Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 85 (2005) 2394–2411

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Efficient semi-blind estimation of multipath channel parameters via a delay decoupling optimization approach

Athanasios A. Rontogiannis[a,*], Kostas Berberidis[b], Alexandra Marava[b], Jacques Palicot[c]

[a]*Institute of Space Applications and Remote Sensing, National Observatory of Athens, Athens, Greece*
[b]*Department of Computer Engineering & Informatics/C.T.I-R&D, University of Patras, Greece*
[c]*Supelec, Campus de Rennes, France*

Received 8 March 2004; received in revised form 18 October 2004
Available online 23 May 2005

Abstract

In this paper a parametric method for estimating the unknown multipath channel impulse response (CIR) in a semi-blind manner is proposed. The main trait of this method is that instead of seeking the whole CIR sequence, only the unknown time delays and attenuation factors of the physical channel multipath components are estimated. The technique is based on a suitable application of the sub-channel response matching (SRM) criterion. The resulting cost function is separable with respect to the two sets of unknown parameters, i.e., time delays and attenuation factors, and thus a two step optimization procedure can be applied. Concerning the first step, which is the most difficult one, it is proven that the resulting non-linear cost function can be decoupled in terms of the respective time delay parameters. Thus, only a small number of simple linear searches needs to be executed in order to estimate the time delays of the multipath channel. The new method offers significant computational savings and a lower mean square estimation error as compared to existing semi-blind channel estimation methods. It performs well even for closely-spaced delays and is quite robust to channel overmodeling.

Keywords: Semi-blind channel estimation; Parametric multipath channel modeling

1. Introduction

Multipath channel estimation is an issue of major importance in reliable wireless communications system design. Due to the multipath phenomenon in wireless applications, the introduced InterSymbol Interference (ISI) may cause a severe

*Corresponding author. Tel.: +30 210 8109188; fax: +30 210 6138343.
E-mail addresses: tronto@space.noa.gr (A.A. Rontogiannis), berberid@ceid.upatras.gr (K. Berberidis), marava@ceid.upatras.gr (A. Marava), jacques.palicot@supelec.fr (J. Palicot).

0165-1684/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2005.02.018

degradation in a system's performance. The problem is much more severe in high-rate applications, since, as the symbol rate becomes higher, the number of symbols spanned by a channel impulse response (CIR) with a certain delay spread is increased. Thus, an efficient and accurate estimation of the CIR is highly desirable, in order to mitigate interference and achieve reliable data detection.

In many high-speed wireless applications, the propagation channel can be modeled as a specular channel consisting of a relatively small number of dominant rays [1], with each one being characterized by its time delay and attenuation factor. Then, provided that the transmitter and receiver filters are known [2], the channel estimation task is reduced to that of estimating the parameters of the multipath channel components, i.e., the time delays and the attenuation factors. Two are the main advantages of using such a parametric approach. First, the number of required data is reduced, resulting in a significant saving in complexity, processing delay and, possibly, in bandwidth. Second, the mean square channel estimation error is expected to be lower since a more parsimonious parameterization of the channel is adopted, compared to conventional CIR estimation.

In a wireless communication channel, bandwidth is a precious resource, therefore the need for training sequence reduction is imperative. Thus, conventional training based channel estimation methods are inappropriate, especially when the channel span is large as in high rate applications. On the other hand, blind channel estimation methods require a high number of data and have very often limited performance. In such cases semi-blind estimation techniques may be employed [3–6]. The main trait of a semi-blind scheme is that a purely blind criterion is suitably modified so as to incorporate information from a short training sequence [3]. The performance of a semi-blind method, if properly designed, can be significantly improved against the corresponding blind method.

The estimation of a multipath channel using a parametric modeling has gained considerable attention in recent years. Blind [7–9] as well as training based [10–12] channel parametric estimation methods have been recently proposed. These methods either rely on preliminary estimates of the CIR [8,10–12] or directly estimate the parameters of the multipath channel components [7,9]. Recently, parametric channel modeling has also been applied to channel estimation in multicarrier (OFDM) [14] and multiuser (CDMA) [15,16] system design. In a blind context, an important implication of parametric channel modeling is that the channel overmodeling problem, which is inherent in second-order statistics blind identification algorithms, can be overcome [9]. That is, channel parameterization with respect to delays and attenuation factors, leads to consistent channel estimates even if the channel order is overestimated. However, such an improvement is traded off with a need to perform a computationally costly multidimensional search, the order of which equals the number of multipath components.

