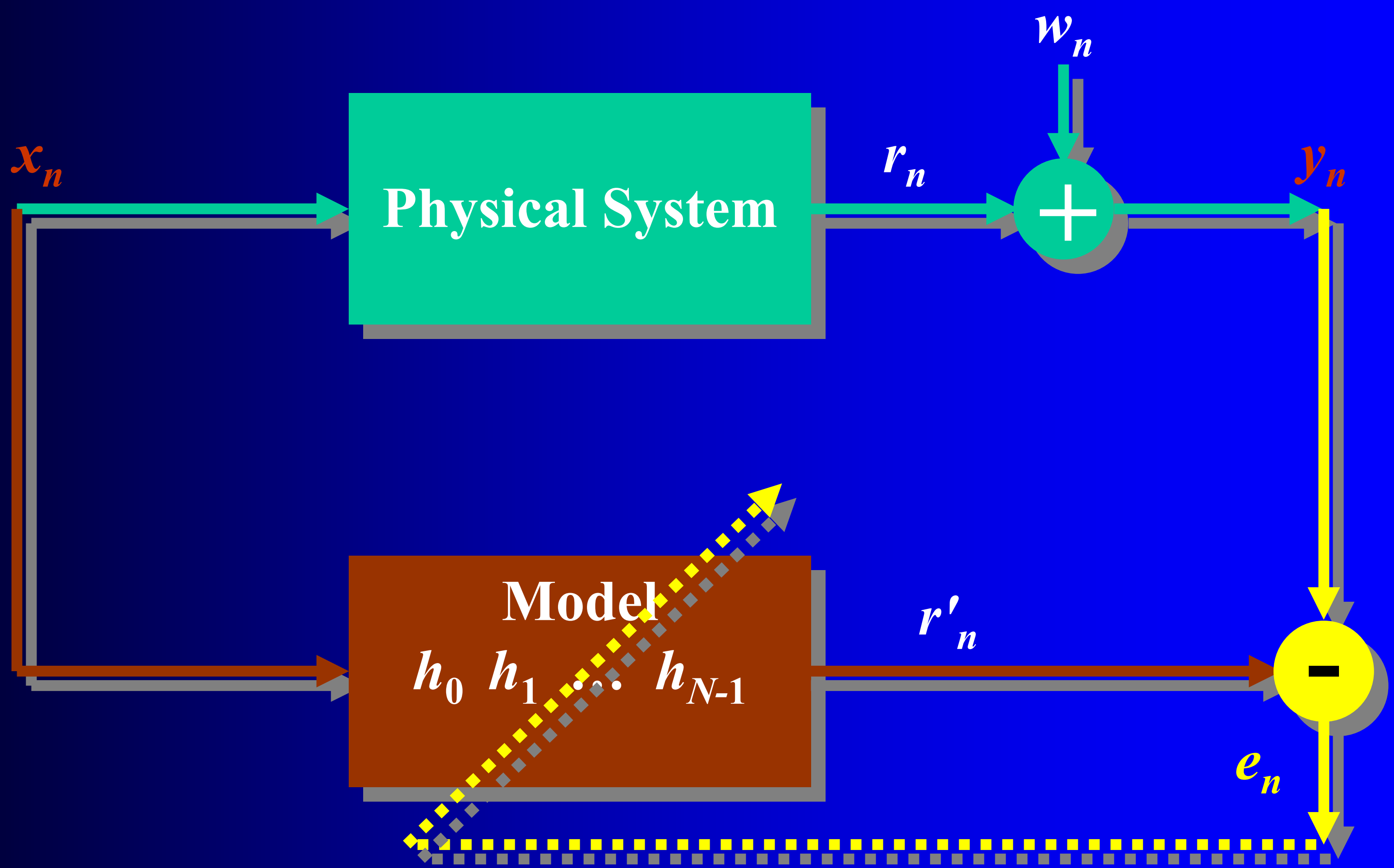
OPTIMUM
ADAPTIVE ALGORITHMS
for
SYSTEM IDENTIFICATION

