FRACTIONALLY-SPACED BLIND CHANNEL EQUALISATION USING HIDDEN MARKOV MODELS

Vikram Krishnamurthy and Kutluyıl Doğançay

Department of Electrical and Electronic Engineering
The University of Melbourne, Parkville, Victoria 3052, Australia
k.dogancay@ee.mu.oz.au

ABSTRACT

The paper presents a maximum likelihood (ML) blind channel equalisation algorithm based on the expectation-maximisation (EM) algorithm. We assume that the channel input sequence is a finite-state Markov chain and the channel output sequence is obtained from the continuous-time channel output by oversampling it at a rate higher than the channel input symbol rate, which leads to a fractionally-spaced channel equalisation problem. The objective of blind channel equalisation is to estimate the channel input symbols without explicit knowledge of the channel characteristics and the requirement of training data. The availability of multi-channel outputs for the same channel input improves the reliability of the estimates. A reduced-cost blind equalisation algorithm which draws on aggregation by stochastic complementation is also proposed. A simulation example is presented to demonstrate the performance of the proposed algorithms.

1. INTRODUCTION

The paper proposes a maximum likelihood (ML) blind channel equalisation algorithm for fractionally-spaced finite impulse response (FIR) channels. We assume that the channel input can be modelled as a first-order, finite-state Markov chain. As different from the previous work [1, 2, 3, 4], we apply hidden Markov model signal processing techniques to the problem of *fractionally-spaced* channel equalisation. The availability of multi-channel output observations helps to improve the reliability of parameter estimates. The developed algorithm yields maximum likelihood (ML) estimates of the channel input sequence and channel parameters. A reduced complexity sub-optimal implementation is also proposed to make the algorithm practical.

This work is supported by the Australian Research Council (ARC) and the Cooperative Research Centre for Sensor, Signal and Information Processing (CSSIP)

Fractionally-spaced blind channel equalisation algorithms invariably assume a channel disparity condition so as to achieve perfect zero-forcing equalisation for FIR channels by means of finite-length parametrisation at the equaliser. ML parameter estimation methods for white channel inputs also require the channel disparity condition to hold. The channel disparity is attained if the subchannels have no zeros common to all of them. The ML algorithms presented in this paper are invariant to the channel disparity condition, which makes them relatively robust and widely applicable. The gradient-descent based blind equalisation algorithms such as the constant modulus algorithm (CMA) are not capable of producing an acceptable equalisation performance at high channel noise, whereas ML estimation methods can yield reasonable estimates even at high noise levels.

The paper is organised as follows. Section 2 defines the channel equalisation problem at hand. In Section 3 we present an EM based blind equalisation algorithm using fixed interval (noncausal) smoothing. In Section 4 an aggregation method based on stochastic complementation is proposed to reduce the computational complexity associated with ML estimation. A computer simulation example for the reduced cost algorithm is presented in Section 5.

2. PROBLEM FORMULATION

If a continuous-time channel output is oversampled at a rate K times higher than the channel input symbol rate T, the resulting channel output sequence can be modelled as shown in Fig. 1. The channel input sequence $\{u(k)\}$ is assumed to be generated by an M-state Markov chain $\boldsymbol{s}(k)$ with state space $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_M\}$ ($\boldsymbol{e}_i$ is the unit column vector of $\mathbb{R}^M$), transition probability matrix $\boldsymbol{A} = [a_{ij}]$ where $a_{ij} = \Pr\{\boldsymbol{s}(k+1) = \boldsymbol{e}_j \mid \boldsymbol{s}(k) = \boldsymbol{e}_i\}$, $i, j \in \{1, \ldots, M\}$ and $\sum_{j=1}^{M} a_{ij} = 1$, and levels $\boldsymbol{g} = [g_1, \ldots, g_M]^T$. The channel input sequence

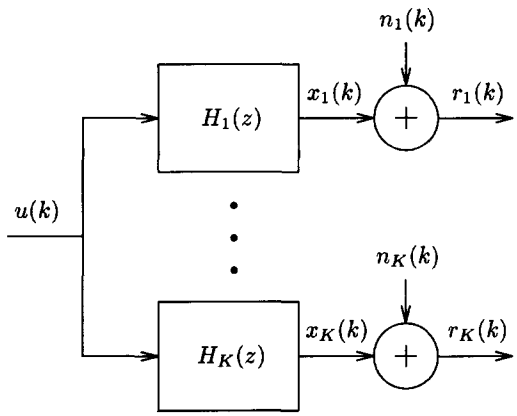

Figure 1: Signal model for fractionally-spaced channel equalisation.

