On the Maximum-Likelihood based data-aided frequency offset and channel estimates

Philippe Ciblat[1] *and Luc Vandendorpe*[2]
[1] Ecole Nationale Supérieure des Télécommunications, Paris, France
(email : philippe.ciblat@enst.fr)
[2] Université catholique de Louvain, Louvain-la-Neuve, Belgium
(email : vandendorpe@tele.ucl.ac.be)

ABSTRACT

We consider transmission over a frequency-selective channel. We focus on the data-aided joint and directed maximum-likelihood estimation of the carrier frequency offset and of the dispersive channel. The directed estimators correspond to the frequency offset estimator assuming the channel known, as well as the channel estimator assuming the frequency offset known. A comparison of directed and non-directed estimators based on asymptotic (large sample) analysis is addressed and shows that the performance of the joint estimates and the directed estimates is not so different.

1 Introduction

In a wireless scenario, the transmitted signal can be affected by two kinds of impairments : Inter Symbol Interference, which occurs because of multi-path propagation, and a carrier frequency offset caused by a Doppler effect or a local oscillator drift. Before applying an offset correction and an equalizer, the carrier frequency offset and the channel, which are unknown, have to be estimated.

In many applications, known data are transmitted to achieve channel estimation at the receiver. Moreover in the new wireless communication applications, the frequency offset cannot be considered as null or negligible. Therefore, the problem of data-aided frequency offset and channel estimation remains of interest.

Many papers treat the problem of data-aided frequency offset estimation for an Additive White Gaussian Noise channel (see [3] and references therein). In a frequency-selective channel context, a few papers rely on the Maximum-Likelihood-like estimator ([2, 4]). The statistical performance of the joint Maximum-Likelihood frequency offset and channel estimators is partially established in [4].

The main concern now seems to be the design of the optimal training sequence associated with the joint Maximum-Likelihood estimators. As this task proves very difficult, it will be interesting to evaluate the loss in performance of the joint estimators carried out with the usual pseudo-random white training sequence with respect to a relevant lower bound. Therefore we consider the so-called directed estimators which are the Maximum-Likelihood channel estimator when the frequency offset is assumed to be known, as well as, the Maximum-Likelihood frequency offset estimator when the channel is assumed to be known. The optimal training sequence design for the both directed estimators [1] is addressed. This allows to see the performance associated with the optimal designed directed estimators as a judicious benchmark.

We focus on a single-carrier and single-user communications scheme. As usually done, we only consider the baud-rate sampled received signal $\{y(k)\}$. Then

$$y(k) = ([h(z)].s_k)\, e^{2i\pi f_0 k} + n(k). \tag{1}$$

The symbol sequence $\{s_k\}_{k\in\mathbb{Z}}$ is assumed to be i.i.d., zero-mean and unit-variance. The notation $[.]$ stands for the convolution operator. The filter $h(z)$, which arises from the convolution of the shaping filter and the propagation channel, is assumed to be time-limited and causal. Therefore an integer value L can be found such that

$$h(z) = \sum_{l=0}^{L} h_l z^{-l}.$$

We stack the channel coefficients in $\mathbf{h}_0 = [h_0, \cdots, h_L]^{\mathrm{T}}$. Superscripts $^{\mathrm{T}}$ and $^{\mathrm{H}}$ stand for the transposition and the complex-conjugate transposition, respectively. Furthermore, $\{n(k)\}$ is a circular white Gaussian noise. Lastly, f_0 denotes the discrete-time equivalent frequency offset.

Our task is to estimate the vector $\mathbf{h}_0$ and the scalar f_0.

In practice, the transmitter sends $(M+L)$ training symbols $\{v_m\}_{-L\leq m\leq M-1}$. On account of the filtering operation, the first L training symbols are fixed to be null. The M remaining non-null training symbols

$$\mathbf{v}_M = [v_0, \cdots, v_{M-1}]^{\mathrm{T}}$$

represent the significant part of the training sequence.

We consider the received samples

$$\mathbf{y}_M = [y(0), \cdots, y(M-1)]^{\mathrm{T}}$$

corresponding to the transmission of the significant part of the training sequence. Then Equation (1) becomes

$$\mathbf{y}_M = D_M(f_0)\mathbf{V}_M\mathbf{h}_0 + \mathbf{n}_M$$

[1] The result is well-known for the directed channel estimator but not for the directed frequency-offset estimator.

where $D_M(f) = \mathrm{diag}([1, \cdots, e^{2i\pi(M-1)f}])$. $\mathrm{diag}(\mathbf{x})$ is the diagonal matrix built from the vector $\mathbf{x}$. As the noise is white, we get $\mathbb{E}[\mathbf{n}_M \mathbf{n}_M^{\mathrm{H}}] = \sigma^2 \mathbf{Id}_M$ where σ^2 and $\mathbf{Id}_M$ represent the variance noise and the identity matrix of size M respectively. The l^{th} column of the matrix $\mathbf{V}_M$ of size $M \times (L+1)$ matrix is defined as $\mathbf{v}_l = [v_{l-1}, \cdots, v_{M-l-1}]^{\mathrm{T}}$.

