Linear and nonlinear independent component analysis by leave-one-out gadaline approximation

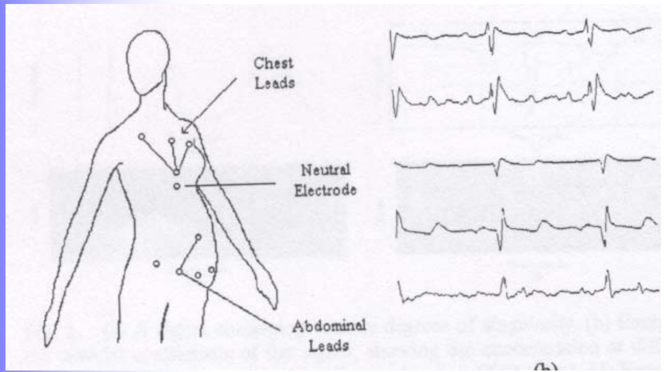
Jiann-Ming Wu
Department of Applied Mathematics
National Dong Hwa University

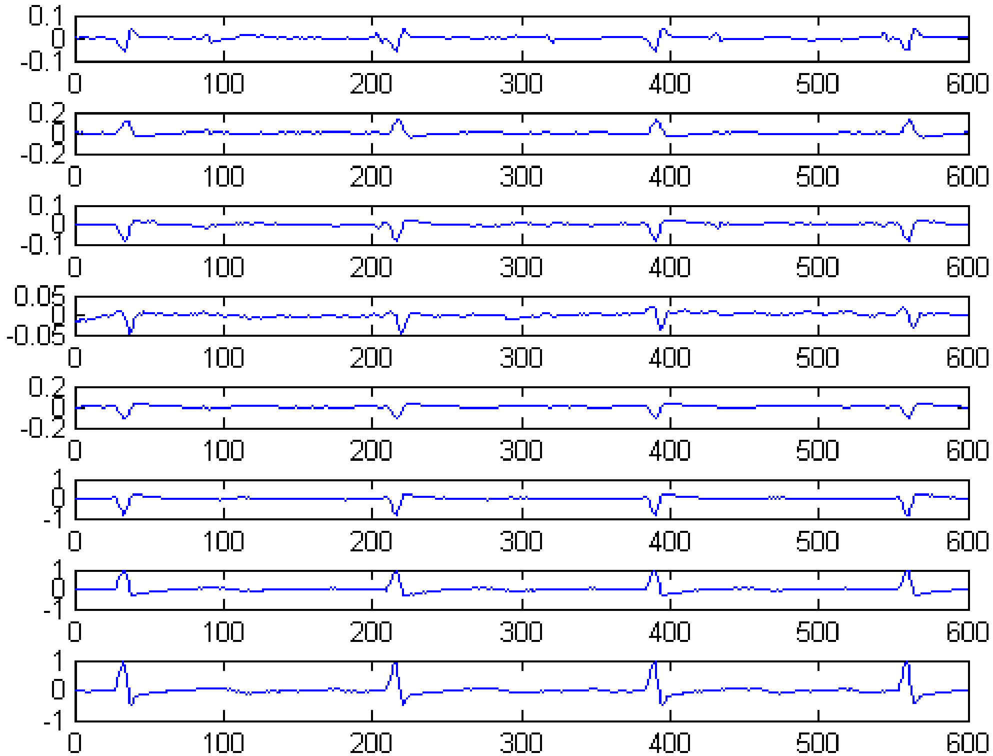
Outline

- Generative models of multi-channel observations
  - Linear mixtures
  - Post-nonlinear mixtures
- Gadaline Networks
  - Generalized adalines
  - Single gadaline approximation
  - Multiple gadalines for PNL mixtures
- Leave-one-out gadaline approximation
  - Estimation of model parameters and independent sources
- Numerical simulations
- Conclusions

Multi-channel observations

- Maternal ECG

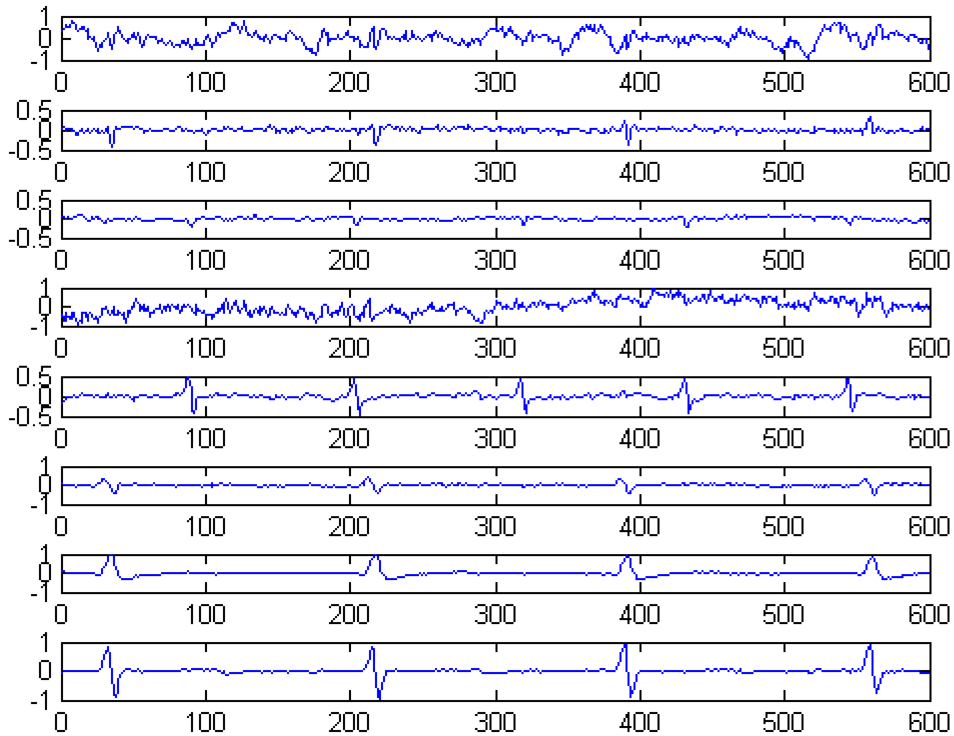




Independent sources

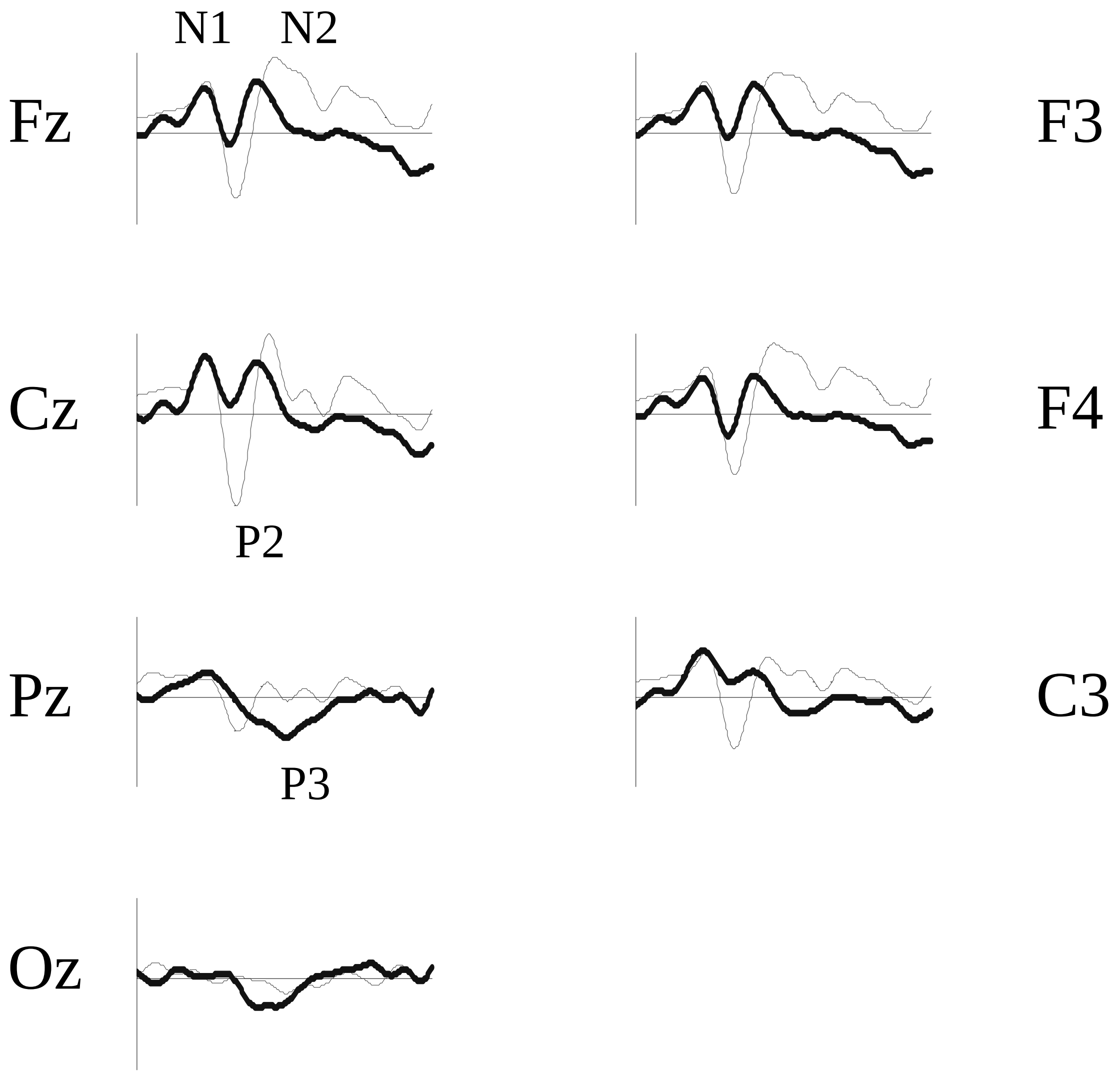
- PottsNICA (Wu & Chiu 2001;Wu 2007)

