LOW POWER STRATEGIES FOR ADAPTIVE FILTERING AND EQUALIZATION

A. O. Hero EECS, University of Michigan, Ann Arbor

3rd Year ARO Review, July 1999
Accomplishments in Years 1-3

1. Register-length, power, and optimal bit allocation
   - (Gupta)

2. Successive weight updating, power, and convergence
   - (Godaevati)

3. Proximal point bundle methods for function optimization
   - (Chertien)
Impact on SP Applications

- Adaptive source separation
- Adaptive nulling and beamforming arrays
- Adaptive multipath combining
- Adaptive anti-jam and noise cancelation
- Space-time processing
- Channel equalization
- Matched filters and correlators
I. Register-Length vs. Power

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|}
& & & & & & MSB & & & LSB \\
1 & & & & & & & B & & 16 \\
\end{array}
\]

Figure 1: Fixed point $B$-bit register

Define:

- $B$ – number of bits available
- $R(k)$ – process auto-correlation $E[d_l d_{l-k}^*]$
- $\eta$ – power dissipated per bit transition

Power dissipation per unit time for $B$ bit register with Gaussian data:

\[
P_B \leq B\eta \cdot \left[ 1 - \frac{1}{2} \text{erf}\left( \frac{2^B \sqrt{2R(0) - 2R(1)}}{-1} \right) \right]
\]
Figure 2: Normalized power versus bit width $b$ as a function of AR parameter $a_1$.

For $|a_1| < 0.8$, power increases approximately linearly as a function of $B$. 
Full Resolution FIR Filter

\[ \hat{Y}_k = W^H X_k \]

Figure 3: Infinite precision FIR filter implemented as tapped delay line
Reduced Resolution FIR Filter

\[ Q_d(\hat{Y}_k) = Q_d \left( Q_c(W)^H Q_d(X_k) \right) \]

Figure 4: Finite precision FIR filter implemented as tapped delay line
Figure 5: Adaptive channel equalizer using LMS with training sequence $y_k$. 
Reduced Resolution FIR Adaptive Filter

Figure 6: Adaptive channel equalizer using LMS with training sequence $s_k$. $Q_d$ and $Q_c$ are uniform scalar quantizers using $B_d + 1$ and $B_c + 1$ bits, respectively. Scaling factor $a$ is used to prevent overflow.
Quantized Algorithm: Mathematical Model

\[ b' + f' - f = (f')^p \mathcal{O} - (f')^p \mathcal{O} = \epsilon \]

\[ b_{\mathcal{M}} + \bar{M} = (\epsilon (X)^p \mathcal{O} \eta) \mathcal{O} + \bar{M} = 1 + \bar{M} \]

Quantized LMS Algorithm

\[ b_{\mathcal{M}H} + \bar{M}X = (\mathcal{M}^3 \mathcal{O}_H (X)^p \mathcal{O}) \mathcal{O} = (f')^p \mathcal{O} \]

Quantized FIR Digital Filter

Quantized Algorithm: Mathematical Model

Strange, ARO-MURI Review: July 1999
(for Quantized FIR Filter)

\[
\frac{g^{\text{opt}}}{d + \frac{1}{2}} = p \alpha, \quad \frac{\text{MSE}_{\text{excess}}}{\alpha} = \frac{g}{d}\frac{1}{2}
\]

(Quantized FIR Filter)

\[
\frac{6}{d + \frac{1}{2}} = p \alpha, \quad \frac{6}{d} = \frac{\text{MSE}_{\text{excess}}}{\alpha}
\]

where

\[
\text{MSE}_{\text{excess}} = b /p = p \alpha
\]

\[
\text{Under white assumption}
\]

\[
\frac{B}{d^p} = d^p \quad \text{data bit allocation factor} \quad \bullet
\]

\[
\frac{\text{total bit width}}{B} = B^p + B
\]

Define:
Filter Power Dissipation

\[ P = \frac{6 d B + 2 d B^2 + 12 - 12 B + 16 B^2 + 8 d B^4 + 6 d B^5}{28 d B^6 + 36 d + 52 d B} \]

Partial product mult.

\[ P = \frac{6 d B + 2 d B^2 + 12 - 12 B + 16 B^2 + 8 d B^4 + 6 d B^5}{28 d B^6 + 36 d + 52 d B} \]

Table lookup mult.

