The Simplified Wiener LMS Algorithm

Lars Lindbom, Mikael Sternad and Anders Ahlén**
Ericsson Infotech AB, P.O. Box 1038, SE–651 15 Karlstad, Sweden.
* Uppsala University, Signals and Systems Group, P.O. Box 528, SE–751 20 Uppsala, Sweden.

Abstract: A new type of tracking algorithm with time-invariant gain is presented. It can be applied for obtaining prediction, filtering or fixed-lag smoothing estimates of time-varying parameters in linear regression models.

The algorithm design constitutes a systematic way of introducing *a priori* information into LMS-like adaptation laws, using the concept of stochastic hypermodelling of the unknown time-varying parameters. The design equations, which provide the structure and adjustment of the tracking algorithms, are derived from a Wiener filtering perspective.

The simplest variant of the novel class of algorithms, denoted Simplified Wiener LMS (SWLMS), is presented here. The SWLMS algorithm is particularly well suited for tracking of parameters of mobile radio channels. The utility of the algorithm will be demonstrated on a mobile radio channel, where channel coefficients are subject to Rayleigh fading. The tracking scenario refers to the D-AMPS 1900MHz standard.

1 Introduction

We often need to construct a linear dynamic model based on data, by first selecting a suitable model structure, and then estimating its parameters. In situations with time-varying parameters, the parameter *estimation* problem becomes a parameter *tracking* problem.

Ad hoc tracking schemes can be obtained by modifying algorithms constructed for the time-invariant case, under the assumption of slow parameter variations. Standard modifications are moving data windows and non-vanishing adaptation gains. The choice of adaptation gain (or data window) is based on a compromise between noise sensitivity and tracking capability. For example, the use of modified stochastic approximation gives the Least Mean Squares (LMS) algorithm, while Gauss-Newton schemes lead to e.g. windowed Recursive Least Squares (RLS) algorithm. These two algorithms are in frequent use. In general, the algorithms work well if the regressors provide sufficiently rich information, and if the time-variations are slow. For fast time-variations, the performance of both LMS and RLS tracking schemes can be poor: We may find no reasonable adjustment of the data window, since both types of algorithms will, in general, have a suboptimal structure.

An efficient way of improving the tracking performance is to utilize any existing *a priori* information about the nature of the time variations. For example, it may be known that the parameters behave approximately as sinusoids. Such information can be included into algorithms based on Kalman filters, either in the form of stochastic models or as functional series. Different approaches for incorporating *a priori* information into tracking schemes can be found in, for example, [2], [4] and [7]–[11].

When we studied adaptive equalizers for the D-AMPS system (IS-54B) a few years ago, existing methods for analysis and design turned out to be unsuitable, for the following reasons:

- The parameter variations of the FIR-channel cannot be regarded as "slow"[1].
- The use of Kalman-based algorithms was out of the question, due to severe limitations on the allowable computational complexity.
- The performance of the LMS and RLS- algorithms was unacceptable.

Taking into consideration that *a priori* information about the nature of the time-variations

[1] "Slow" is here a relative concept. It depends on the speed of parameters drifts as compared to noise power, and on the number of parameters to be estimated.

was available from physical grounds, a response to the problems described above was to develop a novel class of algorithms with time-invariant gains. The algorithms were required to minimize the steady state mean square parameter tracking error for smooth but fast variations in the parameters of linear regression models, in particular FIR models [8].

The proposed algorithms combine low computational complexity with an often significant performance increase as compared to LMS and windowed RLS. The increased performance is attained by introducing stochastic "hypermodels" which describe the second order properties of time-varying parameters.

Another feature of the class of algorithms is that smoothing and prediction parameter tracking can be obtained in a straightforward way. This is valuable, since the use of fixed-lag smoothing will improve the attainable tracking performance whenever future data values are available. Multistep prediction of radio channel properties is of importance in e.g. mobile radio receivers which utilize Viterbi algorithms.

The novel design methodology offers a selection of algorithms of different complexity. There is a direct connection between the generality of the assumed hypermodel and the complexity of the resulting algorithm. As outlined in Section 7, simpler models will result in simpler design equations and in tracking algorithms with a lower computational complexity.

The previous work closest related to our approach is that of Benveniste [2]. It was aimed for the use of (state-space) hypermodels in the design of adaptation algorithms with time-invariant gains. Compared to that approach, the methodology presented here can be claimed to be simpler to use. Moreover, it is not restricted to situations with slowly time-varying parameters.

In this paper, we shall focus the presentation on one of the simplest members of the *a priori* based class of algorithms: the Simplified Wiener LMS (SWLMS) algorithm. It is based on the assumption that all the time-varying parameters have the same statistics, described by second order autoregressive (integrated) processes. The design equations for the steady state MSE-optimal tracking algorithm are particularly simple in this case. The LMS algorithm is a special case of the SWLMS algorithm, and follows by considering random walk hypermodelling and FIR systems with white input data of zero mean.

After introducing the tracking problem in Section 2, we directly present the SWLMS algorithm in Section 3. (Readers more interested in the use of the algorithm than in its derivation could read the Sections 2, 3, 4.2 and 5.2 only.) Section 4 then introduces the two key concepts on which our approach is based:

- The design of adaptation laws with time-invariant gains is reformulated as a linear filter design problem.
- Linear time-invariant "hypermodels" are used for characterizing the assumed dynamics of the time-varying parameters.

The design equations leading to the SWLMS algorithm are then presented in Section 5. We discuss an application of the SWLMS algorithm to the tracking of mobile radio channel coefficients subject to Rayleigh fading in Section 6. Finally, Section 7 outlines how other members of the family of tracking algorithms are obtained by making less restrictive assumptions on the hypermodels.

2 The tracking problem

Our interest will in this presentation be focused on tracking schemes suitable for adapting time-varying parameters of linear regression models, here expressed as

$$\hat{y}_{t|t-1} = \varphi_t^* \hat{h}_{t|t-1} \ . \qquad (1)$$

Above, the column signal vector $\hat{y}_{t|t-1}$ represents an estimate of a sampled signal y_t, based on data up to the time instant $t-1$. It has n_y elements and may, in general, be complex-valued. The column vector $\hat{h}_{t|t-1}$, of dimension n_h, contains adjusted parameters and the regression matrix φ_t^* of dimension $n_y|n_h$ consists of signals which are known or computable at time instant t.

We shall assume that the signal y_t can be described by a dynamic system with the same structure as the model,

$$y_t = \varphi_t^* h_t + v_t \ , \qquad (2)$$

where the zero mean measurement noise v_t is stationary and statistically independent of both the parameter vector h_t and the regressor φ_t^*. The coefficient vector h_t represents the true time-varying parameters to be estimated.

