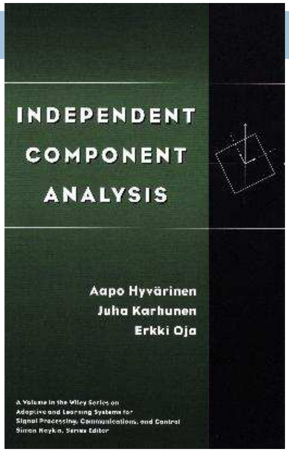


- 15 seconds of EEG data
- 9 (of 100) scalp channels
- 9 ICA components
- Nearby electrodes record nearly identical mixtures of brain and non-brain activity; ICA components are temporally distinct.
- Patient blinking (IC1/3)
- IC12: cardiac pulse
- IC4/7 activity after string of correct responses
- Colored scalps represent ICA unmixing coefficients a_j as heatmap, showing brain or scalp location of the source.

Approaches to ICA

Courtesy: Rob Tibshirani

ICA literature is HUGE. Recent book by Hyvärinen, Karhunen & Oja (Wiley, 2001) is a great source for learning about ICA, and some good computational tricks.



- Mutual Information and Entropy, maximizing non-Gaussianity — FastICA (HKO 2001), Infomax (Bell and Sejnowski, 1995)
- Likelihood methods — ProdDenICA (Hastie +Tibshirani)-later
- Nonlinear decorrelation — Y_1 independent Y_2 iff $\max_{g,f} \text{Corr}[f(Y_1), g(Y_2)] = 0$ (Hérault-Jutten, 1984), KernelICA (Bach and Jordan, 2001)
- Tensorial moment methods

Blackboard discussion

- See lecture notes