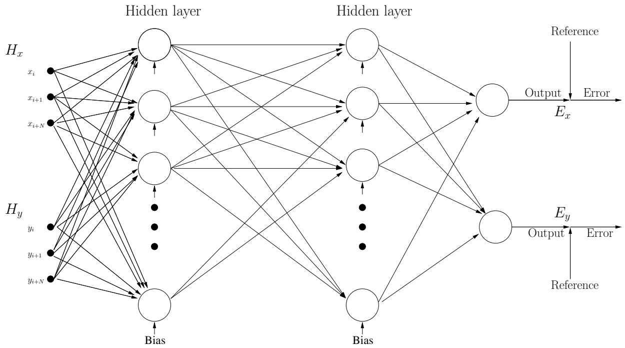
Machine Learning in Geophysics

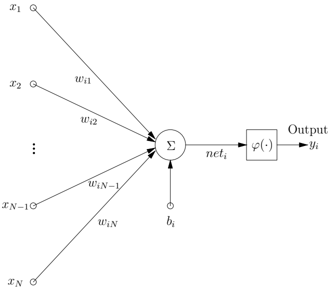
Lecture 7 – Towards neural networks

Neural networks



- Analogy to brain function
- Supervised machine-learning method
- Can be applied to wide range of problems

Sigmoid neuron



Typical NN consists of many individual neurons, e.g.

$$
\begin{aligned}
net_i &= \sum_{j=1}^{N} w_{ij} x_j + b_i \\
\varphi(net_i) &= \tanh\left(\frac{1}{2} net_i\right)
\end{aligned}
$$

Linear adaptive filters

- Powerful NN consist of many neurons, layers
- Difficult to understand what exactly is happening (black box)
- Can perform useful tasks with one elementary unit
- Typically called linear adapative combiner, linear adaptive filter
- Used in signal processing, e.g. noise cancelling headphones

Adaptive noise cancellation

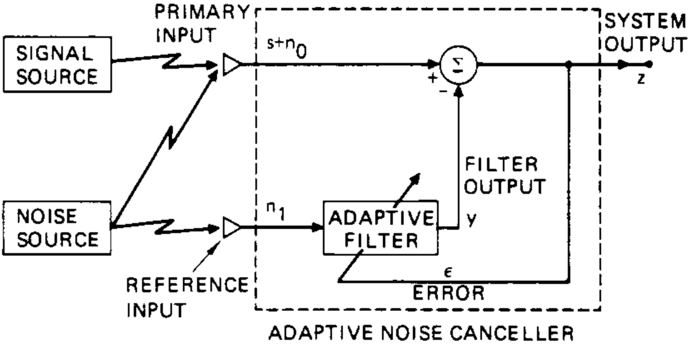
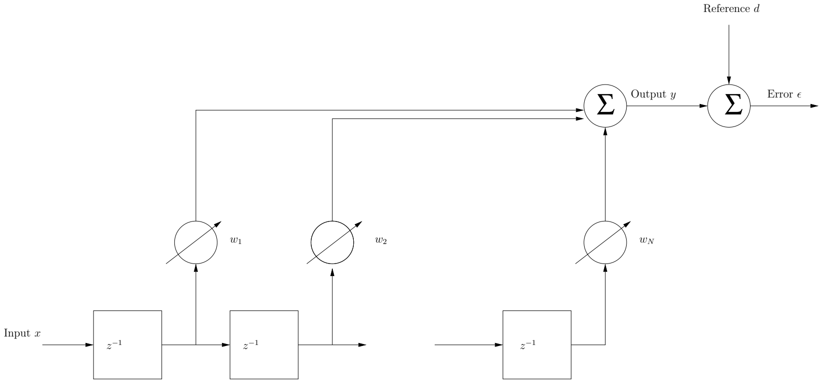


Fig. 1. The adaptive noise cancelling concept.

Widrow et al., 1975

Adaptive filter –schematics



- We have input x, e.g. noise estimate (z^{-1} denotes time delay)
- Output is weighted sum
- Compare with reference d, try to minimize error ϵ

LMS Adaptive filter –equations

Start with $k = 0$, $\mathbf{w} = 0$, at step k

$$\begin{aligned} y_k &= \mathbf{x}_k^T \mathbf{w}_k \\ \epsilon_k &= d_k - \mathbf{x}_k^T \mathbf{w}_k \end{aligned}$$

Question

We want to find **w** that minimize the squared error

$$\epsilon_k^2 = \left(d_k - \mathbf{x_k}^T \mathbf{w_k} \right)^2 \rightarrow \min$$

How can we achieve that?