Total power/iteration of complex FIR filter

\[ \text{vector length (} \# \text{ of filter taps) } = d \]
\[ \text{power per logic gate } = 6 d \]
\[ \text{power per table-lookup per bit } = 2 d \]

Define

Filter Power Dissipation
(Partial product mult.

\[ P_t = \frac{1}{2} P \]

(look up mult.

\[ P_{tol} = \frac{1}{2} P \]

total power/iteration of complex LMS filter)
LMS Power Dissipation

Figure 7: LMS Power Dissipation vs. $B_d$ and $B_c$ with table lookup and partial product accumulation multipliers
FIR Filter Power Dissipation

Figure 8: FIR Filter Power Dissipation vs. \( B_d \) and \( B_c \) with table lookup and partial product accumulation multipliers.
Figure 9: Excess MSE as a function of $B_d$ and $B_c$ for single pole IIR channel.

Increase in MSE due to quantization.
Optimal bit allocation strategy for fixed $PT$

Relation between total bit allocation and power (table lookup)

Optimal bit allocation factor $d$

\[
\begin{align*}
\mathcal{A} + \frac{\frac{d^6\mu + 16\mu^d}{d(\mu + 8\mu^d)}}{\mathcal{Z} + \frac{d^6\mu + 16\mu^d}{d^6\mu + 16\mu^d}} \log_2 \left( \frac{\rho\alpha(\mu + 8\mu^d)}{\mu^d\mu + 8} \right) \end{align*}
\]

where

\[
B_{**d} = \min \left( d^6\mu + (\mu + 8\mu^d)d \right)
\]

\[
\frac{\mathcal{Z} - \mathcal{Z} p\alpha + \mathcal{Z} \rho I}{\mathcal{Z} - \mathcal{Z} p\alpha + \mathcal{Z} \rho I} = b^d
\]
MSE Performance vs. $P_T$

![MSE and Bit Allocation vs. $P_T$](image)

Figure 10: Optimal data bit allocation factor under $P_T$ constraint and MSE as a function of $P_T$. 
Performance under different bit allocations

Figure 11: MSE as a function of $P_T$ for various bit allocation factors.
Figure 12: Quantized LMS (channel identification) learning curve. Complex White Gaussian training sequence $y_k$ with additive noise, Training sequence passed through 31-tap FIR channel. Parameters are: $\sigma_y^2 = 0.1$, $N_0 = 10^{-8}$, $B_c = B_d = 12$, $\mu = 1/32$, $p = 31$, $\xi_{\text{min}} = 10^{-8}$. 
Experimental Results for Blind Equalization (CMA)

Figure 13: CMA Objective function, $J_{CMA}$, vs $B_d, B_c$ and power-optimal bit allocation
Main conclusions for reduced resolution strategies

*Simulations have borne out theoretical results for medium to high performance regimes.*

For FIR matched filters, $B_t = B_p$ is nearly optimal for $B_t$ and $P_t$.

Analysis yields MSE-optimal LMS bit allocation strategies for fixed $B_c = B_d$. Simulations have borne out theoretical results for medium to high performance regimes.

*Significant reductions in $B_t$ and $P_t$ are possible for many DSP applications.*
Future Work on Low-Power Adaptive Filtering

- Extend to Blind Equalization (CMA)
- Extend to Probability of Error determination for typical communications settings
- Increase accuracy of MSE approximations
- Non-white noise models
- Nonlinear quantizer models
II. Partial Update LMS and Power

Requirement: condition on gain to guarantee convergence

\[ P_{gy - LMS} = P_{gy} + e \]

Advantages:
- Computation savings
- Memory savings
- Power savings

Requirement: condition on gain to guarantee convergence
Sequential Partial Update LMS Algorithm

Figure 14: Block diagram of the Sequential Partial Update LMS algorithm
Comparison of Weight Trajectories

Figure 15: Weight Update Trajectories for $\mu = 0.2$ and $\mu = 0.4$
Update Equations: Partial Update LMS Algorithm

\[
\begin{align*}
&\text{for even } k, \quad \begin{bmatrix} X^{\gamma^o L} \\ \gamma^o L \end{bmatrix} \mathcal{E} \begin{bmatrix} \gamma^o L - I \\ 0 & I \end{bmatrix} = \begin{bmatrix} 1 + \gamma^o L \\ 1 + \gamma^o L \end{bmatrix} \mathcal{E} \\
&\text{for odd } k, \quad \begin{bmatrix} X^{\gamma^o L} \\ \gamma^o L \end{bmatrix} \mathcal{E} \begin{bmatrix} I \\ \gamma^o L - I \end{bmatrix} = \begin{bmatrix} 1 + \gamma^o L \\ 1 + \gamma^o L \end{bmatrix} \mathcal{E} 
\end{align*}
\]

