

Training Sequence Optimization for Joint Channel and Frequency Offset Estimation

Philippe Ciblat, Pascal Bianchi, and Mounir Ghogho

Abstract— We consider single-carrier and single-user transmissions over a frequency-selective channel. We address the problem of joint estimation of the channel and the carrier frequency offset using a known training sequence (TS). Since in general the TS that is optimum for CFO estimation is not optimum for channel estimation, here we design the TS to minimize the Mean-Square Error on the soft estimates of the data symbols obtained using a Wiener equalizer after CFO compensation. Our design results apply to the general scenarios of correlated and/or Ricean channel taps. Simulations measuring the bit error rate show an interesting gain when using the proposed training schemes in lieu of standard schemes.

I. INTRODUCTION

In wireless communications, the transmitted signal is usually affected by inter symbol interference due to the multipath channel and carrier frequency offset (CFO) caused by the Doppler effect or/and a local oscillator drift. Before applying a CFO correction and an equalizer, the channel impulse response (CIR) and the CFO have to be estimated. In most practical systems, this estimation task is carried out by transmitting a known training sequence (TS) prior to transmitting information-bearing data. It is therefore of interest to design the TS at the transmitter side so that accurate estimates of the unknown parameters can be obtained at the receiver side.

When the only unknown parameter is the CIR, it is well-known that the best TS (i.e. the TS that minimizes the Cramér-Rao bound (CRB)) is the pseudo-random white sequence [1], [2]. However, when both the CIR and CFO need to be estimated, the design of an optimal TS is still an open issue, in general. Suboptimum training designs were proposed in the literature using different criteria. In [3], [4], the authors used the worst-case asymptotic CRB, i.e., the large-sample CRB associated with the worst channel realization and showed that the white TS is then optimal. In [5], the TS was designed to render the exact (i.e. finite-sample) CRB of the CFO independent of the channel zeros. The developed TS design was shown to outperform the white TS design when the length of the TS is small/moderate. For a large TS, the white sequence is still optimal when considering the criterion in [5]. Here, we

only consider TS designs based on asymptotic performance metrics.

In the above approaches, the channel was considered deterministic and thus any likely correlation between the CIR components was not taken into account when designing the TS. An approach which considers random CIR was proposed in [6], where a channel-independent design criterion was obtained by statistically averaging the CRB of the CFO over the CIR realizations. However, the work in [6] only considered the case where the CIR components are independent and identically distributed (i.i.d.) and their magnitudes are Rayleigh distributed. In such a case, the best TS for CFO estimation was shown to be white [6]. In fact, in this case, the optimal TS for the joint CFO/CIR estimation is still white. Indeed, if the CIR components are i.i.d., the averaged channel frequency response is flat and thus the training power should be equally split between all the frequencies. However, the Rayleigh i.i.d. assumption in [6] may be restrictive in practice (see e.g. [9], [10], [11], [12].) A correlation between the CIR coefficients will render the channel statistically frequency-selective, in which case there is no reason to expect a white TS to be optimal. In such a context (i.e., correlated Rayleigh components of the impulse response), a first characterization of the TS minimizing the CRB of the CFO is available in [13]. Similarly, if the channel is Ricean, the deterministic part of the CIR should play a part in shaping the spectrum of the TS. Of course, in the absence of CFO, the TS optimizing the CRB associated with the CIR estimation is white regardless of the statistics of the channel, when considering asymptotic performance¹.

No training design for the joint CFO and channel estimation problem has been proposed for the case of correlated CIR taps and/or Ricean CIR in the literature. The goal of this paper is to fill this gap. The design difficulty in this case stems from the fact that the optimal TS designs for CFO and CIR estimations are different, unlike the Rayleigh i.i.d. case where the white TS was optimal for both estimations. Therefore, the individual estimation performance metrics are not sufficient to develop a single TS selection strategy for the joint estimation problem. This leads to the need of selecting a criterion that combines these metrics in accordance with their respective impact on the overall system performance. The criterion we propose to use is the Mean Square Error (MSE) between the transmitted and received symbols when the receiver consists of

Philippe Ciblat is with Département Comélec, ENST, Paris, France (philippe.ciblat@enst.fr).

Pascal Bianchi is with Département Télécom, Supélec, Gif-sur-Yvette, France (pascal.bianchi@supelec.fr).

Mounir Ghogho is with School of Electronics and Electrical Engineering, Leeds University, Leeds, United Kingdom (m.ghogho@leeds.ac.uk).

This work has been partly supported by the Royal Society.

Parts of this work have been presented at the Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Helsinki, Finland, June 2007.

¹When the length of the TS is small, a Bayesian approach could be used to estimate the CIR at the receiver, and in this case, the optimal TS is not white. However, as the length of the TS increases, the spectrum of the optimal TS becomes more and more flat

a frequency compensation and a Wiener linear equalizer. This MSE is obtained by statistically averaging the conditional (on the channel) MSE over the channel realizations. This enable us to find an "optimal" TS which is relevant to both channel and CFO estimations.

The paper is organized as follows: in Section II, we introduce the signal model, the channel statistics model, the receiver model as well as the criterion that we wish to optimize. In Section III, we express the proposed criterion in closed-form. Section IV is devoted to the optimization of the criterion derived in Section III. In Section V, we propose numerical illustrations in order to observe the gain in performance provided by our approach.

Notation: Overline $\bar{\cdot}$, superscripts T and H denote complex conjugate, transposition, and conjugate transposition, respectively. $\mathbb{E}[\mathbf{B}|\mathbf{A}]$ and $\mathbb{E}_{\mathbf{A}}[\mathbf{B}]$ denotes the conditional expectation of \mathbf{B} with respect to \mathbf{A} and the average expectation of \mathbf{B} over \mathbf{A} , respectively. $\Re[\mathbf{A}]$ and $\Im[\mathbf{A}]$ denote the real and imaginary parts of \mathbf{A} , respectively. Finally, $\|\mathbf{a}\|$ and $\text{Tr}(\mathbf{A})$ denote the L2-norm of vector \mathbf{a} and the trace of matrix \mathbf{A} , respectively. For any vector $\mathbf{p} = [p(-L_1), \dots, p(L_2)]^{\text{T}}$ of length $(L_1 + L_2 + 1)$, we denote its Fourier transform by $P(\nu) = \sum_{l=-L_1}^{L_2} p(l)e^{-2i\pi\nu n}$.

II. PROBLEM STATEMENT

A. Signal Model

We focus on single-carrier and single-user communications. We assume that the transmission consists of a training mode during which a TS $\mathbf{t} = [t(0), t(1), \dots, t(N_T - 1)]^{\text{T}}$ of length N_T , which is known to the receiver, is transmitted, followed by a data mode during which a data sequence $\mathbf{d} = [d(0), d(1), \dots, d(N_D - 1)]^{\text{T}}$ of length N_D is transmitted. Notice that training and data samples may be complex-valued. The discrete-time baseband received signal, $y(n)$, has the following form:

$$y(n) = e^{2i\pi f n} \sum_{l=0}^{L-1} h(l)s(n-l) + w(n), \quad (1)$$

where f denotes the (normalized) CFO, $\{h(0), \dots, h(L-1)\}$ represent the channel coefficients, $s(n)$ is either $t(n)$ or $d(n)$ depending on the mode of transmission at time instant n , and $w(n)$ denotes a circularly symmetric white Gaussian noise with variance $\sigma^2 = \mathbb{E}[|w(n)|^2]$. Let $\mathbf{h} = [h(0) \dots h(L-1)]^{\text{T}}$ be the unknown channel vector. For the sake of simplicity, \mathbf{d} is assumed to be an i.i.d. sequence with variance $\sigma_d^2 = \mathbb{E}[|d(n)|^2]$. Note however that our results can be generalized to the case where \mathbf{d} is a colored sequence.

In order to obtain an expression for the proposed criterion that is simple and tractable to optimize, we will consider the system performance in the "asymptotic" regime, *i.e.*, we assume that *both* N_T , the size of the TS, and N_D , the size of the data sequence, tend to infinity, while the ratio N_D/N_T converges to a constant, *i.e.*,

$$\lim_{N_T \rightarrow \infty} \frac{N_D}{N_T} = \alpha$$

where α is a constant depending on the system of interest. We also assume that the length of the channel filter L and the length of the equalizer filter L_g tend to infinity, at the same rate. Note that we will first consider that the sizes of the sequences (N_D and N_T) are large, and *then* assume the filters lengths (L and L_g) are large, which means that

$$\lim_{N_T \rightarrow \infty} \frac{L \text{ or } L_g}{N_T} = 0$$

or equivalently $L, L_g \ll N_D, N_T$. In Proof of Lemma 2 reported in Appendix B, we actually need a little bit more restrictive assumption. We will in fact assume that $(L \text{ or } L_g)/\sqrt{N_T}$ is bounded when N_T becomes large. Although for the analysis we assume large values for L and L_g , we will consider realistic values for these parameters in the simulation section.

B. Channel Model

The channel is assumed to be Rice distributed *i.e.*,

$$\mathbf{h} = \sqrt{\frac{K}{K+1}} \mathbf{h}_d + \sqrt{\frac{1}{K+1}} \mathbf{h}_r \quad (2)$$

where \mathbf{h}_d is a deterministic vector normalized in such a way that $\|\mathbf{h}_d\|^2 = 1$, and \mathbf{h}_r is a complex circular Gaussian random vector with zero mean and covariance matrix $\mathbf{\Sigma} = \mathbb{E}[\mathbf{h}_r \mathbf{h}_r^{\text{H}}]$, normalized in such a way that $\text{Tr}(\mathbf{\Sigma}) = 1$. Coefficient K is the so-called Ricean factor. In this paper, we assume that K , $\mathbf{\Sigma}$, and \mathbf{h}_d are known at both the transmitter and the receiver sides.

This channel model can be justified in the following two ways. Firstly, in most wireless applications, it is usual to consider that the channel can be decomposed into two parts: the first part refers to the Line of Sight (LOS), which corresponds to the first term of the righthand side of (2), and the second part is associated with the non-line of sight (NLOS) components of the channel, which corresponds to the second term of the righthand side of (2). The knowledge of K , $\mathbf{\Sigma}$, and \mathbf{h}_d at both sides is motivated by the fact that the coherence times corresponding to K , $\mathbf{\Sigma}$, and \mathbf{h}_d are much larger than that of \mathbf{h}_r [9], [10], [11], [12]. Secondly, even when the Rice model does not hold true, it is often possible to decompose \mathbf{h} as depicted in Eq. (2). For instance, consider the case where the receiver estimates the CIR and the CFO and feeds back these estimates to the transmitter. The latter therefore has a partial knowledge of the CIR, which can be described as in Eq. (2) where the random part represents the uncertainties due to estimation/feedback errors and time-variations of the channel. Notice that the estimation errors are Gaussian asymptotically, *i.e.*, when the length of the TS is large enough, which justifies the Rice model.