Throughout this presentation, we shall assume that the correlation matrix of the regressors,

$$\mathbf{R} = \mathrm{E}\left[\varphi_t \varphi_t^*\right] \ ,$$

is non-singular and time-invariant, with finite third and higher moments. This will in particular be fulfilled in the case of FIR models with stationary input data. For FIR models with scalar outputs, the regression vector is given by

$$\varphi_t^* = (u_t \ \ u_{t-1} \dots u_{t-m}) \ , \tag{3}$$

where u_t denotes the input data at time t.

3 The tracking algorithm

The SWLMS algorithm and its design equations are presented below. The main steps for deriving the algorithm will be outlined in Section 4 and 5. Readers interested in the complete derivation are referred to the PhD thesis [8] by L. Lindbom.

3.1 The algorithm structure

The basic structure of the SWLMS algorithm is given by the following equations

$$\begin{aligned} \varepsilon_t &= y_t - \varphi_t^* \hat{h}_{t|t-1} && (4) \\ \hat{h}_t &= \hat{h}_{t|t-1} + \mu \widehat{\mathbf{R}}^{-1} \varphi_t \varepsilon_t && (5) \\ \hat{h}_{t+1|t} &= -p\, \hat{h}_{t|t-1} + g_0^1 \hat{h}_t + g_1^1 \hat{h}_{t-1} \ , && (6) \end{aligned}$$

where ε_t is the output prediction error, μ is a real-valued step-size ($0 < \mu < 1$) and where $\widehat{\mathbf{R}}$ denotes an estimate of the correlation matrix $\mathbf{R}$. Furthermore, the estimates $\hat{h}_t$ and $\hat{h}_{t|t-1}$ represent a filter parameter estimate and a one-step ahead parameter predictor, respectively.

Specific for SWLMS design is that the real-valued scalar coefficients p, g_0^1 and g_1^1 of the difference equation (6) are determined according to

$$p = \frac{d_1 d_2 (1-\mu)}{1 + d_2(1-\mu)} \tag{7}$$

$$g_0^1 = p - d_1 \qquad g_1^1 = -d_2 \ , \tag{8}$$

in which d_1 and d_2 are required to fulfill

$$|d_1| \leq 1 + d_2 \qquad |d_2| \leq 1 \ . \tag{9}$$

The SWLMS design consists of first selecting the coefficients d_1 and d_2, determining the inverse of $\widehat{\mathbf{R}}$ (or $\mathbf{R}$) and then adjusting the step-size μ. The step-size can be determined by optimal filtering as presented in Section 5.

The choice of d_1 and d_2 reflects what we know, or assume, about the time-variation of the parameters to be estimated. These scalars are coefficients in a second order vector-autoregressive (integrated) model of h_t. (See (31) below.) If no prior knowledge exists about the parameters, a robust choice of d_1 and d_2 turns out to be

$$d_2 \approx 0.9 \ ; \quad d_1 = -(1 + d_2) \ . \tag{10}$$

The intepretation of this particular choice will be discussed in Section 4.2, where also other selections of d_1 and d_2 will be considered.

Remark: LMS is obtained by selecting $d_1 = -1$ and $d_2 = 0$, and replacing $\mu\widehat{\mathbf{R}}^{-1}$ in (5) with a scalar, μ_o. We then have that $\hat{h}_{t+1|t} = \hat{h}_t$ □

3.2 Smoothing and prediction

The SWLMS design also includes a systematic way of updating the parameter estimates $\hat{h}_{t+k|t}$ for arbitrary integers k.

The updating of $\hat{h}_{t+k|t}$ can be performed by extending the filtering (6) according to

$$\hat{h}_{t+k|t} = -p\, \hat{h}_{t+k-1|t-1} + G_k(q^{-1})\, \hat{h}_t \ , \tag{11}$$

where $G_k(q^{-1})$ is an FIR filter,

$$G_k(q^{-1}) = g_0^k + g_1^k q^{-1} + \dots + g_{\max(-k,1)}^k q^{-\max(-k,1)} \tag{12}$$

and where q^{-1} is the unit delay operator[2].

With $k < 0$ we consider fixed lag smoothing, while $k > 0$ represents k-step ahead prediction. We notice from (12) that the filter G_k is of first order as long as $k > -2$. How the coefficients of G_k are determined will be shown in Section 5.2.

We conclude this section by noting from (11) that the filtering of $\hat{h}_t$ can also be expressed as

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{P}}_k(q^{-1})\, \hat{h}_t \ , \tag{13}$$

where

$$\boldsymbol{\mathcal{P}}_k(q^{-1}) = \frac{G_k(q^{-1})}{G_0(q^{-1})}\, \mathbf{I} \tag{14}$$

[2] In the frequency domain, q is replaced by z, or $e^{i\omega}$.

$$G_0(q^{-1}) \quad = \quad 1 + p\, q^{-1} \ , \tag{15}$$

with **I** denoting the identity matrix. Characteristics for SWLMS design is that the matrix transfer function $\boldsymbol{\mathcal{P}}_k$ is diagonal, with equal entries on the diagonal. In the sequel, we shall refer to $\boldsymbol{\mathcal{P}}_k$ as the *coefficient smoothing-prediction filter*. The name originates in Kubin [6], where a one-step ahead prediction filter ($\boldsymbol{\mathcal{P}}_1$) was introduced into a tracking scheme. It was referred to as a coefficient prediction filter.

4 Steps towards SWLMS

In the present section, we introduce two main concepts on which the SWLMS algorithm is based. A key step in the derivation of the algorithm is to rewrite the tracking scheme (4)–(6) as time-invariant filtering.

4.1 Tracking regarded as time invariant filtering

Introducing fictitious measurements

The column vector $\varphi_t \varepsilon_t$ in (5) represents the gradient of the instantaneous negative squared output prediction error $-\varepsilon_t^* \varepsilon_t/2$ with respect to $\hat{h}_{t|t-1}$. By substituting the assumed true system (2) into the output prediction error (4), and introducing the parameter estimation error

$$\tilde{h}_{t|t-1} = h_t - \hat{h}_{t|t-1} \ ,$$

the gradient term $\varphi_t \varepsilon_t$ can be expressed as

$$\begin{aligned} \varphi_t \varepsilon_t &= \varphi_t (y_t - \varphi_t^* \hat{h}_{t|t-1}) \\ &= \varphi_t \varphi_t^* \tilde{h}_{t|t-1} + \varphi_t v_t \ . \end{aligned} \tag{16}$$

Moreover, by adding and subtracting $\mathbf{R}\tilde{h}_{t|t-1}$ on the right-hand side of (16), and defining

$$Z_t = \varphi_t \varphi_t^* - \mathbf{R} \tag{17}$$

$$\eta_t = Z_t \tilde{h}_{t|t-1} + \varphi_t v_t \tag{18}$$

$$f_t = \mathbf{R} h_t + \eta_t \ , \tag{19}$$

the gradient (16) can be formulated as

$$\begin{aligned} \varphi_t \varepsilon_t &= \mathbf{R}\tilde{h}_{t|t-1} + Z_t \tilde{h}_{t|t-1} + \varphi_t v_t \\ &= f_t - \mathbf{R}\hat{h}_{t|t-1} \ . \end{aligned} \tag{20}$$

In this formulation, the gradients can be viewed as the difference between the signal f_t and the predictor $\mathbf{R}\hat{h}_{t|t-1}$. Here, f_t can be regarded as a fictitious measurement, with $\mathbf{R}h_t$ and η_t being the signal and the noise, respectively. The construction of f_t is depicted in Figure 1.