We assume that the known matrix $\mathbf{V}_M$ is tall ($M > L$), full-rank (the inverse of $\mathbf{V}_M^{\mathrm{H}} \mathbf{V}_M$ exists), and power-constrained ($||\mathbf{v}_l||^2 = M$).

2 Maximum-Likelihood Estimators

Joint Maximum-Likelihood estimators Since the noise $\mathbf{n}_M$ is white Gaussian, the maximization of the joint log-Maximum-Likelihood function for frequency offset and channel parameters leads to the following estimators, denoted by $\hat{f}_M$ and $\mathbf{h}_M$, ([2, 4])

$$\hat{f}_M = \arg \max_{f \in [0,1[} J_M(f)$$

$$\hat{\mathbf{h}}_M = (\mathbf{V}_M^{\mathrm{H}} \mathbf{V}_M)^{-1} \mathbf{V}_M^{\mathrm{H}} D_M(\hat{f}_M)^{\mathrm{H}} \mathbf{y}_M. \quad (2)$$

with

$$J_M(f) = \frac{1}{M} \mathbf{y}_M^{\mathrm{H}} D_M(f) \mathbf{P}_M D_M(f)^{\mathrm{H}} \mathbf{y}_M. \quad (3)$$

and

$$\mathbf{P}_M = \mathbf{V}_M (\mathbf{V}_M^{\mathrm{H}} \mathbf{V}_M)^{-1} \mathbf{V}_M^{\mathrm{H}}.$$

We notice that $\mathbf{P}_M$ is a projection matrix and represents the projection on the space spanned by the columns of $\mathbf{V}_M$.

Directed Maximum-Likelihood estimators We assume that f_0 is known. Then the frequency offset can be compensated before estimating the channel. Therefore we can consider the model (1) restricted to the case $f_0 = 0$. The corresponding well-known ML channel estimator can thus be written as follows

$$\hat{\mathbf{h}}_{M|f} = (\mathbf{V}_M^{\mathrm{H}} \mathbf{V}_M)^{-1} \mathbf{V}_M^{\mathrm{H}} \mathbf{y}_M. \quad (4)$$

Obviously, one can notice that Equation (2) is a simple extension of Equation (4) to the case needing a prior frequency offset correction step.

The filter $\mathbf{h}_0$ is now assumed to be known. Although this scheme is quite theoretical, we focus on it in order to measure the performance gap between the frequency offset estimate driven by the knowledge of the filter and that one disturbed by an unknown filter. Furthermore, this scheme can also make sense in a flat-fading channel context since, for such a communication system, the filter is often known at the receiver and, does not need to be evaluated.

According to Equation (1), estimating the carrier frequency offset f_0 is equivalent to estimating the frequency of a sinusoid with time-varying complex-valued amplitude corrupted by an additive noise. In a data-aided estimation context, if the filter $\mathbf{h}_0$ is unknown (resp. known), the time-varying complex-valued amplitude is also unknown (resp. known). In [5], the authors point out that the Maximum-Likelihood function maximization leads to two different frequency estimators according to whether the complex amplitude is known or not. Therefore, the maximization of the Maximum-Likelihood as the filter $\mathbf{h}_0$ is a known parameter provides a frequency offset estimate $\hat{f}_{M|\mathbf{h}}$ associated with a criterion $H_M(f)$ which is slightly different and, even cannot be deduced from the criterion $J_M(f)$ reported in Equation (3). Hence,

$$\hat{f}_{M|\mathbf{h}} = \arg \max_{f \in [0,1[} H_M(f)$$

with

$$H_M(f) = \Re \left[\frac{1}{M} \sum_{m=0}^{M-1} \overline{c(m)} y(m) e^{-2i\pi f m} \right]$$

where $\{c(m)\}_m$ corresponds to the training sequence filtered by $h(z)$ and $\Re[.]$ stands for the real part of a complex-valued number.

3 Large sample behavior

We only focus on the closed-form expression of the large sample covariances of the estimators. Their expressions are valid under very mild standard mixing assumptions on the training sequence. This condition essentially refers to the fact that sufficiently separated samples are approximatively independent.