Fetal ECG



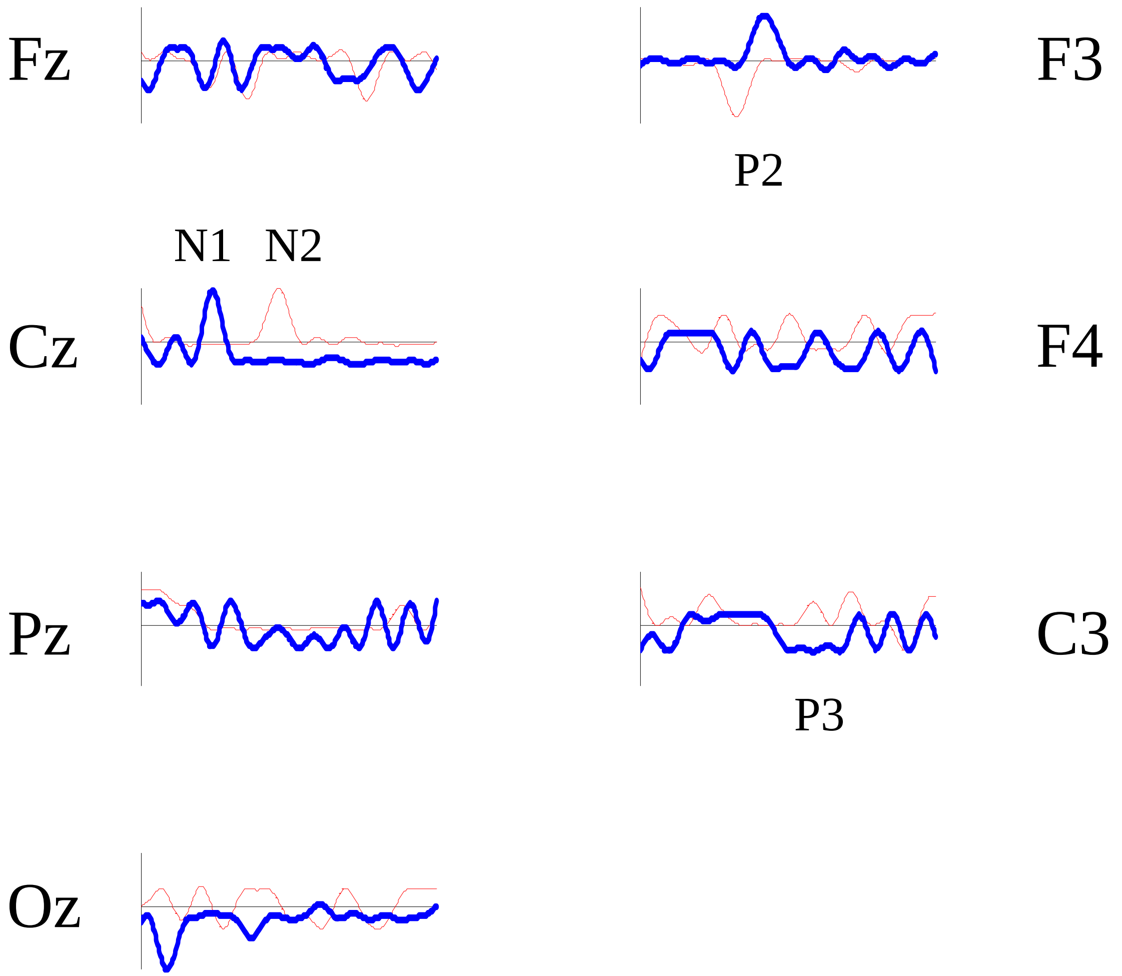
Multi-channel observations

- ERP
- SALK Institute
- (Makeig et al,1997)



Independent components of ERP

- (Wu et al,2005)



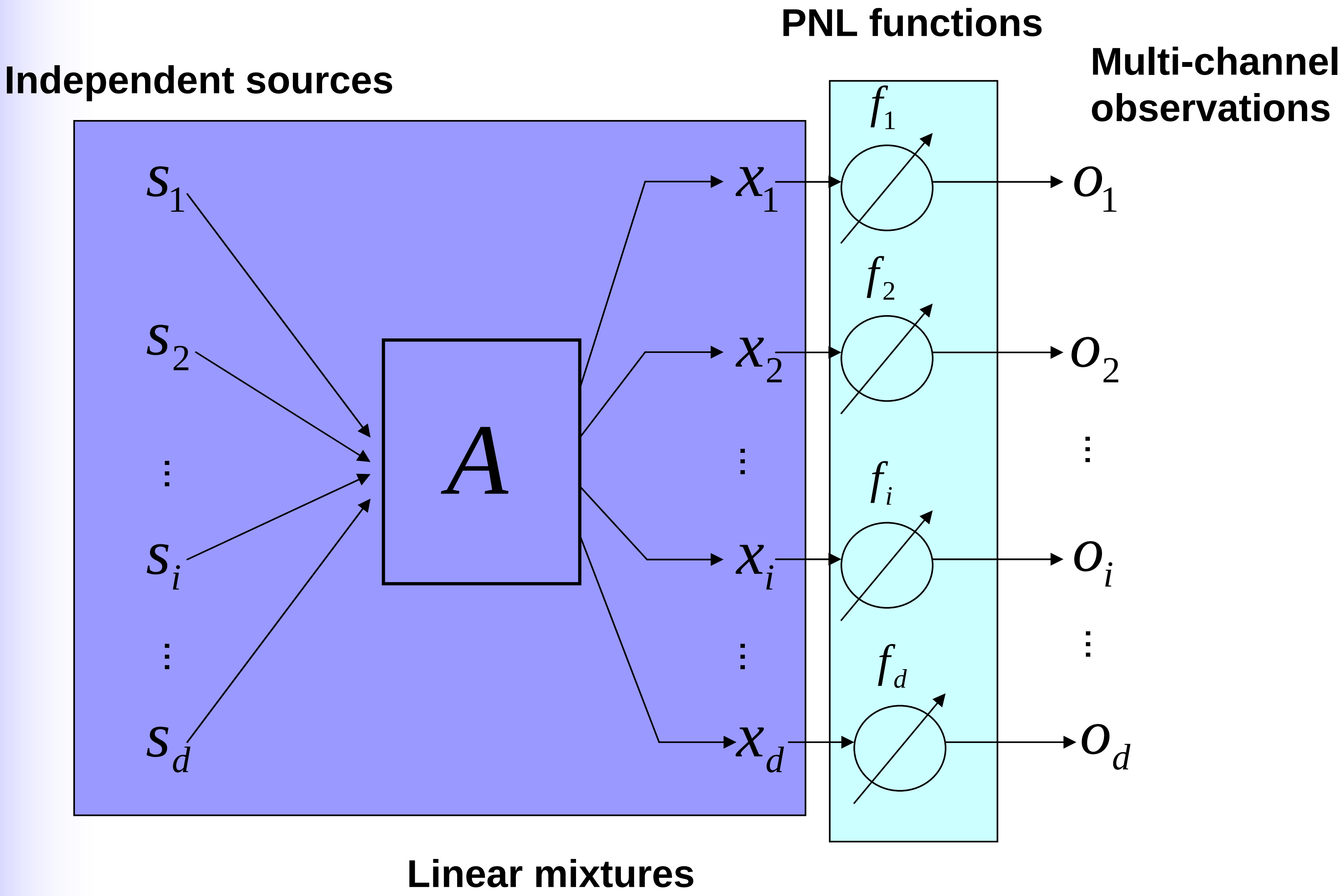
ICA of mixed facial images

(Wu et al,2005)

KL divergence

- Our previous works are based on minimization of Kullback-Leibler divergence.
- The KL divergence is a typical measure for statistical dependency between retrieved components

Pots-nonlinear mixtures



Model parameters

- Linear mixing matrix A
- Post-nonlinear functions $f_1, f_2, \cdots, f_d$

$f_i(x;\mathbf{c})$ is a one-dimensional nonlinear function.

f_i depends on parameters(knots) in vector $\mathbf{c}$

PNL ICA

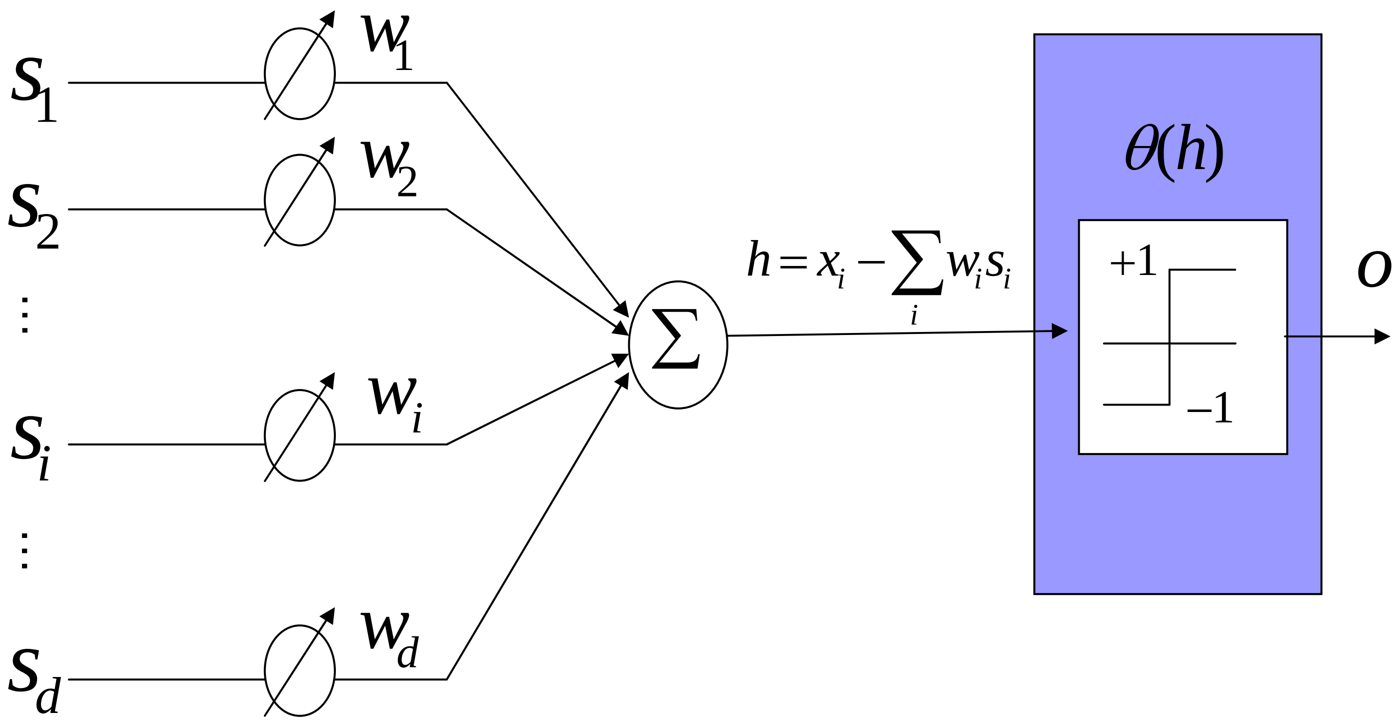
- Given multi-channel observations, PNL ICA aims to recover independent sources
- The by-product includes an estimation to model parameters

Gadalines

- Adalines (Widrow, 1962)
- Generalized adalines
  - Jiann-Ming Wu, Zheng-Han Lin, and Pei-Hsun Hsu
  - IEEE Trans. on Neural Networks, 2006)

Multiple inputs single output(MISO)

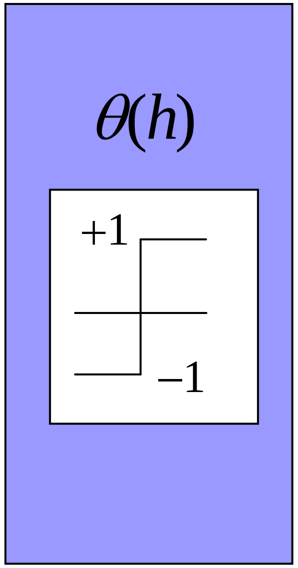
- Adaline
- Threshold PNL



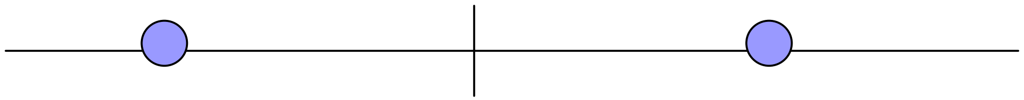
Two-state transfer function

- Threshold function

$$\theta_2(x;c) = [1,0] \;\; if \;\; x \in I_1$$
$$= [0,1] \;\; if \;\; x \in I_2$$