In this paper, we propose a novel semi-blind technique, which exploits the specular channel structure by incorporating the knowledge of the pulse shaping filter. By applying the well-known sub-channel response matching (SRM) criterion [18] to the problem at hand and after incorporating limited input information, we end up with a least squares (LS) problem, which is separable with respect to the unknown parameter sets, i.e. time delays and attenuation factors. Specifically, it is shown that the optimization problem can be separated to two different subproblems. One subproblem which is non-linear with respect to the time delays and another subproblem which is linear with respect to the attenuation parameters. After revealing the special structure of the non-linear problem, a computationally efficient linear search method for the estimation of the unknown time delays is developed. In the sequel, the Gauss–Newton (G–N) algorithm may be applied in order to further improve the accuracy of the estimated values. Finally, the attenuation parameters are estimated by solving a low-order linear LS problem. The new method is very simple to implement and achieves a lower channel estimation error compared to other related methods, e.g. [3], at a reasonable computational cost. A basic

feature of the proposed technique is that, due to the special form of the non-linear cost function, a highly inefficient multidimensional search is avoided and a small number of simple linear searches is executed instead. Moreover, the new method offers the possibility of trading off performance to complexity in an easy manner and yields good estimates even in cases of closely spaced time delays. The performance of the new method has been justified theoretically and tested through extensive simulations. Some preliminary results of this work have been presented in [17].

The paper is organized as follows. In Section 2 the multipath channel model is defined and the problem is formulated. In Section 3 the new method is derived and in Section 4 an efficient version of the proposed method is described. Performance issues of the new semi-blind technique are briefly discussed in Section 5 and simulation results are provided in Section 6. Concluding remarks are presented in Section 7.

2. Problem Formulation and preliminaries

2.1. Channel model

In general, the CIR encountered in wireless communication systems has a form which varies significantly depending on several factors, such as physical environment, transmission rate, mobility etc. However a common trait in most cases, particularly in high-speed applications, is that the multipath channel tends to be of a discrete form (i.e., it consists of a number of dominant multipath components). More specifically, if the CIR is assumed to be time invariant within a small-scale time interval, then it may be written as [1]

$$h_c(t) = \sum_{i=0}^{p-1} \alpha_i \delta(t - \tau_i), \tag{1}$$

where α_i and τ_i are the complex attenuation factor and the delay, respectively, of the ith multipath component. Without loss of generality, it is assumed that $\tau_0 < \tau_1 < \cdots < \tau_{p-1}$. Thus the problem of multipath CIR estimation is reduced to the problem of delay and complex attenuation parameters estimation, which is of smaller dimension.

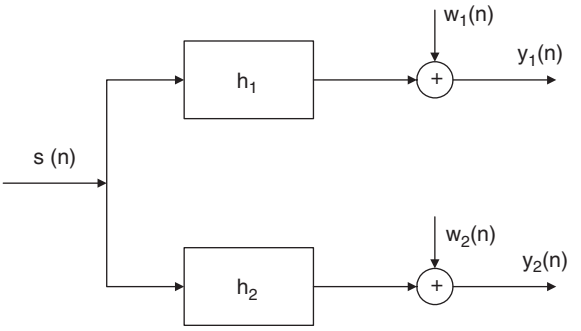

Fig. 1. Multichannel model.

Let $g(t)$ be the pulse shaping filter (convolution of transmitter and receiver filters). The overall impulse response of the communication system is then given as the convolution of $h_c(t)$ with $g(t)$, i.e.

$$h(t) = \sum_{i=0}^{p-1} \alpha_i g(t - \tau_i). \tag{2}$$

In this paper, it is assumed that the pulse shaping filter $g(t)$ is a known raised cosine function[1] and has finite support, i.e. $g(t) = 0$ for $t \notin [0, 2L_g T]$, where T is the symbol period. That is, at time $L_g T$ the raised cosine pulse attains its maximum value. After passing through the receiver filter, the received signal is oversampled by a factor of N_s samples per symbol period. For simplicity, in our derivations we take $N_s = 2$, and hence the channel structure, which can be described by the multichannel model, has the form shown in Fig. 1.[2] Thus, the sampled CIR is expressed by two vectors, one for each sub-channel, i.e. for $i = 1, 2$

$$\boldsymbol{h}_i^{\mathrm{T}} = \left[h(\tfrac{(i-1)}{2}T) h(T + \tfrac{(i-1)}{2}T) \cdots h(LT + \tfrac{(i-1)}{2}T) \right],$$