is obtained from the Markov chain by the inner product $u(k) = \langle \boldsymbol{g}, \boldsymbol{s}(k) \rangle$. The FIR subchannels are defined by

$$H_i(z) = \sum_{j=0}^{L-1} h_j^{(i)} z^{-j}, \quad i = 1, \ldots, K.$$

We shall denote the impulse response of subchannels by $\boldsymbol{h}_i = [h_0^{(i)}, \ldots, h_{L-1}^{(i)}]^T$. The subchannel output disturbances $n_i(k)$ are white Gaussian processes with zero mean and variance σ_n^2 (i.e. $n_i(k) \sim \mathcal{N}(0, \sigma_n^2)$) and are also uncorrelated with each other.

Our objective is to estimate $\{u(k)\}$ and $H_i(z)$ in a blind-folded manner. In the next section we will formulate a method for obtaining ML estimates of these parameters from subchannel output sequences.

3. ML PARAMETER AND STATE ESTIMATION USING THE EM ALGORITHM

The case of $K = 1$ has already been studied in the literature (see e.g. [1]). The process $\boldsymbol{S}(k) = [\boldsymbol{s}(k), \boldsymbol{s}(k-1), \ldots, \boldsymbol{s}(k-L+1)]^T$ associated with the subchannel outputs is also a Markov chain with $N = M^L$ states and state space $\{\boldsymbol{E}_1, \ldots, \boldsymbol{E}_N\}$ where

$$\boldsymbol{E}_i = [\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_L}]_{M \times L}, \quad i \in \{1, \ldots, N\}. \quad (1)$$

The initial state probabilities will be denoted by $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_N]^T$ with $\pi_i = \Pr\{\boldsymbol{S}(1) = \boldsymbol{E}_i\}, i \in \{1, \ldots, N\}$. The transition probabilities of $\boldsymbol{S}(k)$ are

$$\begin{aligned} p_{ij} &= \Pr\{\boldsymbol{S}(k+1) = \boldsymbol{E}_j \mid \boldsymbol{S}(k) = \boldsymbol{E}_i\} \\ &= \Pr\{\boldsymbol{s}(k+1) = \boldsymbol{e}_{j_1} \mid \boldsymbol{s}(k) = \boldsymbol{e}_{i_1}\} \delta(j_2 - i_1) \\ &\quad \cdots \delta(j_L - i_{L-1}) \\ &= a_{i_1 j_1} \delta(j_2 - i_1) \cdots \delta(j_L - i_{L-1}) \end{aligned}$$

for $i, j \in \{1, \ldots, N\}$, where $\delta(\cdot)$ denotes the *Kronecker delta function.*

The noise-free subchannel output sequences can be written as

$$x_i(k) = \boldsymbol{h}_i^T \boldsymbol{S}^T(k) \boldsymbol{g}.$$

Obviously, the noisy subchannel output observations $r_i(k) = x_i(k) + n_i(k)$ represent a hidden Markov model (HMM). Whilst each subchannel $H_i(z)$ and its input sequence can be estimated separately, we will exploit the fact that each subchannel is driven by a common input sequence $\{u(k)\}$ so as to improve the channel input estimates and also to reduce the cost of estimation.

Let us define the conditional symbol probabilities for $\boldsymbol{r}(k) = [r_1(k), \ldots, r_K(k)]^T$ as

$$\begin{aligned} b_i(\boldsymbol{r}(k)) &= f(\boldsymbol{r}(k) \mid \boldsymbol{S}(k) = \boldsymbol{E}_i) \\ &= (2\pi\sigma_n^2)^{-K/2} \\ &\quad \times \exp\left(-\frac{1}{2\sigma_n^2} \sum_{j=1}^{K} \left(r_j(k) - \boldsymbol{h}_j^T \boldsymbol{E}_i^T \boldsymbol{g} \right)^2 \right). \end{aligned}$$

