BAYESIAN ADAPTIVE FILTERING:

PRINCIPLES AND PRACTICAL APPROACHES

Tayeb Sadiki and Dirk Slock

EURECOM Institute, Mobile Communications Department
Sophia Antipolis, France
Outline

- Full bayesian approach
- Adaptive Filtering approach
- General Bayesian Adaptive Filtering (GBAF) approach
- GBAF and Adaptive filtering
- Structured cases of interest
- Low Complexity Bayesian Adaptive Filtering with Independent $AR(1)$ Filter Coefficient Models
- Bayesian Adaptive Filtering At linear cost
- Conclusion
Full Bayesian Approach

- The whole matricial spectrum $S_{H^0} (z) = S_{H^0 H^0} (z)$ of $H^0$ counts: the parameter variation, power delay profile, cross spectra between coefficients...
- Kalman Filtering (KF) allows to do all this in the system identification set-up but ignores the estimation of $S_{H^0} (z)$
Adaptive filtering approach

- The prior error signal $e_k = x_k - Y_k^T H_{k-1}$ is used to adapt the filter coefficients.
- Two well known adaptive filters are introduced:

LMS filtering: $H_{k}^{lms} = H_{k-1}^{lms} + \mu Y_k^* e_k$

RLS filtering: $H_{k}^{rls} = H_{k-1}^{rls} + \hat{R}_k^{-1} Y_k^* e_k$
General Bayesian Adaptive Filtering approach

Let $v_k = x_k - Y_k^T H_k^0$, $R = E[Y_k^* Y_k^T]$. Assume that the input covariance matrix $R$ is well estimated, then

$$G_k = R^{-1} Y_k^* x_k = H_k^0 + R^{-1} Y_k^* v_k + (R^{-1} Y^* Y^T - I) H_k^0$$

\(\tilde{G}_k\)

can be considered as a measurement for the optimal filter sequence $H_k^0$

The General Bayesian Adaptive Filter (GBAF) solution will be based on LMMSE estimation of $H_k^0$ from $G_k$:

$$\hat{H}_k^0 = F(q) G_k$$
General Bayesian Adaptive Filtering approach

In the non-causal WF case:

- \[ F(q) = S_{H^0G}(q)S_{GG}^{-1}(q) = I - S_{\tilde{G}\tilde{G}}(q)S_{GG}^{-1}(q) \]

In the causal WF case:

- \[ F(q) = I - R_{\tilde{G}\tilde{G}}R_{GG}^{-1}P(q), \text{ with } P(z)S_{GG}(z)P^+(z) = R_{\tilde{G}\tilde{G}} \]
GBAF and Adaptive filtering

If we assume the adaptation speed is not too fast, we get approximately:

- $H_{k}^{ls} = (I - (I - \mu R)q^{-1})^{-1} \mu R (H_{k}^{0} + R^{-1}Y^{*}v_{k})$
- $H_{k}^{rls} = \frac{1-\lambda}{1-\lambda q^{-1}}(H_{k}^{0} + R^{-1}Y^{*}v_{k})$

which is very close to the GBAF adaptation structure as $H^{0} + R^{-1}Y^{*}v_{k}$ is related to:

$$G_{k} = R^{-1}Y^{*}x_{k} = H_{k}^{0} + R^{-1}Y^{*}v_{k} + \underbrace{(R^{-1}Y^{*}Y^{T} - I) H_{k}^{0}}_{\text{negligible if } H_{k}^{0} \text{ lowpass}}$$
Performance Analysis

We compare the resulting Excess MSE for different cases:

- LMS with optimized individual stepsize
- Classical LMS with an optimized global stepsize
- Classical RLS with optimized individual forgetting factor
- GBAF
Excess Mean Square Error (EMSE)

The Mean Squared Error can be written as:

\[ MSE = E \left[ e_k^2 \right] = \sigma_v^2 + E \left[ Y_k^H \tilde{H}_k^* \tilde{H}_k^T Y_k \right] \]  

(1)

The Excess Mean Square Error is defined as:

\[ EMSE = E \left[ e_k^2 \right] - MMSE \]

\[ = E \left[ Y_k^H \tilde{H}_k^* \tilde{H}_k^T Y_k \right] \]  

