LMS: PAST, PRESENT AND FUTURE:

Puzzles, Problems and Potentials

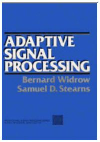
Victor Solo

School of Electrical Engineering
University of New South Wales
Sydney, AUSTRALIA

ICASSP 2019, Brighton, UK

Adaptive Signal Processing
B Widrow USA

Adaptive Control
PC Young UK



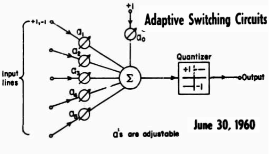
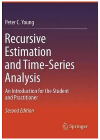
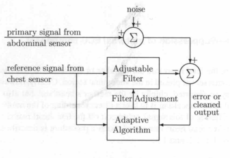


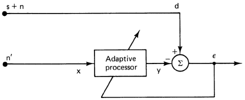
FIG. 1.-- AN ADJUSTABLE NEURON.

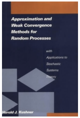


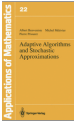
Outline

1. $\begin{matrix} Educated \\ Origins \end{matrix}$ in $\left\{ \begin{matrix} Control \\ SignalProcessing \end{matrix} \right.$
2. LMS does not Converge!;
   But it does Perform!
3. Network LMS.
4. Education Notes
5. The Future is?
   Statistical Signal Processing
   not Machine Learning!
6. Conclusions.

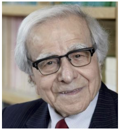
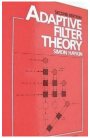


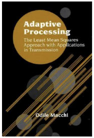


HJ.Kushner 1984 *A.Benveniste et.al.* 1990

S.Haykin 1986 *O.Maachi* 1995

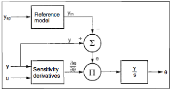


Fig. 5. Block diagram of the parameter adjustment mechanism in a model-reference adaptive system (MRAS).

Whitaker MITRule,1959

dsp HISTORY Bernard Widrow

Thinking About Thinking:

The Discovery of the LMS Algorithm

IEEE SIGNAL PROCESSING MAGAZINE [100] JANUARY 2005

LMS does not Converge!?

Heuristics + Steepest Descent

Problem: Minimize instantaneous squared error $\frac{1}{2}e_t^2(w)$ where the error signal is

$$e_t(w) = y_t - x_t^T w.$$

Steepest Descent gives classic form

$$\underset{\hat{w}_t}{w_{new}} = \underset{\hat{w}_{t-1}}{w_{old}} + \underset{\mu}{gain} * \underset{x_t}{gradient} * \underset{e_t}{error}$$

$$e_t = y_t - x_t^T \hat{w}_{t-1}$$

But $\mu-$scaling is ignored.
To choose μ it has to be scaled: $\mu = \frac{\mu_o}{\sigma_x^2}$ where μ_o is scale free.

Error Analysis

Weight error $= \tilde{w}_t = \hat{w}_t - w_t$. Then

$$\delta \tilde{w}_t = -\mu x_t x_t^T \tilde{w}_{t-1} + \mu x_t n_t + \delta w_t$$

where $\delta w_t = w_t - w_{t-1}$.

- This is a time-variant stochastic difference equation and so its state does not converge.
- But under certain regularity conditions it does settle into a steady state.

LMS Hovers

Averaging Analysis

Need realization-wise analysis.

- Sum the error system

$$\tilde{w}_{T+N} - \tilde{w}_T = -\mu\Sigma_T^{T+N} x_t x_t^T \tilde{w}_{t-1} + \mu\Sigma_T^{T+N} x_t n_t$$

- Change is slow so:

$$\Rightarrow \tilde{w}_{T+N} - \tilde{w}_T \approx -\mu\Sigma_T^{T+N} x_t x_t^T \tilde{w}_{T-1} + \mu\Sigma_T^{T+N} x_t n_t$$

- Now approximate with averages

$$\Rightarrow \tilde{w}_{T+N} - \tilde{w}_N \approx -\mu N R_x \tilde{w}_{T-1} + 0$$

- Now difference

$$\Rightarrow \delta m_t = -\mu R_x m_{t-1} + (\delta w_t)$$

This is the averaged system and is stable if $0 < \mu\lambda_{max}(R_x) < 2$ [a]

[a] ODE, Weak convergence can't give this

Averages

Assume, as $N \to \infty$:

- Stationary Regressors $\frac{1}{N}\Sigma_T^{T+N} x_t x_t^T \to R_x$
- Stationary Noise $\perp$ Regressors $\frac{1}{N}\Sigma_T^{T+N} x_t n_t \to 0$.

Hovering Theorem

What does the original system do?

- It hovers/jitters/fluctuates in the vicinity of the equilibrium points of the averaged system.
- To complete the stability analysis one needs a Hovering Theorem which links the two trajectories.

LMS Performs

Widrow et.al. 1976

Introduced fundamental measures of performance.

$P(\mu) = E(\tilde{w}_t \tilde{w}_t^T)$.

$\mathcal{E}(\mu) = E(e_t^2) = E(y_t - x_t^T \hat{w}_{t-1})^2$

They look simple but are very challenging to calculate.