Let $z_T \triangleq \{\boldsymbol{z}(k) = (\boldsymbol{r}(k), \boldsymbol{S}(k)), k = 1, \ldots, T\}$ denote the "complete" (fully categorised) data for which ML estimation can be computed explicitly. For the incomplete data problem where only $r_T \triangleq \{\boldsymbol{r}(k), k = 1, \ldots, T\}$ is available, explicit computation of an ML estimate is usually not possible. The EM algorithm [5] can be employed to generate a numerical solution to the ML estimate for incomplete data. Starting with an initial estimate $\lambda(0)$ of the Markov chain parameters $\lambda = (\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_K, \sigma_n^2)$, the EM algorithm proceeds as follows:

E step: Evaluate the conditional expectation of the complete data log-likelihood

$$\mathcal{Q}(\lambda(l), \lambda) = E\{\ln f(z_T \mid \lambda) \mid r_T, \lambda(l)\}$$

where $\lambda(l)$ is the parameter estimate on the lth pass.

M step: Find $\lambda = \lambda(l+1)$ that maximises $\mathcal{Q}(\lambda(l), \lambda)$.

These steps are described in some detail next.

3.1. The E Step

The complete data log-likelihood is given by

$$\begin{aligned} &\ln f(z_T \mid \lambda) = \ln \Pr\{\boldsymbol{S}(1) \mid \lambda\} \\ &+ \sum_{t=1}^{T-1} \ln \Pr\{\boldsymbol{S}(t+1) \mid \boldsymbol{S}(t), \lambda\} + \sum_{t=1}^{T} \ln \Pr\{\boldsymbol{r}(t) \mid \boldsymbol{S}(t), \lambda\} \end{aligned}$$

whence, on taking conditional expectation, we obtain

$$\mathcal{Q}(\lambda(l),\lambda) = \sum_{i=1}^{N} \ln \gamma_1(i)\pi_i + \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{j=1}^{N} \xi_t(i,j)\ln p_{ij} + \sum_{t=1}^{T}\sum_{i=1}^{N} \gamma_t(i)\ln b_i(\boldsymbol{r}(t))$$

where the conditional probabilities

$$\gamma_t(i) \triangleq \Pr\{\boldsymbol{S}(t) = \boldsymbol{E}_i \mid z_T, \lambda(l)\}$$
$$\xi_t(i,j) \triangleq \Pr\{\boldsymbol{S}(t) = \boldsymbol{E}_i, \boldsymbol{S}(t+1) = \boldsymbol{E}_j \mid z_T, \lambda(l)\}.$$

are computed using the *forward-backward* algorithm [6]. The effort required for computation of $\gamma_t(i)$ and $\xi_t(i,j)$ is the root cause of high computational complexity of the EM algorithm.

3.2. The M Step

Assuming that we wish to estimate the subchannel parameters, the M step consists of maximising $\mathcal{Q}(\lambda(l),\lambda)$ by setting $\partial\mathcal{Q}(\lambda(l),\lambda)/\partial\boldsymbol{h}_i(l+1) = \boldsymbol{0}$, which leads to

$$\underbrace{\sum_{t=1}^{T}\sum_{j=1}^{N}\gamma_t(j)\left(\boldsymbol{E}_j^T\boldsymbol{g}\right)\left(\boldsymbol{g}^T\boldsymbol{E}_j\right)}_{\boldsymbol{D}_{L\times L}}\boldsymbol{h}_i(l+1) = \underbrace{\sum_{t=1}^{T}\sum_{j=1}^{N}\gamma_t(j)r_i(t)\left(\boldsymbol{E}_j^T\boldsymbol{g}\right)}_{(\boldsymbol{d}_i)_{L\times 1}}$$

whose solution is simply given by $\boldsymbol{h}_i(l+1) = \boldsymbol{D}^{-1}\boldsymbol{d}_i$, $1 \le i \le K$.

3.3. MAP Estimate of Channel Input Symbols

A maximum *a posteriori* (MAP) estimate of $u(t)$ based on the estimated parameters $\lambda(l)$ can be obtained from the state estimate of the E step as follows

$$\begin{aligned} i^* &= \arg\max_{1\le i\le M}\Pr\{\boldsymbol{s}(t) = \boldsymbol{e}_i \mid z_T, \lambda(l)\} \\ &= \arg\max_{1\le i_1\le M}\sum_{i_2,\ldots,i_L}\Pr\{\boldsymbol{s}(t) = \boldsymbol{e}_{i_1}, \boldsymbol{s}(t-1) = \boldsymbol{e}_{i_2}, \\ &\quad \ldots, \boldsymbol{s}(t-L+1) = \boldsymbol{e}_{i_L} \mid z_T, \lambda(l)\} \\ \hat{u}(t) &= \langle \boldsymbol{g}, \boldsymbol{e}_{i^*}\rangle. \end{aligned} \tag{2}$$