George V. Moustakides

Dept. of Computer Engineering & Informatics
University of Patras, GREECE

Definition of the problem



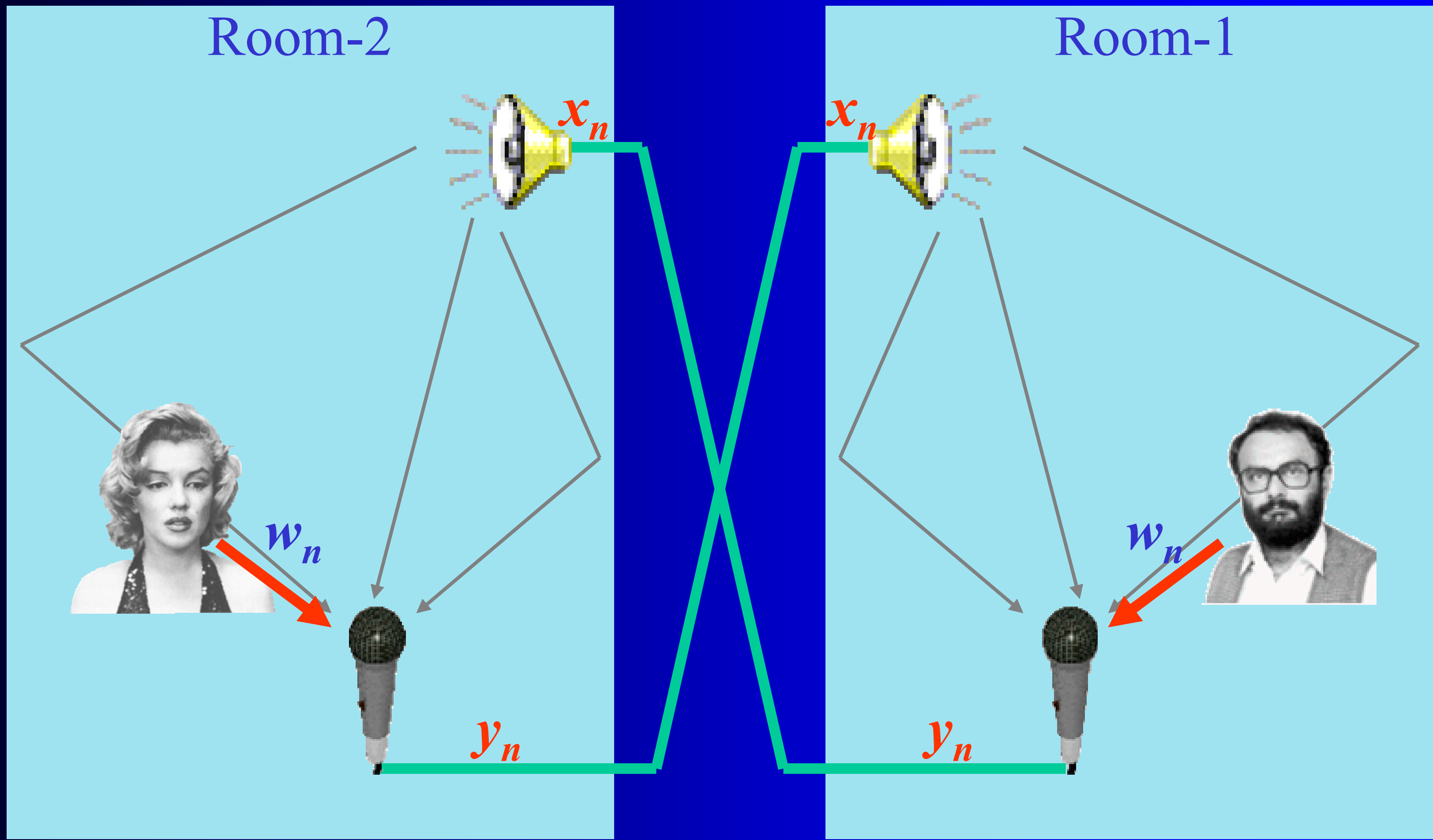
Common Model

Transversal Filter: $r'_n = h_0 x_n + h_1 x_{n-1} + \ldots + h_{N-1} x_{n-N+1}$

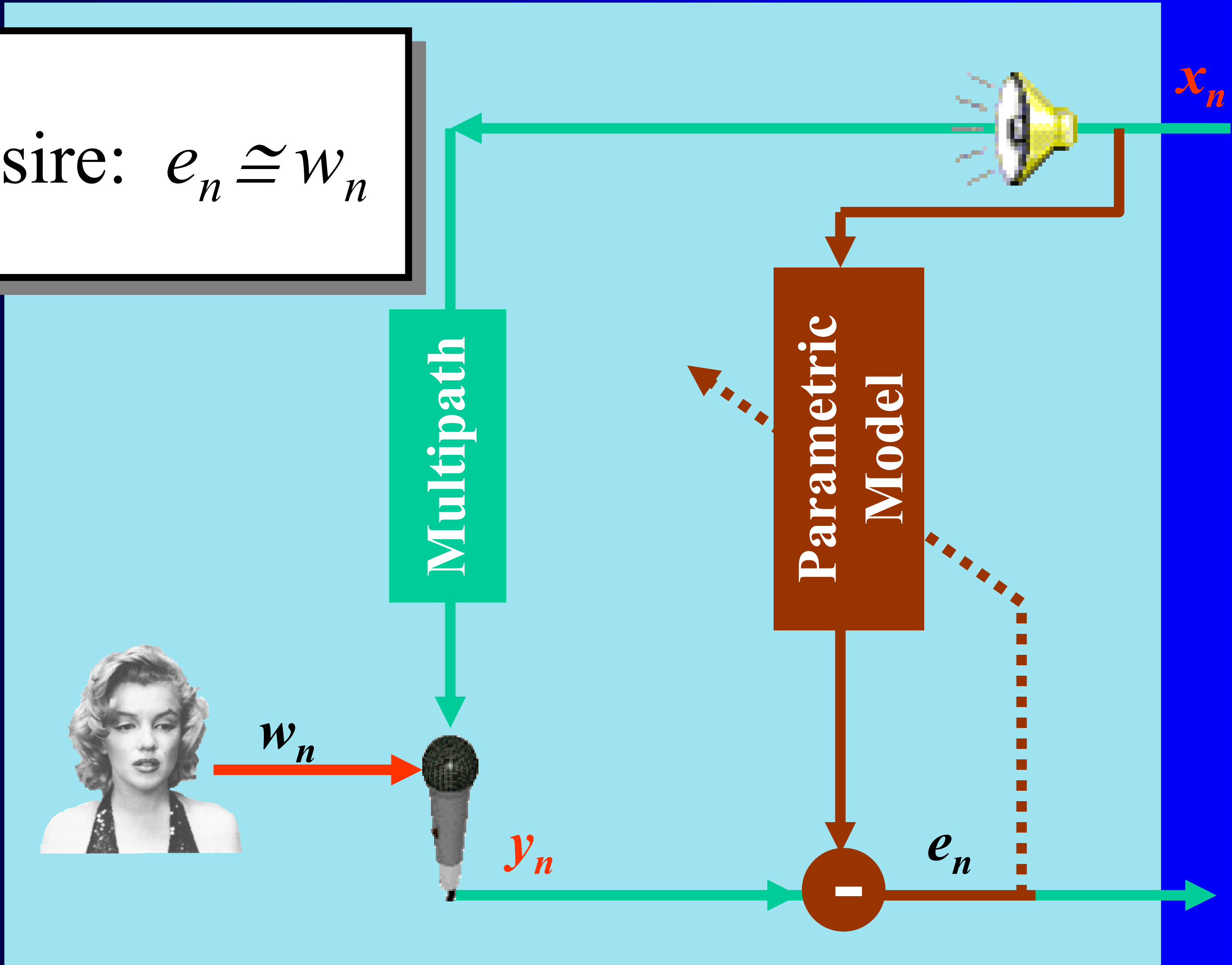
Applications

- Echo Cancellation
- Filtering
- Equalization
- Control
- Seismology
- Array Processing

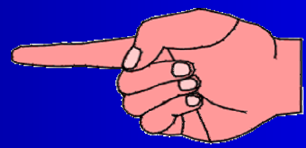
Echo Cancellation in Audio-Conferencing



We desire: $e_n \simeq w_n$



Features of Adaptive Algorithms

- Simplicity & Low Computational Complexity
- Fast Convergence 
- Fast Tracking
- Robustness in Finite Precision