In the sequel, the noise terms η_t and $Z_t \tilde{h}_{t|t-1}$ will be referred to as the *gradient noise* and the *feedback noise*, respectively.

Remark: The matrix Z_t, of dimension $n_h|n_h$, has zero mean by definition. This matrix was first introduced by Gardner [3], and was referred to as the *autocorrelation matrix noise* □

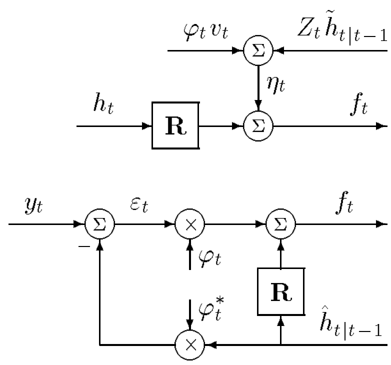

Figure 1: Two equivalent signal flow representations of the fictitious measurement signal f_t. The lower diagram shows how f_t can be generated from available signals, presuming that **R** is known.

The Learning filter

Consider the recursion (5). Introduce the matrix

$$\Upsilon = (\mathbf{I} - \mu \widehat{\mathbf{R}}^{-1} \mathbf{R}) \tag{21}$$

and note from (13) that $\hat{h}_{t|t-1} = \boldsymbol{\mathcal{P}}_1(q^{-1})\hat{h}_{t-1}$. By substituting (20) into (5), we obtain

$$\begin{aligned} \hat{h}_t &= \hat{h}_{t|t-1} + \mu \widehat{\mathbf{R}}^{-1}(f_t - \mathbf{R}\hat{h}_{t|t-1}) \\ &= (\mathbf{I} - \mu \widehat{\mathbf{R}}^{-1}\mathbf{R})\hat{h}_{t|t-1} + \mu \widehat{\mathbf{R}}^{-1} f_t \\ &= \Upsilon \boldsymbol{\mathcal{P}}_1(q^{-1})\hat{h}_{t-1} + \mu \widehat{\mathbf{R}}^{-1} f_t \ . \end{aligned} \tag{22}$$

Now, by moving the term involving $\hat{h}_{t-1}$ to the left-hand side we finally obtain

$$\hat{h}_t = \boldsymbol{\mathcal{L}}_0(q^{-1})\, f_t \tag{23}$$

$$\boldsymbol{\mathcal{L}}_0(q^{-1}) = (\mathbf{I} - q^{-1}\Upsilon \boldsymbol{\mathcal{P}}_1(q^{-1}))^{-1} \mu \widehat{\mathbf{R}}^{-1} \tag{24}$$

Tracking of h_t can thus be regarded as time-invariant filtering of the fictitious measurements f_t introduced in (19), see Figure 2. Note that **R** must be known if the tracking algorithm is expressed in this way.

It is evident by substituting (23) into (13) that $\hat{h}_{t+k|t}$ for all k can be obtained by filtering f_t,

$$\begin{aligned} \hat{h}_{t+k|t} &= \boldsymbol{\mathcal{L}}_k(q^{-1})f_t & (25) \\ \boldsymbol{\mathcal{L}}_k(q^{-1}) &= \boldsymbol{\mathcal{P}}_k(q^{-1})\boldsymbol{\mathcal{L}}_0(q^{-1}) \ . & (26) \end{aligned}$$

The time-invariant filters $\boldsymbol{\mathcal{L}}_k$ shall in the sequel be referred to as the *learning filters.*

Remark: The LMS learning filter is given by

$$\boldsymbol{\mathcal{L}}_{LMS}(q^{-1}) = (\mathbf{I} - q^{-1}(\mathbf{I} - \mu_o\mathbf{R}))^{-1}\mu_o \ .$$

It is readily obtained by replacing $\mu\widehat{\mathbf{R}}^{-1}$ with μ_o and by setting $\boldsymbol{\mathcal{P}}_1$ to unity in (24). Evidently, this learning filter is of first order □

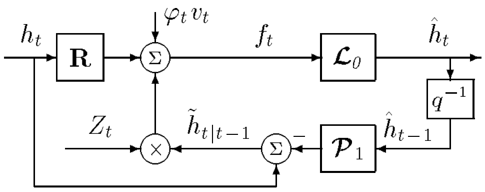

Figure 2: The tracking algorithm formulated as a time-invariant filter. If a parameteric stochastic model of h_t is introduced, and if the non-linear feedback via $\tilde{h}_{t|t-1}$ introduces negligable correlation, then the filter $\mathcal{L}_0$ can be designed analytically.

Under mild conditions, discussed in Section 5, the correlation introduced by the non-linear feedback via $\tilde{h}_{t|t-1}$ in Figure 2 can be neglected. The design of the learning filter $\boldsymbol{\mathcal{L}}_0$ can then be regarded as a linear open-loop Wiener filter design problem, in which the feedback noise $Z_t\tilde{h}_{t|t-1}$ is considered as an additional disturbance.

Remark: If the gradient expression (16) had been used instead of (20), the filtering expression would have been

$$\begin{aligned} \hat{h}_t &= \boldsymbol{\mathcal{L}}_{0,t}(q^{-1})(\varphi_t\varphi_t^* h_t + \varphi_t v_t) \\ \boldsymbol{\mathcal{L}}_{0,t}(q^{-1}) &= (\mathbf{I} - q^{-1}\Upsilon_t\boldsymbol{\mathcal{P}}_1(q^{-1}))^{-1}\mu\widehat{\mathbf{R}}^{-1} \end{aligned}$$

in which $\Upsilon_t = (\mathbf{I} - \mu\widehat{\mathbf{R}}^{-1}\varphi_t\varphi_t^*)$. It is far from obvious how to perform a filter synthesis from such expressions, since the regression matrix φ_t appears in the filter in addition to its presence in the signal to be filtered. It can be seen that substituting $\varphi_t\varphi_t^*$ by $\mathbf{R}$ would correspond to a total neglection of the feedback noise $Z_t\tilde{h}_{t|t-1}$ □