Update Equations for expected weight error vector
\[
\begin{bmatrix}
1.3333 - 0.3280i \\
3.8022i & 1.3333 + 0.3280i \\
-0.5733 + 0.6381i & 3.8022i \\
-0.5733 - 0.6381i & 1.3333 + 0.3280i
\end{bmatrix}
= \mathbf{R}
\]

\[
\begin{bmatrix}
7.1354 \\
3.8022i & 7.1354 \\
3.8022i & 7.1354
\end{bmatrix}
= \mathbf{R}
\]

Example: cyclo-stationary with period 2 having autocorrelation matrices \(\{\mathbf{s}\}\) ·

\(\mathbf{u} + \mathbf{s} = \mathbf{x}\)

\(\mathbf{u} + \mathbf{s}^{M} + \mathbf{s}^{M} = \mathbf{p}\)

Model ·

2-tap adaptive filter ·

\[\text{Example}\]
Example contd.

Eigenvalues of the update equation \( \eta = 0.33 \) have magnitudes 1.0481 and 0.4605.

Sufficient condition derived here gives \( \eta = 0.0254 \).

Regular LMS condition gives \( \eta = 0.33 \).
Large Step Size Weight Trajectories

Figure 16: Trajectory of $w_{1,k}$ and $w_{2,k}$ for $\mu = 0.33$
Small Step Size Weight Trajectories

Figure 17: Trajectory of $w_{1,k}$ and $w_{2,k}$ for $\mu = 0.0254$
Conclusions:

- Partial Update LMS algorithm can attain significant power savings with appreciable loss in the mean square error.

For future work:

- Extension of current work to update of arbitrary subsets of weights.
- Derivation of theoretical results for mean square error convergence.
- Sufficient conditions for selecting a Standard LMS condition for selecting does not guarantee convergence of the Partial Update LMS algorithm.
- Derivation of theoretical results for mean square error convergence.
\[ 0 \leftarrow \gamma \chi \quad \text{and} \quad \chi > 0. \]

sequence of relaxation parameters

\[ I = \hat{h}_p(\theta : \hat{h}) \mathbb{E} \int 0 < (\theta : \hat{h}) \hat{h}_p \frac{d(\gamma \theta : \hat{h})}{d(\gamma : \hat{h})} \mathbb{E} (\theta : \hat{h}) \mathbb{E} = \left( \gamma \theta , \theta \right) F. \]

\[ \{ (\gamma \theta , \theta) \chi \mathbb{E} \gamma \chi - (\theta) F \}^{\theta_{\text{max}}} = I + \gamma \theta \]

(Chretien & Hero: SIAM 1999)

PPA with Kullback Penalty (Cerețian & Hero: SIAM 1999)

\[ \cdots = 1 \leftarrow \chi \quad \{ \gamma \theta - \theta \mathbb{E} \gamma \chi - (\theta) F \}^{\theta_{\text{max}}} = I + \gamma \theta \]

(Rockafellar: SIAM 1976)

Proximal Point Algorithm (PPA)

Proximal Point Methods
Advantages:

1. Superlinear convergence rates for smooth $f(\theta)$.  
2. Can be applied to non-differentiable $f$, e.g. l_1 CMA.  
3. Obtain EM-ML algorithm for $f(\theta, X)$, \( \theta \in \mathcal{W} \).  
4. Obtain new class of accelerated EM algorithms for $f(\theta, X)$.  
5. Successive iterates produce increasing set \{\(g(\theta)\}\}.  
6. Under local quadratic approximation to in $\mathcal{W}$, $f(\theta, X)$, \( \theta, \mathcal{W} \), and \( \mathcal{W} \) becomes hybrid EM/Newton algorithm.  
Example: Maximum likelihood sequence estimation

\[ y_k = \sum_{i=1}^{k} a_{k-i} \theta_i + n_k \quad k = 1, \ldots, n \]

Figure 18: Likelihood trajectory comparisons for ML sequence estimation
Publications citing ARO-MURI


2. Generalized proximal point algorithms and bundle implementations, S. Chretien and A. Hero, accepted.


Technology Transfer Activities

1. ISIS CDR on Telecommunications, Telecom Paris, "Bit allocation strategies for reduced power mobile"


4. First Biennial Communications Symposium, Kingston ON, "EM, absorption and beyond" (Plenary Presentation), June 1998, Contacts: Steve Blostein, Jerry Hayes

5. ARO/-MURI Review/- July 1999