Let us express $\gamma_t(i)$ explicitly as $\gamma_t(i_1, i_2, \ldots, i_L)$ where the relationship between the indices is governed by (1). Then (2) can be rewritten as

$$\begin{aligned} i^* &= \arg\max_{1\le i_1\le M}\sum_{i_2,\ldots,i_L}\gamma_t(i_1,i_2,\ldots,i_L) \\ \hat{u}(t) &= \langle \boldsymbol{g}, \boldsymbol{e}_{i^*}\rangle. \end{aligned} \tag{3}$$

4. REDUCING THE COST BY AGGREGATION

We use the concept of stochastic complementation [7] to aggregate the Markov chain $\boldsymbol{S}(k)$. For FIR filtered Markov chains aggregated transition probability matrix $\tilde{P}$ is simply given by the channel input transition probability matrix $\boldsymbol{A}$ [8]. In addition to aggregating the transition probability matrix $\boldsymbol{P}$, we also need to aggregate the conditional symbol probabilities. The aggregated symbol probabilities are given by [3]

$$\tilde{b}_i(\boldsymbol{r}(k)) = \frac{\sum_{j\in\mathcal{S}_i}\zeta_j b_j(\boldsymbol{r}(k))}{\sum_{j\in\mathcal{S}_i}\zeta_j}, \quad i = 1,\ldots,M \tag{4}$$

where $\boldsymbol{\zeta} = [\zeta_1,\ldots,\zeta_N]$ is the *Perron-Frobenius* eigenvector of $\boldsymbol{P}$ and $\mathcal{S}_i$ is the state index set for the ith aggregated state.

The EM algorithm can be applied to the aggregated chain in the same way as before. Neglecting the terms that do not depend on $\boldsymbol{h}_i$, the aggregated conditional expectation of the complete log-likelihood is given by

$$\tilde{\mathcal{Q}}(\lambda(l),\lambda) = \sum_{t}^{T}\sum_{i=1}^{M}\tilde{\gamma}_t(i)\ln\tilde{b}_i(\boldsymbol{r}(t)).$$

The M step for the aggregated chain does not have an analytical solution. Gradient search or Newton-Raphson methods may be used to maximise $\tilde{\mathcal{Q}}(\lambda(l),\lambda)$ recursively. The gradient of $\tilde{\mathcal{Q}}(\lambda(l),\lambda)$ with respect to $\boldsymbol{h}_i(l+1)$ is

$$\frac{\partial\tilde{\mathcal{Q}}(\lambda(l),\lambda)}{\partial\boldsymbol{h}_i(l+1)} = \frac{1}{\sigma_n^2}\sum_{t=1}^{T}\sum_{j=1}^{M}\tilde{\gamma}_t(j) \times \frac{\sum_{m\in\mathcal{S}_j}\zeta_m b_m(\boldsymbol{r}(t))\left(r_i(t) - \boldsymbol{h}_i^T(l+1)\boldsymbol{E}_m^T\boldsymbol{g}\right)\boldsymbol{E}_m^T\boldsymbol{g}}{\sum_{m\in\mathcal{S}_j}\zeta_m b_m(\boldsymbol{r}(t))}.$$

We use the following gradient descent algorithm to find the maximising $\boldsymbol{h}_i(l+1)$

$$\boldsymbol{h}_i^{(j+1)}(l+1) = \boldsymbol{h}_i^{(j)}(l+1) + \mu\frac{\partial\tilde{\mathcal{Q}}(\lambda(l),\lambda)}{\partial\boldsymbol{h}_i^{(j)}(l+1)}, \quad i = 1,\ldots,K$$

where μ is a possibly time-varying stepsize and $\boldsymbol{h}_i^{(j)}(l+1)$ is the value of $\boldsymbol{h}_i(l+1)$ at the jth recursion.

