Proc. NUTEK Workshop on Digital Communication, Uppsala, Sweden, May 25-26 1992, p122

Simplified Kalman Estimation of Fading Mobile Radio Channels: High Performance at LMS Computational Load

Lars Lindbom
Automatic Control Group, Uppsala University
P.O. Box 27, S–751 03 Uppsala, Sweden

Abstract
A low-complexity algorithm for channel estimation in Rayleigh fading environments is presented. The channel estimator is presumed to operate in conjunction with a Viterbi detector. The algorithm is based on simplified internal modelling of time-variant channel coefficients and approximation of a Kalman estimator. A novel averaging approach is used to replace the on-line update of the Riccati equation. Compared to RLS tracking, both a significantly lower bit error rate and a much lower computational complexity is attained.

1. Introduction

An inherent difficulty associated with adaptive equalization/detection is that unknown transmitted data are needed for channel or filter adaptation, but are not available. They can be replaced by decisioned data in decision directed mode. In severe fading environments, there is a potential risk of losing tracking ability when incorrect decisions are used.

Parameter tracking in severe fading environments (high Doppler frequencies and low signaling rates) with an RLS algorithm, often requires the forgetting factor to be chosen rather small ($< .8$). With a small effective memory, the algorithm becomes sensitive to noise and to incorrect decisions. The "classical" RLS algorithm is therefore not suitable in such tracking problems. Algorithms with longer memory are required.

In fading environments, the coefficients of a FIR channel model exhibit typical trend or quasiperiodic behaviour . By utilizing this information, adaptive algorithms with longer memories can be designed. One way of building in this a priori information is discussed in Section 3 and 4. For more details, see [1].

When the transmitted sequences of symbols are *white* and the symbols have *constant modulus*, an algorithm with high performance and low complexity can be obtained. In this paper, derivation of such a channel estimator is outlined.

2. Statement of the problem

Consider a received sampled sequence $\{y(n)\}_1^N$. It is generated by transmission of one data burst $\{d(n)\}_1^N$ over a HF channel, represented by an equivalent discrete-time complex baseband channel model

$$\begin{aligned} y(n) &= \varphi^H(n)\theta(n) + v(n) \\ \varphi^H(n) &= (d(n), \dots, d(n-m)) \\ \theta(n) &= (h_0(n), \dots, h_m(n))^T \end{aligned} \tag{1}$$

The channel coefficients, $h_k(n)$, $k = 0, \dots, m$, are subject to *independent* Rayleigh fading. This is a realistic assumption for urban mobile radio channels. The measurement noise, $v(n)$, is white, with zero mean and variance σ_v^2. It is uncorrelated with the symbol sequence $\{d(n)\}$. All signals are complex valued.

The objective is to reconstruct $\{d(n)\}_{N_{tr}+1}^N$ from $\{y(n)\}_1^N$ and a known training sequence $\{d(n)\}_1^{N_{tr}}$. Figure 1 depicts the over-all structure of the combined adaptation and detection. Here, we will focus on the channel estimation.

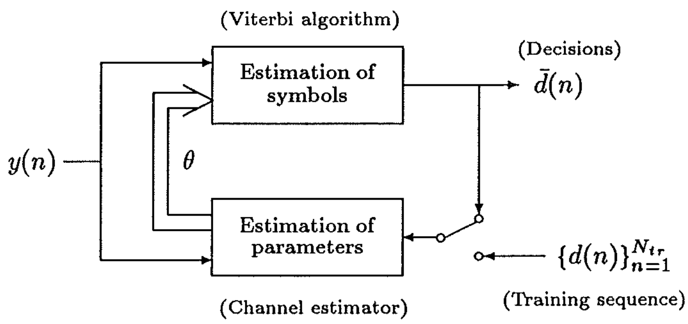

Figure 1 The adaptation/detection structure.