$c_1 = -1$ $\qquad$ $c_2 = 1$



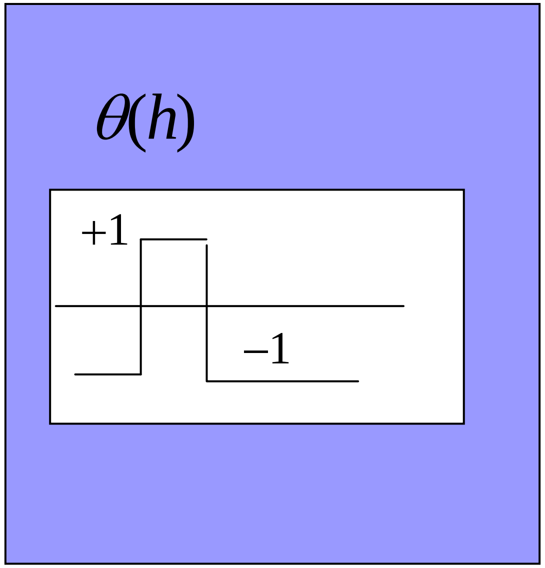
Threshold=0

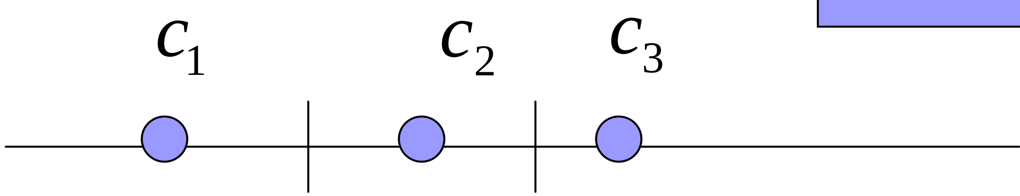
Non-overlapping intervals I_1 $\qquad$ I_2

Triple state transfer function

- Threshold function

$$\theta_3(x;c) = [1,0,0] \ \ if \ \ x \in I_1$$
$$= [0,1,0] \ \ if \ \ x \in I_2$$
$$= [0,0,1] \ \ if \ \ x \in I_3$$





Non-overlapping intervals I_1 I_2 I_3

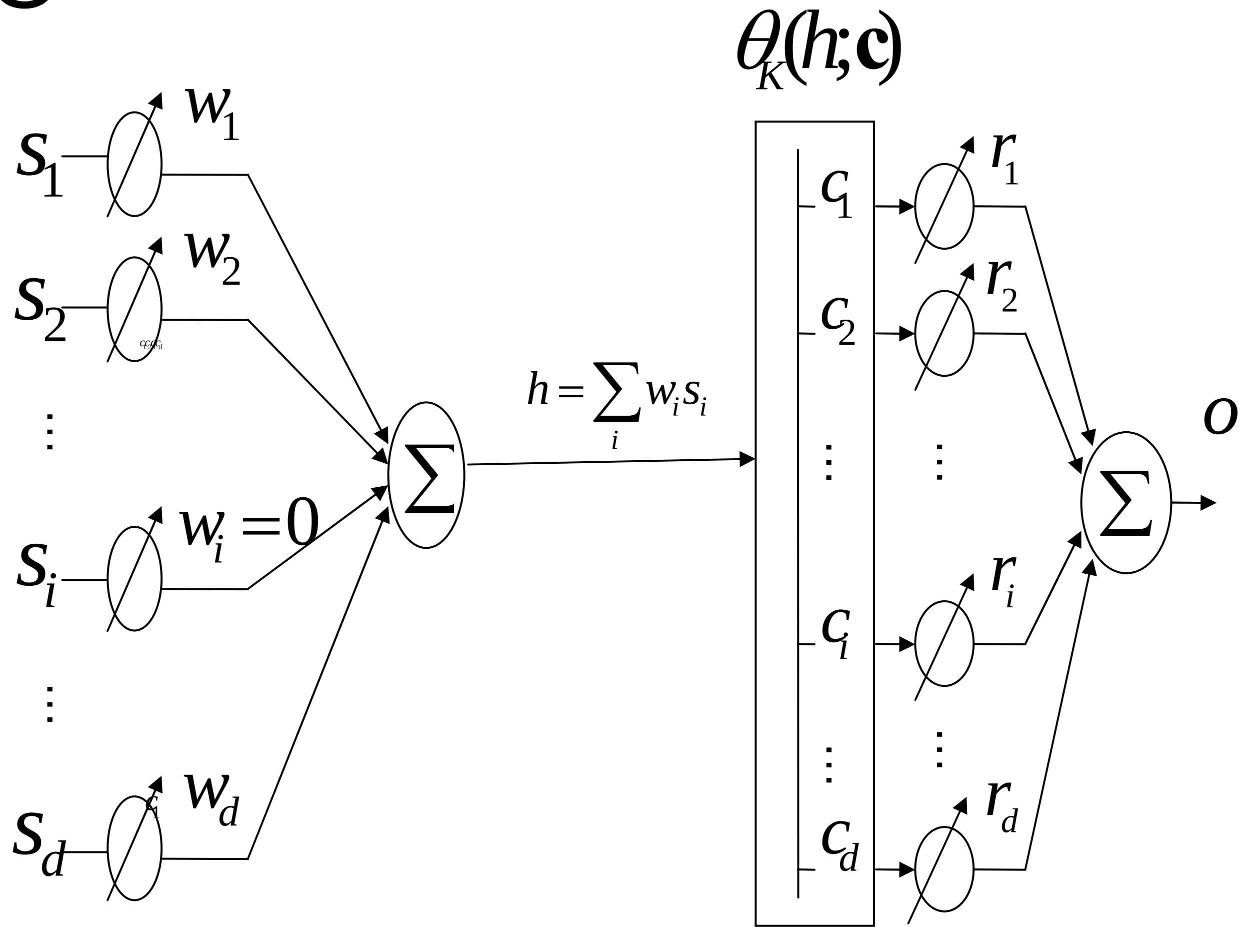
K-state transfer function

$$\theta_3(x;h) = [1,0,0] = \mathbf{e}_1^3 \quad if \quad h \in I_1$$
$$= [0,1,0] = \mathbf{e}_2^3 \quad if \quad h \in I_2$$
$$= [0,0,1] = \mathbf{e}_3^3 \quad if \quad h \in I_3$$

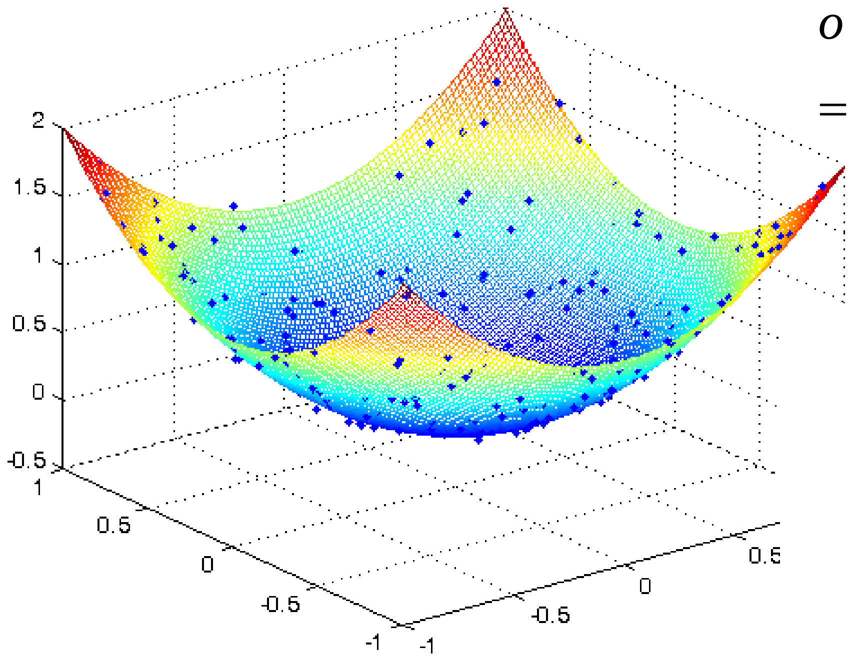
$$\theta_K(h;\mathbf{c}) = \left\{ \begin{array}{c} \mathbf{e_1^K} \text{ if } \mathbf{h} \in \mathbf{I_1} \\ \vdots \\ \mathbf{e_i^K} \text{ if } \mathbf{h} \in \mathbf{I_i} \\ \vdots \\ \mathbf{e_K^K} \text{ if } \mathbf{h} \in \mathbf{I_K} \end{array} \right\},$$