C. Receiver model

First, the receiver compensates for the CFO using the estimated value \hat{f} of f . It then generates the signal $y_c(n) = e^{-2i\pi \hat{f} n} y(n)$. For each $n \geq N_T + L$, we have that

$$y_c(n) = e^{-2i\pi \Delta f n} \sum_{l=0}^{L-1} h(l)d(n - N_T - l) + w(n). \quad (3)$$

where $\Delta f = \hat{f} - f$. Note that CFO compensation does not change the statistics of the white noise. Thus, to simplify notations, the same variable is used to denote the noise before and after CFO compensation. The receiver then compensates for the channel distortion using a standard Wiener filter with coefficients $\hat{\mathbf{g}} = [\hat{g}(-L_g), \dots, \hat{g}(L_g)]$. The linear equalizer $\hat{\mathbf{g}}$ should be calculated using the estimated value $\hat{\mathbf{h}}$ of \mathbf{h} . The output equalizer $z(n)$ is defined by

$$z(n) = \sum_{k=-L_g}^{L_g} \hat{g}(k)y_c(n-k).$$

Finally, a detector is used on the equalizer output in order to recover the transmitted data symbols.

When the receiver has statistical knowledge on the parameters to be estimated, one can use the bayesian approach to estimate the parameters of interest. Nevertheless, this approach has the drawback of complicating the performance analysis. Furthermore, when the SNR is large and/or when the number of available training samples, N_T , is large, which is the case in the asymptotic regime, the best Bayesian estimator and the best deterministic estimator (Maximum Likelihood - ML) lead to the same performance since the a priori contribution is bounded whereas the likelihood grows with the number of available observations. Therefore, in this paper, we concentrate on the standard deterministic approach based on the ML estimator². As a consequence, it is reasonable to consider that the estimation errors on the CIR and CFO are well described by the deterministic CRB in the asymptotic regime.

D. The proposed criterion

A natural approach would be to design a training strategy which leads to the minimum bit error rate at the detector output. Unfortunately, such a criterion is very difficult to express as a simple function of the training strategy. Here, we propose to minimize the MSE at the equalizer output. In the sequel, we define

$$\text{MSE}_n(\mathbf{t}) = \mathbb{E} [|z(n) - d(n - N_T)|^2 | \mathbf{t}].$$

It is worth noting that $z(n)$ is a non stationary sequence due to the presence of factor $e^{-2i\pi\Delta f n}$. Therefore, the above expression of the MSE depends on index n . It is obviously impractical to minimize the MSE for all possible values of n . Here, we propose to minimize the average MSE, i.e. MSE averaged over all data symbols:

$$\text{MSE}(\mathbf{t}) = \frac{1}{N_D} \sum_{n=N_T}^{N_T+N_D-1} \text{MSE}_n(\mathbf{t}). \quad (4)$$

The above criterion depends on the TS, \mathbf{t} , via the estimation errors on parameters \mathbf{h} and f . The objective of the next subsection is to express $\text{MSE}(\mathbf{t})$ in closed-form.

Our TS design strategy basically consists of searching for the TS that minimizes the MSE on the data symbols at the output of the Wiener filter. Our TS will be ‘optimal’ when the receiver is as described above.

²Notice that ML estimator can be easily carried out by means of correlation and periodogram [1], [2]

III. CRITERION DERIVATIONS

To derive the MSE in eq. (4), we firstly focus on $\text{MSE}(\mathbf{t}|\mathbf{h})$ which is the MSE given a realization of the channel and which satisfies

$$\text{MSE}(\mathbf{t}) = \mathbb{E}_{\mathbf{h}}[\text{MSE}(\mathbf{t}|\mathbf{h})].$$

The main steps of the derivations are as follows

- In Subsection III-A, we provide a simple expression for $\text{MSE}(\mathbf{t}|\mathbf{h})$ as a function of the estimation errors on parameters \mathbf{h} and f . Results are drawn in Lemma 2.
- In Subsection III-B, we relate the estimation errors to the training strategy. Results are given in eq. (22).
- In Subsection III-C, we finally average $\text{MSE}(\mathbf{t}|\mathbf{h})$ over the channel using our statistical model in eq. (2). Results are given in eqs. (27) and (32).

A. $\text{MSE}(\mathbf{t}|\mathbf{h})$ as a function of the estimation error

Let $\Delta \mathbf{h} = \hat{\mathbf{h}} - \mathbf{h}$ and $\Delta \mathbf{g} = \hat{\mathbf{g}} - \mathbf{g}$ where \mathbf{g} denotes the Wiener filter associated with \mathbf{h} . We recall that $\hat{\mathbf{g}}$ is the Wiener filter based on the estimate filter $\hat{\mathbf{h}}$.

After straightforward but tedious algebraic manipulations reported in Appendix A, we obtain the following lemma.

Lemma 1: Let $\text{MSE}(\mathbf{t}|\mathbf{h}, \hat{\mathbf{h}}, \hat{f})$ denote the MSE for a given realization of the channel, a realization of the channel estimate and a realization of the CFO estimate. We get

$$\text{MSE}(\mathbf{t}|\mathbf{h}) = \mathbb{E}_{\hat{\mathbf{h}}, \hat{f}}[\text{MSE}(\mathbf{t}|\mathbf{h}, \hat{\mathbf{h}}, \hat{f})]$$

where

$$\begin{aligned} \text{MSE}(\mathbf{t}|\mathbf{h}, \hat{\mathbf{h}}, \hat{f}) &= \sigma_d^2 + \sigma^2 \int_0^1 |(G + \Delta G)(\nu)|^2 d\nu \\ &+ \sigma_d^2 \int_0^1 |H(\nu)|^2 |(G + \Delta G)(\nu - \Delta f)|^2 d\nu \\ &- 2\sigma_d^2 \Re \left[\mathcal{S}(\Delta f) \int_0^1 (G + \Delta G)(\nu - \Delta f) H(\nu) d\nu \right] \end{aligned}$$

with

$$\mathcal{S}(\nu) = \frac{1}{N_D} \sum_{n=N_T}^{N_T+N_D-1} e^{-2i\pi\nu n}.$$

In order to obtain a simple link between $\text{MSE}(\mathbf{t}|\mathbf{h})$ and the estimation errors, we consider the ‘asymptotic’ regime described in Subsection II-A. We recall (cf. [1], [2]) that the MSE on channel estimation is of order $1/N_T$ while the MSE on CFO estimation is of order $1/N_T^3$. Consequently, function $\mathcal{S}(f)$ can be decomposed as follows

$$\begin{aligned} \mathcal{S}(\Delta f) &= 1 - i\pi(2 + \alpha)N_T\Delta f \\ &- 2\pi^2(1 + \alpha + \alpha^2/3)N_T^2(\Delta f)^2 + o_p(1/N_T) \end{aligned} \quad (5)$$

where $o_p(1/N_T)$ refer to terms that are negligible compared to $1/N_T$ in probability. Based on the above decomposition, we obtain the following lemma whose proof is given in Appendix B.

Lemma 2: Assuming the asymptotic regime, we have that

$$\text{MSE}(\mathbf{t}|\mathbf{h}) = e_0 + e_1 + e_2 + e_3 + o_p(1/N_T)$$

where

$$e_0 = \sigma_d^2 + \int_0^1 (\sigma_d^2 |H(\nu)|^2 + \sigma^2) |G(\nu)|^2 d\nu - 2\sigma_d^2 \Re \left[\int_0^1 H(\nu) G(\nu) d\nu \right] \quad (6)$$

$$e_1 = \int_0^1 \frac{(\sigma_d^8 |H(\nu)|^4 + \sigma_d^4 \sigma^4)}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^3} \gamma_{h,h}(\nu) d\nu - 2\Re \left[\int_0^1 \frac{\sigma_d^6 \sigma^2 \overline{H(\nu)}^2 \tilde{\gamma}_{h,h}(\nu)}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^3} d\nu \right] \quad (7)$$

$$e_2 = 2\sigma_d^2 (2 + \alpha) \pi N_T \Im \left[\int_0^1 H(\nu) \frac{\sigma_d^2 \sigma^2 \overline{\gamma_{h,f}(\nu)}}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^2} - h(\nu) \frac{\sigma_d^4 \overline{H(\nu)}^2 \gamma_{h,f}(\nu)}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^2} d\nu \right] \quad (8)$$

$$e_3 = 4\sigma_d^2 \pi^2 (1 + \alpha + \alpha^2/3) N_T^2 \times \gamma_{f,f} \Re \left[\int_0^1 G(\nu) H(\nu) d\nu \right] \quad (9)$$

with

$$\gamma_{h,h}(\nu) = \mathbb{E}[|\Delta H(\nu)|^2] \quad (10)$$

$$\tilde{\gamma}_{h,h}(\nu) = \mathbb{E}[\Delta H(\nu)^2] \quad (11)$$

$$\gamma_{h,f}(\nu) = \mathbb{E}[\Delta H(\nu) \Delta f] \quad (12)$$

$$\gamma_{f,f} = \mathbb{E}[(\Delta f)^2]. \quad (13)$$

Notice that the term e_0 represents the error at the output of the Wiener filter when \mathbf{h} and f are perfectly known. Error e_1 and e_3 are the extra term associated with the mis-estimation of \mathbf{h} and of f respectively. Finally e_2 is an extra error caused by the mis-estimation of both \mathbf{h} and f .

B. MSE($\mathbf{t}|\mathbf{h}$) as a function of the TS statistics

In the sequel, we express $\gamma_{h,h}(\nu)$, $\tilde{\gamma}_{h,h}(\nu)$, $\gamma_{h,f}(\nu)$ and $\gamma_{f,f}(\nu)$ as functions of the TS statistics. Indeed, since we focus on the asymptotic regime, the design of the TS is equivalent to the design of its statistics, as it is justified by the CRB expressions below. Therefore, we treat the TS as a realization of a zero-mean stationary random sequence, whose statistics, namely its covariance matrix, will be designed to minimize the total averaged MSE defined above.