Mathematical Setting

We are given sequentially two sets of data

x_n: Input sequence

y_n: Measured sequence

We would like to express y_n as

$$y_n = h_0 x_n + h_1 x_{n-1} + \ldots + h_{N-1} x_{n-N+1} + e_n$$

and identify the filter coefficients h_i adaptively using algorithms of the form

$$H_n = H_{n-1} + \mu F(H_{n-1}, Y_n, X_n)$$

where

$H_n = [\, h_0 \ \ h_1 \ \ h_2 \ \ldots \ h_{N-1} \,]^T$, $X_n = [\, x_n \ \ x_{n-1} \ \ldots \,]^T$, $Y_n = [\, y_n \ \ y_{n-1} \ \ldots \,]^T$

An Important Algorithmic Class

$$e_n = y_n - H_{n-1}^T X_n$$
$$H_n = H_{n-1} + \mu e_n Z_n$$

$X_n = [\, x_n \;\; x_{n\text{-}1} \;\ldots\, x_{n\text{-}N+1}]^T$

Z_n: Regression Vector function of the input data $x_n \;\; x_{n\text{-}1} \;\ldots$

Well known algorithms in the class:

LMS: $Z_n = X_n$

RLS: $Z_n = \boldsymbol{Q}_n^{-1} X_n$ with $\boldsymbol{Q}_n = (1\text{-}\mu)\, \boldsymbol{Q}_{n\text{-}1} + \mu X_n X_n^T$
(exponentially windowed sample covariance matrix)

FNTF: $Z_n = \boldsymbol{Q}_n^{-1} X_n$ with $\boldsymbol{Q}_n$ the covariance matrix of the x_n sequence assuming that it is AR(M).

Generalization

$$e_{n,i} = y_{n-i} - H_{n-1}^{T} X_{n-i}, \quad i = 0, \dots, p-1$$

$$H_n = H_{n-1} + \mu \sum_{i=0}^{p-1} e_{n,i} Z_{n,i}$$

$X_n = [\, x_n \;\; x_{n-1} \; \dots x_{n-N+1}]^T$

$Z_{n,i}$: vector functions of the input data $x_n \;\; x_{n-1} \; \dots$

Well known algorithms in the class:

SWRLS: $Z_{n,i} = \boldsymbol{Q}_n^{-1} X_{n-i}$ with $\boldsymbol{Q}_n = X_n X_n^T + \cdots + X_{n-p+1} X^T_{n-p+1}$

UDRLS: …

$$e_{n,i} = y_{n-i} - H_{n-1}^T X_{n-i}, \quad i = 0, \ldots, p-1$$

$$H_n = H_{n-1} + \mu \sum_{i=0}^{p-1} e_{n,i} Z_{n,i}$$

By selecting different Regression Vectors $Z_{n,i}$ we obtain different adaptive algorithms. We need to compare them in order to select the optimum!