Single gadaline

- MISO



Single adaline approximation

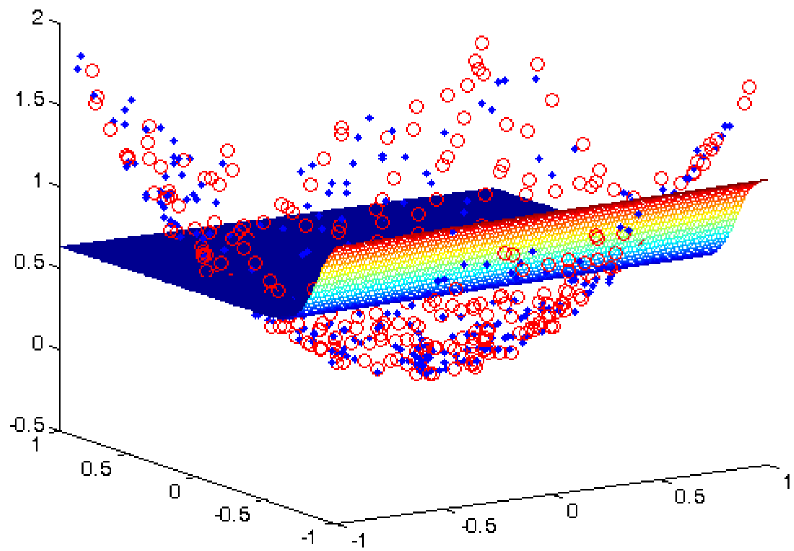


target function

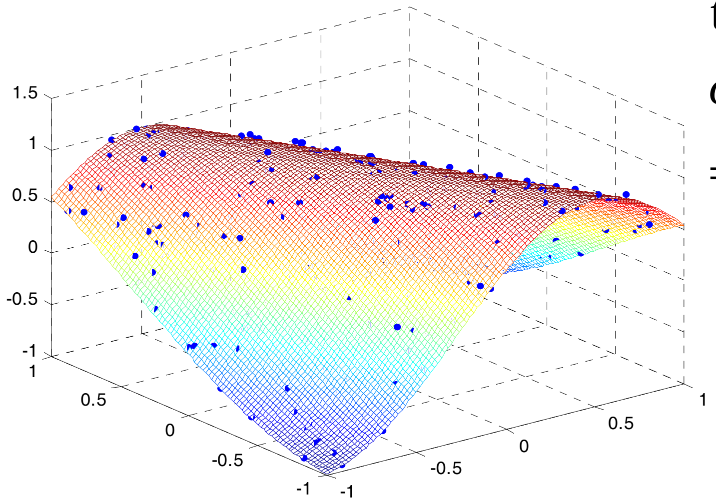
$$o = f(x_1, x_2) = x_1^2 + x_2^2$$

Noisy Sample

Approximation by single perceptron

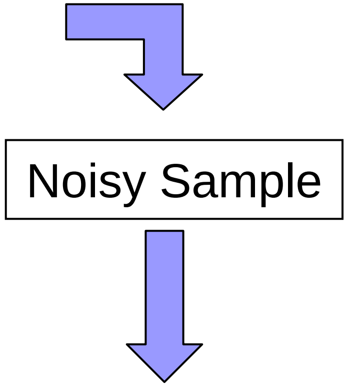


Single adaline approximation

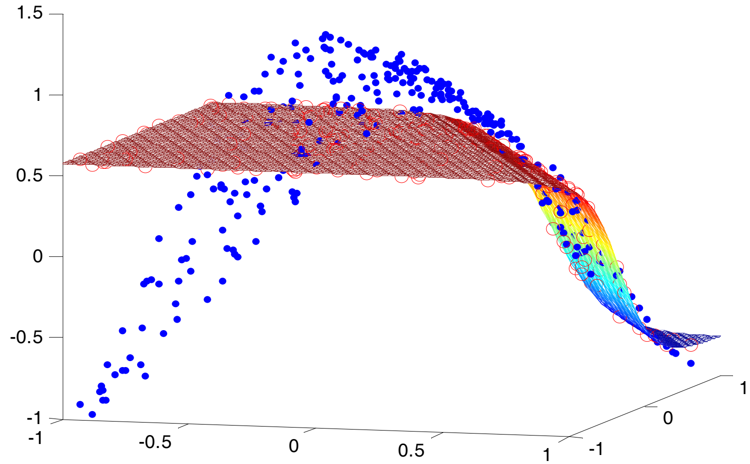


target function

$$o = f(x_1, x_2)$$
$$= \cos(x_1 + 2x_2)$$



Approximation by single perceptron

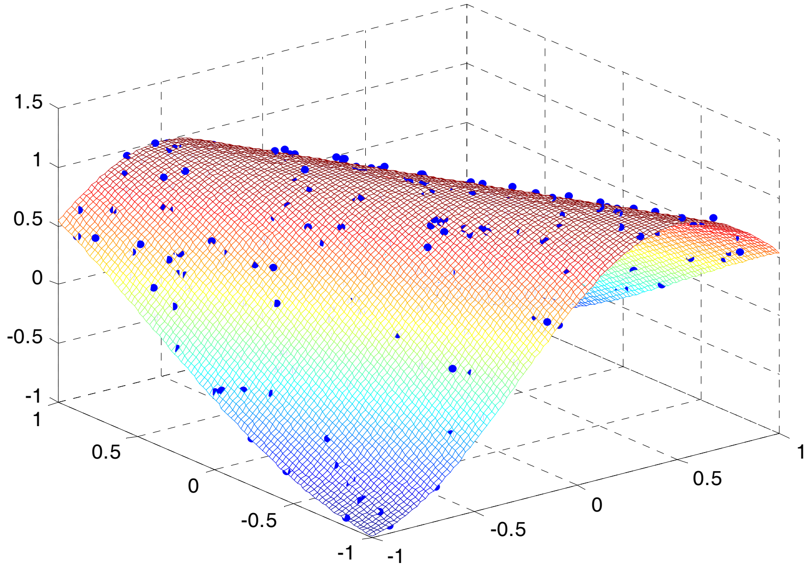


Single gadaline approximation

target function

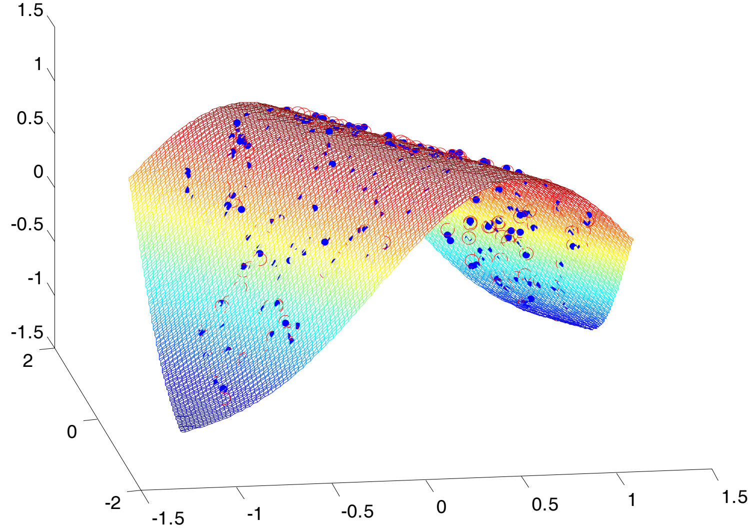
$o = f(x_1, x_2)$

$= \cos(x_1 + 2x_2)$



Noisy Sample

Approximation by single gadaline



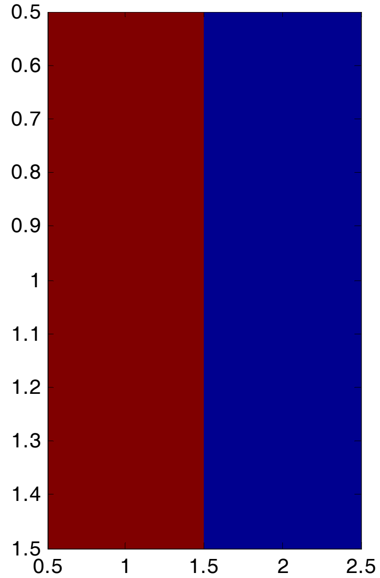
Function Composition

target function

$$o = f(x_1, x_2)$$

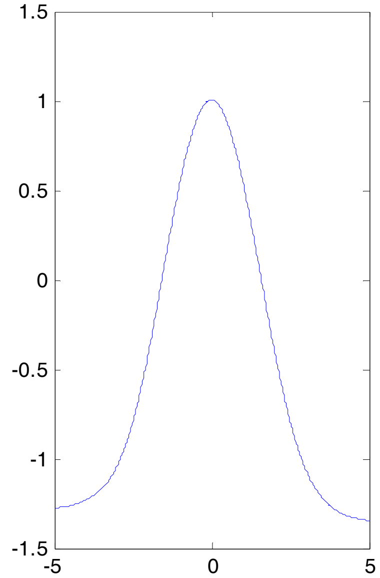
$$= \cos(x_1 + 2x_2)$$

Inner: Linear projection



w=[1 2]

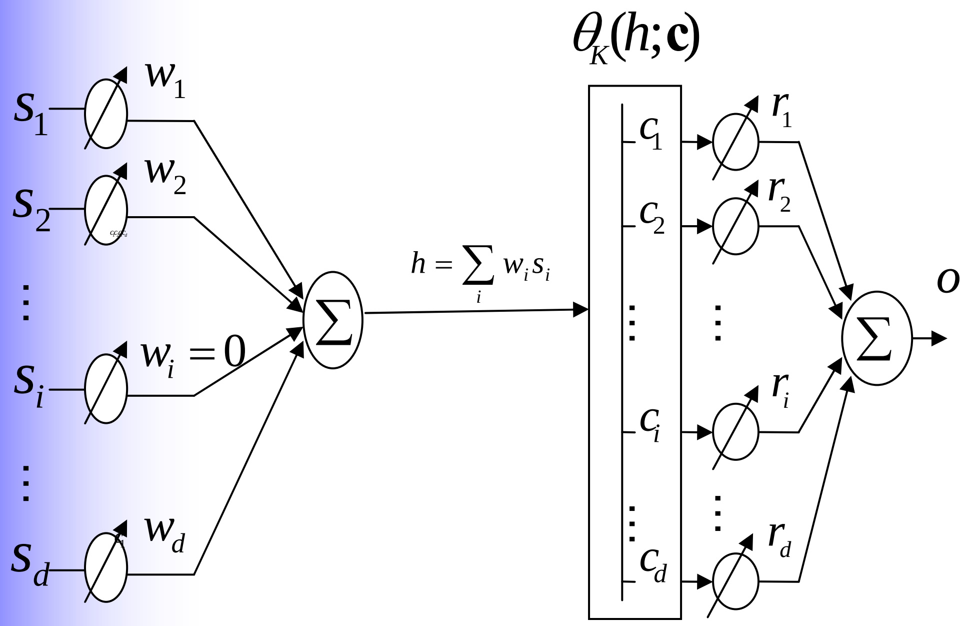
Outer: post-nonlinear



cos(h)

Single gadaline

- MISO



Approximating function

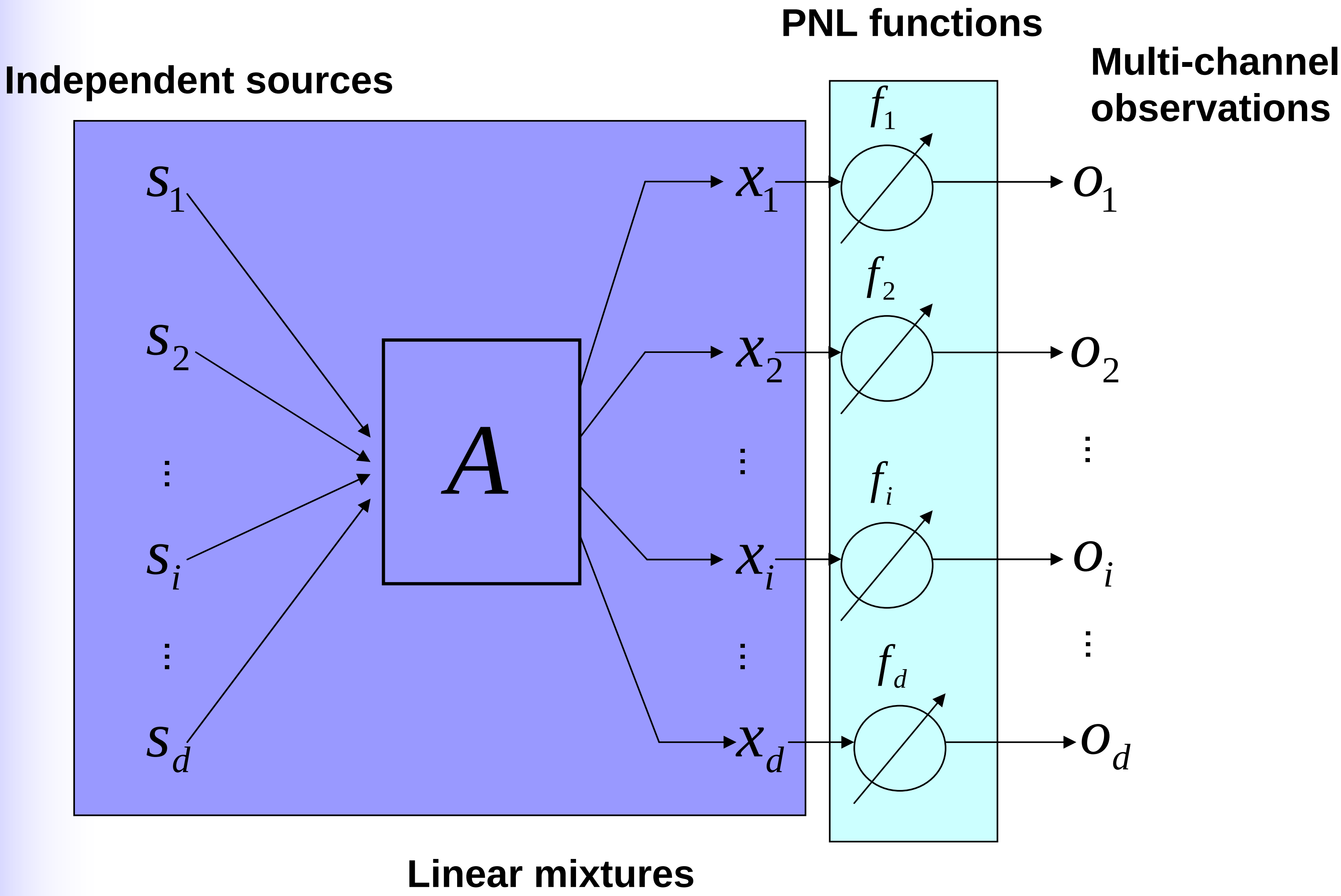
$o = g(\mathbf{s};\mathbf{w},\mathbf{c},\mathbf{r})$