Let $\{x_M\}$ be a stochastic or deterministic sequence. $\mathrm{val}(x_M)$ corresponds to the evaluation of x_M as M is large enough, i.e., we have neglected the deterministic (resp. stochastic) terms converging (resp. almost surely) to zero.

Joint estimation performance Let

$$\begin{cases} \gamma_f & = & \mathrm{val}\left(M^3 \mathbb{E}[(\hat{f}_M - f_0)(\hat{f}_M - f_0)^{\mathrm{H}}]\right) \\ \gamma_{\mathbf{h}} & = & \mathrm{val}\left(M \mathbb{E}[(\hat{\mathbf{h}}_M - \mathbf{h}_0)(\hat{\mathbf{h}}_M - \mathbf{h}_0)^{\mathrm{H}}]\right). \end{cases}$$

Before going further, we introduce the following $(L+1) \times (L+1)$ matrices which will play a great role :

$$\mathbf{W}_K = \frac{\mathbf{V}_M^{\mathrm{H}} \Delta_M^K \mathbf{V}_M}{M^{(K+1)}}$$

where $\Delta_M = \mathrm{diag}([0, 1, \cdots, M-1])$.

Since $\mathbf{V}_M$ is power-constrained, we obtain that

$$\mathbf{W}_K = \mathcal{O}(1) \quad \text{for} \quad K \in \mathbb{N} \quad (5)$$

In the sequel, we express γ_f and $\gamma_{\mathbf{h}}$. Due to the lack of space, we do not report the derivations.

The expression of the large sample covariance of the joint frequency offset estimate performance has been already done in [4] by means of the derivation of the joint Cramer-Rao Bound. Hence,

$$\gamma_f = \frac{\sigma^2}{8\pi^2 \mathbf{h}_0^{\mathrm{H}} (\mathbf{W}_2 - \mathbf{W}_1 \mathbf{W}_0^{-1} \mathbf{W}_1) \mathbf{h}_0}. \quad (6)$$

In contrast, the expression of $\gamma_{\mathbf{h}}$ was not performed. After straightforward manipulations based on either the complete calculation of the joint Cramer-Rao bound or the direct derivations connected to related works [1], we show that

$$\gamma_{\mathbf{h}} = \sigma^2 \mathbf{W}_0^{-1} + 4\pi^2 \mathbf{W}_0^{-1} \mathbf{W}_1 \mathbf{h}_0 \gamma_f \mathbf{h}_0^{\mathrm{H}} \mathbf{W}_1 \mathbf{W}_0^{-1}.$$

This finally means that

$$\gamma_{\mathbf{h}} = \sigma^2 \mathbf{W}_0^{-1} + \frac{\sigma^2 \mathbf{W}_0^{-1} \mathbf{W}_1 \mathbf{h}_0 \mathbf{h}_0^{\mathrm{H}} \mathbf{W}_1 \mathbf{W}_0^{-1}}{2\mathbf{h}_0^{\mathrm{H}} (\mathbf{W}_2 - \mathbf{W}_1 \mathbf{W}_0^{-1} \mathbf{W}_1) \mathbf{h}_0}. \quad (7)$$

According to Equation (5), γ_f and $\gamma_{\mathbf{h}}$ are bounded. These terms thus correspond to the large sample covariances of the estimators. Finally the convergence rate of the frequency offset (resp. channel) estimate is equal to M^3 (resp. M).

One can notice that designing training sequence optimizing the joint Maximum-Likelihood cost function cannot be managed on account of the specific form of the expressions (6) and (7).

Directed estimation performance As the carrier frequency is assumed to be perfectly known, the Maximum-Likelihood filter estimator is well-known. Indeed, numerous analyses can be found in the literature. The result is

$$\begin{aligned} \gamma_{\mathbf{h}|f} &= \mathrm{val}\left(M \mathbf{E}[(\hat{\mathbf{h}}_{M|f} - \mathbf{h}_0)(\hat{\mathbf{h}}_{M|f} - \mathbf{h}_0)^{\mathrm{H}}] \right) \\ &= \sigma^2 \mathbf{W}_0^{-1}. \end{aligned}$$

Consequently, the directed filter estimator performance can be obtained by setting $\gamma_f = 0$ (which means that we perfectly know the frequency offset) in Equation (7) providing the joint filter estimator performance. Moreover we can also proved that $\gamma_{\mathbf{h}} \geq \gamma_{\mathbf{h}|f}$, i.e., the filter estimation error associated with the directed problem is always smaller than that one associated with the joint problem. Lastly, it is well-known that the optimal training sequence has to be a white pseudo-random sequence.