3. Stochastic internal modelling of the channel coefficients

Channel coefficients, with statistical properties as predicted by the Rayleigh fading model, can be generated by feeding white complex Gaussian noise through a filter of high order (> 5). The cutoff frequency of the filter is controlled by the maximum Doppler frequency (f_m), see e.g [1]. The channel coefficients, $h_k(n)$, can thus be well described by linear stochastic models expressed as

$$\begin{aligned} x_k(n+1) &= Fx_k(n) + Ge_k(n) \\ h_k(n) &= Hx_k(n) \qquad k = 0, 1, \dots, m \end{aligned}$$

Here, the sequences $\{e_k(n)\}$, $k = 0, \dots, m$ are mutually independent white *complex* Gaussian noises, with zero means and variances $\sigma_{e_k}^2$. The matrices F, G and H are real-valued. An accurate model would depend on (f_m).

4. State space description of the time-variant model and a Kalman estimator

By introducing the state vector

$$X(n) \triangleq (x_0^T(n), \ldots, x_m^T(n))^T$$

a state space description of the time-variant model (1) can be expressed as

$$\begin{aligned} X(n+1) &= \mathcal{F}X(n) + \mathcal{G}e(n) \qquad \theta(n) = \mathcal{H}X(n) \\ y(n) &= \Phi^H(n)X(n) + v(n) \end{aligned}$$

where $e(n) \triangleq (e_0(n), \ldots, e_m(n))^T$ and

$$\begin{aligned} \mathcal{F} &\triangleq diag(F, \ldots, F) \qquad \mathcal{G} \triangleq diag(G, \ldots, G) \\ \mathcal{H} &\triangleq diag(H, \ldots, H) \\ \Phi^H(n) &\triangleq (d(n)H, \ldots, d(n-m)H) \end{aligned}$$

Given the a priori information, the optimal adaptive algorithm for Gaussian noise would be the Kalman predictor

$$\begin{aligned} \widehat{X}(n+1) &= \mathcal{F}\widehat{X}(n) + K(n)\varepsilon(n) \\ \hat{\theta}(n) &= \mathcal{H}\widehat{X}(n) \\ K(n) &= \frac{P(n-1)\Phi(n)}{\sigma_v^2 + \Phi^H(n)P(n-1)\Phi(n)} \\ \varepsilon(n) &= y(n) - \varphi^H(n)\hat{\theta}(n) \end{aligned}$$

$$\begin{aligned} P(n) &= \mathcal{F}\{P(n-1) - P(n-1)Q(n) \\ &\quad P(n-1)\}\mathcal{F}^T + \mathcal{G}R_e\mathcal{G}^T \qquad (2) \\ Q(n) &\triangleq \frac{\Phi(n)\Phi^H(n)}{\sigma_v^2 + \Phi^H(n)P(n-1)\Phi(n)} \end{aligned}$$

where $R_e = diag(\sigma_{e_0}^2, \ldots, \sigma_{e_m}^2)$.[1] From a practical point of view, a Kalman predictor based on the accurate high order model is not attractive. This model depends on (f_m), which has to be accurately estimated since the power spectrum of a channel coefficient, subject to Rayleigh fading, has a pronounced peak at f_m. The quality of the estimate would be sensitive to errors in estimates of f_m. Furthermore, the computational load would be high, since the matrices in the state space model would be of high order[2]. Fortunately, the high order model can be replaced with a model of low order, without loosing much in performance.

It should also be noted that when possibly incorrect decisions are used in the predictor, there is no guarantee that a Kalman predictor, based on a model of high order, will give better performance than a low order based predictor[3]. We will therefore regard the matrices in the state space model as design variables, rather than very accurate models of the channel coefficient dynamics.