When N_T is large, it is known that the asymptotic covariance of the ML channel coefficients and CFO estimates coincides with the CRB, which is reported below (cf. [3])

$$\mathbb{E}[\Delta \mathbf{h} \Delta \mathbf{h}^H] = \frac{\sigma^2}{N_T} \left(\mathbf{R}_t^{-1} + \frac{3}{2} \frac{\mathbf{h} \mathbf{h}^H}{\mathbf{h}^H \mathbf{R}_t \mathbf{h}} \right) \quad (14)$$

$$\mathbb{E}[\Delta \mathbf{h} \Delta \mathbf{h}^T] = -\frac{3\sigma^2}{2N_T} \frac{\mathbf{h} \mathbf{h}^T}{\mathbf{h}^H \mathbf{R}_t \mathbf{h}} \quad (15)$$

$$\mathbb{E}[\Delta \mathbf{h} \Delta f] = -i \frac{3\sigma^2}{2\pi N_T^2} \frac{\mathbf{h}}{\mathbf{h}^H \mathbf{R}_t \mathbf{h}}$$

$$\mathbb{E}[(\Delta f)^2] = \frac{3\sigma^2}{2\pi^2 N_T^3} \frac{1}{\mathbf{h}^H \mathbf{R}_t \mathbf{h}} \quad (16)$$

where \mathbf{R}_t is the L -dimensional covariance matrix defined by the entries $\{r(k-l)\}_{k,l=0,\dots,L-1}$ with $r(k-l) = \mathbb{E}[t(n+k)t(n+l)]$. When the CFO is known and does not need to be

estimated, the second term in the RHS of Eq. (14) vanishes. One can remark that the estimation error only depends on the second order statistics of the TS and not on each realization of the TS. If we assume that the sequence $\{r(k); k=0, \pm 1, \dots\}$ is absolutely summable, then we can define the following spectrum associated with the TS

$$S_{tt}(\nu) = \sum_{k \in \mathbb{Z}} r(k) e^{-2i\pi k \nu}$$

It is easy to check that

$$r(k) = \int_0^1 S_{tt}(\nu) e^{2i\pi k \nu} d\nu$$

This implies that

$$\mathbf{h}^H \mathbf{R}_t \mathbf{h} = \int_0^1 S_{tt}(\nu) |H(\nu)|^2 d\nu$$

Thanks to Eq. (11), we have

$$\tilde{\gamma}_{h,h}(\nu) = \mathbf{d}_L(\nu) \mathbb{E}[\Delta \mathbf{h} \Delta \mathbf{h}^T] \mathbf{d}_L(\nu)^T$$

where $\mathbf{d}_L(\nu) = [1, \dots, e^{-2i\pi(L-1)\nu}]$. Using Eq. (15) leads to

$$\tilde{\gamma}_{h,h}(\nu) = -\frac{3\sigma^2}{2N_T} \frac{(H(\nu))^2}{\int_0^1 S_{tt}(\nu) |H(u)|^2 du} \quad (17)$$

Similar algebraic manipulations can be done to obtain

$$\gamma_{h,f}(\nu) = -i \frac{3\sigma^2}{2\pi N_T^2} \frac{H(\nu)}{\int_0^1 S_{tt}(\nu) |H(u)|^2 du} \quad (18)$$

$$\gamma_{f,f} = \frac{3\sigma^2}{2\pi^2 N_T^3} \frac{1}{\int_0^1 S_{tt}(\nu) |H(u)|^2 du}. \quad (19)$$

Deriving a closed-form expression for $\gamma_{h,h}(\nu)$ is a complicated task since we have to handle matrix \mathbf{R}_t^{-1} . More precisely, $\gamma_{h,h}(\nu)$ depends on $\mathbf{d}_L(\nu) \mathbf{R}_t^{-1} \mathbf{d}_L(\nu)^H$. In order to express $\gamma_{h,h}(\nu)$ as a simple function of $q(\nu)$, we notice that when the channel length L is large enough, \mathbf{R}_t becomes a large Toeplitz matrix, which implies that its inverse \mathbf{R}_t^{-1} can be well approximated by the circulant matrix (cf. [7], [14]) whose first row is

$$\left[\int_0^1 \frac{1}{S_{tt}(\nu)} e^{2i\pi k \nu} d\nu \right]_{k=0,\dots,L-1} \quad (20)$$

Based on this approximation, we obtain

$$\gamma_{h,h}(\nu) = \frac{\sigma^2}{N_T} \left(\frac{L}{S_{tt}(\nu)} + \frac{3}{2} \frac{|H(\nu)|^2}{\int_0^1 S_{tt}(\nu) |H(u)|^2 du} \right)$$

When both N_T and L are large, putting Eqs. (17), (18), (19), and (20) into Eqs. (6), (7), (8), and (9) leads to the following expression for MSE($\mathbf{t}|\mathbf{h}$)

$$\text{MSE}(\mathbf{t}|\mathbf{h}) = \text{MSE}_{\text{known}}(\mathbf{h}) + \text{MSE}_{\text{unknown}}(\mathbf{t}|\mathbf{h}) \quad (21)$$

where $\text{MSE}_{\text{known}}(\mathbf{h})$ represents the MSE that one would have observed if parameter estimation was perfect, and is given by

$$\text{MSE}_{\text{known}}(\mathbf{h}) = \int_0^1 \frac{\sigma_d^2 \sigma^2}{\sigma_d^2 |H(\nu)|^2 + \sigma^2} d\nu$$

and where the ‘excess MSE’, $\text{MSE}_{\text{unknown}}(\mathbf{t}|\mathbf{h})$, due to the mis-estimation of CIR and CFO, has the following form

$$\text{MSE}_{\text{unknown}}(\mathbf{t}|\mathbf{h}) = \frac{\sigma_d^2 \sigma^2}{N_T} \left(\int_0^1 \frac{c_1(\nu|\mathbf{h})}{S_{tt}(\nu)} d\nu + \beta \frac{1}{\int_0^1 c_2(\nu|\mathbf{h}) S_{tt}(\nu) d\nu} \right) \quad (22)$$

with $\beta = 27/2 + 9\alpha + 2\alpha^2$ and

$$c_1(\nu|\mathbf{h}) = L \sigma_d^2 \frac{\sigma_d^4 |H(\nu)|^4 + \sigma^4}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^3} \quad (23)$$

$$c_2(\nu|\mathbf{h}) = \frac{|H(\nu)|^2}{\int_0^1 \frac{\sigma_d^2 |H(u)|^2}{\sigma_d^2 |H(u)|^2 + \sigma^2} du}. \quad (24)$$

In Eq. (22), notice that the MSE depends on the color of the TS. Now we would like to find the color that minimizes the MSE but independently of the channel realization \mathbf{h} . Therefore the last step of our criterion derivations consists of averaging the MSE in eq. (21) over \mathbf{h} .

C. Averaging MSE($\mathbf{t}|\mathbf{h}$) over \mathbf{h}

Under the assumption that N_T and L are large, the averaged MSE can, using the above equations, be expressed as

$$\text{MSE}(\mathbf{t}) = \text{MSE}_{\text{known}} + \frac{\sigma_d^2 \sigma^2}{N_T} J(S_{tt}) \quad (25)$$

where $\text{MSE}_{\text{known}} = \mathbb{E}_{\mathbf{h}}[\text{MSE}_{\text{known}}(\mathbf{h})]$ is a constant representing the average MSE in the absence of mis-estimation of the channel and CFO, and where the ‘excess MSE’ is proportional to $J(S_{tt})$, which is given by

$$J(S_{tt}) = J_h(S_{tt}) + J_f(S_{tt}) \quad (26)$$

where

$$J_h(S_{tt}) = \int_0^1 \frac{c_1(\nu)}{S_{tt}(\nu)} d\nu, \quad (27)$$

$$J_f(S_{tt}) = \mathbb{E}_{\mathbf{h}} \left[\frac{1}{\int_0^1 c_2(\nu|\mathbf{h}) S_{tt}(\nu) d\nu} \right], \quad (28)$$

with $c_1(\nu) = \mathbb{E}_{\mathbf{h}}[c_1(\nu|\mathbf{h})]$.

The function $S_{tt} \mapsto J(S_{tt})$ is the proposed criterion that we would like to minimize in order to select a relevant power spectrum $S_{tt}(\nu)$. Notice that $J(S_{tt})$ is a linear combination of two terms. The first term, $J_h(S_{tt})$, is associated with the channel estimation error. Nonetheless, this term is not directly related to the CRB for the CIR. Indeed, the asymptotic CRB for the CIR (when CFO is known) is equal to the integral of the inverse of the power spectrum of the TS (cf. Eqs. (14) and (20)). It is the function $c_1(\nu)$ that prevents us from having a direct relationship between the CRB for the CIR and the impact of channel estimation error on the MSE. The second term, $J_f(S_{tt})$, is associated with the CFO estimation error. Once again, $J_f(S_{tt})$ is not proportional to the averaged CRB associated with CFO estimation due to the presence of $c_2(\nu|\mathbf{h})$. Nevertheless, at high SNR, one can remark that $c_2(\nu|\mathbf{h})$ tends to $|H(\nu)|^2$ which implies that $J_f(S_{tt})$ becomes proportional to the CRB for the CFO described in Eq. (19).

We now would like to derive a closed-form expression for the expectations in $J_h(S_{tt})$ and $J_f(S_{tt})$. Let us start with $J_h(S_{tt})$. Thanks to Eq. (2), it is easy to check that $H(\nu)$, for any ν , is Gaussian distributed with mean $\sqrt{K/(K+1)}H_d(\nu)$ and variance

$$\sigma_H^2(\nu) = \mathbb{E}[|H_r(\nu)|^2]/(K+1). \quad (29)$$

Besides, $H(\nu)$ minus its mean is circularly symmetric. Therefore the probability density function of $|H(\nu)|^2$ takes the following form

$$p_{|H(\nu)|^2}(x) = \frac{K+1}{\mathbb{E}[|H_r(\nu)|^2]} e^{-\frac{(K+1)x + K|H_d(\nu)|^2}{\mathbb{E}[|H_r(\nu)|^2]}} \times I_0 \left(\sqrt{2K \frac{|H_d(\nu)|^2}{\mathbb{E}[|H_r(\nu)|^2]}} x \right) \mathbf{1}_{x \geq 0}$$

where $I_0(\cdot)$ is the modified Bessel function of first kind and where $\mathbf{1}_{x \geq 0}$ is equal to 1 when x is positive and equal to 0 otherwise. Consequently, we get

$$c_1(\nu) = \frac{L \sigma_d^2 (K+1)}{\mathbb{E}[|H_r(\nu)|^2]} e^{-\frac{K|H_d(\nu)|^2}{\mathbb{E}[|H_r(\nu)|^2]}} \times \int_0^\infty \frac{\sigma_d^4 x^2 + \sigma^4}{(\sigma_d^2 x^2 + \sigma^2)^3} e^{-\frac{(K+1)x}{\mathbb{E}[|H_r(\nu)|^2]}} \times I_0 \left(\sqrt{2K \frac{|H_d(\nu)|^2}{\mathbb{E}[|H_r(\nu)|^2]}} x \right) dx \quad (30)$$

Although this integral cannot be expressed in closed-form, a numerical computation can be easily carried out. Once $c_1(\nu)$ is evaluated, another numerical integration is carried out to evaluate $J_h(S_{tt})$.