We notice that the learning filter (24) is a matrix transfer function, and that its elements will in general have different impulse responses. However, if $\widehat{\mathbf{R}}$ coincides with $\mathbf{R}$, so that the matrix Υ in (21) reduces to $(1-\mu)\mathbf{I}$, and if the coefficient predictor $\boldsymbol{\mathcal{P}}_1$ is diagonal with equal entries,

$$\boldsymbol{\mathcal{P}}_1(q^{-1}) = P_1(q^{-1})\mathbf{I} \ ,$$

then the learning filter $\boldsymbol{\mathcal{L}}_0$ follows as

$$\begin{aligned} \boldsymbol{\mathcal{L}}_0(q^{-1}) &= S_0(q^{-1})\mathbf{R}^{-1} & (27) \\ S_0(q^{-1}) &= \frac{\mu}{(1 - q^{-1}(1-\mu)P_1(q^{-1}))} \ . & (28) \end{aligned}$$

Substitution of (27) and (19) into (23) yield

$$\begin{aligned} \hat{h}_t &= S_0(q^{-1})\mathbf{R}^{-1}(\mathbf{R}h_t + \eta_t) \\ &= S_0(q^{-1})\mathbf{I}\,h_t + S_0(q^{-1})\mathbf{R}^{-1}\eta_t \ . & (29) \end{aligned}$$

We note that all elements of the vector h_t will in this case be equally affected by the learning filter, while forming the estimate $\hat{h}_t$.

A *necessary* condition for obtaining bounded estimation errors $\tilde{h}_{t|t-1}$ is that both the learning filter $\boldsymbol{\mathcal{L}}_0$ and the coefficient predictor $\boldsymbol{\mathcal{P}}_1$ are stable, i.e. that they have all their poles located inside the unit circle. In the SWLMS design, $\boldsymbol{\mathcal{P}}_k$ will, by construction, always be stable. This is also the case for $\boldsymbol{\mathcal{L}}_0$ when $\widehat{\mathbf{R}}$ coincides with $\mathbf{R}$.

Due to the presence of the feedback noise $Z_t\tilde{h}_{t|t-1}$, filter stability is *not a sufficient* condition for boundedness of $\tilde{h}_{t|t-1}$. To guarantee boundedness, it becomes necessary to take the impact of the feedback noise into consideration. For more details see Chapter 4 in [8].

4.2 Hypermodelling

By regarding tracking as time-invariant filtering, we obtain an intuitive understanding of the nature of the problem as well as of the desired properties of tracking schemes. For example, if the elements of h_t behave as narrow band signals (e.g. sinusoids in (29)), the corresponding learning filter should have the property of a bandpass filter.

One way of incorporating e.g. bandpass properties into learning filters is to introduce hypermodels of order two, or higher, in the design.

A large class of parameter dynamics can be described by linear time-invariant hypermodels of the form

$$h_t = \boldsymbol{\mathcal{H}}(q^{-1})e_t \ , \tag{30}$$

where $\boldsymbol{\mathcal{H}}$ is a transfer function matrix of dimension $n_h|n_h$ and where e_t is a white sequence of random vectors. We will here assume e_t to be stationary with zero mean. The elements of $\boldsymbol{\mathcal{H}}$ may have poles inside or on the stability border.

The SWLMS design, outlined in Section 3, is based on second order hypermodels,

$$h_t + d_1 h_{t-1} + d_2 h_{t-2} = e_t \ , \tag{31}$$

or in a transfer function description by

$$h_t = \frac{\mathbf{I}}{D(q^{-1})} e_t \ , \tag{32}$$

with

$$D(q^{-1}) = 1 + d_1 q^{-1} + d_2 q^{-2} \ .$$

This simple type of stochastic hypermodels captures the essential behavior of the time variability in a wide range of practical situations.

The zeros of the polynomial $z^2 D(z^{-1})$ are not allowed to be outside the unit circle. This is fulfilled if d_1 and d_2 are selected according to (9).

Filtered random walk hypermodelling

Consider the case in which the parameter dynamics is assumed to be given by

$$h_t = h_{t-1} + \frac{\mathbf{I}}{1 - d_2 q^{-1}} e_t \ . \tag{33}$$

This model is obtained by setting $d_1 = -(1 + d_2)$ in (31), and is commonly referred to as filtered random walks. An example is the coefficient choice suggested in (10) in Section 3.1. Simulations have shown (10) to be a reasonable and robust first approximation in a wide variety of situations.

The "design" parameter d_2 can here be used to match the degree of smoothness of the parameter variations. With d_2 close to one, the elements of h_t are modeled as if they have a strong trend behavior, while moving the value of d_2 towards zero decreases the correlation between consecutive vector increments $h_t - h_{t-1}$, until random walk modelling is obtained for $d_2 = 0$.

Remark: We noticed in Section 3.1 that LMS was obtained by setting $d_1 = -1$ and $d_2 = 0$ in (7) and in (8) (as well as replacing $\mu \widehat{\mathbf{R}}^{-1}$ with μ_o). Hence, LMS design indirectly corresponds to random walk hypermodelling □

Remark: With $d_1 = -2$ and $d_2 = 1$, we obtain integrated random walk hypermodelling. The coefficient predictor filter $\boldsymbol{\mathcal{P}}_1$ in (13), based on this hypermodel, closely relates to the so-called "degree-1 least squares fading-memory prediction", presented in [4] □

Remark: When some, or all, elements of the parameter vector h_t are time-invariant, bias will be introduced in the estimates, unless the estimator includes an integrating term. Integration is introduced when D contains $1 - q^{-1}$ as a factor. The polynomial D of the filtered random walk model (33) contains this factor:

$$D(q^{-1}) = (1 - q^{-1})(1 - d_2 q^{-1})$$

□

Lightly damped AR(2) models

The simplest stochastic hypermodel describing oscillatory parameter behavior is a lightly damped second order AR model. The coefficients of these models can be expressed as

$$d_1 = -2 r_o \cos(\Omega_o) \qquad d_2 = r_o^2 < 1 \ , \tag{34}$$

where $z = r_o e^{\pm j\Omega_o}$ represents the pole locations of the transfer function of (32). The pole radius r_o reflects the damping while Ω_o captures the dominating frequency of the coefficient variation. If this frequency is not exactly known, the spectral peak of the amplitude response of the transfer function should be well damped, to obtain a design which is robust to errors in the assumed dominant frequency. (For a general discussion on robust design, see Section 3.5 in [8].)

Example: *Jakes' model*

Consider a digital mobile radio channel coefficient with zero mean, unit variance and subject to Rayleigh fading according to Jakes' model [5]. Then, a covariance function of a channel coefficient, denoted by h_t^i, can be expressed as

$$\mathrm{E}\,[h_{t+\ell}^i h_t^{i*}] = J_0(\Omega_{\mathrm{D}} \ell) \qquad \ell = 0, \pm 1, \ldots, \tag{35}$$

where $J_0(\cdot)$ is the Bessel function of first kind and zero order, Ω_{D} is the maximum normalized Doppler frequency and ℓ is the time lag.