(2)

where \( \tilde{H}_k = H_k^o - H_k \).
The excess MSE can be expressed in the following form:

\[ \text{EMSE} = tr\{E(\tilde{H}_k \tilde{H}_k^H R)\} \]

\[ = tr\{R \int_{-\frac{1}{2}}^{\frac{1}{2}} S \tilde{H} \tilde{H}^H (e^{j2\pi f}) df\} \tag{3} \]

\[ \tilde{H}_k = H_k^o - H_k \]

\[ = (I - F(q)) H_k^o - R^{-1} F(q)Y_k^*v_k \]

\[ = (I - F(q)) H_k^o - F(q)\tilde{G}_k \]

and becomes:

\[ \text{EMSE} = tr R\{\int_{-\frac{1}{2}}^{\frac{1}{2}} F(e^{j2\pi f}) R \tilde{G} \tilde{G}^H (e^{j2\pi f}) df\} \]

\[ + tr R\{\int_{-\frac{1}{2}}^{\frac{1}{2}} (I - F(e^{j2\pi f})) S_{H^o H^o} (e^{-j2\pi f})(I - F^H e^{j2\pi f}) df\} \]
Remark that the EMSE can be broken up into two terms:

- \( E_{\text{noise}} = \text{tr} R \{ \int_{-\frac{1}{2}}^{\frac{1}{2}} F(e^{j2\pi f}) R_{\tilde{G}} \tilde{G} F^H(e^{j2\pi f}) df \} \) characterizing the noise contribution; and can be interpreted as the estimation accuracy under stationary conditions.

- \( E_{\text{lag}} = \text{tr} R \{ \int_{-\frac{1}{2}}^{\frac{1}{2}} (I - F(e^{j2\pi f})) S_{H^o H^o}(e^{j2\pi f}) (I - F^H(e^{j2\pi f})) df \} \) representing the estimation error resulting from the system variations (Lag noise).
Structured cases of interest

The analysis is easiest when the input is white, $R = \sigma_y^2 I$. We consider the following structured models for the optimal Doppler spectrum:

1. (i) subspace model: $H_k^0 = A W_k$ where $A$ is tall and $S_{W W}(z)$ is diagonal

2. (ii) decoupled coefficient dynamics: $S_{H^0 H^0}(z)$ diagonal

3. (iii) uniform dynamics plus power delay profile:

   $S_{H^0 H^0}(z) = S_{h h}(z) D$ where $S_{h h}(z)$ is scalar and $D$ is a constant diagonal.
Special case of uniform dynamics plus power delay profile

Consider (iii) above with the scalar spectrum $S_{hh}$ being a flat low-pass spectrum; i.e.

$$S_{hh}(e^{j2\pi f}) = \begin{cases} 
1 & \text{if } |f| < f_0 \\
0 & \text{otherwise}
\end{cases}$$
Performance analysis for different algorithms

The use of a general filter $F(q)$ will lead to an estimation error

$$\tilde{H} = H_k^0 - H_k$$

with covariance matrix

$$R_{\tilde{H}\tilde{H}} = \int \frac{dz}{2\pi jz} (I - F) S_{H^0H^0}(I - F)^H + \int \frac{dz}{2\pi jz} FR\tilde{G}\tilde{G}F^H$$

where $F = F(z)$ and $S_{H^0H^0} = S_{H^0H^0}(z)$ and results in Excess MSE

$$EMSE = tr(R_{\tilde{H}\tilde{H}}R)$$
**EMSE in the LMS case**

\[ F(z) = \frac{\mu R}{I - (I - \mu R)z^{-1}} \]  

\[ EMSE^{LMS} = N\sigma_v^2 \frac{\mu \sigma_y^2}{2 - \mu \sigma_y^2} \]

\[ + 2\sigma_y^2(1 - \mu \sigma_y^2)tr(D) \left( f - \frac{\mu \sigma_y^2}{\pi(2 - \mu \sigma_y^2)} \left( \text{arctan} \frac{\mu \sigma_y^2}{2 - \mu \sigma_y^2} \tan(\pi f) \right) \right) \]
**EMSE in the RLS case**

\[ F(z) = \frac{1 - \lambda}{1 - \lambda z^{-1}} I, \]  

(5)

the EMSE expression for RLS case:

\[
EMSE^{RLS} = N\sigma_v^2 \frac{1 - \lambda}{1 + \lambda}
+ 2\sigma_y^2 tr(D) \left( \lambda f_o - \frac{\lambda}{\pi} \frac{1 - \lambda}{1 + \lambda} \arctan \left( \frac{1 + \lambda}{1 - \lambda} \tan(\pi f_o) \right) \right)
\]