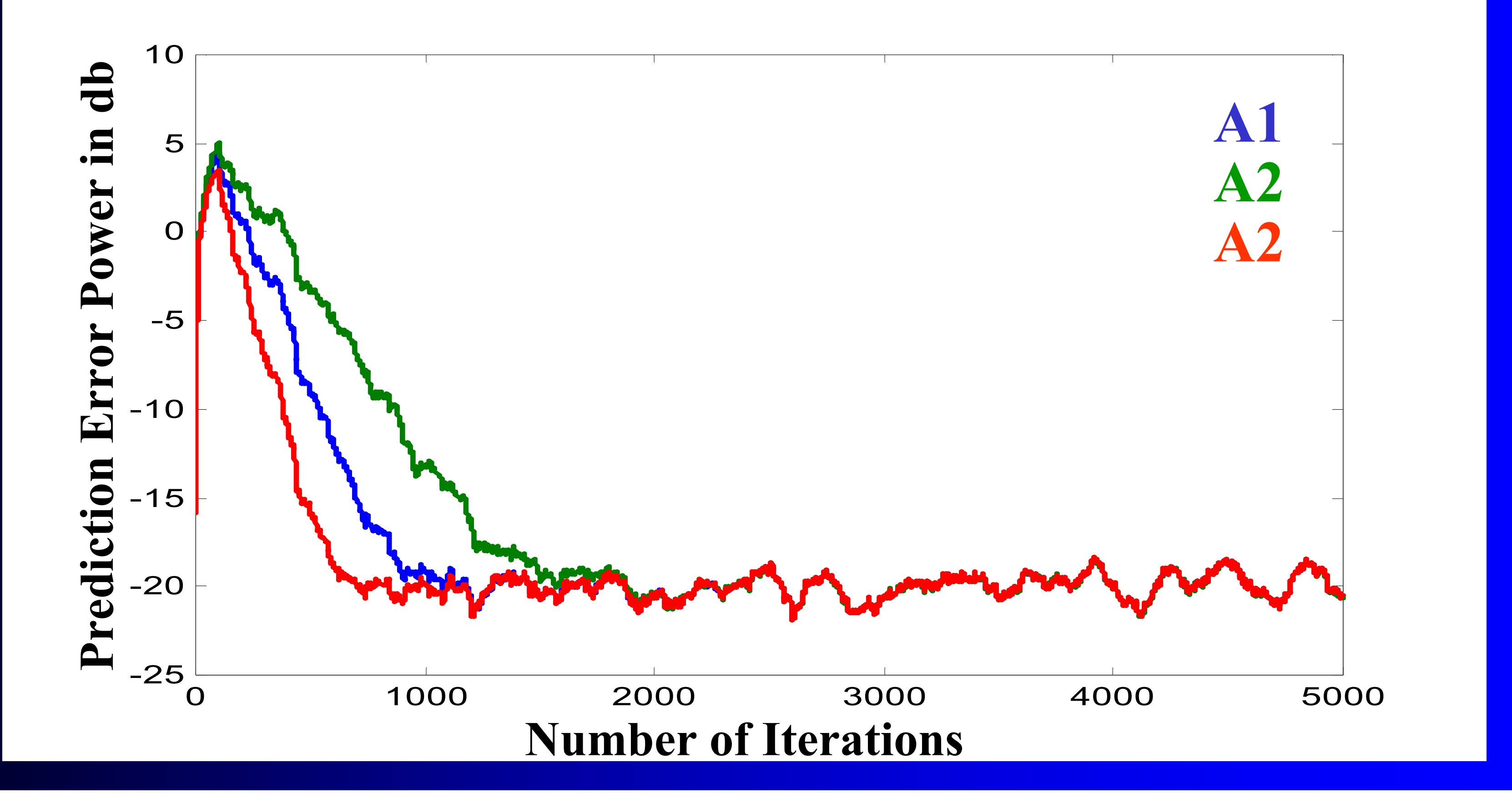
For simplicity we assume EXACT MODEL!!

$$y_n = H_*^T X_n + w_n$$

and w_n white noise

Comparing the Prediction Error

Classically the transient phase refers to stationary data. Let us assume that the noise w_n has power 20db and we have two competing algorithms A1 and A2. We observe their prediction errors e_n:

