Denoising source separation

Jaakko Särelä
Neural networks research centre
Helsinki university of technology
P.O. Box 5400, 02015-HUT, Finland
jaakko.sarela@hut.fi, http://www.cis.hut.fi

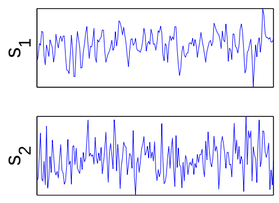
Harri Valpola
Artificial Intelligence Laboratory
University of Zurich
Andreasstrasse 15, 8050 Zurich, Switzerland
valpola@ailab.ch, http://ailab.ch/people/valpola/

Comprehensive article available at Cogprints `http://cogprints.ecs.soton.ac.uk/archive/00003493/`

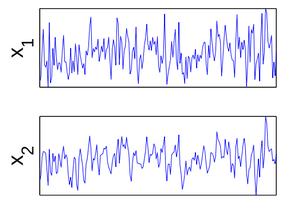
Source separation is often used in signal processing but it can be exploited in feature extraction as well.

Denoising is frequently used for enhancing signal characteristics.
We show that denoising combined with sphering produces effective source separation algorithms.

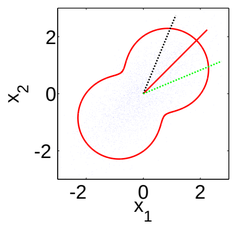
Original sources
independent
1st signal Gaussian with time structure
2nd signal Gaussian i.i.d noise



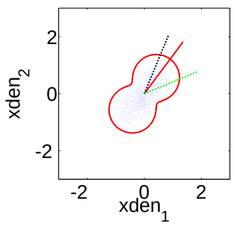
Observed signals
linear mixtures



Scatterplot
shows the correlation structure



Denoising
enhances the desired signals



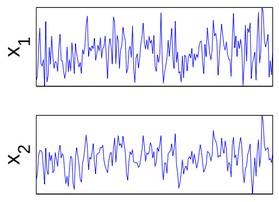
standard deviation in each direction — maximum variance direction — 1st signal direction — 2nd signal direction

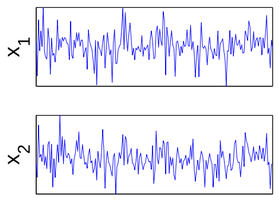
The aim is to use Hebbian learning (PCA, correlation based) to identify the sources: biologically plausible, fast algorithms exist, online versions possible. Denoising enhances the desired signals, but the original variance structure dominates if the denoising is not very precise. Sphering removes the variance structure. $\longrightarrow$ Even slight denoising renders the correct direction identifiable with Hebbian learning. Denoising can exploit for instance the time structure or the non-Gaussianity of the sources (which leads to ICA).

Time-structure source separation

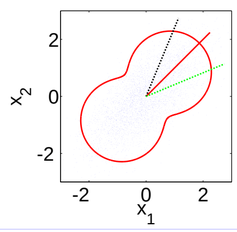
Independent component analysis

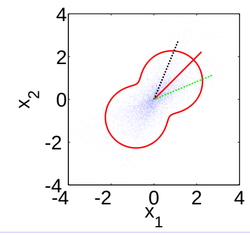
Observed signals



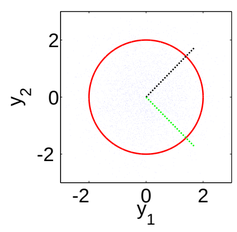


Scatterplot
shows the correlation structure

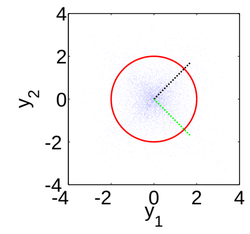


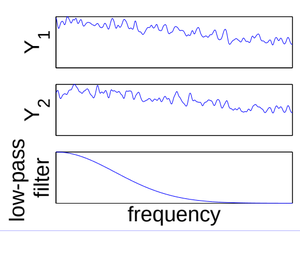


↓ Sphering ↓



Scatterplot of the sphered signals
removes the correlation structure
implementation: $\mathbf{Y} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{X}$
$\mathbf{\Lambda}$ = eigenvalues, $\mathbf{E}$ = eigenvectors

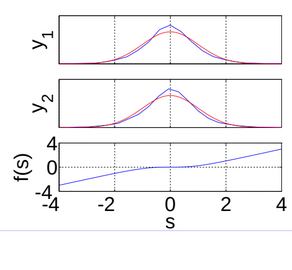




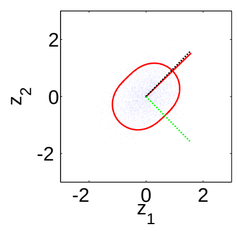
Characteristics

Spectra
low-pass filtering as denoising

Distributions
shrinkage function as denoising
Gaussian distribution



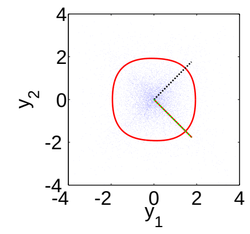
↓ Denoising ↓



Denoised signals

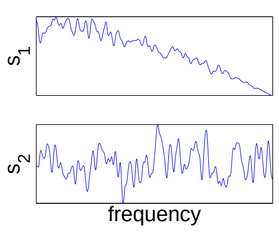
Linear denoising
$\mathbf{Z} = \mathbf{YD}$, $\mathbf{D} = \mathbf{V\Gamma V}^T$
$\mathbf{V}$ = DCT, $\mathbf{\Gamma}$ = filter amplitude response
denoising $\mathbf{D}$ applies separately to the signals $\mathbf{y}_i$

Nonlinear denoising
projection variance nonlinearly dependent on denoising
$\longrightarrow$ denoising $\mathbf{f}(\mathbf{s})$ is embedded in source estimation
by power method
$\mathbf{s}^+ = \mathbf{f}(\mathbf{w}^T\mathbf{Y})$, $\mathbf{f}(\mathbf{s}) = \mathbf{s} - \tanh\mathbf{s}$



↓ PCA ↓

↓ power method ↓



Extracted sources

Spectra
1st signal Gaussian with time structure
2nd signal Gaussian i.i.d noise

Distributions
Laplace-distributed signals