The simple idea with the hypermodelling approach is here to approximate the Bessel function with a covariance function corresponding to an AR(2) process. This is achieved by first determining Yule-Walker equations related to the process, and then utilize (35) in the so obtained set of equations. In other words, the goal is to adjust the coefficients of $D(q^{-1})$ such that

$$J_0\big(\Omega_{\mathrm{D}}\ell\big)+d_1 J_0\big(\Omega_{\mathrm{D}}(\ell-1)\big)+d_2 J_0\big(\Omega_{\mathrm{D}}(\ell-2)\big) \approx 0 \ .$$

With the coefficients of the hypermodel specified by (34), an approximate solution is given by the following parameter choices

$$\Omega_o = \frac{\Omega_{\mathrm{D}}}{\sqrt{2}} \qquad r_o = 0.998 \ . \tag{36}$$

In Figure 3, the Bessel function is compared to the covariance function of this AR(2) process, for normalized Doppler frequency $\Omega_{\mathrm{D}} = 0.02$.

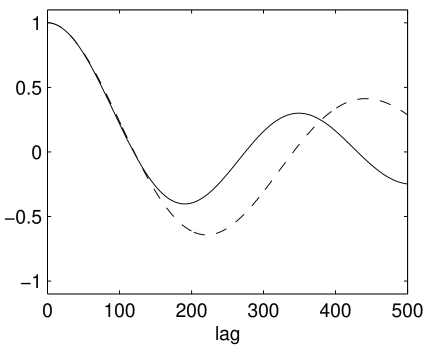

Figure 3: Comparison between $J_0(0.02\,\ell)$ (solid) and a covariance function based on the AR(2) model (34) and (36) (dashed) $\square$

5 Optimal filter design

Here, the objectives are to state conditions, under which the structure of the SWLMS algorithm will be optimal, as well as outlining the final steps to obtain the algorithm.

5.1 Multivariable estimation

In Figure 4, we have combined the hypermodel discussed in Section 4.2 with the time-invariant filtering formulation introduced in Section 4.1. We can then formulate a filter design problem, see Figure 4, in which the goal is to determine the learning filter $\boldsymbol{\mathcal{L}}_k$ such that the steady state Mean Squared tracking Error (MSE)

$$\lim_{t\to\infty} \mathrm{E}\,\|\tilde{h}_{t+k|t}\|_2^2 \ , \tag{37}$$

is minimized. Here the estimation errors $\tilde{h}_{t+k|t}$ can be expressed, by using (25) and (19), as

$$\begin{aligned} \tilde{h}_{t+k|t} &= h_{t+k} - \hat{h}_{t+k|t} = h_{t+k} - \boldsymbol{\mathcal{L}}_k\, f_t \\ &= (q^k \mathbf{I} - \boldsymbol{\mathcal{L}}_k \mathbf{R}) h_t - \boldsymbol{\mathcal{L}}_k \eta_t \ . \end{aligned} \tag{38}$$

By minimizing (37), the initial transient response of $\tilde{h}_{t+k|t}$ is not taken into account in the filter optimizations.

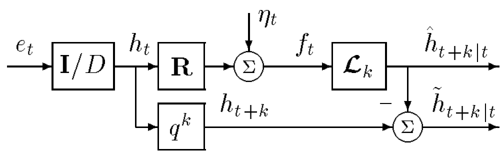

Figure 4: A multivariable estimation problem. The vector h_t is to be estimated from f_t, such that the steady state MSE $\lim_{t\to\infty} \mathrm{E}\,|\tilde{h}_{t+k|t}\|_2^2$ is minimized.

Design assumptions and constraints

The expression (37) will be minimized under the constraint of stability and causality of the filter $\boldsymbol{\mathcal{L}}_k$, and under the following *design* assumptions:

A1: The dynamics of the parameter vector h_t is described by (32), where e_t has zero mean and is stationary and white.

A2: The gradient noise η_t has zero mean and is assymptotically stationary and white.

A3: The random vector sequence e_t and the gradient noise η_t are mutually uncorrelated.

A4: The matrix $\mathbf{R}$ is known and invertible.

In addition to the above constraints, the learning filter is also structurally constrainted as

$$\boldsymbol{\mathcal{L}}_k(q^{-1}) = S_k(q^{-1})\mathbf{R}^{-1} \ , \tag{39}$$

where $S_k(q^{-1})$ is a scalar transfer function. The main motivation for constraining the structure of $\boldsymbol{\mathcal{L}}_k$ is that this leads to simpler design equations.

Remarks: It can be noted from (18) that stationarity of η_t requires stationary and bounded moments for both the noise v_t and the feedback noise $Z_t\tilde{h}_{t|t-1}$. The whiteness assumption of η_t will in most situations not be fulfilled exactly, but it becomes a very good approximation for FIR systems (2),(3) with white inputs. The assumption A3 implies that e_t and the feedback noise will also be uncorrelated. This makes it possible to discard the feedback loop via $\tilde{h}_{t|t-1}$. In general, A3 will hold only approximately, unless Z_t and $\tilde{h}_{t|t-1}$ are independent. A case of practical importance where A2–A4 above hold exactly is considered in Section 6. For details see [8] □

Realizable Wiener filter design

We are now ready to state the following result, obtained by applying a polynomial approach to the design of realizable Wiener filters [1].

Result: Under the design Assumptions A1–A4, the constrained optimal linear estimator minimizing (37) can be expressed as

$$\begin{aligned} \epsilon_t &= \frac{D(q^{-1})}{\beta(q^{-1})}\mathbf{R}^{-1} f_t & (40)\\ \hat{h}_{t+k|t} &= \frac{Q_k(q^{-1})}{D(q^{-1})}\mathbf{I}\epsilon_t \ , & (41) \end{aligned}$$

where the polynomial $\beta(q^{-1})$ is a monic and stable spectral factor, determined by solving a polynomial spectral factorization

$$r\beta(z^{-1})\beta(z) = \gamma + D(z^{-1})D(z) \ , \qquad (42)$$

in which r is a positive scalar (> 1) and γ is the parameter drift-to-noise ratio, defined as

$$\gamma \triangleq \frac{\mathrm{tr}\,\mathrm{E}\,[e_t e_t^*]}{\mathrm{tr}\,\mathrm{E}\,[\eta_t \eta_t^*]} \ . \qquad (43)$$

The polynomial $Q_k(q^{-1})$ is, together with $L_k(q)$, obtained by solving a Diophantine equation

$$q^k\gamma = rQ_k(q^{-1})\beta(q) + qD(q^{-1})L_k(q) \ , \qquad (44)$$

where Q_k and L_k have the polynomial degrees

$$n_Q = \max(-k, n_D - 1) \qquad n_L = \max(k, n_D) - 1 \ ,$$

with n_D being the degree of $D(q^{-1})$ □

Proof. See Section 3.6 in [8] ■

In traditional realizable Wiener filtering theory, the signal ϵ_t represents the innovations of f_t, while the transfer function of (40) is the so-called whitening filter. The transfer function of (41) corresponds to the casual factor of the realizable Wiener filter solution.