5. Steps towards a low-complexity estimator

The dominating computational load of a Kalman predictor is the recursive update of the Riccati equation (2). One conceivable way to avoid this update would be to replace $P(n)$ by the solution to an algebraic Riccati equation. However, since $P(n)$ depends on $\Phi^H(n)$, a stationary $P(n)$ does *not* exist. Another way could be to compute $\{P(n)\}_{n=1}^N$ in advance, and store this sequence for subsequent use. However, the sequence $\{P(n)\}_{n=1}^N$ will be different from data batch to data batch.

Let us instead regard $\{P(n)\}$ as a matrix-valued stochastic process and each data burst as an independent realization of this process. Assume the ensemble mean $E[P(n)]$ to exist. Taking expectation of (2) gives

$$\begin{aligned} E[P(n)] &= \mathcal{F}\{E[P(n-1)] - E[P(n-1)Q(n) \\ &\quad P(n-1)]\}\mathcal{F}^T + \mathcal{G}R_e\mathcal{G}^T \end{aligned}$$

After the initial transient phase, the time-variations of $P(n)$ are caused by the time-variant $\Phi^H(n)$. Suppose that the mean-value sequence $\{E[P(n)]\}$ does not vary much with n for $n > n_0$. Then, we can replace the mean value sequence with a constant matrix.

The reasonableness of these assumptions is verified below. In Figure 2, the trajectories of one element in $P(n)$, for ten realizations of a symbol sequence, is depicted. The system is the one described in the example in Section 7.

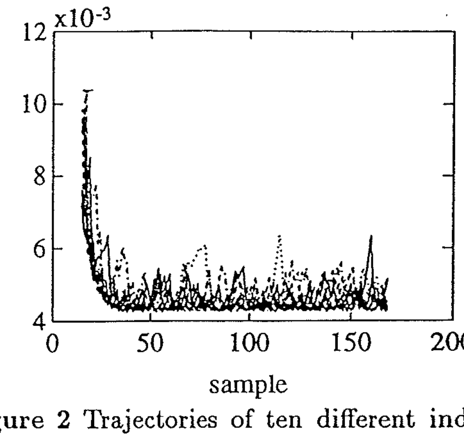

Figure 2 Trajectories of ten different independent realizations for one element in $P(n)$ (SNR= 20dB).

An approximation of the asymptotic value of $E[P(n)]$, which can be calculated at the beginning of each data burst and utilized in the filtering thereafter, is sought.

[1] Here, we have assumed that $e(n)$ is a stationary sequence. This is a reasonable assumption over one data burst.

[2] If the dimension of F is $n_f|n_f$, the dimension of $P(n)$ is $(m+1)n_f|(m+1)n_f$.

[3] Actually, an accurate model description of the channel coefficients gives substantially better performance only when the probability of incorrect decisions is low, ie. when the signal to noise ratio is high.

If the deviation of the matrix $P(n)$ from $E[P(n)]$, $n > n_0$, is not too large, a constant approximation P of the mean value sequence $\{E[P(n)]\}$ can be computed by solving the algebraic equation

$$\begin{aligned} P &= \mathcal{F}[P - PQP]\mathcal{F}^T + \mathcal{G}R_e\mathcal{G}^T \qquad (3) \\ Q &= E[Q(n)] = E\left[\frac{\Phi(n)\Phi^H(n)}{\sigma_v^2 + \Phi^H(n)P\Phi(n)}\right] \end{aligned}$$

Now, $Q(n)$ is a random matrix with a finite number of possible values. The evaluation of the expectation $E[Q(n)]$ is then relatively straightforward. If the symbol sequence is white, if $d(n)$ has constant modulus and if the channel coefficients are subjected to independent fading, there exist a blockdiagonal solution $P = diag(P_{00}, \ldots, P_{mm})$ to (3). Equation (3) can then be written as a system of *coupled* equations

$$\begin{aligned} P_{kk} &= F\left\{P_{kk} - P_{kk}\frac{H^TH}{\rho_k + HP_{kk}H^T}P_{kk}\right\}F^T \\ &+ GG^T\sigma_{e_k}^2 \qquad k = 0, \ldots, m \qquad (4) \\ \rho_k &\triangleq \frac{\sigma_v^2}{\sigma_d^2} + \sum_{i \neq k}^{m} HP_{ii}H^T \end{aligned}$$