$h = \mathbf{s}^T\mathbf{w}$

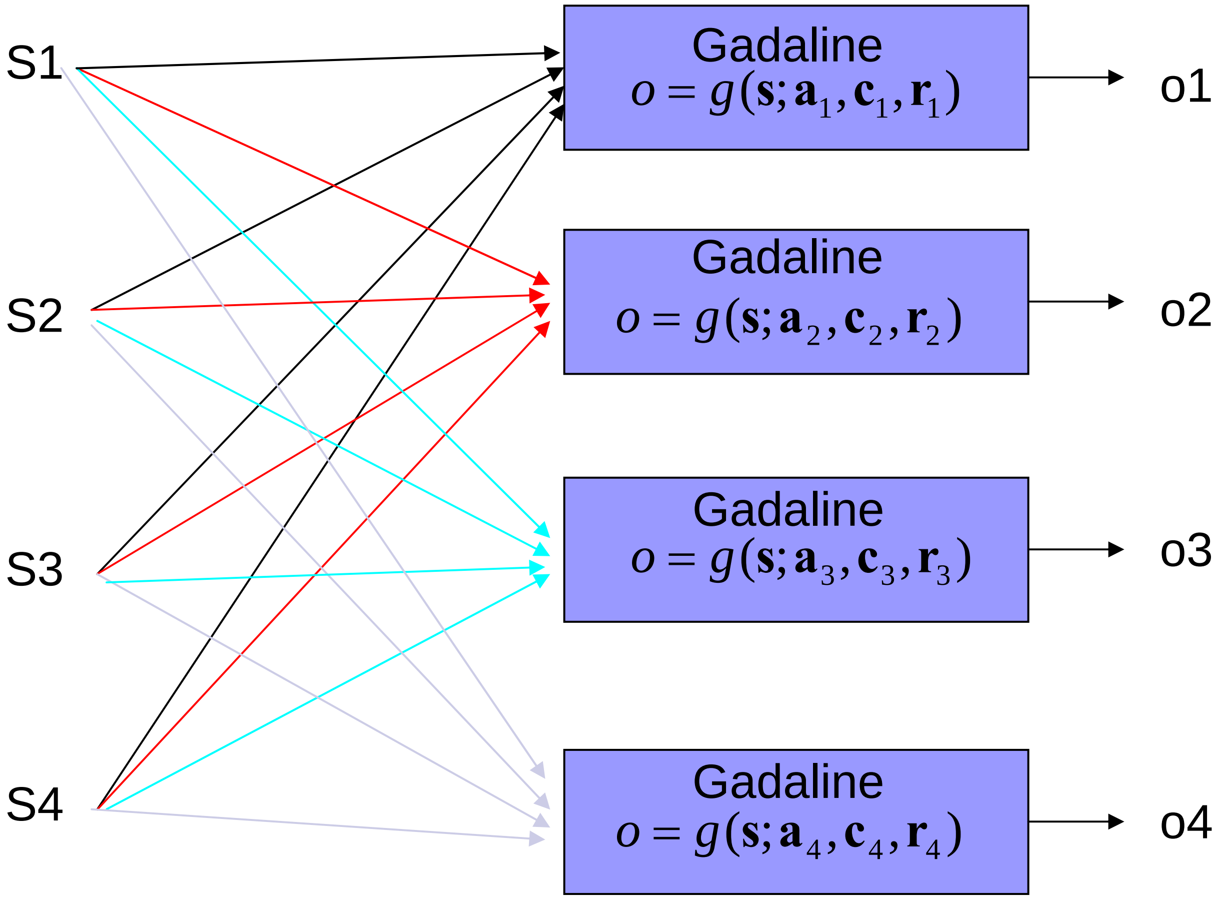
$\boldsymbol{\delta} = \theta_K(h;\mathbf{c})$