The Diophantine equation (44) corresponds to a linear system of equations, with equal number of unknowns and equations. There exists a closed-form solution to (44). For one-step ahead parameter prediction $\hat{h}_{t+1|t}$, the solution is given by

$$\begin{aligned} Q_1(q^{-1}) &= q(\beta(q^{-1}) - D(q^{-1})) & (45)\\ L_1(q) &= r\beta(q) - D(q) \ , \end{aligned}$$

which is verified by direct substitution into (44). The polynomials Q_k can all be obtained from β, r and D as described in Corollary 3.4 in [8].

For second order spectral factors, there exists an analytical solution also to the spectral factorization (42) Define the non-negative scalar μ as

$$\mu = 1 - 1/r \ . \qquad (46)$$

The coefficients of the stable spectral factor

$$\beta(q^{-1}) = 1 + \beta_1 q^{-1} + \beta_2 q^{-2} \qquad (47)$$

can then (see Result 3.5 in [8])) be expressed as

$$\beta_1 = \frac{d_1(1+d_2)(1-\mu)}{1+d_2(1-\mu)} \qquad \beta_2 = d_2(1-\mu). \quad (48)$$

If we now substitute (47) into (45) it can, by straightforward calculations, be shown that

$$\begin{aligned} Q_1(q^{-1}) &= q((\beta_1 - d_1)q^{-1} + (\beta_2 - d_2)q^{-2})\\ &= \mu(g_0^1 + g_1^1 q^{-1}) = \mu G_1(q^{-1}) \ , & (49) \end{aligned}$$

where g_0^1 and g_1^1 are given by (8). We observe that Q_1 has μ as a factor, which can be shown to be the case for all Q_k. Note also that μ is an optimal choice of determining p in (7).

Remark: The ratio γ in (43) will rarely be known *a priori*. Furthermore, the variance of η_t will depend on the selected algorithm, via $\tilde{h}_{t|t-1}$ in (18). Thus, in practice, the scalar μ will constitute a remaining design parameter □

The optimized learning filters

The optimized $\boldsymbol{\mathcal{L}}_k$ readily follows from (40) and (41), and with μG_k substituted for Q_k, as

$$\boldsymbol{\mathcal{L}}_k(q^{-1}) = \frac{\mu G_k(q^{-1})}{\beta(q^{-1})}\mathbf{R}^{-1} \ . \qquad (50)$$

Since the stable spectral factor β constitutes the denominator of $\boldsymbol{\mathcal{L}}_k$, the SWLMS learning filter (50) is thus stable by construction.

Given the learning filter (50), the optimal coefficient smoothing-prediction filters can be obtained directly from (26) as

$$\boldsymbol{\mathcal{P}}_k(q^{-1}) = \boldsymbol{\mathcal{L}}_k(q^{-1})\boldsymbol{\mathcal{L}}_0(q^{-1})^{-1} = \frac{G_k(q^{-1})}{G_0(q^{-1})}\mathbf{I} \quad (51)$$

which is the filter presented in (14). The result (51) provides an explicit expression for the optimal design of e.g. the coefficient prediction filter $\boldsymbol{\mathcal{P}}_1$, which appeared in the work by Kubin [6].

5.2 The SWLMS design equations for smoothing and prediction

In Section 3.2, the SWLMS structure of the difference equations for updating the estimates $\hat{h}_{t+k|t}$ for arbitrary k was presented. Below, we shall explain how the coefficients of the FIR filters $G_k(q^{-1})$ in (12) are determined. The derivations are based on Corollary 3.4 in [8], the use of the spectral factor (47) and the identity (46).

For k-step ahead prediction, the coefficients of $G_k(q^{-1})|_{k\geq 0}$ can be determined through

$$\begin{pmatrix} g_0^k \\ g_1^k \end{pmatrix} = \begin{pmatrix} -d_1 & 1 \\ -d_2 & 0 \end{pmatrix}^k \begin{pmatrix} 1 \\ p \end{pmatrix} \quad k \geq 0 \quad , \quad (52)$$

where p is given by (7). For fixed lag smoothing, the coefficients of $G_k(q^{-1})|_{k<0}$ can be obtained through the backward recursion

$$G_{k-1}(q^{-1}) = q^{-1}G_k(q^{-1}) + D(q^{-1})(1-\mu)L_0^k, \quad (53)$$

where the scalars L_0^k are determined through

$$\begin{pmatrix} L_0^k \\ L_1^k \end{pmatrix} = \begin{pmatrix} -\beta_1 & 1 \\ beta_2 & 0 \end{pmatrix}^{|k|} \begin{pmatrix} g_0^1 \\ g_1^1 \end{pmatrix} \quad k \leq 0, \quad (54)$$

and where β_1 and β_2 are given by (48).

6 A case study

In this section, we shall compare the attainable tracking performances of the SWLMS algorithm and the LMS algorithm, in a situation where the time variations of the parameters are known to differ substantially from random walks.

Channel tracking for mobile radio systems

In present TDMA systems, digital data are transmitted in bursts of fixed length. Within each burst, a small fraction of the transmitted data (the training sequence) is known to the receiver. The channel can be estimated by correlating the known channel input with the received output signal. The resulting model is, however, invalid during the remainder of the burst, if the fading is severe. The coefficients of the channel (or an equalizer) must then be tracked during the burst. This problem can be solved through decision-directed adaptation, where the outputs from the detector are used as regressors in place of the unknown transmitted symbols.

The considered system

In the D-AMPS 1900 system, the wireless channel is subject to severe fading, so channel coefficients will be rapidly time-varying. An adequate description of the wireless channel is given by

$$\begin{aligned} y_t &= h_t^0 u_t + h_t^1 u_{t-1} + v_t \\ &= \varphi_t^* h_t + v_t \qquad\qquad (55) \\ \varphi_t^* &= (u_t \; u_{t-1}) \qquad h_t = (h_t^0 \; h_t^1)^T \quad , \end{aligned}$$

where the channel coefficients h_t^0 and h_t^1 are subject to Rayleigh fading, u_t are $\pi/4$ DQPSK modulated symbols of variance $\sigma_u^2 = 2$ ($\mathbf{R} \approx \sigma_u^2\mathbf{I}$) and v_t represents noise and co-channel interference. All signals in (55) are complex valued and stationary with zero mean, and v_t, φ_t^* and h_t are assumed to be mutually independent.