where LT is the span of the overall CIR (i.e. $L = 2L_g + \lceil \frac{\tau_{p-1}}{T} \rceil - 1$). It is straightforward to write the sub-channels' impulse responses in the form

$$\boldsymbol{h}_i = G_i(\boldsymbol{\tau})\mathbf{a}, \quad i = 1, 2, \tag{3}$$

where $\boldsymbol{\tau} = [\tau_0 \tau_1 \cdots \tau_{p-1}]^{\mathrm{T}}$ and $\mathbf{a} = [\alpha_0 \alpha_1 \cdots \alpha_{p-1}]^{\mathrm{T}}$. Finally, $G_i(\boldsymbol{\tau})$ is an $(L+1) \times p$ matrix whose

[1] Note that other pulse shaping functions can be used as well.

[2] The proposed method can easily be generalized for $N_s > 2$.

columns are delayed versions of $g(t)$ depending on the unknown parameters τ_i. Specifically this matrix has the form

$$G_i(\tau)=\begin{bmatrix} g(\frac{i-1}{2}T-\tau_0) & \cdots & g(\frac{i-1}{2}T-\tau_{p-1}) \\ g(T+\frac{i-1}{2}T-\tau_0) & \ldots & g(T+\frac{i-1}{2}T-\tau_{p-1}) \\ \vdots & \ddots & \vdots \\ g(LT+\frac{i-1}{2}T-\tau_0) & \ldots & g(LT+\frac{i-1}{2}T-\tau_{p-1}) \end{bmatrix}.$$

The goal of the proposed method is the estimation of the unknown vectors $\boldsymbol{\tau}$ and $\mathbf{a}$, using a small number of information symbols, which are assumed to be known at the receiver. After estimating $\boldsymbol{\tau}$ and $\mathbf{a}$, the overall CIR can be directly computed from the parametric channel model given in (3). In the following section, this channel model will be exploited in order to derive an appropriate parametric cost function for the solution of the problem at hand.

2.2. The semi-blind parametric cost function

Let us write the sub-channels' output samples as follows

$$y_i(n)=\boldsymbol{h}_i^{\mathrm{T}}\boldsymbol{s}_L(n)+w_i(n), \quad i=1,2, \tag{4}$$

where $\boldsymbol{s}_L^{\mathrm{T}}(n)=[s(n),s(n-1),\ldots,s(n-L)]$ is the common input data vector and $w_i(n)$ stands for the output noise of sub-channel i at time n. The input sequence is assumed to be i.i.d. and independent of the noise sequences.

The starting point of our derivation is the SRM concept, which wisely exploits the multichannel structure [18]. More specifically, in the two channels, noise-free case, the following relation holds

$$\{y_1\}\star\{h_2\}=\{y_2\}\star\{h_1\}, \tag{5}$$

where $\star$ stands for convolution and $\{y_i\}$, $i=1,2$, denotes the output sequence of sub-channel i. Considering that output samples $y_i(k)$, $y_i(k+1),\ldots,y_i(k+N-1)$, $i=1,2$, $N>L$, are available at the receiver, (5) leads to the following system of equations

$$[Y_2 \;\; -Y_1]\begin{bmatrix}\boldsymbol{h}_1\\ \boldsymbol{h}_2\end{bmatrix}=\mathbf{0}, \tag{6}$$

where

$$Y_i=\begin{bmatrix} y_i(k+L) & \cdots & y_i(k) \\ y_i(k+L+1) & \cdots & y_i(k+1) \\ \vdots & \vdots & \vdots \\ y_i(k+N-1) & \cdots & y_i(k+N-L-1) \end{bmatrix} \tag{7}$$

for $i=1,2$. Let us further assume that the information symbols $s(k-L),s(k-L+1),\ldots,s(k+M)$ are known at the receiver, where M is of the order of L and $M\ll N$. Then, if we take into account the channel input–output equation (4) and after ignoring for the moment the involved noise, the system of equations in (6) can be augmented as

$$\begin{bmatrix} Y_2 & -Y_1 \\ S_{ML} & \mathbf{0} \\ \mathbf{0} & S_{ML} \end{bmatrix}\begin{bmatrix}\boldsymbol{h}_1\\ \boldsymbol{h}_2\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\ \boldsymbol{y}_{1M}\\ \boldsymbol{y}_{2M}\end{bmatrix}, \tag{8}$$

where $\boldsymbol{y}_{iM}^{\mathrm{T}}=[y_i(k),y_i(k+1),\ldots,y_i(k+M)]$ for $i=1,2$ and

$$S_{ML}=\begin{bmatrix} s(k) & \cdots & s(k-L) \\ s(k+1) & \cdots & s(k+1-L) \\ \vdots & \vdots & \vdots \\ s(k+M) & \cdots & s(k+M-L) \end{bmatrix}. \tag{9}$$