We now assume that the filter is known. Following the same approach as that developed in [1] leads to

$$\begin{aligned} \gamma_{f|\mathbf{h}} &= \mathrm{val}\left(M^3 \mathbb{E}[(\hat{f}_{M|\mathbf{h}} - f_0)(\hat{f}_{M|\mathbf{h}} - f_0)] \right) \\ &= \frac{\sigma^2}{8\pi^2 \mathbf{h}_0^{\mathrm{H}} \mathbf{W}_2 \mathbf{h}_0}. \end{aligned} \quad (8)$$

Once again one can easily check that $\gamma_f \geq \gamma_{f|\mathbf{h}}$.

Moreover, Equation (8) enables us to design an optimal training sequence, namely, a sequence minimizing the large sample covariance $\gamma_{f|\mathbf{h}}$. In fact, we get

$$\mathbf{h}_0^{\mathrm{H}} \mathbf{W}_2 \mathbf{h}_0 = \mathbf{v}_M^{\mathrm{H}} \mathbf{Q}_M \mathbf{v}_M$$

where

$$\mathbf{Q}_M = \frac{\mathcal{T}_M(\mathbf{h}_0)^{\mathrm{H}} \Delta_M^2 \mathcal{T}_M(\mathbf{h}_0)}{M^3}$$

with $\mathcal{T}_M(\mathbf{h}_0)$ a $M \times M$ square matrix of which the m^{th} column is defined as $[\mathbf{0}_{1,m-1}, h_0 \cdots h_{\min(L,M-m)}, \mathbf{0}_{1,\max(0,M-L-m)}]^{\mathrm{T}}$. The optimal training sequence of length M is thus provided by the eigenvector of square norm M associated with the largest eigenvalue of the known matrix $\mathbf{Q}_M$.

Finally, we notice that optimizing the covariance of the directed frequency offset and channel estimators leads to two different kinds of training sequence. This remark likely justifies the fact that one cannot succeed to finding an optimal training sequence for the joint estimation problem,

4 Numerical Comparisons

The filter $\mathbf{h}_0$ results from the convolution of a shaping filter (which is a square-root raised cosine with roll-off $\rho = 0.2$) with a multi-path channel (which is modeled by five paths with delays in the interval $[0, 3T_s]$). The noise $n(k)$ is assumed to be white. In each simulation, we average the result over 200 Monte-Carlo noise sequence trials. Finally, the sought frequency offset f_0 is equal to 0.075.

We now focus on the choice of the training sequences. $\{v_{n,1}\}$ denotes a white binary pseudo-random sequence. $\{v_{n,2}\}$ denotes the eigenvector of square norm M and length M associated with the largest eigenvalue of $\mathbf{Q}_M$. The training sequence $\{v_{n,1}\}$ (resp. $\{v_{n,2}\}$) is used for computing the directed channel (resp. frequency offset) estimate because such a sequence is optimum for directed estimation. The choice for the joint estimation problem is more difficult since there does not exist an optimal training sequence of such a problem. Nevertheless, as classically done, we use the training sequence $\{v_{n,1}\}$ for carrying out the joint frequency offset and channel estimates.

These choices of training sequences will enable us to evaluate the loss in performance due the joint presence of the frequency offset and filter estimation step, in comparison to each optimal partial step corresponding to the estimation of one of the parameters when the other is known. According to whether the gap is wide or not, we will show if we need to design sub-optimal training sequences for the joint estimation problem or not.

A theoretical frequency offset (resp. channel) estimation mean square error (MSE) is obtained by performing the appropriate large sample covariance divided by M^3 (resp. M). An empirical mean square error is obtained by averaging the square norm between the sought parameter and the estimate over the number of Monte-Carlo trials. In order to obtain the joint (resp. directed) frequency offset estimate, we proceed in two steps : at first, a *coarse* search step performing the criterion $J_M(f)$ (resp. $H_M(f)$) by means of a Fast Fourier Transform (FFT) of size M. Then, we compute the *fine* search step by means of a gradient algorithm initialized at the estimate provided by the coarse search step. Finally, the joint (resp. directed) filter estimate is obtained by plugging the frequency offset estimated value into Equation (2) (resp. by applying Equation (4)).

Figures 1 and 2 depict the theoretical and empirical mean square error versus the SNR for the frequency offset estimator and the channel estimator, respectively. We put $M = 32$.