For $J_f(S_{tt})$, one can remark that the expression (24) can be greatly simplified when the high SNR regime is considered, as follows

$$J_f(S_{tt}) \stackrel{\text{high SNR}}{\approx} \mathbb{E}_{\mathbf{h}} \left[\frac{1}{\int_0^1 |H(\nu)|^2 S_{tt}(\nu) d\nu} \right]$$

Next, we will derive a closed-form expression for the high-SNR approximation of $J_f(S_{tt})$, given the above equation. Since Eqs. (16) and (19) represent the same term, we deduce that

$$J_f(S_{tt}) = \mathbb{E}_{\mathbf{h}} \left[\frac{1}{\mathbf{h}^H \mathbf{R}_t \mathbf{h}} \right] \quad (31)$$

The next theorem is proved in Appendix C.

Theorem 1: Let $J_f(S_{tt})$ given by Eq. (31). If L is large, we have that

$$J_f(S_{tt}) = (K+1) \frac{\ell_1}{\ell_1^2 - \ell_2} \quad (32)$$

where

$$\begin{aligned} \ell_1 &= \text{Tr}(\mathbf{R}_t(\boldsymbol{\Sigma} + K\mathbf{h}_d\mathbf{h}_d^H)) \\ \ell_2 &= \text{Tr}(\mathbf{R}_t\boldsymbol{\Sigma}\mathbf{R}_t(\boldsymbol{\Sigma} + 2K\mathbf{h}_d\mathbf{h}_d^H)). \end{aligned}$$

Using the above expressions for $J_h(S_{tt})$ and $J_f(S_{tt})$, we next address the issue of training optimization.

IV. CRITERION OPTIMIZATION

As mentioned in the previous section, our training design is based on the minimization of $J(S_{tt})$, which is a weighted sum of $J_h(S_{tt})$ and $J_f(S_{tt})$. This optimization problem is, in general, intractable in closed-form, due to the highly nonlinear relationship between $J(S_{tt})$ and the power spectrum of the TS. Nevertheless, we show in this section that this optimization problem is convex, which implies that any gradient-type descent algorithm will provide the global optimum solution. Further, in the case of high Rice factor ($K \rightarrow \infty$), we are able to obtain closed-form solutions using the Lagrange multiplier-based method.

A. General case

First, we will simplify $J_f(S_{tt})$ given in Eq. (32). We recall that $J_f(S_{tt})$ can be written as follows

$$J_f(S_{tt}) = (K + 1) \frac{\text{Tr}(\mathbf{R}_t \tilde{\mathbf{A}})}{(\text{Tr}(\mathbf{R}_t \tilde{\mathbf{A}}))^2 - \text{Tr}(\mathbf{R}_t \tilde{\mathbf{B}} \mathbf{R}_t \tilde{\mathbf{C}})}$$

with

$$\begin{aligned} \tilde{\mathbf{A}} &= \mathbf{\Sigma} + K \mathbf{h}_d \mathbf{h}_d^H \\ \tilde{\mathbf{B}} &= \mathbf{\Sigma} \\ \tilde{\mathbf{C}} &= \mathbf{\Sigma} + 2K \mathbf{h}_d \mathbf{h}_d^H \end{aligned}$$

This expression is an extension of that reported in [6] for the iid Rayleigh channel case (obtained when $K = 0$ and $\mathbf{\Sigma} = (1/L) \mathbf{I}_L$).

To solve our optimization problem, i.e. minimization of $J(S_{tt})$ in Eq. (26), we again assume that the channel length L is large. We recall that this assumption has already been made in the previous section to obtain closed-form expression of $J(S_{tt})$. Hence, the $L \times L$ Toeplitz matrix \mathbf{R}_t can be almost diagonalized in the Fourier basis for large L . Consequently, we get

$$\mathbf{R}_t = \mathbf{F} \mathbf{\Lambda}_{tt} \mathbf{F}^H$$

where \mathbf{F} is an $L \times L$ Fourier matrix and $\mathbf{\Lambda}_{tt}$ is a diagonal matrix whose ℓ^{th} diagonal element is $\lambda_\ell = S_{tt}(\ell/L)$. As a consequence, $J_f(S_{tt})$ is completely captured by $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_L]$ and can be rewritten as

$$J_f(\boldsymbol{\lambda}) = (K + 1) \frac{\sum_{l=1}^L \lambda_l a_{ll}}{(\sum_{l=1}^L \lambda_l a_{ll})^2 - \sum_{k,l=1}^L \lambda_k \lambda_l b_{kl} c_{lk}} \quad (33)$$

where a_{kl} , b_{kl} , and c_{kl} are the components of the k^{th} row and l^{th} column of matrices $\mathbf{A} = \mathbf{F}^H \tilde{\mathbf{A}} \mathbf{F}$, $\mathbf{B} = \mathbf{F}^H \tilde{\mathbf{B}} \mathbf{F}$, and $\mathbf{C} = \mathbf{F}^H \tilde{\mathbf{C}} \mathbf{F}$, respectively.

As to $J_h(S_{tt})$, using the fact that for large L , the integral in Eq. (27) can be approximated by a Riemann sum with a grid equal to $1/L$, we obtain

$$J_h(\boldsymbol{\lambda}) = \sum_{l=1}^L \frac{s_l}{\lambda_l} \quad (34)$$

with $s_l = c_1(l/L)/L$.

In the next theorem, we state that the criterion (26) is convex with respect to $\boldsymbol{\lambda}$ when J_h and J_f satisfy Eqs. (34) and (33) respectively. The proof is reported in Appendix D.

Theorem 2: When L is large, minimizing $J(S_{tt})$ is equivalent to minimizing $J_g(\boldsymbol{\lambda})$ defined as follows

$$J_g(\boldsymbol{\lambda}) = \sum_{l=1}^L \frac{s_l}{\lambda_l} + \beta \frac{\sum_{l=1}^L \lambda_l a_{ll}}{(\sum_{l=1}^L \lambda_l a_{ll})^2 - \sum_{k,l=1}^L \lambda_k \lambda_l b_{kl} c_{lk}} \quad (35)$$

The above function is convex with respect to $\boldsymbol{\lambda}$ if $\lambda_l \geq 0$ for all $l \in \{1, \dots, L\}$.

Since the λ_l 's are positive by definition, $J_g(\boldsymbol{\lambda})$ is convex. Hence, our optimization problem is convex since the function to be minimized is convex and the constraints

$$\text{Tr}(\mathbf{R}_t) = \sum_{l=1}^L \lambda_l = LP \quad \text{and} \quad \lambda_l \geq 0, \quad \forall l \quad (36)$$

are also convex. To obtain numerical values for the optimal $\boldsymbol{\lambda}$, we can use a standard convex optimization toolbox.

B. Large K case

When K , the Rice factor, is large, we can replace $H(\nu)$ in (26) with the transfer function $H_d(\nu)$. This approximation is supported by the fact that for any continuous mapping \mathcal{F} of the channel coefficients \mathbf{h} , $\mathbb{E}[\mathcal{F}(\mathbf{h})]$ converges to $\mathcal{F}(\mathbf{h}_d)$ as the Ricean factor K tends to infinity.

The simplified TS design criterion is then defined by

$$J_d(S_{tt}) = \int_0^1 \frac{c_1^d(\nu)}{S_{tt}(\nu)} d\nu + \beta \frac{1}{\int_0^1 c_2^d(\nu) S_{tt}(\nu) d\nu}. \quad (37)$$

where $c_1^d(\nu)$ is obtained by removing the mathematical expectation and by replacing $H(\nu)$ with $H_d(\nu)$ in $c_1(\nu)$, and where $c_2^d(\nu)$ is obtained by replacing $H(\nu)$ with $H_d(\nu)$ in Eq. (24). Of course, the above criterion $J_d(S_{tt})$ is likely to be a valid approximation of $J(S_{tt})$ provided that the Ricean factor K is large enough. The training strategy proposed is therefore appropriate when K is large. However, we will show in the simulation part that, even for moderate and realistic values of K , the proposed training strategy based on the minimization of $J_d(S_{tt})$ outperforms classical training strategies.

Once again, we replace the integrals with their Riemann sums. Then

$$J_d(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{n=1}^N \frac{\alpha_n}{\lambda_n} + \beta \frac{1}{\frac{1}{N} \sum_{n=1}^N \beta_n \lambda_n} \quad (38)$$

with $\alpha_n = c_1^d(n/N)$, $\beta_n = c_2^d(n/N)$, and N being an integer.

In the following theorem, we show that the minimization of $J_d(\boldsymbol{\lambda})$ w.r.t. $\boldsymbol{\lambda}$ under the following constraint

$$\text{Tr}(\mathbf{R}_t) = L \int_0^1 S_{tt}(\nu) d\nu = L \frac{1}{N} \sum_{n=1}^N \lambda_n = LP \quad (39)$$

reduces to a convex optimization problem. Unlike the general case, we are here able to develop the Lagrange optimization method and thus obtain a closed-form expression for the optimal $\boldsymbol{\lambda}$. This result, which is summarized in Theorem 3, is proven in Appendix E.

Theorem 3: The criterion J_d defined by Eq. (38) is convex and minimum under the power constraint (39) for

$$\lambda_m = P \frac{\sqrt{\alpha_m/(\mu - \beta_m)}}{\frac{1}{N} \sum_{n=1}^N \sqrt{\alpha_n/(\mu - \beta_n)}} \quad (40)$$

where μ is such that

$$\frac{1}{N} \sum_{n=1}^N \beta_n \sqrt{\alpha_n/(\mu - \beta_n)} = \sqrt{\beta}.$$

As N can be taken arbitrarily large, we finally have that

$$S_{tt}(\nu) = P \frac{\sqrt{c_1^d(\nu)/(\mu - c_2^d(\nu))}}{\int_0^1 \sqrt{c_1^d(u)/(\mu - c_2^d(u))} du} \quad (41)$$

where μ is such that $\int_0^1 c_2^d(u) \sqrt{c_1^d(u)/(\mu - c_2^d(u))} du = \sqrt{\beta}$.