Weight Error Variance

Under S⊥F :

stationarity$+ \underset{regressors}{noise\perp} +$fixed w

it can be shown that

$P(\mu) = P_o + o(\mu)$ where

$R_x P_o + P_o R_x = F_{xn}(0) = \Sigma_{-\infty}^{\infty} \gamma_k^x \gamma_k^n$

MSE

Under S⊥F it can be shown

$\mathcal{E}(\mu) = \mu tr(F_{xn}(0)) + o(\mu)$.

White Noise Fallacy

With either white regressors or white noise (or both) the formulae reduce to the <u>all</u> white noise formulae with

$F_{xn}(0) = \gamma_0^x \gamma_0^n$.

This explains mistaken claims that the <u>all</u> white noise formulae are always correct.

Network LMS Stability

Node-wise Measurements

Each node records measurements related to a common weight vector.

$y_{k,t} = x_{k,t}^T w_e + n_{k,t}$
for $k = 1, \cdots, N$.

LMS has various NW extensions; but all use local information.

Network LMS does not Converge

$\hat{w}_t =$
$\mathcal{A}_2^T(\mathcal{A}_o^T - \mu\mathcal{R}_t)\mathcal{A}_1^T \hat{w}_{t-1} + \mu\mathcal{A}_1^T \sigma_t^{xy}$

$\mathcal{A}_i$ are adjacency matrices.
$\mathcal{R}_t = bdiag(x_{k,t}x_{k,t}^T)$
$\sigma_t^{xy} = [x_{k,t}y_{k,t}]$

Error System

Unexpectedly, the error system is a two-time scale system

$\delta\theta_t = \mu f(t, \theta_{t-1}, \xi_{t-1})$ ←*slow*
$\xi_t = S\xi_{t-1} + \mu g(t, \theta_{t-1}, \xi_{t-1})$ ←*fast*

Under S⊥F +
$M = \mathcal{A}_1\mathcal{A}_o\mathcal{A}_2$ is primitive e.g. strongly connected & ≥ 1 self loop.
$\Rightarrow M$ is left stochastic ($1^T M = 1^T$) and so has a Perron right eigenvector with unit eigenvalue.

- Set $A = \Sigma_1^N \alpha_k R_{x,k}$ where α_k depend on the Perron eigenvector.
- Then the averaged system is stable if $\mu \times spec.rad.(A) < 2$.

Network LMS Performance

Network Weight Error Variance

Under S⊥F and M primitive.

$P(\mu) = P_o\mu + o(\mu)$ where

$AP_o + P_oA = F_{xn}(0) = \Sigma_1^N \alpha_k^2 F_{xnk}(0)$

$F_{xnk}(0) = \Sigma_{-\infty}^{\infty} \gamma_r^{x_k} \gamma_r^{n_k}$

Network MSE

Under S⊥F and M primitive.

$\mathcal{E}(\mu) = \mu tr(F_{xn}(0)) + o(\mu)$.

White Noise Fallacy

With either white regressors or white noise (or both) the formulae reduce to the all white noise formulae with

$F_{xnk}(0) = \gamma_0^{n_k} \gamma_0^{x_k}$.

This explains mistaken claims that the all white noise formulae are always correct.

Education Notes*

Hardware

Emerging opportunities for simple physical demos.

Averaging

First order averaging is easy to teach but extremely powerful.
Two-time scale phenomena are crucial but poorly treated (Kokotovic).

Emerging Applications

Networks
Internet of Things
NextG Communications (Quantum?)
Cyber Security (from DSP/Control angle)

Simulation

Badly done (Ripley).
Design is poorly motivated.
Tracking is ignored.
Scaling is ignored.
Visualization is uneven (Tufte).

New Approaches

IEEE Magazines provide a superb source of projects/implicit teaching approaches + offline via $\frac{\textit{forwards}}{\textit{backwards}}$ ASP.

* Beware the two 'adaptive' imposters in: Statistics; Spatial Signal Processing

The Future of ASP

Strengths

- Cheap tracking in real time.

⇒ Cheap Tracking offline.
(noncausal methods are costly).

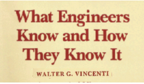
- TV-parameters are ubiqitos
  but widely ignored *online& offline*

Opportunities

Streaming Data
Event Triggered Data (Point Processes)
Biomedical → Neuroimaging
and Neuroscience
Internet of Things

Weaknesses

FIR.
⇒<u>Causal</u> basis systems
(e.g. Laguerre) - big potential
Kernel versions?

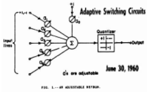


DESIGN

Challenges

- Machine Learning (McL) and AI communities shows little awareness of Adaptive Signal Processing/Control.
- Their algorithms reflect no training in: Physics/Dynamics/ Stability/Autocorrelation.
- Push back: go to McL conferences!

Conclusions

Design

LMS is the 'gift that keeps giving'.
Why? Because it is:
Linear, Adaptive, Design flexible

Analysis

Averaging is simple but can handle any kind of Adaptive algorithm in any scenario. It can be developed heuristically as well as rigorously.

Education

Hands on + Simulation taken seriously + analysis via Averaging + emerging applications

The Future

Big data provides huge opportunities for adaptive algorithms both online and offline.

Physics/Dynamics/Stability/Autocorrelation are central to real time data analysis. Uninformed by these knowledge realms, McL/AI solutions will fall (even catastrophically) short†

Vivat ASP/DSP!*

†'Universal' algorithms too conservative
* and Control!