$o = \boldsymbol{\delta}^T\mathbf{r}$

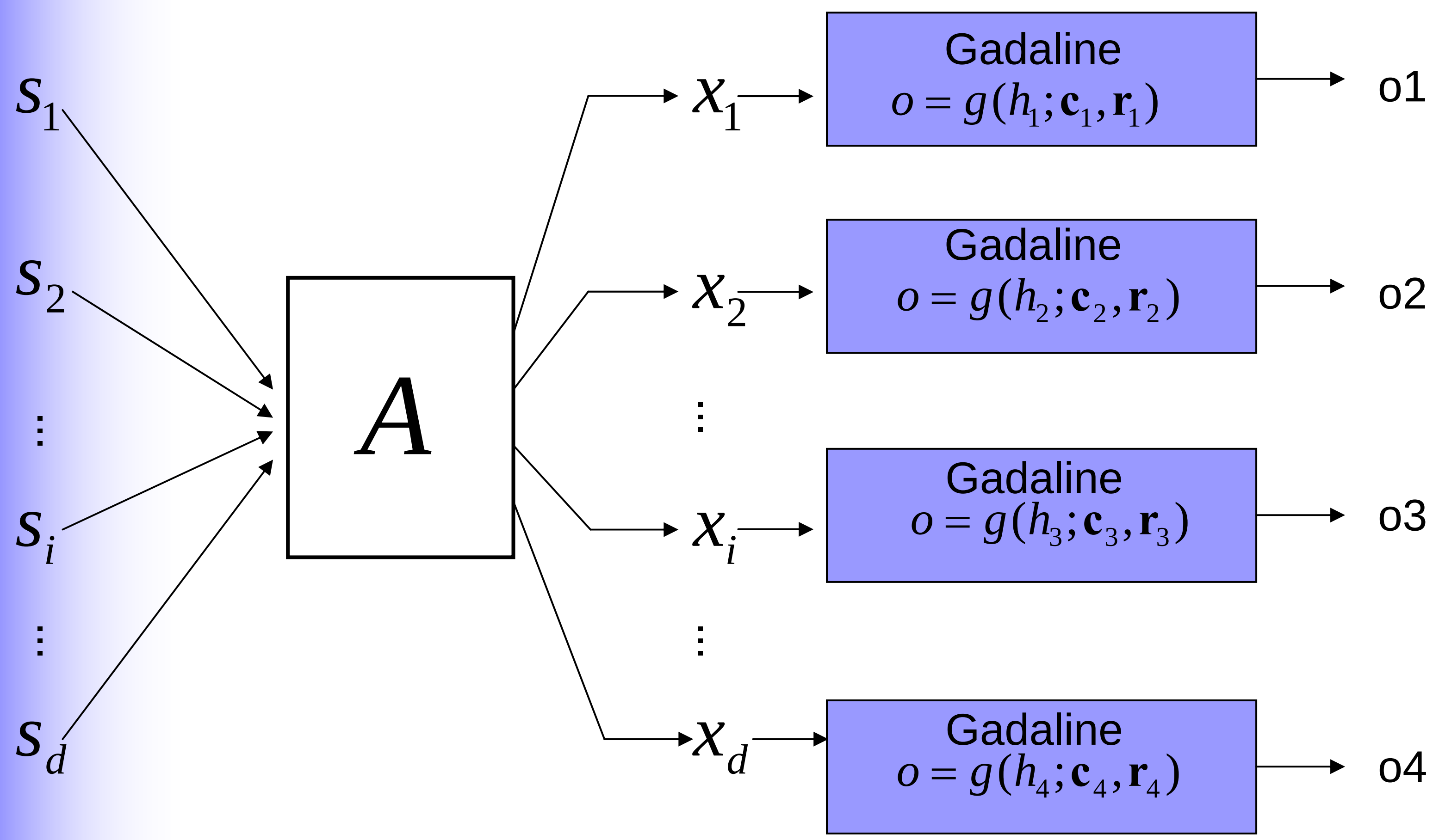
Pots-nonlinear mixtures



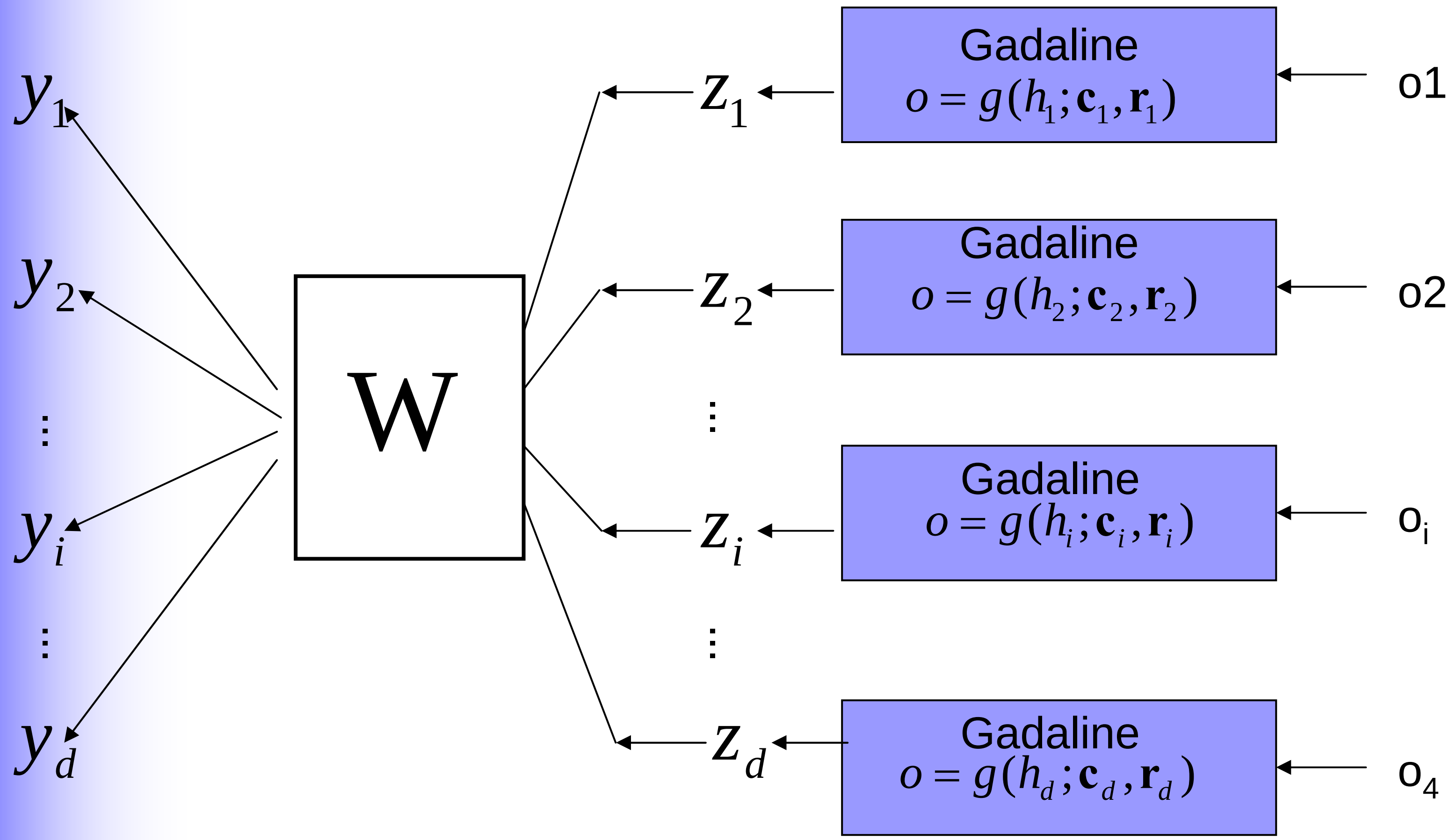
Gadaline network



Gadaline network



Gadaline network for demixing



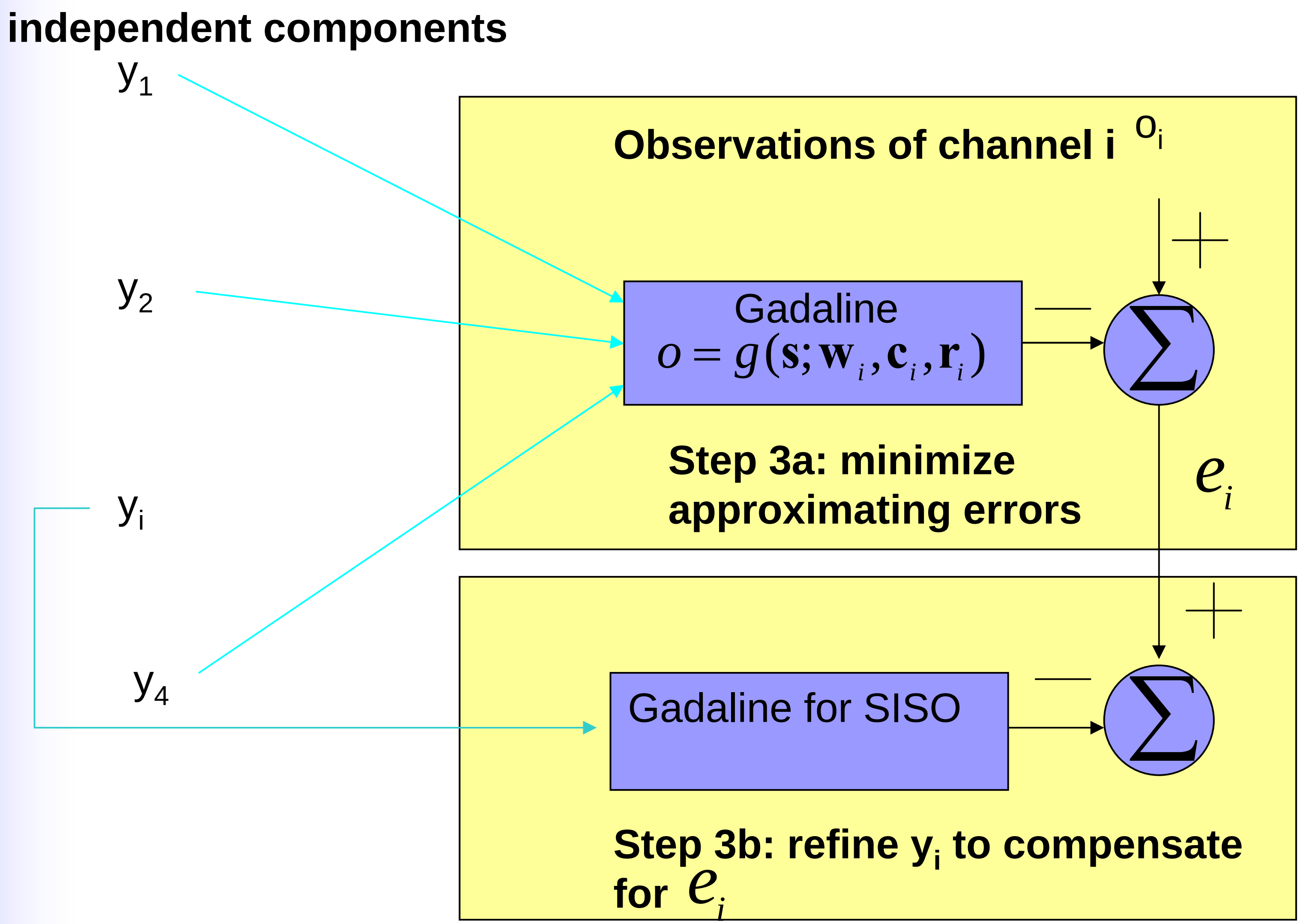
Goal

- $\{x_i[t]\}_t$ denotes observations from an output channel
- Given multi-channel observations, find independent sources and network parameters
- Outputs of learning gadalines
  - Independent sources $\{y_i[t]\}_t$
  - Network parameters

PNLICA (IJCNN 2007)

1. Input multi-channel observations
2. Set independent components to given observations
3. For each y_i
   a. Train the ith gadaline
      - Approximate x_i using current independent components other than y_i
   b. Use the approximating error to refine y_i
4. Schedule a temperature-like parameter to emulate physical annealing
5. Goto step 3 until a halting holds

Leave-one-out approximation



Strategy of leave-one-out approximation

- Step 3
  - For a selected channel, the dominant independent component is assumed absent to contribute its correspondent observations
  - The dominant component is refined to compensate for the error of approximating the selected channel in terms of the remaining independent components

Linear ICA

Lineare gadaline

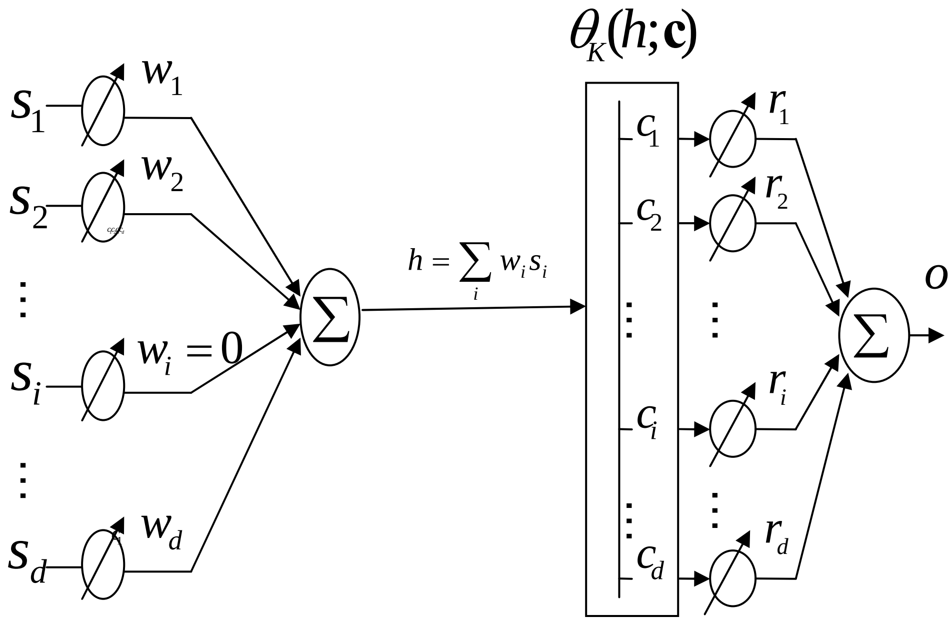
$\mathbf{c}$ contains equally spaced knots

$$\mathbf{r} = \mathbf{c}$$

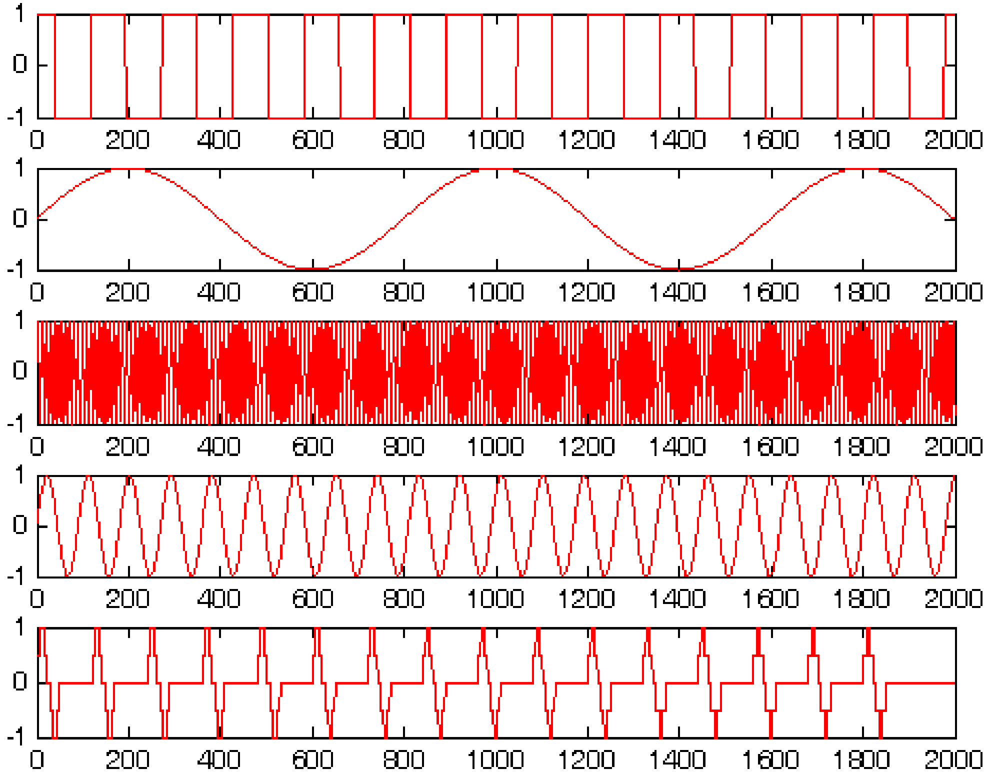
Step 3b

$$\phi(z) = \sum_{k=1}^{K} v_k[t] c_k,$$

$$\begin{cases} u_k[t] = (z - c_k)^2 \\ v_k[t] = \dfrac{\exp(-\beta u_k[t])}{\sum_{l=1}^{K} \exp(-\beta u_l[t])} \end{cases}$$