C. Comments

We recall that our approach holds when the TS size N_T , the data size N_D , and channel length L are large. In practice, a TS with power spectrum provided by means of Theorems 2 or 3 can be very simply generated as the output of a digital filter with relevant coefficients excited by a pre-generated pseudo-random sequence. Further, this filtering does not have to be performed frequently since the channel statistics, unlike the channel state information, are likely to change very slowly with time. Therefore, the additional computational complexity of the proposed training design is low.

It is worth pointing out that our derivations enable us to analyze the scenario where the channel has to be estimated but no CFO is present. In this case, criteria $J_g(\boldsymbol{\lambda})$ and $J_d(\boldsymbol{\lambda})$ can be simplified by removing the second term in the RHS of Eq. (35) and of Eq. (38) respectively. The optimization problem is still convex in this case. By adapting the derivations in Appendix E, the power allocation at FFT bin n , denoted by λ_n , is found to be $\sqrt{s_n}$ in the general case and to $\sqrt{\alpha_n}$ in the high Ricean factor case. Consequently the best TS may be colored (if the sequence s or α is not constant). We recall that the TS that minimizes the CRB of the channel (when no CFO is present) is white regardless of the channel statistics. This means that the minimization of the channel CRB and of the symbol MSE may lead to two different kinds of TS designs. This leads to the following question: which criterion is the most relevant? When a Wiener-type equalizer is carried out, the MSE of the output of the Wiener filter represents better the true performance of the overall system than the CRB of the channel since the CRB of the channel does not take into account the receiver structure. Nevertheless, by simulations, we observed that the improvement due to the color of the TS is negligible when no CFO is present. Therefore, in this case, the white TS remains a good candidate for channel estimation.

We can also address the CFO estimation problem in frequency-selective channel when the channel is not a parameter of interest as it was the case in [4], [6]. This problem may occur when synchronisation is carried out more frequently than channel estimation. As mentioned in Section III-C, at high SNR, the MSE degradation due to CFO estimation is proportional to the CRB for the CFO. Unlike for channel estimation, for CFO estimation, the CRB and MSE criteria

lead to the same TS design. Minimizing the CRB with respect to the TS has been partially treated in the literature. [4] selects the TS which minimizes the worst CRB (i.e., the maximum of the CRB over all normalized channel realizations), [5] selects the TS that renders the finite-sample CRB for the CFO channel-independent, and [6] characterizes the TS which minimizes the average CRB when the channel coefficients are assumed i.i.d. and Rayleigh (i.e., $K = 0$ and $\boldsymbol{\Sigma}$ is proportional to the identity matrix). Here, by adapting Theorems 2 or 3, we design the TS which minimizes the average CRB when the channel is Rice whatever the Ricean factor K and the color of the random part. In the general case, criterion $J_g(\boldsymbol{\lambda})$ can be simplified by removing the first term in the RHS of Eq. (35). Obviously, the optimization problem remains convex and a standard optimization tool can be applied to find the optimum TS spectrum. In the case where the channel has a high Ricean factor, K , $J_d(\boldsymbol{\lambda})$ can be simplified by ignoring the first term in the RHS of Eq. (38). In this case, $\boldsymbol{\lambda}$ has one non-zero element (i.e. a TS with one nonzero FFT bin only), which is associated with the maximum β_n .

V. SIMULATIONS

Unless otherwise stated, we set $N_t = 50$, $L = 5$ and $\alpha = 10$. The CFO is fixed to $f = 0.1$. All simulated points are averaged over 1000 Monte-Carlo runs. In each run, the channel was independently generated according to model (2). We assume that the correlation between two channel taps $h(k)$ and $h(l)$ of the CIR is given by $\rho^{|k-l|}$ where ρ belongs to $[0, 1)$.

In Figure 1, we plot the spectrum of one realization of \mathbf{h}_d and the normalized spectrum of the associated optimal TS when $K = 3$, $\rho = 0$ and SNR=10dB. We observe that the TS spectrum gives an important role to the best frequencies of the deterministic part of the channel to ensure a good estimation of the CFO but also is spread over a large enough band to ensure an accurate estimation of the random part of the channel.

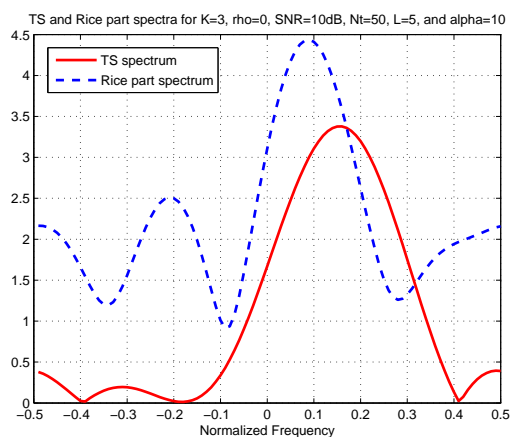


Fig. 1. Spectrum shape for one realization of \mathbf{h}_d and the optimal TS (for $K = 3$ and $\rho = 0$)

In Figure 2, we plot the theoretical and empirical data MSE. The MSE when the parameters (i.e. CFO and CIR) are known

is also depicted for comparison purposes. We have considered $\mathbf{R}_t = P\mathbf{I}_{d_L}$, $K = 10$ and $\rho = 0$. When the parameters are estimated, we display the theoretical MSE depicted in Eq. (26) and the one drawn in Eq. (27) and Eq. (32). We thus observe that all our approximations are accurate since the empirical MSE is close to the approximated theoretical MSE.

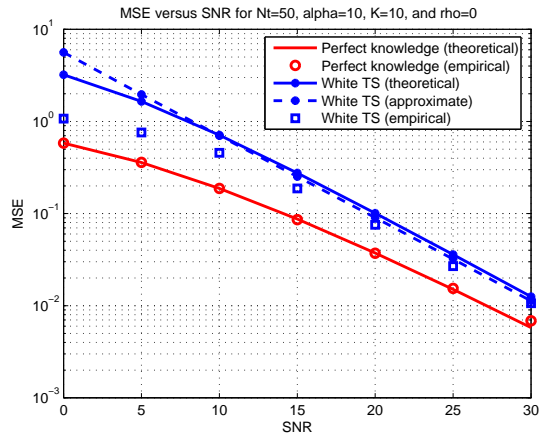


Fig. 2. Approximation MSE versus SNR

In Figure 3, we display the theoretical data MSE versus SNR (with $K = 10$ and $\rho = 0$) when the parameters are perfectly known and when the parameters are estimated with either a white TS or the colored TS suggested by Theorem 3 for large K . We observe that the colored TS exhibits a gain in terms of MSE but this gain seems to be small.

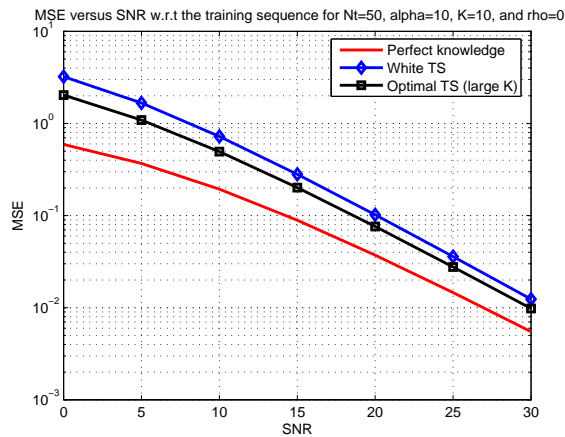


Fig. 3. MSE versus SNR (for $K \neq 0$ and $\rho = 0$)

In Figure 4, we plot the theoretical data MSE versus K (with $\text{SNR}=20\text{dB}$ and $\rho = 0$). We observe that the optimization carried out under the assumption of large K provides non-negligible gain as soon as $K > 3$. In many practical situations, it is reasonable to assume the Ricean factor to be larger than 3 (cf. [9], [10], [11], [12].) Therefore, Theorem 3 can be of great interest in realistic scenarios.

In Figure 5, we plot the Bit Error Rate (BER) versus SNR (with $K = 10$ and $\rho = 0$) when a frequency compensation

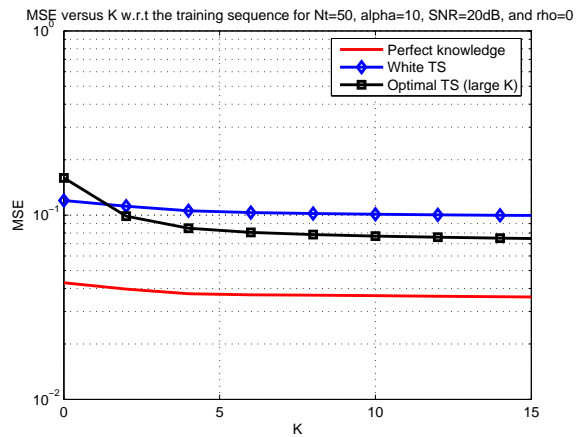


Fig. 4. MSE versus K (for $\text{SNR}=20\text{dB}$ and $\rho = 0$)

and a Wiener equalizer based on the estimated values of the parameters are employed. The training is either white or colored as suggested by Theorem 3. We remark that the gain in performance is significant, which advocates TS optimization at the transmitter.

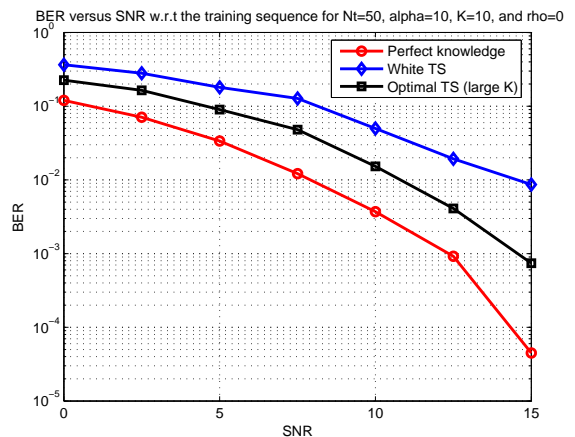


Fig. 5. BER versus SNR (for $K \neq 0$ and $\rho = 0$)

In Figure 6, we display the MSE versus ρ (for $K = 0$) when optimal training or white training are employed and when the parameters are assumed perfectly known. The gain in MSE obtained by selecting the best TS is very small. One can remark that the gap between the case of known parameters and that of unknown parameters is small as well. Actually the correlation between the channel taps needs to be strong in order to observe a difference between the optimal TS and the white TS. Nevertheless, as we will see in the next figures, a little difference in the MSE may lead fortunately to an important gain in BER, which is the ultimate criterion.