By imposing the channel parametric structure, Eq. (8) can be written as

$$Y_S G(\boldsymbol{\tau})\mathbf{a}=\boldsymbol{z}, \tag{10}$$

where

$$Y_S=\begin{bmatrix} Y_2 & -Y_1 \\ S_{ML} & \mathbf{0} \\ \mathbf{0} & S_{ML} \end{bmatrix}, \quad \boldsymbol{z}=\begin{bmatrix}\mathbf{0}\\ \boldsymbol{y}_{1M}\\ \boldsymbol{y}_{2M}\end{bmatrix},$$

$$G(\boldsymbol{\tau})=\begin{bmatrix}G_1(\boldsymbol{\tau})\\ G_2(\boldsymbol{\tau})\end{bmatrix}. \tag{11}$$

When the channel is corrupted by noise, we can estimate the channel parameters $\mathbf{a},\boldsymbol{\tau}$ by solving the following LS problem

$$\min_{\mathbf{a},\boldsymbol{\tau}} \|\boldsymbol{z}-\Phi(\boldsymbol{\tau})\mathbf{a}\|^2, \quad \text{where } \Phi(\boldsymbol{\tau})=Y_S G(\boldsymbol{\tau}). \tag{12}$$

In the next section the form of the non-linear cost function of (12) will be explored, in order to reveal and take advantage of its possible salient characteristics.

3. Derivation of the algorithm

In this section an efficient method for estimating the multipath channel parameters is described, by exploiting the special form of the non-linear cost function given in (12). In particular, the cost function is shown to possess the following two properties. First, the optimization problem can be split up into two subproblems in terms of the delay and attenuation parameters, respectively. Second, the delays are shown to be decoupled from each other, allowing for their efficient estimation.

3.1. Separable form of the problem

It is readily seen that the non-linear LS problem in (12) is separable with respect to the unknown parameters $\boldsymbol{\tau}$ and $\mathbf{a}$. In particular, the LS cost function is non-linear with respect to the delay parameters $\boldsymbol{\tau}$ and linear with respect to the attenuation parameters $\mathbf{a}$. Thus, it can be proven [19, Chapter 9] that the optimization process can be conducted separately with respect to the distinct parameter sets $\boldsymbol{\tau}$ and $\mathbf{a}$. More specifically

- The delay parameters $\boldsymbol{\tau}$ are obtained via the solution of the following non-linear optimization problem

$$\boldsymbol{\tau}_{\text{opt}} = \arg\min_{\boldsymbol{\tau}} \{f(\boldsymbol{\tau})\}, \tag{13}$$

where

$$f(\boldsymbol{\tau}) = \|(I - \Phi(\boldsymbol{\tau})\Phi^{\dagger}(\boldsymbol{\tau}))\mathbf{z}\|^2 \tag{14}$$

and $\dagger$ denotes the pseudoinverse of a matrix.
- The attenuation parameters $\mathbf{a}$ are determined by the linear LS method as (assuming that the tall matrix $\Phi(\boldsymbol{\tau}_{\text{opt}})$ is full rank)

$$\mathbf{a}_{\text{opt}} = \Phi^{\dagger}(\boldsymbol{\tau}_{\text{opt}})\mathbf{z}. \tag{15}$$

Note, from (13) and (14), that $\boldsymbol{\tau}_{\text{opt}}$ is the value of the delay vector, which minimizes the projection of vector $\mathbf{z}$ to the orthogonal complement of the space spanned by the columns of $\Phi(\boldsymbol{\tau})$. It is readily understood that the same vector maximizes the projection of $\mathbf{z}$ to the column space of $\Phi(\boldsymbol{\tau})$. Consequently, the optimization problem of (13) and (14) can be written in the following equivalent form

$$\boldsymbol{\tau}_{\text{opt}} = \arg\max_{\boldsymbol{\tau}} \{F(\boldsymbol{\tau})\},$$
$$\text{where } F(\boldsymbol{\tau}) = \|\Phi(\boldsymbol{\tau})\Phi^{\dagger}(\boldsymbol{\tau})\mathbf{z}\|^2. \tag{16}$$

We see that the most difficult part of the initial optimization problem (12), is actually solved by maximizing $F(\cdot)$ in (16). A non-linear cost function of the type of (16) could be optimized either by performing a multidimensional search in the space of the parameter set $\boldsymbol{\tau}$ or by applying a non-linear optimization search method, e.g. a Newton type method. In the former case the computational burden may be prohibitive whereas in the latter the procedure may be trapped in a local minimum, away from the global solution. In the following section, we show that the cost function $F(\cdot)$ possesses a special form and derive an efficient and accurate method for the detection of the optimum delay vector.