Adapting **w**

Answer

Calculate derivative

$$\frac{\partial}{\partial \mathbf{w}} \left(d_k - \mathbf{x}_k^T \mathbf{w}_k \right)^2 = 2 \left(d_k - \mathbf{x}_k^T \mathbf{w}_k \right) \mathbf{x}_k = 2\epsilon \mathbf{x}_k$$

Choose step-length μ, adapt weights

$$\mathbf{w}_{k+1} = \mathbf{w}_k + 2\mu\epsilon\mathbf{x}_k$$

Not the most efficient algorithm, but computationally simple

Step size

- As we saw, the step size is a critical parameter and needs to be chosen "right".
- Can estimate step size

$$\mu = \frac{\tilde{\mu}}{\delta + \|\mathbf{x}_k\|^2}$$

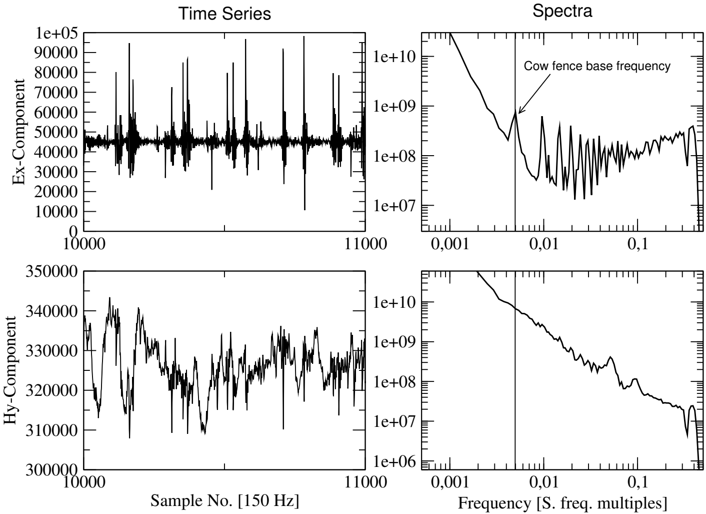
  with δ a small number to avoid division by zero and adaptation constant $\tilde{\mu} \approx 0.5$.
- Makes algorithm more robust

Magnetotellurics

$$\begin{pmatrix} E_x(\omega) \\ E_y(\omega) \end{pmatrix} = \frac{1}{\mu_0} \underbrace{\begin{pmatrix} Z_{xx}(\omega) & Z_{xy}(\omega) \\ Z_{yx}(\omega) & Z_{yy}(\omega) \end{pmatrix}}_{\text{complex impedance tensor}} \begin{pmatrix} B_x(\omega) \\ B_y(\omega) \end{pmatrix}$$

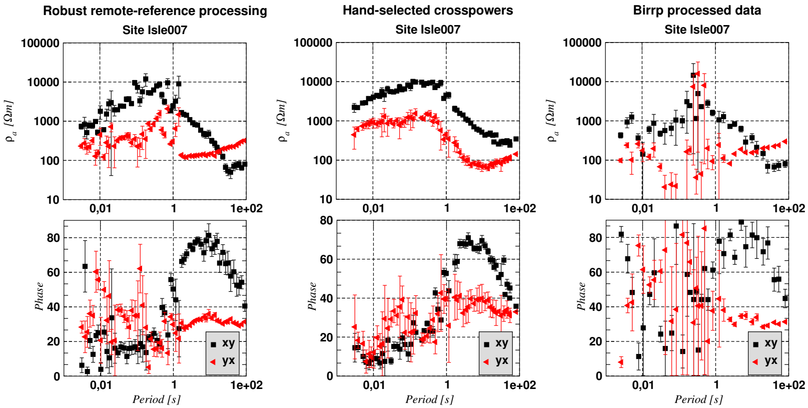
- Inductive, passive electromagnetic method
- Electric field and magnetic field are measured simultaneously
- Recorded data is windowed and transformed into frequency domain
- Local noise sources make far-field assumption invalid