We regard (55) as the "true system", from which we (by simulation) obtain the received signal y_t. To generate the Rayleigh fading coefficients of h_t, Jakes' model is used. Hence, the covariance function of the channel coefficients is given by (35).

We shall here study a case with a mobile traveling at a speed of 50 km/h. In the D-AMPS 1900 system, with a sampling period of about 40μs and 1900 MHz carrier frequency, this corresponds to a Doppler frequency Ω_{D} of approximately 0.02.

The selected hypermodel

We describe h_t by an AR(2) hypermodel (31), with d_1 and d_2 determined from (34) and (36). With $\Omega_o = \Omega_{\text{D}} = 0.02$, the coefficients of $D(q^{-1})$ are then given by

$$\begin{aligned} d_1 &= -2r_o\cos(\Omega_o/\sqrt{2}) = -1.9956 \\ d_2 &= r_o^2 = 0.9960 \quad . \end{aligned}$$

This hypermodel is motivated by the quasi-periodic behavior of channel coefficients subject to Rayleigh fading according to Jakes' model.

The optimized trackers

Consider the SWLMS algorithm (4)–(6) with $\widehat{\mathbf{R}} = 2\mathbf{I}$, its design equations (7) and (8), and the above specified coefficients d_1 and d_2.

For a signal-to-noise ratio of 15 dB and a Ω_{D} of 0.02, the SWLMS algorithm, with optimized step-size[3] $\mu = 0.09$, will then be given by

$$
\begin{aligned}
\varepsilon_t &= y_t - \varphi_t^* \hat{h}_{t|t-1} \\
\hat{h}_t &= \hat{h}_{t|t-1} + 0.045\, \varphi_t \varepsilon_t \\
\hat{h}_{t+1|1} &= -p\, \hat{h}_{t|t-1} + g_0^1 \hat{h}_t + g_1^1 \hat{h}_{t-1} \ ,
\end{aligned}
$$

where $p = -0.95$, $g_0^1 = 1.045$ and $g_1^1 = -0.996$. The LMS algorithm with step-size μ_o optimized for the above case is given by

$$
\hat{h}_{t+1|t} = \hat{h}_{t|t-1} + 0.1 \varphi_t \varepsilon_t \ .
$$

The properties of these tracking algorithms and their influence on the symbol error rate when used in conjunction with a Viterbi detector in an adaptive equalizer, is depicted in Figure 5 below. In the left-hand figures, the tracking performance is illustrated by simulations of 600 symbol times, using known transmitted symbols as regressors.

As can be seen from the upper right-hand figure, the attainable MSE tracking performance is improved almost three times by the SWLMS algorithm. The improvement of the bit error rate obtained by using the SWLMS algorithm with estimated symbols as regressors is approximately 2 times at 15 dB and 7 times at 25 dB. This corresponds to a gain between 3 and 5 dB. We conclude that a considerable improvement can be obtained by simple means in cases when the parameter variations differ substantially from random walk behavior.

The reason for the improved performance can be understood by inspecting the frequency responses of the learning filters. As illustrated in Figure 6, the SWLMS learning filter (50) (with $\mathbf{R} = 2\mathbf{I}$) possesses band-pass character around the dominating frequencies of the time-variations, and it has lower gain at high frequencies.

[3] For the considered system, Result 8.1 in [8] contains an exact analytical expression of the steady state MSE. The selected step-size minimizes this MSE. The optimal value of μ depicted in Figure 5 then gives p in (7).

7 A class of algorithms

The presented SWLMS algorithm (4)–(6) is one of the simplest members of a class of Wiener filter designed tracking algorithms [8]. The basic structure of this class of algorithms is given by

$$
\begin{aligned}
\varepsilon_t &= y_t - \varphi_t^* \hat{h}_{t|t-1} \\
\hat{h}_{t+1|t} &= \boldsymbol{F}(q^{-1}) \hat{h}_{t|t-1} + \boldsymbol{\mathcal{G}}_1(q^{-1}) \varphi_t \varepsilon_t
\end{aligned}
\quad (56)
$$

where $\boldsymbol{F}(q^{-1})$ is a polynomial matrix, which is related directly to the hypermodel $\boldsymbol{\mathcal{H}}$ in (30), and where $\boldsymbol{\mathcal{G}}_1(q^{-1})$ is a rational matrix, here referred to as the *gain filter*. In this formulation, the SWLMS algorithm corresponds to

$$
\begin{aligned}
\boldsymbol{F}(q^{-1}) &= (1 - D(q^{-1}))\mathbf{I} \\
\boldsymbol{\mathcal{G}}_1(q^{-1}) &= \mu (g_0^1 + g_1^1 q^{-1}) \widehat{\mathbf{R}}^{-1} \ ,
\end{aligned}
$$

obtained by substituting (5) and (8) into (6).

Characteristics for our Wiener filtering approach to the design of tracking algorithms is that different assumptions on the structure of $\boldsymbol{\mathcal{H}}$ in (30) and on the properties of the gradient noise η_t in (18) will result in different levels of complexity in the design equations and in the resulting algorithm. The following cases deserve to be mentioned.

1. If no restrictions are placed on $\boldsymbol{\mathcal{H}}$ and if η_t may be colored, then the gain filter $\boldsymbol{\mathcal{G}}_1$ will be a rational matrix. A polynomial matrix spectral factorization and bilateral Diophantine equation will then, in general, have to be solved to optimize the tracking scheme.

2. If $\boldsymbol{\mathcal{H}}$ is general, while η_t is assumed to be *white*, then the gain filter $\boldsymbol{\mathcal{G}}_1$ will be a polynomial matrix (multivariable FIR). In this case, *no Diophantine equation* needs to be solved. An algebraic Riccati equation will determine the optimal estimator.

3. To further reduce the complexity of the algorithm, the optimization can be performed under the constraint that

$$
\boldsymbol{\mathcal{G}}_1(q^{-1}) = \boldsymbol{\mathcal{S}}_1(q^{-1}) \mathbf{R}^{-1}
$$

   where $\boldsymbol{\mathcal{S}}_1$ is a *diagonal* rational matrix. The design equations will then consist of uncoupled sets of scalar spectral factorizations and polynomial Diophantine equations, one for each parameter. This type of algorithm was in [8] called *Generalized Wiener LMS*.

4. We may assume that the elements of the parameter vector h_t are uncorrelated but have the same dynamics,

$$\boldsymbol{\mathcal{H}}(q^{-1}) = C(q^{-1})/D(q^{-1})\,\mathbf{I} \;\;,$$

and also assume that η_t is white. The optimal gain filter will then reduce to $\mu G_1 \mathbf{R}^{-1}$, where G_1 is a polynomial. The design equations will consist of a single polynomial spectral factorization. No Diophantine equations are required. The resulting algorithm was in [8] called *Wiener LMS* (WLMS).