We notice that the empirical curves are in good agreement with the theoretical asymptotic analysis except at low SNR. At low SNR, the so-called *outliers effect* appears for the frequency offset estimation problem ([5, 4]). We also observe that the failure of the frequency offset estimation caused by the outliers effect is not very prejudicial for the filter estimation step. The gap between optimal directed frequency estimate and the joint frequency offset estimate performances is quite significant but not redhibitory. Furthermore the performance of both filter estimators are very similar which means that the frequency offset estimation step does not really degrade the filter identification step Therefore, before seeking optimal or sub-optimal training sequence in order to decrease a little the non-dramatic gap, we guess that mitigating the outliers effect, namely, developing frequency offset estimator robust to low SNR, is a more relevant task for the future because it is the most important drawback of such a method.

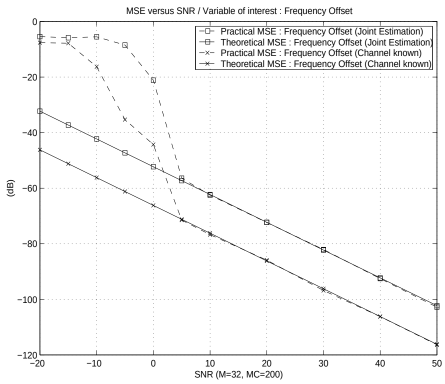

Figure 1: Frequency Offset MSE versus SNR

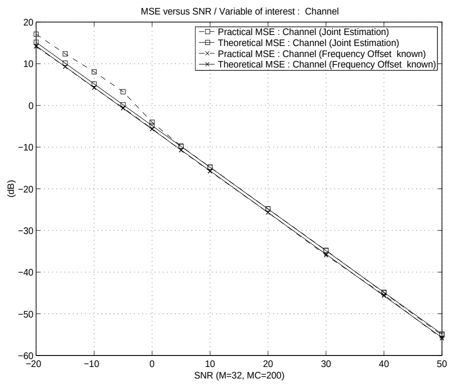

Figure 2: Channel MSE versus SNR

In Figures 3 and 4, we have plotted the theoretical and empirical mean square error for the frequency offset and channel versus M. The SNR is fixed to be 30 dB. We notice that the agreement between the theoretical and practical performance is good even for very small M. These curves confirm that the frequency offset and channel estimation mean square errors are proportional to $1/M^3$ and $1/M$ respectively.

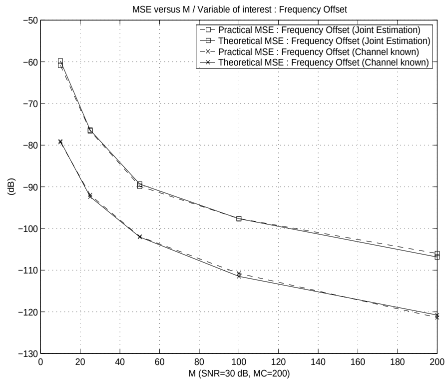

Figure 3: Frequency Offset MSE versus M

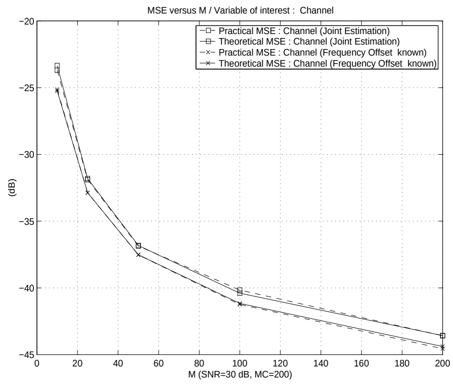

Figure 4: Channel MSE versus M

5 Conclusion

The gain in performance for the optimized directed estimators with respect the sub-optimal joint ones is not very important (especially for the channel). Therefore the main further task may be to counteract the outliers effect rather than to design optimal training sequence for the joint issue.

References

[1] Ph. Ciblat, Ph. Loubaton, E. Serpedin, and G.B. Giannakis. Performance of blind carrier offset estimation for non-circular transmissions through frequency-selective channels. *IEEE Trans. on Signal Processing*, 50(1):130–140, January 2002.

[2] S.A. Fechtel and H. Meyr. Improved frame synchronisation for spontaneous packet transmission over frequency-selective radio channels. In *PIMRC*, pages 353–357, 1994.

[3] M. Morelli and U. Mengali. Feedforward frequency estimation for PSK : a tutorial review. *European Transactions on Telecommunications*, 9(2):103–115, March 1998.

[4] M. Morelli and U. Mengali. Carrier frequency estimation for transmissions over selective channels. *IEEE Trans. on Communications*, 48(9):1580–1589, September 2000.

[5] D.C. Rife and R.R. Boorstyn. Single-tone parameter estimation from discrete-time observations. *IEEE Trans. on Information Theory*, 20(5):591–598, September 1974.