In Figure 7, we plot the MSE versus SNR in a correlated Rayleigh channel environment. Once again, the gain is small whatever the SNR.

In Figure 8, we plot the BER versus SNR in a correlated Rayleigh channel environment. Unlike the MSE criterion, we observe a non-negligible gain between the white TS and the

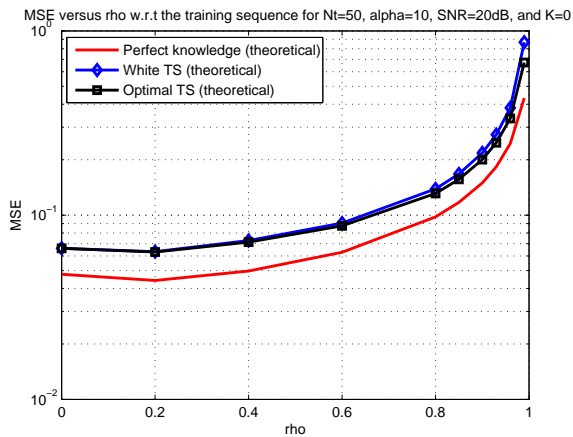


Fig. 6. MSE versus ρ (for SNR=20dB and $K = 0$)

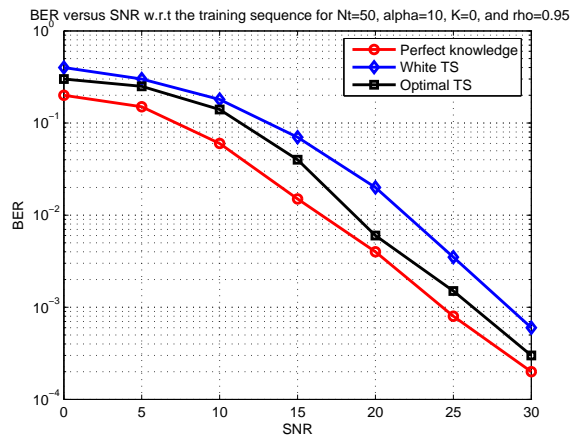


Fig. 8. BER versus SNR (for $K = 0$ and $\rho = 0.95$)

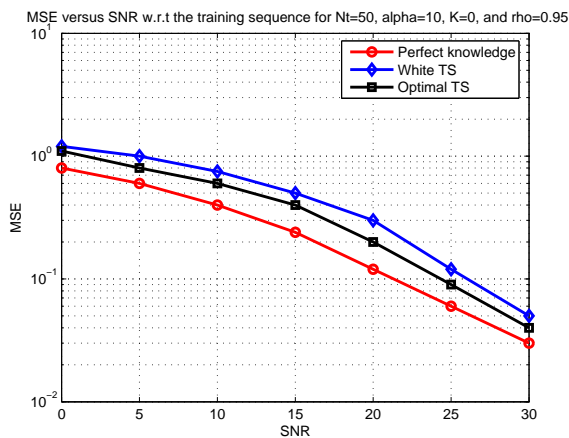


Fig. 7. MSE versus SNR (for $K = 0$ and $\rho = 0.95$)

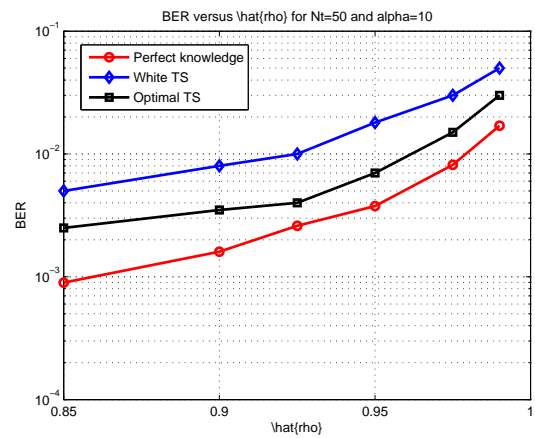


Fig. 9. BER versus $\hat{\rho}$ (for $K = 0$ and $\rho = 0.95$)

optimal TS. Consequently, in a correlated Rayleigh environment, it is worth performing an optimization of the TS at the transmitter.

In Figure 9, we plot the BER versus $\hat{\rho}$ where $\hat{\rho}$ is an estimation of the true correlation factor which is fixed to $\rho = 0.95$ as before. The SNR is 20dB. The other parameters are as in Figure 8. We remark that the design of the TS is not so sensitive to the perfect knowledge of the correlation factor.

VI. CONCLUSION

In this paper, we develop a new method to design optimum training sequences which efficiently account for the errors due to channel and carrier frequency offset estimation. This method is based on the minimization of the mean square error on data estimation at the output of the Wiener-type equalizer averaged over the channel statistics. The obtained criterion is simplified by invoking an asymptotic regime. Then we prove that the criterion is convex which enables us to find numerically the optimal training sequence. Closed-form solutions are obtained in the case of channels with high Ricean factors. Simulations show that the gain in bit error rate can

be significant in many scenarios. In recent systems, multi-carrier modulations are being considered instead of single carrier modulations. Our work may be extended to such a scenario. We guess that the MSE to be minimized will offer a similar expression (if the number of subcarriers is assumed to be large) but the part associated with a mis-estimation of the CFO should have a stronger contribution because of the high sensitivity to the CFO in multicarrier systems.

APPENDIX

A. Proof of Lemma 1

We have that $\text{MSE}_n(\mathbf{t}|\mathbf{h}, \hat{\mathbf{h}}, \hat{f}) = \mathbb{E}[|z(n) - d(n - N_T)|^2|\mathbf{t}, \mathbf{h}, \hat{\mathbf{h}}, \hat{f}]$ which can be easily decomposed as follows

$$\begin{aligned} \text{MSE}_n(\mathbf{t}|\mathbf{h}, \hat{\mathbf{h}}, \hat{f}) &= \sigma_d^2 + \sum_{l, l'=-L_g}^{L_g} \hat{g}(l)\overline{\hat{g}(l')}r_{y_c y_c}(l' - l) \\ &\quad - 2\Re \left[\sum_{l=-L_g}^{L_g} \hat{g}(l)r_{dy_c}(n, -l) \right] \end{aligned} \quad (42)$$

where $r_{y_c y_c}(\tau) = \mathbb{E}[y_c(n + \tau)\overline{y_c(n)}]$ and $r_{dy_c}(n, \tau) = \mathbb{E}[y_c(n + \tau)d(n - N_T)]$. Thanks to Eq. (3), one can prove

that

$$\begin{aligned} r_{y_c y_c}(\tau) &= \sigma_d^2 \sum_{k=0}^{L_h} h(k+\tau) \overline{h(k)} e^{2i\pi\Delta f\tau} + \sigma^2 \delta(\tau) \\ &= \sigma_d^2 \int_0^1 |H(\nu)|^2 e^{2i\pi(\nu+\Delta f)\tau} d\nu + \sigma^2 \delta(\tau) \end{aligned}$$

We also remark that

$$r_{dy_c}(n, \tau) = \sigma_d^2 h(\tau) e^{2i\pi\Delta f(n+\tau)}.$$

The second term of the RHS of Eq. (42) can then be simplified as follows

$$\begin{aligned} \sum_{l, l'=-L_g}^{L_g} \hat{g}(l) \overline{\hat{g}(l')} r_{y_c y_c}(l' - l) &= \sigma_d^2 \int_0^1 |H(\nu)|^2 |\hat{G}(\nu + \Delta f)|^2 d\nu \\ &\quad + \sigma^2 \int_0^1 |\hat{G}(\nu)|^2 d\nu \end{aligned}$$

By similar algebraic manipulations, we obtain

$$\begin{aligned} \sum_{l=-L_g}^{L_g} \hat{g}(l) r_{dy_c}(n, -l) &= \sigma_d^2 e^{2i\pi\Delta f n} \sum_{l=-L_g}^{L_g} \hat{g}(l) h(-l) e^{-2i\pi\Delta f l} \\ &= \sigma_d^2 e^{2i\pi\Delta f n} \int_0^1 \hat{G}(\nu + \Delta f) H(\nu) d\nu \end{aligned}$$

plugging both previous equations into Eq. (42) and using the fact that

$$\text{MSE}(\mathbf{t} \parallel \mathbf{h}, \hat{\mathbf{h}}, \hat{f}) = \frac{1}{N_D} \sum_{n=N_T}^{N_T+N_D-1} \text{MSE}_n(\mathbf{t} \parallel \mathbf{h}, \hat{\mathbf{h}}, \hat{f})$$

conclude the proof.

B. Proof of Lemma 2

We assume that the Wiener filter length, denoted by L_g , is large and of the same order as the filter length L . Since L is assumed to be negligible with respect to N_T , we get $L_g \Delta f \ll 1$ when N_T is large. More precisely, we have that $L_g/\sqrt{N_T}$ is bounded when N_T is large. By definition, we obtain that

$$\hat{G}(\nu + \Delta f) = \sum_{-L_g}^{L_g} \hat{g}(l) e^{2i\pi(\nu+\Delta f)l}$$

As $L_g/\sqrt{N_T}$ is bounded, we have $L_g \Delta f \propto L_g/N_T^{3/2} \propto 1/N_T$. Consequently, for every l , we can write that

$$e^{2i\pi\Delta f l} = 1 + 2i\pi\Delta f l + \mathcal{O}(1/N_T^2).$$

where $\mathcal{O}(1/N_T^2)$ is bounded in $1/N_T^2$. This implies that

$$\begin{aligned} \hat{G}(\nu + \Delta f) &= G(\nu) + \Delta G(\nu) \\ &\quad + 2i\pi\Delta f \left(\sum_{-L_g}^{L_g} l \hat{g}(l) \right) + \mathcal{O}_p(L_g/N_T^2) \end{aligned}$$

where $\mathcal{O}_p(1/N_T^2)$ is bounded in $1/N_T^2$ in probability. As $L_g/\sqrt{N_T}$ is bounded, we finally get

$$\hat{G}(\nu + \Delta f) = G(\nu) + \Delta G(\nu) + o_p(1/N_T). \quad (43)$$

Recall that

$$\text{MSE}(\mathbf{t} \parallel \mathbf{h}, \hat{\mathbf{h}}, \hat{f}) = A + B + C$$

where A , B and C are the first, the second and the third row of the RHS of the equation provided in lemma 1 respectively and given by the following expressions

$$A = \sigma_d^2 + \sigma^2 \int_0^1 |(G + \Delta G)(\nu)|^2 d\nu$$

$$B = \sigma_d^2 \int_0^1 |H(\nu)|^2 |(G + \Delta G)(\nu - \Delta f)|^2 d\nu$$

$$C = -2\sigma_d^2 \Re \left[\Im(\Delta f) \int_0^1 (G + \Delta G)(\nu - \Delta f) H(\nu) d\nu \right].$$