where $E|d(n)|^2 = \sigma_d^2$. The coupling, here isolated in ρ_k, turns out to be quite loose. To simplify the calculations of $\{P_{kk}\}$, we regard $\{\rho_k\}$ as design variables. The system of equations then becomes *decoupled* and the matrices $\{P_{kk}\}$ can be calculated separately. The equations (4) are ordinary algebraic Riccati equations, solvable under mild conditions.

We can expect the approximations above to be reasonable when $\|Q(n)\|$ changes slightly with time. This is the case when the symbols have constant modulus ($\|\Phi(n)\Phi^H(n)\|$ =constant) and $P(n)$ is nearly blockdiagonal.

6. The low-complexity algorithm

The on-line computions required are given by

$$\begin{aligned} \varepsilon(n) &= y(n) - \varphi^H(n)\hat{\theta}(n) \\ \hat{x}_k(n+1) &= F\hat{x}(n) + \mu(n)L_kd^*(n-k)sat[\varepsilon(n)] \\ \hat{h}_k(n) &= H\hat{x}_k(n) \qquad k = 0, 1, \ldots, m. \qquad (5) \end{aligned}$$

The gains L_k $k = 0, \ldots, m$ can be pre-computed from the solutions to

$$\begin{aligned} \bar{P}_{kk} &= F\left\{\bar{P}_{kk} - \frac{\bar{P}_{kk}H^TH\bar{P}_{kk}}{1 + H\bar{P}_{kk}H^T}\right\}F^T + \gamma_kGG^T \qquad (6) \\ L_k &= \frac{F\bar{P}_{kk}H^T}{\sigma_d^2(1 + H\bar{P}_{kk}H^T)} \quad \bar{P}_{kk} \triangleq \frac{P_{kk}}{\rho_k} \quad \gamma_k \triangleq \frac{\sigma_{e_k}^2}{\rho_k} \end{aligned}$$

(A *time-varying positive scalar* $\mu(n)$ has been included to increase the gain during the transient phase, imitating the behaviour of $\{E[P(n)]\}$). Saturation of $\varepsilon(n)$ is used to robustify the algorithm. Here, γ_k can be used to tune the gain to a specific signal to noise ratio. Either, $\{\gamma_k\}$ can be changed for each burst or they can be kept constant over several bursts. Tuning of $\{\gamma_k\}$ can also be used to account for possible non-stationary behaviour of R_e.

With second order models $\{F, G, H\}$, simple *analytical expressions* for the gains L_k exist. Two different second order models will be discussed next.

An integrated random walk model

In fading environments, the channel coefficients exhibit typical trend behaviour, ie. they continue in some direction for a while. A simple way to incorporate such behaviour is to model the coefficients as integrated random walks [4]

$$h_k(n) = h_k(n-1) + \frac{1}{1 - q^{-1}}e_k(n)$$

This model can be represented in a state-space form by the following matrices

$$F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad G = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad H = (1\,0)$$

Analytic expressions of the gains L_k are then given by

$$L_k = \frac{F\bar{P}_{kk}H^T}{\sigma_d^2(1 + \bar{p}_{1,1})} = \frac{1}{\sigma_d^2(1 + \bar{p}_{1,1})}\begin{pmatrix} \bar{p}_{1,1} + \bar{p}_{1,2} \\ \bar{p}_{1,2} \end{pmatrix}$$

where $\bar{p}_{1,1}$ and $\bar{p}_{1,2}$ are elements of $\bar{P}_{kk}$ and given by

$$\begin{aligned} \zeta &\triangleq \frac{4 + \gamma_k + \sqrt{\gamma_k(16 + \gamma_k)}}{2} \\ \bar{p}_{1,1} &= \frac{\zeta + \sqrt{\zeta^2 - 4}}{2} - 1 \quad \bar{p}_{1,2} = \sqrt{\gamma_k(1 + \bar{p}_{1,1})} \end{aligned}$$