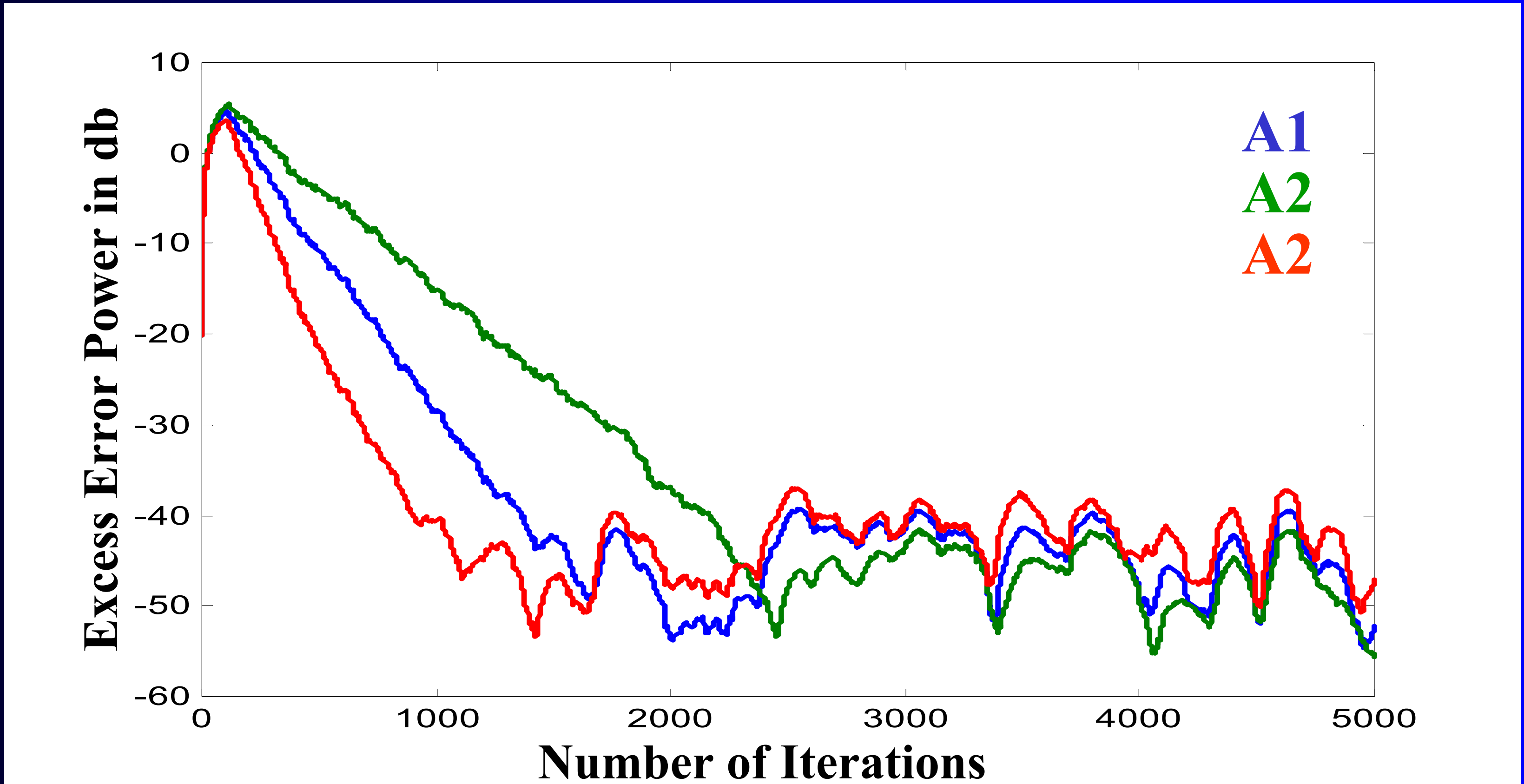


Comparing the Excess Error

$$e_n = w_n - (H_{n-1} - H_*)^T X_n$$

$$\varepsilon_n = (H_{n-1} - H_*)^T X_n$$

Excess Error



A Fair Comparison Method

To fairly compare adaptive algorithms

- We first select the step size μ in each algorithm so that all algorithms under comparison have the same steady state excess error power (Excess Mean Square Error **EMSE**).
- The algorithm that converges faster is considered as the "best".

Can we select the step size μ analytically in order to achieve a predefined value for the EMSE at steady state?

Can we characterize the speed of convergence analytically during the transient phase?

Analysis of the EMSE

$$\textbf{EMSE} = \mathrm{E}\{[\,(H_{n\text{-}1}-H_*)^T X_n]^2\} = \gamma_n + \pi_n$$

γ_n: Starts from an $O(1)$ value and tends exponentially fast to zero for stable algorithms (due to initial conditions).

π_n: Starts from an $O(\mu^2)$ value and tends to an $O(\mu)$ value at steady state (due to the additive noise).

$$\mu << 1$$

Assumptions

$$e_{n,i} = y_{n-i} - H_{n-1}^T X_{n-i}, \quad i = 0,\ldots, p-1$$

$$H_n = H_{n-1} + \mu \sum_{i=0}^{p-1} e_{n,i} Z_{n,i}$$

- **The input vector process $\{X_n\}$ is stationary.**
- **The Regression vector processes $\{Z_{n,i}\}$ are stationary.**
- **The noise process $\{w_n\}$ is zero mean white stationary and independent of the process $\{X_n\}$.**

RLS: $Z_n = \boldsymbol{Q}_n^{-1} X_n$ with $\boldsymbol{Q}_n = (1-\mu)\, \boldsymbol{Q}_{n-1} + \mu X_n X_n^T$