Using Eqs.(5) and (43) leads to

$$\begin{aligned} A &= \sigma_d^2 + \sigma^2 \int_0^1 |G(\nu)|^2 d\nu + \sigma^2 \int_0^1 |\Delta G(\nu)|^2 d\nu \\ &\quad + \text{one order term} + o_p(1/N_T) \end{aligned}$$

$$\begin{aligned} B &= \sigma_d^2 \int_0^1 |H(\nu)|^2 |G(\nu)|^2 d\nu + \sigma_d^2 \int_0^1 |H(\nu)|^2 |\Delta G(\nu)|^2 d\nu \\ &\quad + \text{one order term} + o_p(1/N_T) \end{aligned}$$

and

$$\begin{aligned} C &= -2\sigma_d^2 \Re \left[\int_0^1 H(\nu) G(\nu) d\nu \right] \\ &\quad + 2\sigma_d^2 \pi (2 + \alpha) N_T \Im \left[i \int_0^1 \Delta G(\nu) \Delta f H(\nu) d\nu \right] \\ &\quad + 4\sigma_d^2 \pi^2 (1 + \alpha + \alpha^2/3) N_T^2 \Re \left[\int_0^1 G(\nu) (\Delta f)^2 H(\nu) d\nu \right] \\ &\quad + \text{one order term} + o_p(1/N_T) \end{aligned}$$

where "one order term" is a sum of one-order monomials associated with variables Δf or Δg . By averaging A , B , and C over the channel estimation error and the carrier frequency error, the "one-order term" vanishes when N_T is large since the ML estimators are asymptotically unbiased. The remaining terms lead to the following expressions.

$$\begin{aligned} e_0 &= \sigma_d^2 + \int_0^1 (\sigma_d^2 |H(\nu)|^2 + \sigma^2) |G(\nu)|^2 d\nu \\ &\quad - 2\sigma_d^2 \Re \left[\int_0^1 H(\nu) G(\nu) d\nu \right] \end{aligned}$$

$$e_1 = \int_0^1 (\sigma_d^2 |H(\nu)|^2 + \sigma^2) \gamma_{g,g}(\nu) d\nu$$

$$e_2 = 2\sigma_d^2 (2 + \alpha) \pi N_T \Im \left[\int_0^1 H(\nu) \gamma_{g,f}(\nu) d\nu \right]$$

$$e_3 = 4\sigma_d^2 \pi^2 (1 + \alpha + \alpha^2/3) N_T^2 \gamma_{f,f} \Re \left[\int_0^1 G(\nu) H(\nu) d\nu \right]$$

where $\gamma_{g,g}(\nu)$, $\gamma_{g,f}(\nu)$ are defined similarly to Eq. (10) and Eq. (12).

Our aim now is to express the error on the Wiener filter as a function of the error on the channel filter. Under the assumption of large Wiener filter length, it is well known that

$$G(\nu) = \frac{\sigma_d^2 \overline{H(\nu)}}{\sigma_d^2 |H(\nu)|^2 + \sigma^2}.$$

Differentiating previous expression with respect to $H(\nu)$ leads to

$$\Delta G(\nu) = \frac{-\sigma_d^4 \overline{H(\nu)}^2 \Delta H(\nu) + \sigma_d^2 \sigma^2 \overline{\Delta H(\nu)}}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^2}.$$

Then we have that

$$\gamma_{g,g}(\nu) = \frac{(\sigma_d^8 |H(\nu)|^4 + \sigma_d^4 \sigma^4) \gamma_{h,h}(\nu)}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^4} \quad (44)$$

$$- \frac{2\Re \left[\sigma_d^6 \sigma^2 \overline{H(\nu)}^2 \tilde{\gamma}_{h,h}(\nu) \right]}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^4}$$

$$\gamma_{g,f}(\nu) = \frac{\sigma_d^2 \sigma^2 \overline{\gamma_{h,f}(\nu)} - \sigma_d^4 \overline{H(\nu)}^2 \gamma_{h,f}(\nu)}{(\sigma_d^2 |H(\nu)|^2 + \sigma^2)^2} \quad (45)$$

which leads to the expressions given in the lemma.

C. Proof of Theorem 1

First of all, we write $\mathbf{h}^H \mathbf{R}_t \mathbf{h}$ as a linear combination of non-central chi-square distributions. Let

$$\mathbf{x}_d = \mathbf{R}_t^{1/2} \sqrt{\frac{K}{K+1}} \mathbf{h}_d \quad \text{and} \quad \mathbf{x}_r = \mathbf{R}_t^{1/2} \sqrt{\frac{1}{K+1}} \mathbf{h}_r$$

Let $\mathbf{R}_x = \mathbb{E}[\mathbf{x}_r \mathbf{x}_r^H]$ be the autocorrelation matrix of \mathbf{x}_r . The matrix can be diagonalized as follows

$$\mathbf{R}_x = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$$

where \mathbf{U} is a unitary matrix made of the eigenvectors of \mathbf{R}_x and where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_L)$ with λ_k being the k^{th} eigenvalue of \mathbf{R}_x . Let $\tilde{\mathbf{x}}_d$ and $\tilde{\mathbf{x}}_r$ be the following vectors

$$\tilde{\mathbf{x}}_d = \mathbf{\Lambda}^{-1/2} \mathbf{U}^H \mathbf{x}_d \quad \text{and} \quad \tilde{\mathbf{x}}_r = \mathbf{\Lambda}^{-1/2} \mathbf{U}^H \mathbf{x}_r.$$

$\tilde{\mathbf{x}}_d$ is a deterministic vector whereas $\tilde{\mathbf{x}}_r$ is a circularly Gaussian distributed vector with zero-mean and unit-variance. Then we get

$$\xi = \mathbf{h}^H \mathbf{R}_t \mathbf{h} = \sum_{k=1}^L \lambda_k |\tilde{x}_d(k) + \tilde{x}_r(k)|^2$$

where $\tilde{x}_d(k)$ and $\tilde{x}_r(k)$ are the k^{th} component of $\tilde{\mathbf{x}}_d$ and $\tilde{\mathbf{x}}_r$ respectively.

Secondly, it is well known ([8] and references therein) that a weighted sum of non-central chi-square distribution of two degrees of freedom can be well approximated by a central Gamma distribution. Let $p_G(t)$ be a Gamma distribution with standard parameters (p_1, p_2) . Then we get

$$p_G(t) = \frac{p_1^{p_2}}{\Gamma(p_2)} t^{p_2-1} e^{-p_1 t} \mathbf{1}_{t \geq 0}$$

Notice that the mean and the variance of a Gamma distribution are given by p_2/p_1 and p_2/p_1^2 respectively. When L is large enough, the distribution of the random variable ξ is close to the Gamma distribution of parameters (p_1, p_2) given by

$$p_2/p_1 = \mathbb{E}[\xi]$$

$$p_2/p_1^2 = \mathbb{E}[(\xi - \mathbb{E}[\xi])^2]$$

After straightforward but tedious algebraic manipulations, we obtain

$$p_1 = \frac{\sum_{k=1}^L \lambda_k (1 + |\tilde{x}_d(k)|^2)}{\sum_{k=1}^L \lambda_k^2 (1 + 2|\tilde{x}_d(k)|^2)} \quad (46)$$

$$p_2 = \ell_1 \sum_{k=1}^L \lambda_k (1 + |\tilde{x}_d(k)|^2) \quad (47)$$

Thirdly, it remains to evaluate the expectation of $1/\xi$ when ξ is assumed to be Gamma distributed. One can easily check that

$$\mathbb{E} \left[\frac{1}{\xi} \right] = \int_0^\infty \frac{1}{t} p_G(t) dt = \frac{p_1}{p_2 - 1}$$

Finally, we need to evaluate p_1 and p_2 with respect the channel statistics $(K, \mathbf{h}_d, \mathbf{\Sigma})$ and the training correlation matrix \mathbf{R}_t . Towards this objective, recall that

$$\tilde{\mathbf{x}}_d = \mathbf{\Lambda}^{-1/2} \mathbf{U}^H \mathbf{R}_t^{1/2} \sqrt{\frac{K}{K+1}} \mathbf{h}_d$$

which implies that

$$\sum_{k=1}^L \lambda_k |\tilde{x}_d(k)|^2 = \frac{K}{K+1} \mathbf{h}_d^H \mathbf{R}_t \mathbf{h}_d$$

and

$$\sum_{k=1}^L \lambda_k^2 |\tilde{x}_d(k)|^2 = \frac{K}{(K+1)^2} \mathbf{h}_d^H \mathbf{R}_t \mathbf{\Sigma} \mathbf{R}_t \mathbf{h}_d.$$

As $\mathbf{\Lambda}$ represents the diagonal matrix whose elements are the eigenvalues of \mathbf{R}_x , we have

$$\sum_{k=1}^L \lambda_k = \text{Tr}(\mathbf{R}_x) = \frac{1}{K+1} \text{Tr}(\mathbf{R}_t \mathbf{\Sigma})$$

and

$$\sum_{k=1}^L \lambda_k^2 = \text{Tr}(\mathbf{R}_x^2) = \frac{1}{(K+1)^2} \text{Tr}(\mathbf{R}_t \mathbf{\Sigma} \mathbf{R}_t \mathbf{\Sigma})$$

which concludes the proof by setting $\ell_1 = (K+1)p_2/p_1$ and $\ell_2 = (K+1)^2 p_2/p_1^2$.

D. Proof of Theorem 2

Since s_l is positive or null, the function $\lambda_l \mapsto s_l/\lambda_l$ is convex when λ_l is positive. As a sum of convex functions is convex, we have that $\boldsymbol{\lambda} \mapsto J_h(\boldsymbol{\lambda})$ is convex if all the components of $\boldsymbol{\lambda}$ are positive.