LMS = RLS if \( 1 - \lambda = \mu \sigma_y^2 \).
EMSE In the Non-Causal WF Case

\[ F(q) = I - S_{\tilde{G}\tilde{G}}(q)S_{GG}^{-1}(q) \]  \hspace{1cm} (6)

the EMSE expressions being:

\[ EMSE_{ncc} = \sum_{i=1}^{N} \frac{1}{j2\pi} \int \frac{dz}{z} \left( \frac{1}{\sigma_v^2} + \frac{1}{\sigma_y^2} D_{ii} S_{hh}(z) \right)^{-1} \]  \hspace{1cm} (7)

\[ = \sigma_v^2 2f_o \sum_{i=1}^{N} \frac{1}{1 + \frac{\sigma_v^2}{\sigma_y^2} D_{ii} \frac{1}{2f_o}} \]
EMSE comparison

We compare the minimum EMSE achieved by LMS, RLS (with optimized parameters $\lambda; \mu$) and GBAF. Figure 1 plots the minimum EMSE curves (as a function of the bandwidth $f_0$).
Figure 1: $EMSE_{opt}$ curves for flat low-pass variations
Low Complexity Bayesian Adaptive Filtering:

with Independent $AR(1)$ Filter Coefficient Models

\[
H_{k+1} = AH_k + W_k \\
y_k = X_k^H H_k + e_k
\]  

(8)

where: $E[W_k W_i^H] = Q\delta_{ki}$.

The problem is: how to estimate $\theta = \{A, Q\}$ with the state sequence i.e. $H_k$?

One solution is to use the EM algorithm.
Low Complexity Bayesian Adaptive Filtering:

with Independent $AR(1)$ Filter Coefficient Models(2)

The EM algorithm consists of the following steps:

- Initialisation: select $\hat{\theta}_0$, set $i = 0$
- The E step:
  Compute: $Q(\theta, \hat{\theta}_i) = E_{\hat{\theta}_i}[\log p_{\theta}(Y, H)|Y]$
- The M step:
  solve: $\hat{\theta}_{i+1} = \arg \max_{\theta} Q(\theta, \hat{\theta}_i)$
- If converged, terminate, otherwise increment $i$ and return to E-step. $Y$ is the incomplete data and $Z = \{Y, H\}$ is a complete data.
Low Complexity Bayesian Adaptive Filtering:

with Independent $AR(1)$ Filter Coefficient Models(3)

The negative log-likelihood function of the observed data is given by:

$$L \sim \frac{M}{2} (\log \sigma_v^2 + \log \text{det}Q + TrQ^{-1} \sum_{k=1}^{M} (H_k - AH_{k-1})(H_k - AH_{k-1})^H)$$

$$+ \sigma_v^{-2} \sum_{k=1}^{M} |y_k - X_k^H H_k|^2$$

Henceforth, $\text{Tr}(.)$ is the trace operator.
Low Complexity Bayesian Adaptive Filtering:

with Independent $AR(1)$ Filter Coefficient Models

and:

$$Q(\theta, \hat{\theta}_k) = \frac{M}{2} (\log \sigma_v^2 + \log \det Q + Tr Q^{-1} \sum_{k=1}^{M} E_{\tilde{\theta}_i} (H_k - AH_{k-1})(H_k - AH_{k-1})^H )$$

$$+ \sigma_v^{-2} \sum_{k=1}^{M} E_{\tilde{\theta}_i} |y_k - X_k^H H_k|^2$$
Low Complexity Bayesian Adaptive Filtering:

with Independent $AR(1)$ Filter Coefficient Models

The expectation step in this scenario requires the computation of the quantities:

$\hat{H}_k|_M = E_{\hat{\theta}_k}\{H_k|Y\}$, $E_{\hat{\theta}_k}\{H_kH_k^H|Y\}$ and $E_{\hat{\theta}_k}\{H_kH_k^H_{k-1}|Y\}$. These quantities are easily calculated using a Kalman smoother, where $Z$ is the complete data $Z = \{H_k, Y_k\}$, and $H_k$ is the missing data.
Algorithm