3.2. Decoupling with respect to delays

Let us start with an example and assume that the multipath channel contains two unknown delays, namely τ_0 and τ_1. In Fig. 2, the function $f(\boldsymbol{\tau})$ is plotted versus τ_0, τ_1 for a two-ray multipath channel. It is clear from this figure that the optimum point is at the intersection of two perpendicular trenches corresponding to the delays τ_0 and τ_1. Thus, instead of performing a 2-dimensional search, two 1-dimensional searches are adequate for locating the global minimum. More specifically, one can initially assign any value to τ_1, within a range, and then minimize the cost function by varying parameter τ_0. In the sequel, the same procedure can be repeated for parameter τ_1 after fixing the value of τ_0 at the optimum value obtained from the first search. Note that, as expected, function $F(\cdot)$ has a similar (actually complementary) form with $f(\cdot)$ and hence a similar approach can be followed.

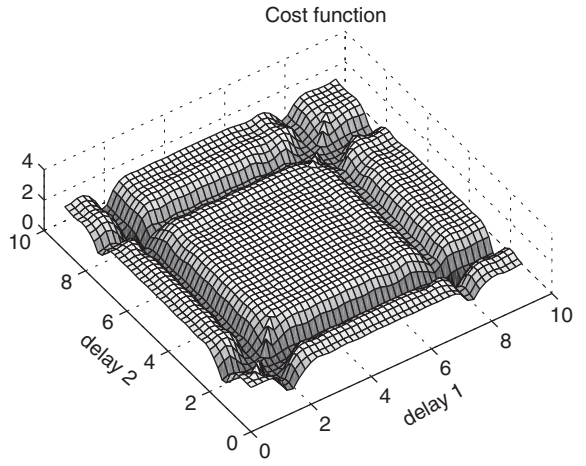

Fig. 2. Cost function, $\boldsymbol{\tau} = [1.2\ \ 7.4]^{\mathrm{T}}$ and $\mathbf{a} = [0.5\ \ 0.5]^{\mathrm{T}}$.

In the lines to follow, we generalize the concept described above and show that under certain conditions, which are satisfied in the applications of interest, the cost function can be decoupled with respect to the delay parameters. Indeed, since $\Phi(\boldsymbol{\tau})\Phi^{\dagger}(\boldsymbol{\tau})$ is a projection operator, the cost function $F(\cdot)$ is written as follows

$$F(\boldsymbol{\tau}) = \|\Phi(\boldsymbol{\tau})\Phi^{\dagger}(\boldsymbol{\tau})\boldsymbol{z}\|^2 = \boldsymbol{z}^{\mathrm{H}}\Phi(\boldsymbol{\tau})(\Phi^{\mathrm{H}}(\boldsymbol{\tau})\Phi(\boldsymbol{\tau}))^{-1}\Phi^{\mathrm{H}}(\boldsymbol{\tau})\boldsymbol{z}, \tag{17}$$

where $(\cdot)^{\mathrm{H}}$ denotes the conjugate transpose operation. By setting $\boldsymbol{v}(\boldsymbol{\tau}) = \Phi^{\mathrm{H}}(\boldsymbol{\tau})\boldsymbol{z}$, (17) is written as

$$F(\boldsymbol{\tau}) = \boldsymbol{v}^{\mathrm{H}}(\boldsymbol{\tau})C^{-1}(\boldsymbol{\tau})\boldsymbol{v}(\boldsymbol{\tau}) \tag{18}$$

with

$$C(\boldsymbol{\tau}) = G^{\mathrm{T}}(\boldsymbol{\tau})Y_S^{\mathrm{H}}Y_S G(\boldsymbol{\tau}) \tag{19}$$

and

$$\boldsymbol{v}^{\mathrm{H}}(\boldsymbol{\tau}) = [\boldsymbol{z}^{\mathrm{H}}Y_S\boldsymbol{g}(\tau_0)\ \ \boldsymbol{z}^{\mathrm{H}}Y_S\boldsymbol{g}(\tau_1)\cdots\boldsymbol{z}^{\mathrm{H}}Y_S\boldsymbol{g}(\tau_{p-1})] = [v(\tau_0)\ \ v(\tau_1)\ \ \cdots\ \ v(\tau_{p-1})], \tag{20}$$