Instead of proving directly the convexity of $J_f(\boldsymbol{\lambda})$, we will prove the convexity of $\phi(\boldsymbol{\lambda}) = -(K+1)/J_f(\boldsymbol{\lambda})$, because it is easier. As J_f is positive by construction whatever the matrix \mathbf{R}_t and as the inverse function is also convex, the convexity of ϕ implies the convexity of J_f . Therefore we now concentrate on $\boldsymbol{\lambda} \mapsto \phi(\boldsymbol{\lambda})$ which can be written as follows

$$\phi(\boldsymbol{\lambda}) = \frac{\sum_{k,l=1}^L \lambda_k \lambda_l b_{kl} c_{lk}}{\sum_{l=1}^L \lambda_l a_{ll}} - \sum_{l=1}^L \lambda_l a_{ll}$$

We now calculate the Hessian matrix of ϕ , which is defined as follows

$$\mathcal{H} = \left[\frac{\partial^2 \phi(\lambda_1, \dots, \lambda_L)}{\partial \lambda_m \partial \lambda_n} \right]_{m,n=1, \dots, L}$$

One can find that

$$\begin{aligned} \frac{\partial \phi}{\partial \lambda_m} &= \frac{2t_{mm}^2 \lambda_m + \sum_{k=1, k \neq m}^L \lambda_k (b_{km} c_{mk} + b_{mk} c_{km})}{\sum_{l=1}^L \lambda_l a_{ll}} \\ &- a_{mm} \frac{\sum_{k,l=1}^L \lambda_k \lambda_l b_{kl} c_{lk}}{(\sum_{l=1}^L \lambda_l a_{ll})^2} - a_{mm} \end{aligned}$$

Then we have that

$$\begin{aligned} \frac{\partial^2 \phi}{\partial \lambda_m \partial \lambda_n} &= \frac{1}{S} [b_{nm} c_{mn} + b_{mn} c_{nm}] \\ &- \frac{1}{S^2} [2b_{mm} c_{mm} a_{nn} \lambda_m + 2b_{nn} c_{nn} a_{mm} \lambda_n \\ &+ a_{mm} \sum_{\substack{k=1 \\ k \neq n}}^L \lambda_k (b_{kn} c_{nk} + b_{nk} c_{kn}) \\ &+ a_{nn} \sum_{\substack{k=1 \\ k \neq m}}^L \lambda_k (b_{km} c_{mk} + b_{mk} c_{mk})] \\ &+ \frac{1}{S^3} [2a_{mm} a_{nn} \sum_{k,l=1}^L \lambda_k \lambda_l b_{kl} c_{lk}] \end{aligned}$$

with $S = \sum_{l=1}^L \lambda_l a_{ll}$.

To prove convexity, we need to prove that the Hessian matrix \mathcal{H} is positive. Therefore we focus on the following term

$$c = \frac{S}{2} \mathbf{x}^T \mathcal{H} \mathbf{x}$$

where $\mathbf{x} = [x_1, \dots, x_L]^T$ is a real-valued vector of length L . After straightforward but tedious derivations, we obtain

$$\begin{aligned} c &= \sum_{k,l=1}^L x_k b_{kl} c_{lk} x_l \\ &- \frac{(\sum_{k,l=1}^L \lambda_k (b_{kl} c_{lk} + b_{lk} c_{kl}) x_l) (\sum_{l=1}^L a_{ll} x_l)}{S} \\ &+ \frac{(\sum_{k,l=1}^L \lambda_k b_{kl} c_{lk} \lambda_l) (\sum_{l=1}^L a_{ll} x_l)^2}{S^2} \end{aligned}$$

Let $\mathbf{T} = \mathbf{B} \odot \mathbf{C}^T$ where \odot stands for the term-by-term Hadamard product. Since \mathbf{B} and \mathbf{C} are hermitian positive, \mathbf{T} is hermitian positive as well [16]. As a consequence, we can apply the following Schwartz inequality as follows

$$|\boldsymbol{\lambda}^T \mathbf{T} \mathbf{x}|^2 \leq (\boldsymbol{\lambda}^T \mathbf{T} \boldsymbol{\lambda}) (\mathbf{x}^T \mathbf{T} \mathbf{x})$$

which is equivalent to

$$\left| \sum_{k,l=1}^L \lambda_k b_{kl} c_{lk} x_l \right| \leq \sqrt{\sum_{k,l=1}^L \lambda_k b_{kl} c_{lk} \lambda_l} \sqrt{\sum_{k,l=1}^L x_k b_{kl} c_{lk} x_l}$$

Thus we have

$$\begin{aligned} c &\geq \sum_{k,l=1}^L x_k b_{kl} c_{lk} x_l \\ &- 2 \frac{\sqrt{\sum_{k,l=1}^L \lambda_k b_{kl} c_{lk} \lambda_l} \sqrt{\sum_{k,l=1}^L x_k b_{kl} c_{lk} x_l} \left| \sum_{l=1}^L a_{ll} x_l \right|}{S} \\ &+ \frac{(\sum_{k,l=1}^L \lambda_k b_{kl} c_{lk} \lambda_l) (\sum_{l=1}^L a_{ll} x_l)^2}{S^2} \\ &= \left[\sqrt{\sum_{k,l=1}^L x_k b_{kl} c_{lk} x_l} - \frac{\sqrt{\sum_{k,l=1}^L \lambda_k b_{kl} c_{lk} \lambda_l} \left| \sum_{l=1}^L a_{ll} x_l \right|}{S} \right]^2 \\ &\geq 0 \end{aligned}$$

which concludes the proof.

E. Proof of Theorem 3

The function $J_d(\boldsymbol{\lambda})$ is convex since i) the inverse function is convex, ii) the composition of a convex function and an affine function is convex, and iii) the sum of convex functions is convex.

We define the Lagrange function as follows

$$\begin{aligned} \mathcal{L}(\boldsymbol{\lambda}, \mu_1) &= \frac{1}{N} \sum_{n=1}^N \frac{\alpha_n}{\lambda_n} + \beta \frac{1}{\frac{1}{N} \sum_{n=1}^N \beta_n \lambda_n} \\ &+ \mu_1 \left(\frac{1}{N} \sum_{n=1}^N \lambda_n - P \right) \end{aligned}$$

Thanks to the KKT condition, we have that

$$\frac{\partial \mathcal{L}}{\partial \lambda_m} = -\frac{1}{N} \frac{\alpha_m}{\lambda_m^2} - \beta N \frac{\beta_m}{\left(\sum_{n=1}^N \beta_n \lambda_n \right)^2} + \frac{1}{N} \mu_1 \quad (= 0)$$

thus

$$\mu_1 = \frac{\alpha_m}{\lambda_m^2} + \beta N^2 \frac{\beta_m}{\left(\sum_{n=1}^N \beta_n \lambda_n \right)^2}$$

Let $\mu_2 = \left(\sum_{n=1}^N \beta_n \lambda_n \right)^2$ and $\mu_3 = \mu_1 \mu_2$. We obtain

$$\lambda_m = \sqrt{\mu_2} \sqrt{\frac{\alpha_m}{\mu_3 - N^2 \beta \beta_m}}$$

As $\frac{1}{N} \sum_{n=1}^N \lambda_n = P$ and by setting $\mu = \mu_3 / (\beta N^2)$, we have that

$$\lambda_m = P \frac{\sqrt{\frac{\alpha_m}{\mu - \beta_m}}}{\frac{1}{N} \sum_{n=1}^N \sqrt{\frac{\alpha_n}{\mu - \beta_n}}}$$

As $\sqrt{\mu_2}$ is also equal to $\sum_{n=1}^N \beta_n \lambda_n$, it is easy to prove that

$$\frac{1}{N} \sum_{n=1}^N \beta_n \sqrt{\frac{\alpha_n}{\mu - \beta_n}} = \sqrt{\beta}$$

which concludes the proof.

REFERENCES

- [1] M. Morelli and U. Mengali, "Carrier frequency estimation for transmissions over selective channels", *IEEE Trans. on Comm.*, vol. 48, pp. 1580-1589, Sep. 2000.
- [2] P. Ciblat and L. Vandendorpe, "On the Maximum-Likelihood based data-aided frequency offset and channel estimates", in *European Signal Processing Conference (EUSIPCO)*, Sep. 2002.
- [3] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation", *IEEE Trans. on Comm.*, vol. 51, pp. 1910-1917, Nov. 2003.
- [4] O. Besson and P. Stoica, "Training Sequence Selection for Frequency Offset estimation in frequency selective channels", *Digital Signal Processing A Review Journal*, vol. 13(1), pp.106-127, Jan. 2003.
- [5] M. Ghogho and A. Swami, "Training Design for Multipath Channel and Frequency-Offset Estimation in MIMO Systems," *IEEE Trans. on Signal Processing*, vol. 54(10), pp. 3957-65, Oct. 2006.
- [6] H. Minn, X. Fu, and V. Bhargava, "Optimal Periodic Training signal for frequency offset estimation in frequency-selective fading channels", *IEEE Trans. on Comm.*, vol. 54, pp. 1081-1096, Jun. 2006.
- [7] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, Univ. California (Berkeley) Press, 1958.
- [8] Q.T. Zhang and D.P. Liu, "A simple capacity formula for correlated diversity Ricean fading", *IEEE Communications Letters*, vol. 6, issue 11, pp. 481-483, Nov. 2002.
- [9] V. Erceg et al., "Channel models for fixed wireless applications", *IEEE 802.16 broadband wireless access working group report*, Jul. 2001.
- [10] M. Haardt and Q. Spencer, "Smart antennas for wireless communications beyond the third generation", *Computer Communications*, vol. 26, pp.41-45, 2003.
- [11] M. Iskander and S. Shimamoto, "Channel characterization and performance evaluation of mobile communication employing stratospheric platforms", *IEICE Trans. on Communications*, vol. E89-B(3), pp. 937-944, Mar. 2006.
- [12] H. Suzuki, C.D. Wilson, and K. Ziri-Castro, "Time variation characteristics of wireless broadband channel in urban area", *First European Conference on Antennas and Propagation*, Nice, France, Nov. 2006.
- [13] Y.-D. Kim, J.K. Lim, C. Suh, and Y.H. Lee, "Designing Training Sequences for Carrier Frequency estimation in frequency-selective channels", *IEEE Trans. on Vehicular Technology*, vol. 5 (1), pp. 151-157, Jan. 2006.
- [14] R.M. Gray, "Toeplitz and circulant matrices: a review", *Foundations and Trends in Communications and Information Theory*, vol.2 (3), pp. 155-239, 2006.
- [15] H.L. Van Trees, "Detection, Estimation and Modulation", *John Wiley and Sons*, 1968.
- [16] R.A. Horn and C.R. Johnson, "Topics in Matrix Analysis", *Cambridge University Press*, 1991.