Exponential Convergence

$$\lim_{n\to\infty} \frac{\log\left(\gamma_n^{-1}\right)}{n} = \mu 2\lambda_{\min}(\mathbf{A}) + o(\mu)$$

$$\mathbf{A} = \sum_{i=0}^{p-1} E\left\{Z_{n,i} X_{n-i}^T\right\} = \sum_{i=0}^{p-1} E\left\{Z_{n+i,i} X_n^T\right\} = E\left\{\bar{Z}_n X_n^T\right\}$$

$$\bar{Z}_n = \sum_{i=0}^{p-1} Z_{n+i,i}$$

$$\lambda_{\min}(\mathbf{A}) = \min_i \left\{\mathrm{Re}\left(\lambda_i\right)\right\}$$

λ_i : eigenvalues of the matrix A

Steady State EMSE

$$\lim_{n\to\infty} \pi_n = \mu\sigma_w^2 trace\{\mathbf{QP}\} + o(\mu)$$

$$\sigma_w^2 = E\{w_n^2\}$$

$$\mathbf{Q} = E\{X_n X_n^T\}$$

$$\mathbf{AP} + \mathbf{PA}^T = \mathbf{R}$$

$$\mathbf{A} = \sum_{i=0}^{p-1} E\{Z_{n,i} X_{n-i}^T\} = \sum_{i=0}^{p-1} E\{Z_{n+i,i} X_n^T\} = E\{\overline{Z}_n X_n^T\}$$

$$\overline{Z}_n = \sum_{i=0}^{p-1} Z_{n+i,i}$$

$$\mathbf{R} = E\{\overline{Z}_n \overline{Z}_n^T\}$$

Analytic Local Measure: Efficacy

$$\pi \approx \mu \sigma_w^2 trace\{\mathbf{QP}\}$$

$$\mu \approx \frac{\pi}{\sigma_w^2 trace\{\mathbf{QP}\}}$$

$$Rate \approx 2\mu\lambda_{\min}(\mathbf{A}) \approx \frac{4\pi}{\sigma_w^2} \frac{\lambda_{\min}(\mathbf{A})}{2trace\{\mathbf{QP}\}}$$

$$EFF = \frac{\lambda_{\min}(\mathbf{A})}{2trace\{\mathbf{QP}\}}$$

An algorithm A_1 is better than an algorithm A_2 if

$EFF_1 > EFF_2$

Goal: Maximization of the Efficacy with respect to the regression vectors $Z_{n,i}$.

Optimum Algorithm

Theorem: The maximum value of the Efficacy is $1/N$. The maximum is attained if and only if:

$$\overline{Z}_n = \alpha \boldsymbol{Q}^{-1} X_n$$

where $\boldsymbol{Q}=E\{X_n X_n^T\}$, α is any positive scalar.

The algorithm is called: **LMS-Newton** (LMS-N)

Corollary: If $\boldsymbol{Q}=E\{X_n X_n^T\}=\sigma_x^2 \boldsymbol{I}$ then the optimum algorithm in the class is the LMS.

Practically Optimum Algorithms

RLS:

$Z_n = \boldsymbol{Q}_n^{-1} X_n$

$\boldsymbol{Q}_n = (1-\mu)\, \boldsymbol{Q}_{n-1} + \mu X_n X_n^T$

Under steady state $\boldsymbol{Q}_n$ is a good approximation to $\boldsymbol{Q}=E\{X_n X_n^T\}$

So RLS is expected to match the optimum performance.

SWRLS:

$Z_{n,i} = \boldsymbol{Q}_n^{-1} X_{n-i}$

$\boldsymbol{Q}_n = X_n X_n^T + X_{n-1} X^T_{n-1} + \ldots + X_{n-p+1} X^T_{n-p+1}$

If the window p is small then $\boldsymbol{Q}_n$ does approximate well $\boldsymbol{Q}$. If however p is large then due to the LLN the approximation can be good.