5. If it is assumed that the elements of the true parameter vector are uncorrelated but have the same second order autoregressive (integrated) dynamics and if η_t is white, then no design equations at all will be required. The optimal algorithm is then the here considered *Simplified Wiener LMS* (SWLMS). This algorithm is also optimal when elements of h_t are correlated, if the constraint (39) is placed upon the structure of $\boldsymbol{\mathcal{L}}_k$.

The algorithms based on case (3),(4), and (5) above can all be seen as generalizations of LMS, with various complexity and structure of the design equations. It should be noted that the designs (3),(4) and (5) do not utilize possible correlations among the parameters. On the other hand, the resulting algorithms require a small number of computations at each time step. If the matrix $\mathbf{R}$ is known and diagonal, then the number of required multiplications will grow linearly with the dimension of the parameter vector.

If the parameters are known to be significantly mutually correlated, and if η_t is colored, then the design (1) provides the lowest steady state MSE. Any of the other alternatives can, of course, be utilized, regardless of what is known about the nature of h_t. If a mismatching structure is deliberately chosen, the user will be aware of what kind of approximation it corresponds to.

References

[1] A. Ahlén and M. Sternad, "Wiener filter design using polynomial equations," *IEEE Trans. Signal Processing*, vol. 39, 1991.

[2] A. Benveniste, "Design of adaptive algorithms for tracking of time–varying systems," *Int. J. Adaptive Control and Signal Processing*, vol. 2, pp. 3–29, 1987.

[3] W. A. Gardner, "Nonstationary learning characteristics of the LMS algorithm," *IEEE Trans. Circuits and Systems*, vol. 34, pp. 1199–1207, 1987.

[4] S. Hariharan and A. P. Clark, "HF channel estimation using a fast transversal filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1353–1362, 1990.

[5] W. C. Jakes, *Microwave Mobile Communications.* Wiley, New York, NY, 1974.

[6] G Kubin, "Adaptation in rapidly time–varying environments using coefficient filters," *Proc. of ICASSP*, April 1991, Toronto, Ontario, Canada, vol. 3, pp. 2097–2100.

[7] J. Lin, J. G. Proakis, F. Ling and H. Levi-Ari, "Optimal tracking of time-varying channels: a frequency domain approach for known and new algorithms," *IEEE Trans. Selected Areas in Communications*, vol. 13, pp. 141–154, 1995.

[8] L. Lindbom, *A Wiener filtering approach to the design of tracking algorithms: with application in mobile radio communications.* Phd Thesis, Dep. of Technology, Uppsala University, Sweden, 1995.

[9] L. Lindbom, "Simplified Kalman estimation of fading mobile channels: high performance at LMS computational load," ICASSP 1993, Minneapolis, MN, USA, vol. 3, pp. 352–355.

[10] S. McLaughlin, B. Mulgrew and C. F. N. Cowan, "The use of a priori knowledge for HF channel estimation," *Signal Processing IV*, pp. 371–373, 1988.

[11] M. Niedźwiecki, "Recursive functional series modeling estimators for identification of time-varying plants–More bad news than good?" *IEEE Trans. Automatic Control*, vol. 35, pp. 610–616, 1990.

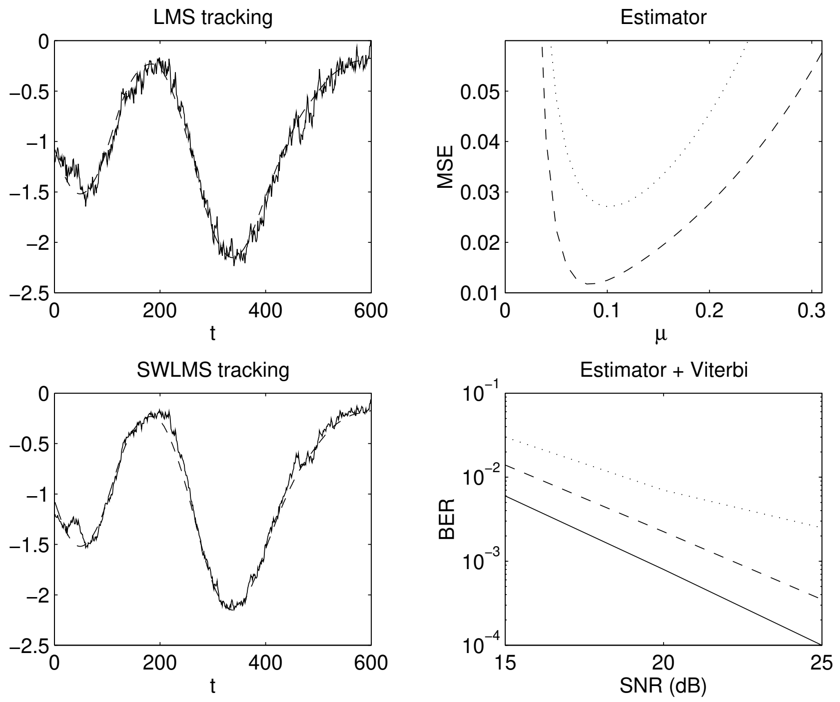

Figure 5: The result of LMS and SWLMS parameter estimation. Top and bottom left figures, estimates (solid) and the true parameter variation (dashed). Top right, the MSE tracking performance for the LMS algorithm (dotted) and for the SWLMS algorithm (dashed), evaluated for 15 dB SNR and $\Omega_{\mathrm{D}} = 0.02$ (50 km/h). Bottom right, bit error rate for estimator LMS (dotted) and SWLMS (dashed), concatenated with a Viterbi algorithm, and the bit error rate obtained by using true channel parameters (solid).

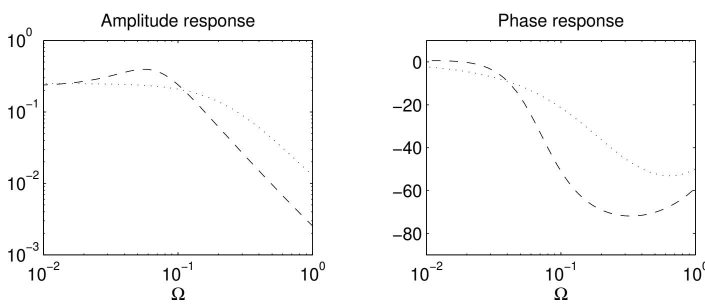

Figure 6: Amplitude and phase response of the learning filters, corresponding to SWLMS tracking (dashed) and to LMS tracking (dotted), as a function of the normalized frequency Ω