Lightly damped AR(2) models

In Rayleigh fading environments, the channel coefficients behave as narrow band noises with a spectral peak. (The autocovariance function of a channel coefficient, subject to Rayleigh fading, is given by a Bessel function of the first kind and zero order, $J_0(2\pi f_m/f_s\,\tau)$, where f_s is the symbol rate and τ is a time lag.) The simplest models, which describes such oscillatory behaviour, are lightly damped second order AR models with real coefficients

$$\begin{aligned} h_k(n) &= \frac{1}{1 + a_1q^{-1} + a_2q^{-2}}e_k(n) \qquad (7) \\ a_1 &= -2r_d\cos(\omega_d) \quad a_2 = r_d^2 \end{aligned}$$

The pole locations are $r_de^{\pm j\omega_d}$. The pole radius (r_d) reflects the damping and (ω_d) is the dominating frequency of the coefficient variation. If f_m is unknown, the spectral peak should be well damped, to obtain a robust model. We use $r_d = 0.995$, $\omega_d = 0.02$ below.

The model (7) can be represented in an observable canonical state-space form by the following matrices

$$F = \begin{pmatrix} -a_1 & 1 \\ -a_2 & 0 \end{pmatrix} \quad G = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad H = (1\,0)$$

[4] The integrated random walk model is a natural and simple extension of the random walk model when the coefficients have lowpass character. Kalman filtering based on an integrated random walk model constitutes a simple form of so called *multistep algorithms*, see eg [2].

Analytic expressions of the gains L_k are given by

$$L_k = \frac{F\bar{P}_{kk}H^T}{\sigma_d^2(1+\bar{p}_{1,1})} = \frac{1}{\sigma_d^2(1+\bar{p}_{1,1})}\begin{pmatrix} -a_1\bar{p}_{1,1}+\bar{p}_{1,2} \\ -a_2\bar{p}_{1,1}\end{pmatrix}$$

where $\bar{p}_{1,1}$ and $\bar{p}_{1,2}$ are elements in $\bar{P}_{kk}$ and given by

$$\zeta \triangleq \frac{\alpha_0 - 2a_2 + \sqrt{(\alpha_0+2a_2)^2 - 4a_1^2(1+a_2)^2}}{2}$$

$$\alpha_0 \triangleq 1 + a_1^2 + a_2^2 + \gamma_k$$

$$\bar{p}_{1,1} = \frac{\zeta + \sqrt{\zeta^2 - 4a_2^2}}{2} - 1 \quad \bar{p}_{1,2} = \frac{a_1 a_2}{1+a_2+\bar{p}_{1,1}}\bar{p}_{1,1}$$

Given γ_k, the gains L_k can be calculated once, for each data burst. It might even be sufficient to have one set of pre-calculated gains for *all* data bursts.

7. An example

Consider transmission of data bursts of 168 symbols over a two taps Rayleigh fading channel ($m = 1$). The first 14 symbols constitute the training sequence. The taps change independently. The maximum Doppler frequency is 83Hz and the symbol rate is 25kHz. The symbols take the values $\{1+j, -1+j, -1-j, 1-j\}$, equally likely, and correspond to the set of bits $\{00, 01, 11, 10\}$. [5] The variance of the sequence is then $\sigma_d^2 = |d(n)|^2 = 2$. The symbols are differentially encoded [6].

The channel estimator is used in conjunction with a Viterbi detector. In Figures 3. and 4., simulations of two extremes of a range of situations, to be expected in practice, are summarized. The performance is compared for the low-complexity algorithm (dashed), the Kalman predictor (dashed-dotted), both based on second order oscillative models, the "classical" RLS (dotted), and using known channel states in the Viterbi algorithm (solid).