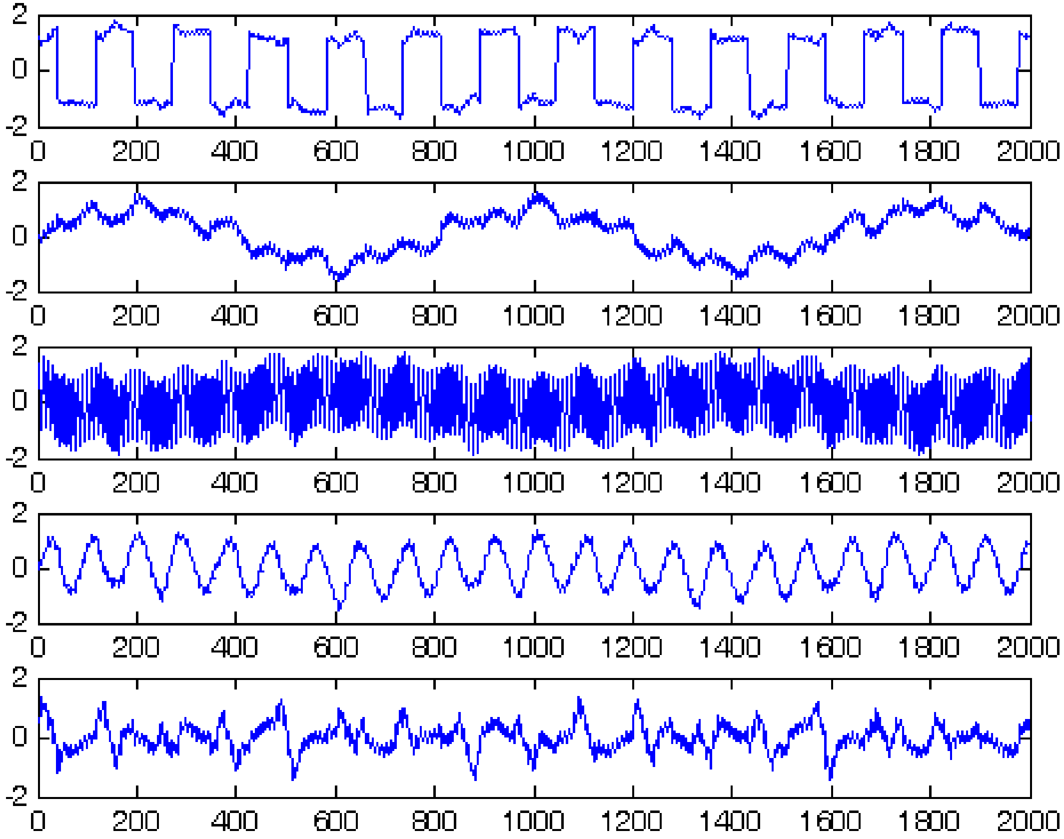


Numerical simulations



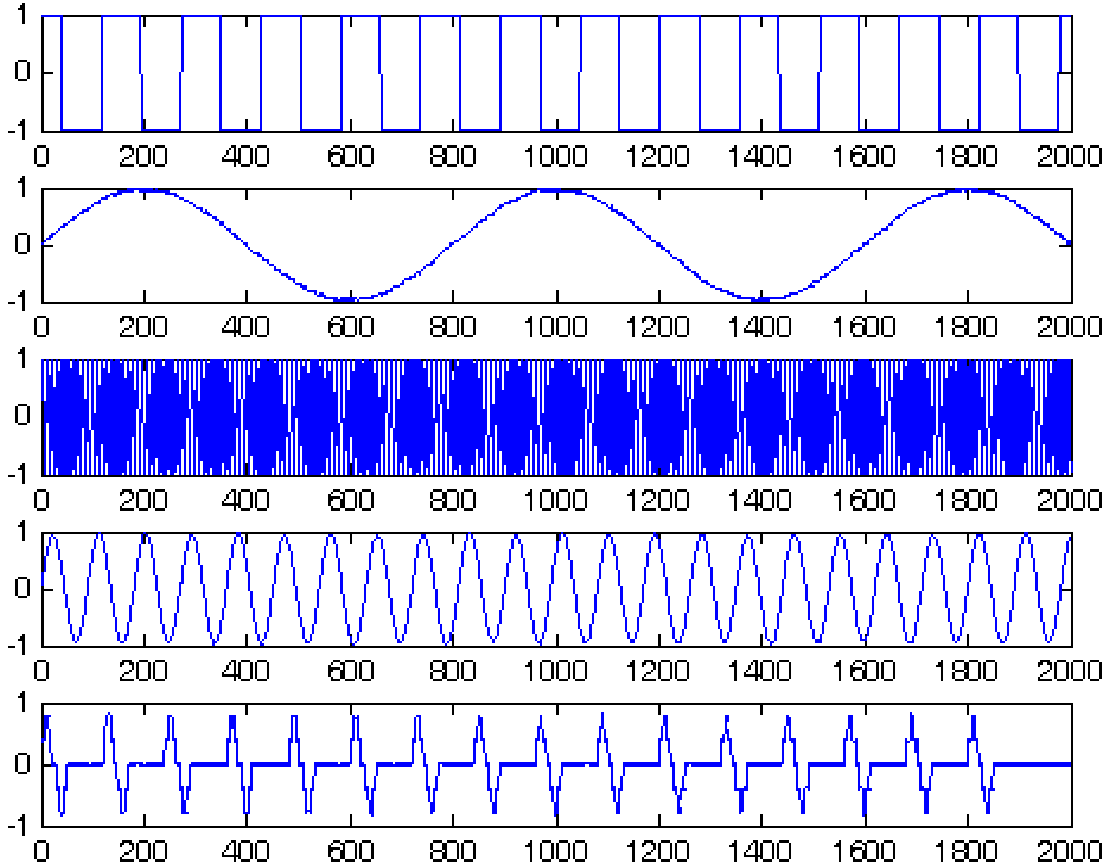
Independent sources

Multi-channel observations



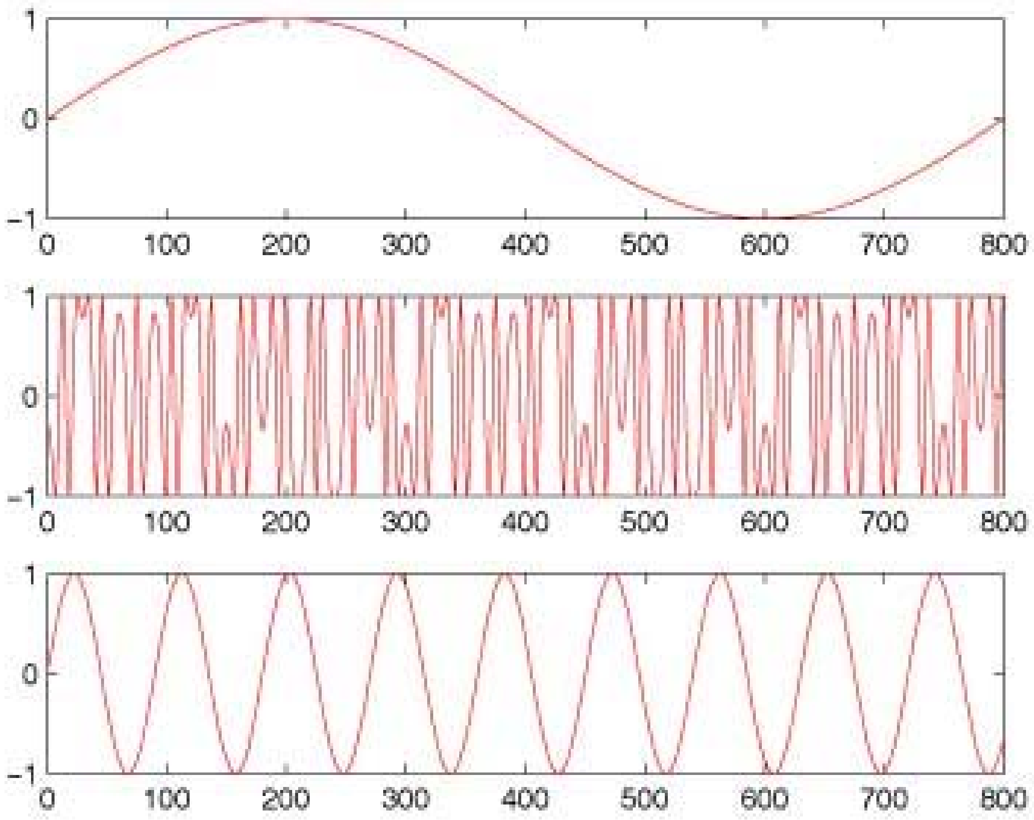
Linear mixtures of five independent sources

Recovered independent components



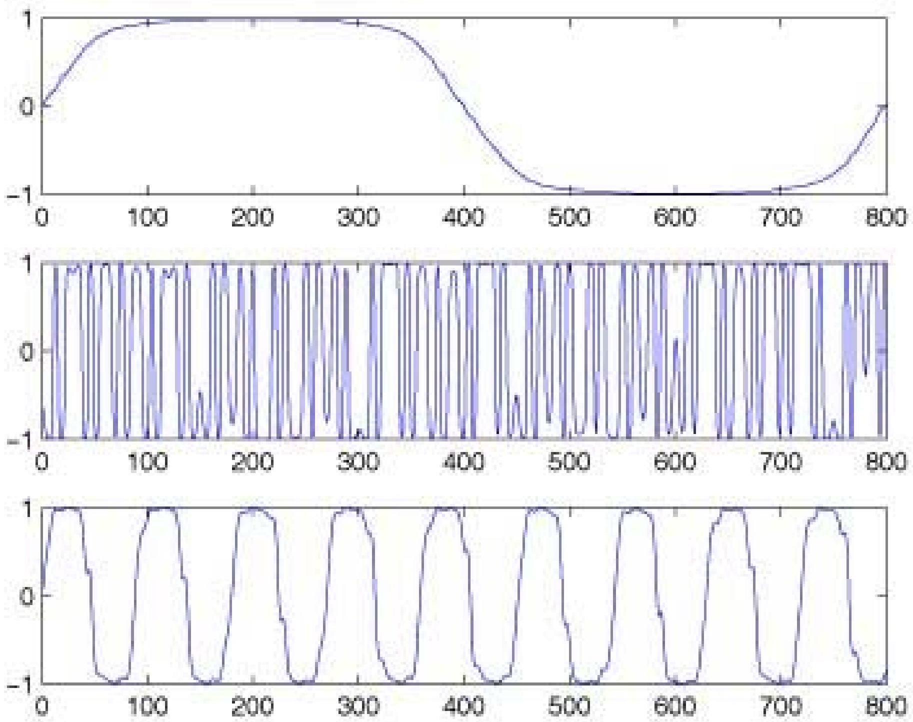
Leave-one-out linear gadaline approximation

Independent sources



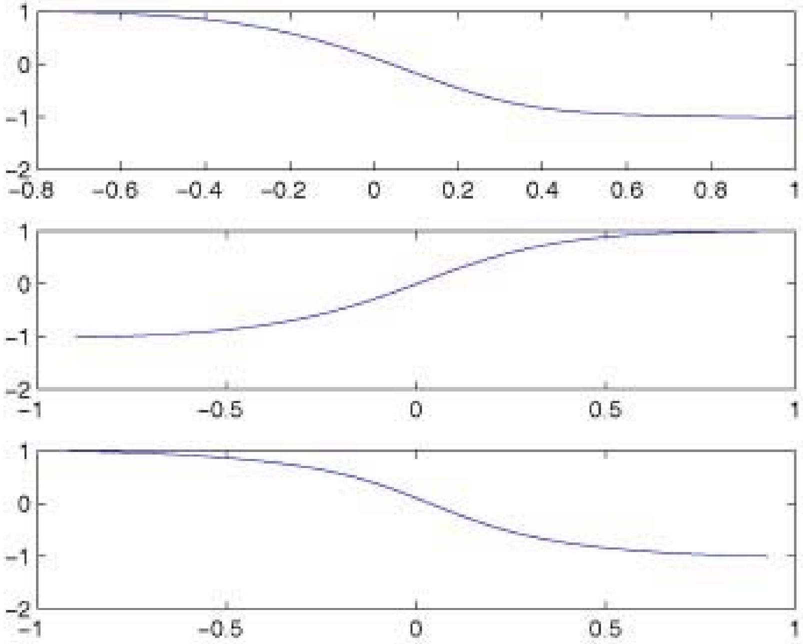
Multi-channel observations

- PNL functions: hyper-tangent functions



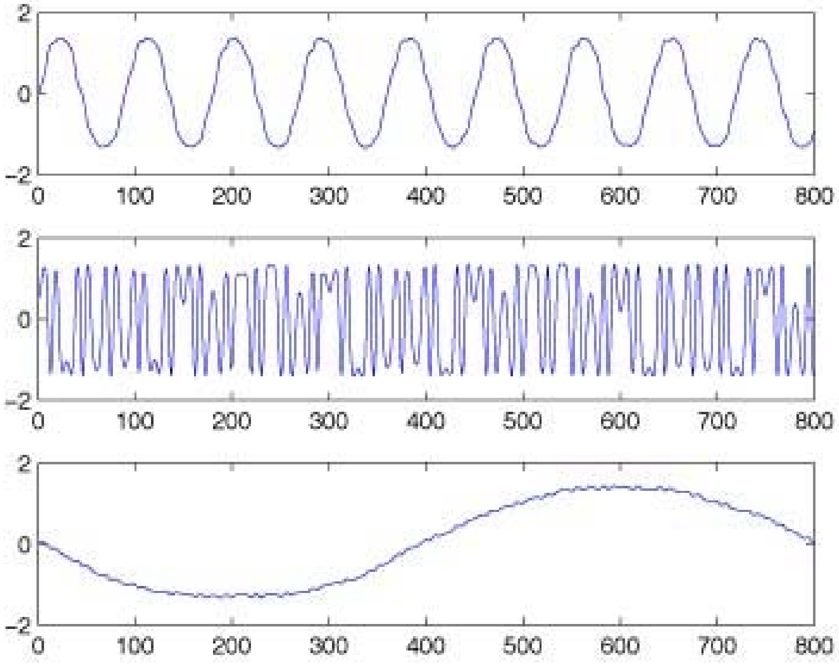
Derived PNL functions

- Leave-one-out gadaline approximation



Independent components

- Leave-one-out gadaline approximation



Conclusions

- Leave-one-out gadaline approximation has been shown effective for linear ICA and potential for post-nonlinear ICA.
- PNL ICA is translated to concurrent estimation of the gadaline network and independent components.
  - The idea is simple but its implementation needs accurate and reliable collective decisions to optimize tremendous discrete and continuous unknowns.
  - Interactive dynamics derived for leave-one-out gadaline approximation executed under the mean field annealing process are potential to fit the computation requirement.