4.1. MAP Estimation

Similarly to Section 3.3, MAP estimate of the channel input symbols can be obtained from

$$\begin{aligned} i^* &= \arg\max_{1\le i\le M}\tilde{\gamma}_t(i) \\ \hat{u}(t) &= \langle \boldsymbol{g}, \boldsymbol{e}_{i^*}\rangle. \end{aligned} \tag{5}$$

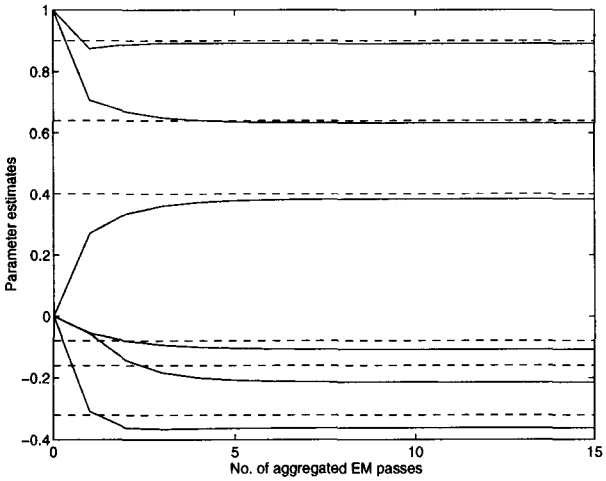

Figure 2: Subchannel parameter estimates.

5. SIMULATION EXAMPLES

To emphasise the advantage of the proposed ML estimation method over other ML methods which require the channel disparity condition to be satisfies, we consider the following multichannel model ($K = 2$)

$$H_1(z) = 0.4 + 0.64z^{-1} - 0.32z^{-2}$$
$$H_2(z) = 0.9 - 0.16z^{-1} - 0.08z^{-2}$$

where the subchannels have a common zero at $z = 0.4$ and therefore the channel disparity condition is not satisfied. The multichannel model is assumed to be driven by a binary input sequence with $u(k) \in \{-1, 1\}$ ($M = 2$), which is generated by a Markov chain with transition probability matrix

$$\boldsymbol{A} = \begin{bmatrix} 0.65 & 0.35 \\ 0.35 & 0.65 \end{bmatrix}$$

and equiprobable initial states.

We have simulated the reduced cost aggregation method described in Section 4. The aggregated transition probability matrix is given by $\breve{\boldsymbol{P}} = \boldsymbol{A}$. For $T = 1000$ and $\sigma_n^2 = 0.3$, the convergence of the aggregated EM is shown in Fig. 2 where the dashed lines indicate the true subchannel parameters. In the M step, we used 50 iterations per EM pass and a constant stepsize of $\mu = 2 \times 10^{-4}$ for the gradient descent maximisation. The symbol error of the resultant MAP estimates is 4.8%.

6. DISCUSSION AND CONCLUSION

We have presented an EM based ML estimation method for fractionally-spaced blind channel equalisation. A reduced cost aggregation method based on stochastic complementation was also proposed. The existence of local maxima and the local convergence of the EM algorithm prompt the need of a good initialisation, which may be achieved by other non-optimal blind equalisation algorithms.

7. REFERENCES

[1] V. Krishnamurthy and L. B. White, "Blind equalization of FIR channels with Markov inputs," in *Proc. IFAC Int. Symposium on Adaptive Systems in Control and Signal Processing*, (Grenoble, France), pp. 633–638, July 1992.

[2] L. B. White and V. Krishnamurthy, "Adaptive blind equalisation of FIR channels using hidden Markov models," in *Proc. IEEE Int. Conference on Communications, ICC '93*, (Geneva, Switzerland), May 1993.

[3] V. Krishnamurthy, "Adaptive estimation of hidden nearly completely decomposable Markov chains with applications in blind equalization," *Int. J. Adaptive Control and Signal Processing*, vol. 8, pp. 237–260, 1994.

[4] M. Shao and C. L. Nikias, "An ML/MMSE estimation approach to blind equalization," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, ICASSP '94*, vol. 4, (Adelaide, Australia), pp. 569–572, Apr. 1994.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B*, vol. 39, pp. 1–38, 1977.

[6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–286, Feb. 1989.

[7] C. D. Meyer, "Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems," *SIAM Rev.*, vol. 31, pp. 240–272, June 1989.

[8] K. Doğançay and V. Krishnamurthy, "Quick aggregation of Markov chain functionals via stochastic complementation," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, ICASSP '97*, (Munich, Germany), Apr. 1997.