With "classical" RLS-tracking, the best performance, shown in the figures, was achieved with forgetting factor 0.7. The difference between the Kalman predictor and the simplified algorithm is almost negligible. In Figure 5, the performance of the low-complexity algorithm based on the AR(2)-model (dotted) is compared with the integrated random walk model (dashed).

The number of real-valued arithmetic operations per iteration in the low-complexity ("KLMS") algorithm, based on the integrated random walk, is summarized in the table below. It is compared to LMS tracking of $h_k(n)$.

[5] All results are based on simulations of 1000 bursts (350 000 unknown bits) at 15, 20 and 25 dB SNR. Time-varying coefficients were generated by filtering white complex noise. The filters used were a second order oscillative filter, followed by a 5'th order lowpass filter. We generated 10 batches of 100 bursts each. It was ascertained that the level-crossing statistics for *each* batch corresponded closely to the Rayleigh fading model. This reduced the variance of the results significantly.

[6] This specification is similar to the proposed North American digital mobile radio standard. We disregard the $\pi/4$-shift in $\pi/4$-DQPSK.

	×	+
KLMS	8(m+1)+2	7(m+1)
LMS	4(m+1)+2	4(m+1)

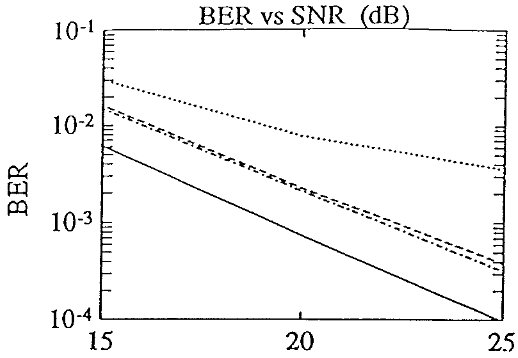

Figure 3 Performance of adaptive Viterbi when $E|h_0(n)|^2 = E|h_1(n)|^2$

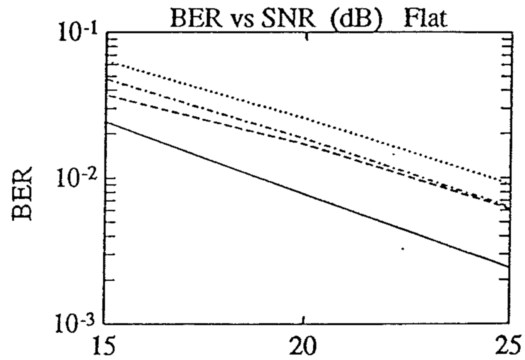

Figure 4 As Fig 3., for Flat fading, $E|h_1(n)|^2 = 0$.

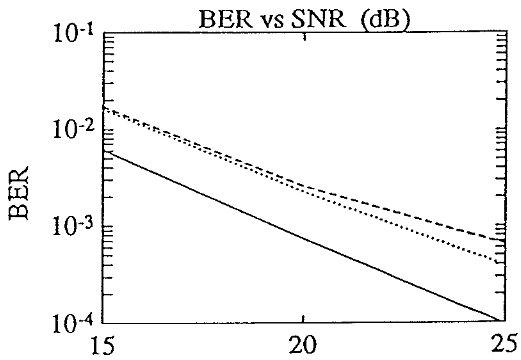

Figure 5 Comparsion between the use of a damped oscillatory model (dotted) and an integrated random walk model (dashed), $E|h_0(n)|^2 = E|h_1(n)|^2$.

References

[1] **L. Lindbom** Adaptive equalization for fading mobile radio channels. Licentiate Thesis, Department of Technology, Uppsala University, Sweden, June 1992.

[2] **A Benveniste, M Métivier and P Priouret** (1990) Adaptive Algorithms and Stochastic Approximations. Springer-Verlag, Berlin.