Set initial condition

Kalman filtering and one-step smoothing

\[
\begin{align*}
\hat{H}_{k|k-1} &= \hat{A}_k \hat{H}_{k-1|k-1} \\
\hat{y}_{k|k-1} &= X_k^H \hat{H}_{k|k-1} \\
K_k &= P_{k|k-1} X_k (X_k^H P_{k|k-1} X_k + \sigma_v^2)^{-1} \\
C_{k-1} &= P_{k-1|k-1} \hat{A}_k^H P_k^{-1} \\
\hat{H}_{k|k} &= \hat{H}_{k|k-1} K_k (y_k - \hat{y}_{k|k-1}) \\
\hat{H}_{k-1|k} &= \hat{H}_{k-1|k-1} + C_{k-1} (\hat{H}_{k|k} - \hat{H}_{k|k-1}) \\
P_{k|k} &= (I - K_k X_k^H) P_{k|k-1} \\
P_{k-1|k} &= P_{k-1|k-1} + C_{k-1} (P_{k|k} - P_{k|k-1}) 
\end{align*}
\]

(9)
Algorithm

model Parameters adaptation

\[
B_{1}^{(k)} = \lambda B_{1}^{(k-1)} + diag\{ \hat{H}_{k|k} \hat{H}_{k|k}^{H} \} \\
B_{2}^{(k)} = \lambda B_{2}^{(k-1)} + diag\{ \hat{H}_{k-1|k-1} \hat{H}_{k-1|k-1}^{H} \} \\
B_{12}^{(k)} = \lambda B_{12}^{(k-1)} + diag\{ \hat{H}_{k|k} \hat{H}_{k-1|k}^{H} \} \\
\gamma^{(k)} = \lambda \gamma^{(k-1)} + 1 \\
\hat{Q}_{k+1} = \frac{1}{\gamma^{(k)}}(B_{1}^{(k)} - B_{12}^{(k)} B_{2}^{-1}(B_{12}^{(k)})^{H}) \\
\hat{A}_{k+1} = B_{12}^{(k)} B_{2}^{-1}(k) \\
\mathbf{P}_{k+1|k} = \hat{A}_{k+1}^{(k+1)} \mathbf{P}_{k|k} \hat{A}_{k+1}^{H} + \hat{Q}_{k+1}
\]
Simulations

![Graph showing MSE over discrete time for different filters: MSE Kalman filter, MSE adaptive Kalman filter, MMSE.](image)

- MSE Kalman filter
- MSE adaptive Kalman filter
- MMSE
Bayesian Adaptive Filtering At linear cost

Diagonal $AR(1)$ state space model:

$$H_{k+1} = AH_k + W_k$$
$$Y_k = X_k^H H_k + e_k$$

(15)

We can write the following Component-wise model:

$$h_{k+1,n} = a_n h_{k,n} + w_{k,n}$$

(16)

$$y_k = h_{k,n} x_k,n + \sum_{j \neq n} h_{k,n} x_j,n + v_k$$

(17)

for $n = 1 \ldots N$, where $N$ is the length of the filter.
Bayesian Adaptive Filtering At linear cost

We can write

\[ y_k - \sum_{j \neq n}^{N} \hat{h}_{k|k-1,n} x_k,n = h_{k,n} x_k,n + \sum_{j \neq n}^{N} \tilde{h}_{k,n} x_k,n + v_k \]

For each filter component, \( y_k \) and \( v_k \) are updated as follows

\[ y'_k = y_k - \sum_{j \neq n}^{N} \hat{h}_{k|k-1,n} x_k,n \]

and

\[ v'_k = \sum_{j \neq n}^{N} \tilde{h}_{k,n} x_k,n + v_k \]
Bayesian Adaptive Filtering At linear cost

![Graph showing MSE (dB) vs discrete time for different filtering methods: Adaptive Kalman filter via CW-EM, Adaptive Kalman filter via EM, Kalman filter MMSE.](image)
Conclusion and Perspectives

- The convergence speed of the proposed algorithm in a random time-varying environment is approximately as fast as the one shown by conventional Kalman filtering (known parameters).
- Considering $A$ and $Q$ a diagonal matrices, the complexity of the Kalman filter is limited to $O(N^2)$ and the Adaptive Kalman filter has the same order of complexity.
- The convergence of the Component-Wise adaptive Kalman filter speed is slower than for the adaptive Kalman filter or conventional Kalman filter algorithms.
- The complexity of the Component-Wise adaptive Kalman filter is linear in $N$, the adaptive filter order
- Perspectives: to analyze the convergence properties of adaptive Kalman filter and Component-Wise adaptive Kalman filter.
References