Simulation

$X_n=[x_n\ x_{n-1}\ \ldots\ x_{n-19}]$, $H_*=[h_0\ h_1\ \ldots\ h_{19}]$

$x_n=0.9x_{n-1}+v_n$, Gaussian AR process

w_n: Gaussian additive white noise 20db

$N=20$

We select μ such that the steady state EMSE=35db

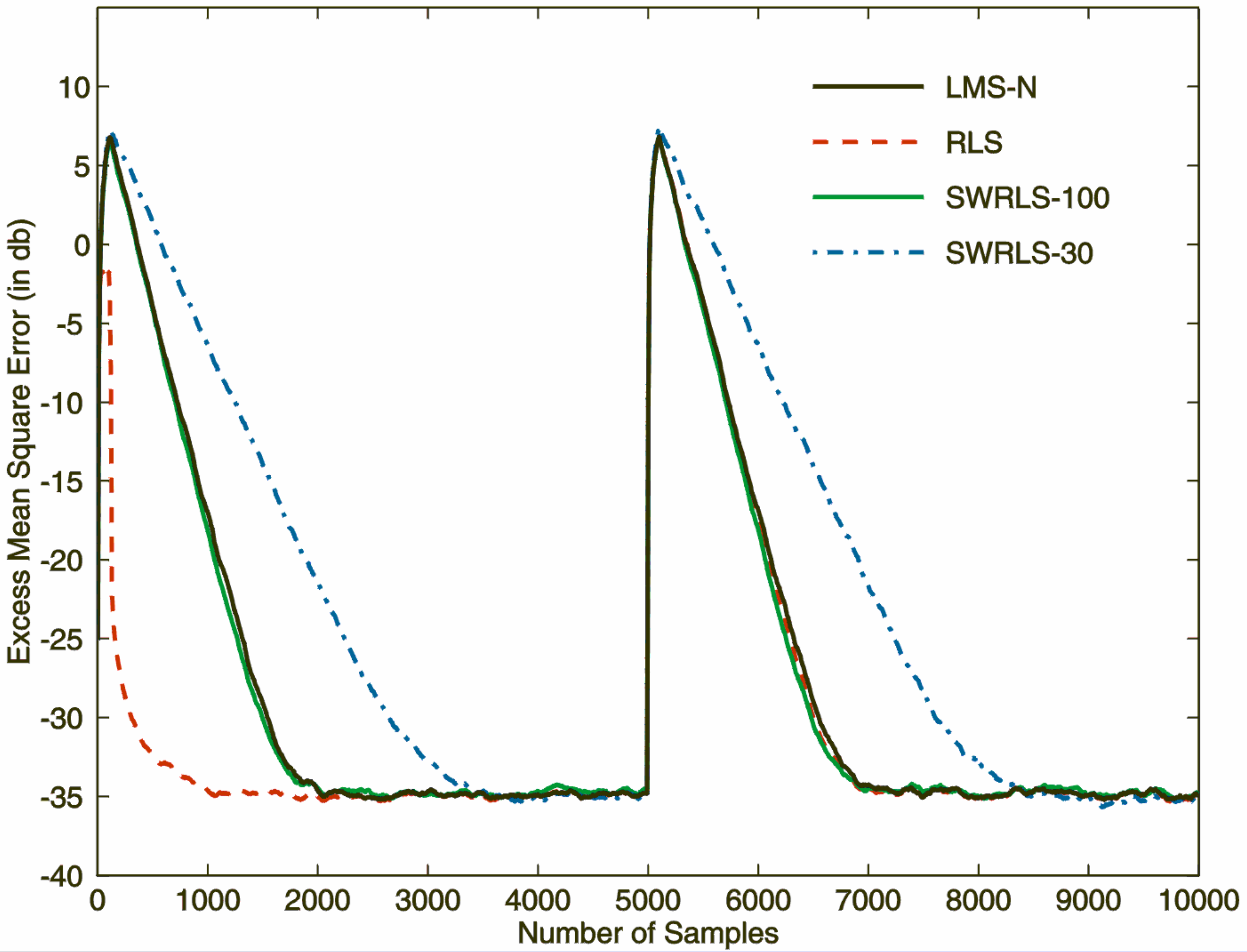
At time n=5000 abrupt change $H_*= -[h_0\ h_1\ \ldots\ h_{19}]$

LMS-N (Optimum)

RLS

SWRLS-30 (Window size p=30)

SWRLS-100 (Window size p=100)



LMS-N
RLS
SWRLS-100
SWRLS-30
Excess Mean Square Error (in db)
Number of Samples
10
5
0
-5
-10
-15
-20
-25
-30
-35
-40
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

END