Conclusions

- Effective PNLICA could extend application domain of blind separation
  - Traditional linear ICA algorithms are impractical for blind separation of PNL mixtures of independent sources.
  - The PNL mixture assumption which is more general for modeling formation emulation of real world signals than the linear mixture assumption.
- Properties of leave-one-out gadaline approximation
  - The proposed learning method translates PNLICA to individual sub-tasks of single gadaline optimization
  - The learning process operates under the mean field annealing process to pursuit for accurate neural computations.
  - Its derivation involves without complicate statistical criteria, such as the Kurtosis or Kullback Leibler divergence, for measuring statistical dependency of multivariates that are PNL mixtures of independent sources.

Conclusions

- Under the PNL mixture assumption, effective reduction of statistical dependency of independent components through direct minimization of the KL divergence is a complicate task that is still challenging researchers in the field of neural networks for nonlinear independent component analysis.
- Application of PNLICA to blind separation of real world signals
  - Electrocardiograms(ECG)
  - Electroencephalograms(EEG)
  - Event related potential(ERP)
  - Magnetic resonance images(MRI)

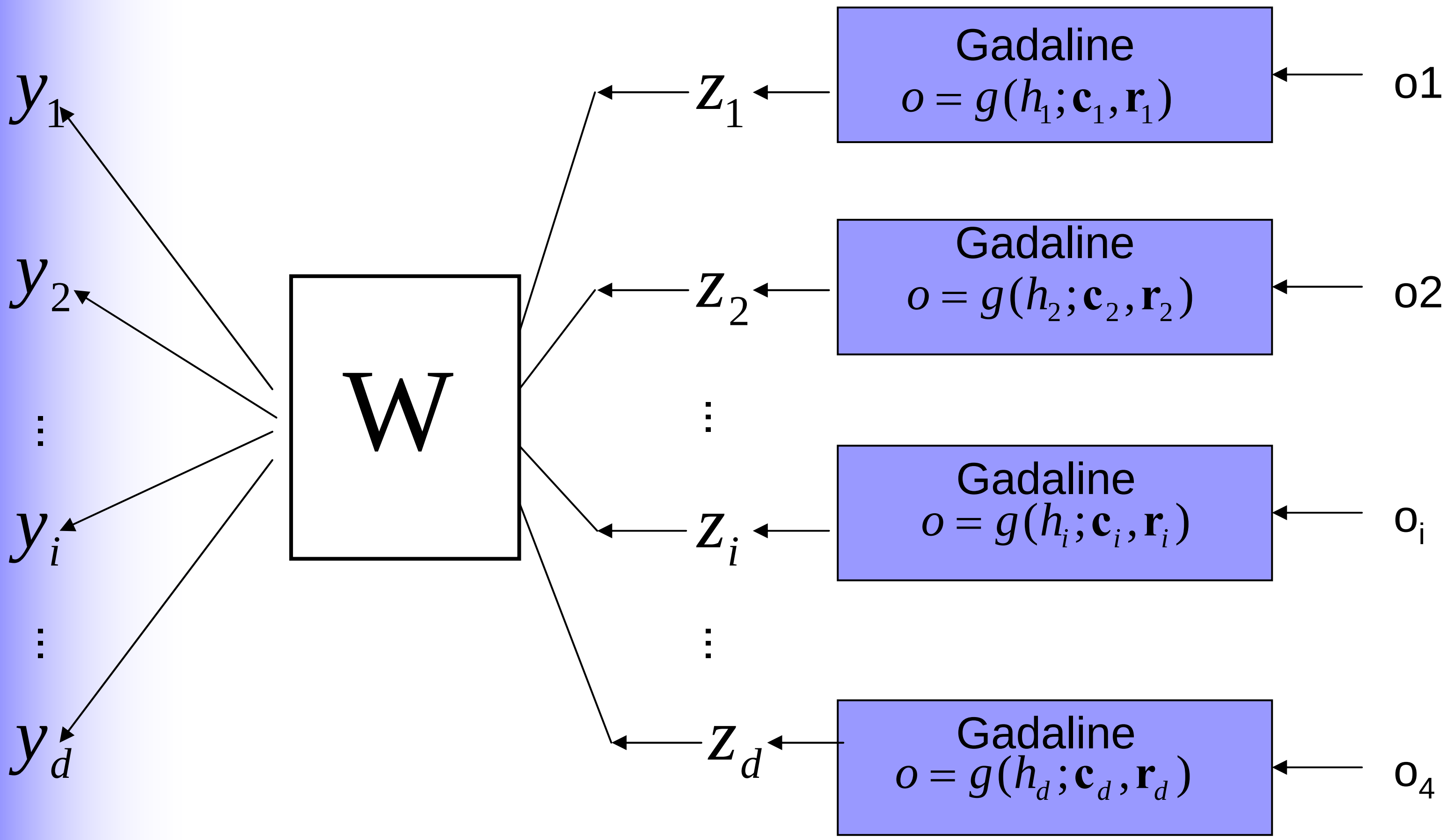
Reference

- Jiann-Ming Wu & Chiu, Independent component analysis using Potts models, IEEE Trans. On Neural Networks, Vol. 12, No. 2, March (2001)
- Jiann-Ming Wu, M.H. Chen, Lin Z.H., Independent component analysis based on marginal density estimation using weighted Parzen windows, revised for Neural Networks, 2005/11
- Jiann-Ming Wu , Z. H. Lin, and P. H. Hsu, Function approximation using generalized adalines, IEEE Trans. Neural Networks., vol. 17, no. 3, pp. 541-558, May 2006.
- Jiann-Ming Wu, Yi-Cyun Yang, Nonlinear independent component analysis by learning generalized adalines, accepted by IJCNN 2007.

PART II

- Learning gadalines for demixing

Gadaline network for demixing

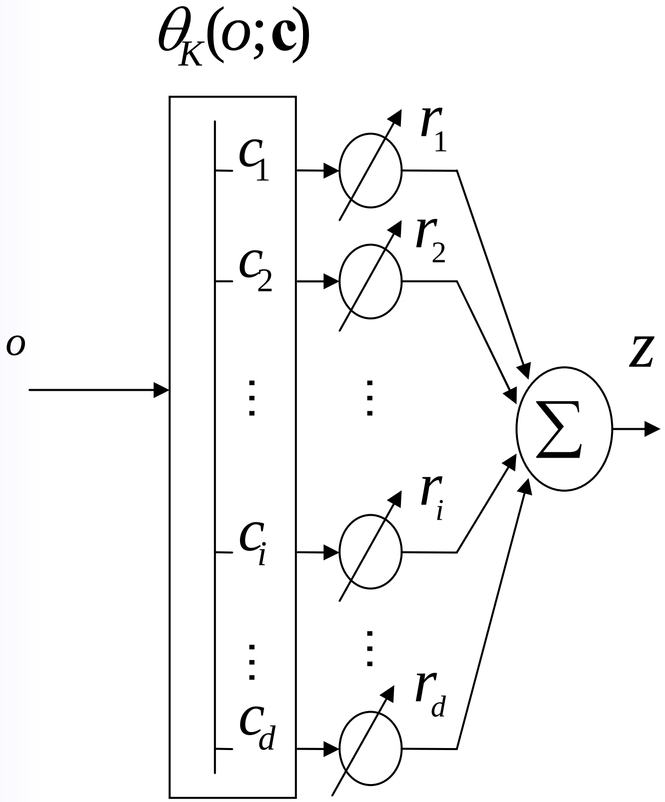


K-state transfer function

$$\theta_3(x;h) = [1,0,0] = \mathbf{e}_1^3 \quad if \quad h \in I_1$$
$$= [0,1,0] = \mathbf{e}_2^3 \quad if \quad h \in I_2$$
$$= [0,0,1] = \mathbf{e}_3^3 \quad if \quad h \in I_3$$

$$\theta_K(h;\mathbf{c}) = \left\{ \begin{array}{c} \mathbf{e_1^K} \text{ if } \mathbf{h} \in \mathbf{I_1} \\ \vdots \\ \mathbf{e_i^K} \text{ if } \mathbf{h} \in \mathbf{I_i} \\ \vdots \\ \mathbf{e_K^K} \text{ if } \mathbf{h} \in \mathbf{I_K} \end{array} \right\},$$

Inverse function

- 

$$\boldsymbol{\delta} = \theta_K(o;\mathbf{c})$$

$\boldsymbol{\delta}$ is a unitary vector

$$z = \mathbf{r}^T \boldsymbol{\delta}$$

Multiple inverse functions

- $$\boldsymbol{\delta}_{\mathrm{i}}[\mathrm{t}] = \theta_K(o[t];\mathbf{c}_i)$$

  $\boldsymbol{\delta}_{\mathrm{i}}[\mathrm{t}]$ is a unitary vector

  $$z_i[t] = \mathbf{r}_i^{\,T}\boldsymbol{\delta}_i[t]$$

  $$\mathbf{z}[t] = (z_1[t],\ldots,z_d[t])^T$$

Math framework

- $$E = \frac{1}{2}\sum_t \sum_i \sum_k \delta_{ik}[t] \left\| o_i[t] - c_k \right\|^2$$

  $$\mathrm{H} = \mathrm{E} + \mathrm{L}$$

  $\mathbf{x}[t]$ in L is replaced with $\mathbf{z}[t]$

Math programming for minimization of KL divergence

$$L(\xi, \mathbf{W}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\xi_{ik}[t]|\mathbf{w}_i\mathbf{x}[t] - h_k|^2$$

$$- c\log|\det(\mathbf{W})|$$

$$- \frac{c}{T}\sum_{k=1}^{K}\sum_{t=1}^{T}\xi_{ik}[t]\ln\left(\sum_{t=1}^{T}\xi_{ik}[t]\right)$$

$$\sum_{k=1}^{K}\xi_{ik}[t] = 1 \quad \text{for all } i, t,$$

$$\xi_{ik}[t] \in \{0, 1\} \quad \text{for all } i, k, t.$$

Hopfield-like energy function

Hopfield-like energy function:

$$L(\xi, \mathbf{W}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} \xi_{ik}[t] |\mathbf{w}_i \mathbf{x}[t] - h_k|^2 - c \log |\det(\mathbf{W})| - \frac{c}{T} \sum_{k=1}^{K} \sum_{t=1}^{T} \xi_{ik}[t] \ln \left(\sum_{t=1}^{T} \xi_{ik}[t] \right)$$