Example of contaminated time series and spectra



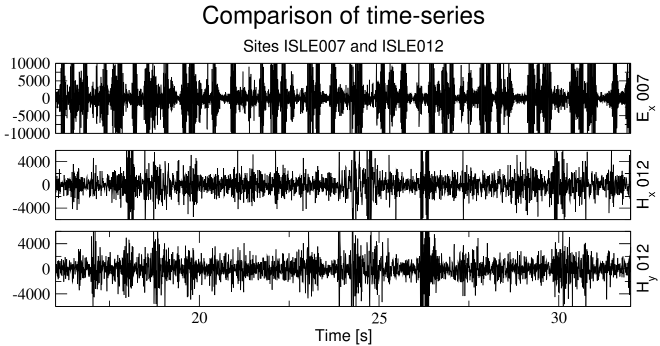
Strong cow-fence spikes in time series, spectra show strong

Remote-reference processing of contaminated data



Curves should be smooth with small error bars

Basic idea

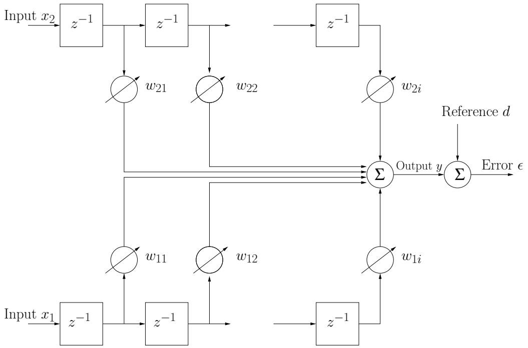


- We have several synchronously recording sites
- Noise free magnetic field is highly correlated but not identical
- Everything is affected by noise to a different degree
- Need to extract the correlated parts

Noise reduction using adaptive filters

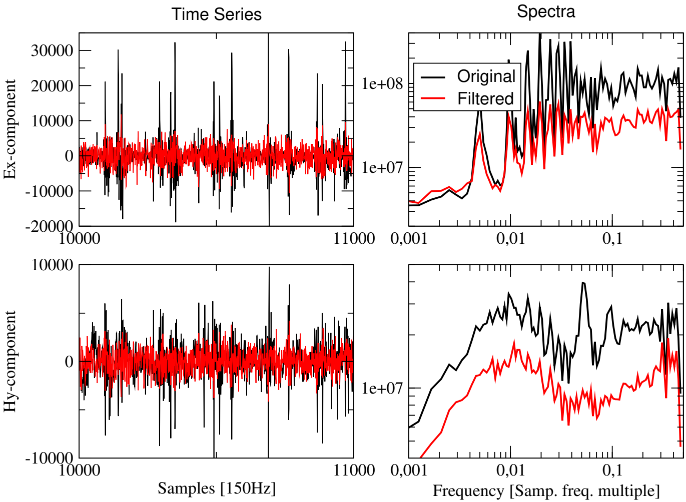
- Adaptive filters are widely used in a range of signal-processing tasks from electrical engineering to medicine
- Provide flexible design that can adjust to problem
- LMS-adaptive linear combiner simplest adaptive filter
- Assume stationary or slowly varying non-stationarity
- Can be extended to non-linear domain $\Rightarrow$ neural networks

Conceptional diagram of a two-channel LMS-Filter



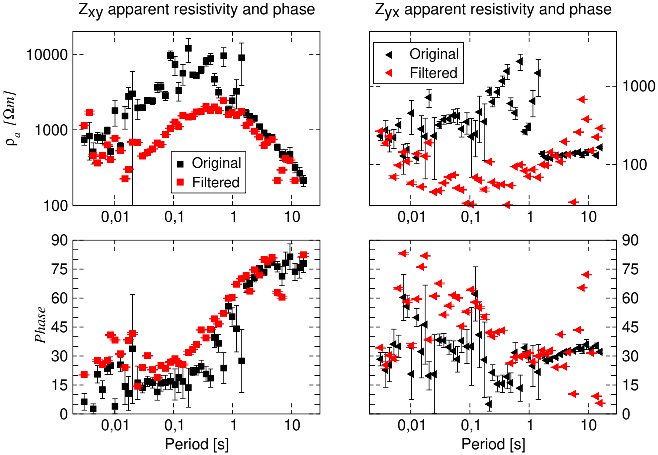
H_x and H_y of remote site used to clean up local data (E_x, E_y, H_x and H_y).

Application to a noise contaminated site



Filter removes spikes and reduces spectral signature

Comparison of transfer functions



Improvement in Z_{xy}, maybe apparent resistivity Z_{yx}, suspicious phase Z_{yx}

Summary

- Adaptive combiner is linear part of neural network neuron
- Used as adaptive filter to remove noise
- Implementation relatively simple, some important tuning parameters
- Can be used in noise removal for geophysical time-series