where $\boldsymbol{g}(\tau_i)$ is the $(i-1)$th column of $G(\boldsymbol{\tau})$. That is, the ith element of vector $\boldsymbol{v}$ is a function of the $(i-1)$th delay parameter only, for $i = 1, 2, \ldots, p$. Due to this form of vector $\boldsymbol{v}$, we deduce from (18) that $F(\boldsymbol{\tau})$ would be decoupled with respect to the delay parameters, if matrix $C^{-1}(\boldsymbol{\tau})$ was diagonal and its ith diagonal element was a function of τ_{i-1} only. Although, such an ideal case is not valid in general, it will be shown that, under reasonable assumptions, matrix $C^{-1}(\boldsymbol{\tau})$ is strongly diagonally dominant, i.e., its diagonal part is much 'heavier' than the off-diagonal part and its ith diagonal element is approximately a function of τ_{i-1} only. The derivation is based on the following general proposition, which is proven in Appendix A.

Proposition 1. *Let A be an $n \times n$ matrix with complex elements in general. Let us also assume that its diagonal elements are all equal to one whereas the off-diagonal entries are considered as uncorrelated complex gaussian random variables with zero mean and variance σ_a^2 (i.e., the real and the imaginary part have variance $\sigma_a^2/2$). We also assume that $(n-1)\sigma_a^2 < 1$, in other words, at each row the sum of the off-diagonal elements has a variance less that unity. Let us next define*

$$B = Q^{\mathrm{H}}AQ, \tag{21}$$

where Q is an $n \times m$ unitary matrix, i.e., $Q^{\mathrm{H}}Q = I$, and the ratios

$$r_i^A = \frac{\sum_{\substack{j=1\\ j\neq i}}^{n} |a_{i,j}|}{|a_{i,i}|} \tag{22}$$

and

$$r_i^B = \frac{\sum_{\substack{j=1\\ j\neq i}}^{m} |b_{i,j}|}{|b_{i,i}|}, \tag{23}$$

where $a_{i,j}$ and $b_{i,j}$ are the (i,j)th elements of matrices A and B, respectively. Then, we maintain that

$$\mathrm{mean}\{r_i^B\} \leqslant \frac{m}{n}\mathrm{mean}\{r_i^A\} \tag{24}$$

and

$$\mathrm{var}\{r_i^B\} \leqslant \frac{m^2}{n}\mathrm{var}\{r_i^A\} \tag{25}$$

where mean$\{\cdot\}$ *and* var$\{\cdot\}$ *denote mean value and variance, respectively.*

Comments: (a) If the ratio r_i^A is smaller than or equal to unity, for $i = 1, \ldots, n$, then matrix A is diagonally dominant. Moreover, the smaller than unity these ratios are the more dominant matrix A is. Since the off-diagonal elements of A are taken as random variables we can only make statistical considerations about the diagonal dominance of this matrix. How often matrix A is dominant and its degree of dominance depend on n and σ_a^2. The

important implication of relations (24) and (25) is that for $m^2 < n$ the ratios r_i^B, $i = 1, \ldots, m$, have a much smaller mean value and a smaller variance as compared to r_i^A and therefore matrix B has in general a higher degree of dominance than matrix A. In fact, for $m \ll n$, matrix B tends to have a diagonal form.

(b) In the above proposition, $a_{i,j}$ are assumed to be uncorrelated gaussian random variables. However, in many matrices encountered in practice, $a_{i,j}$ should also be bounded. For instance, if A represents an autocorrelation matrix then its off-diagonal elements should have a magnitude less than or equal to unity (which is the magnitude of the diagonal elements). In such a case it is suitable to consider the off-diagonal elements as truncated gaussian random variables. Although the proposition can easily be extended to include this case, it turns out that such an extension is not really necessary since it has also been assumed that $a_{i,j}$ have a variance much smaller than unity, and therefore the probability to take a value larger than unity tends to zero. Another possible extension of the proposition is concerned with the assumption that $a_{i,j}$ are uncorrelated. Indeed, the proposition can be extended to include the case of correlated $a_{i,j}$ as well as that of matrices with a particular structure (i.e., Toeplitz). In such an extension, it is easy to show that, the relations for the mean values remain the same and inequality (24) is still valid. It is, however, a more intricate task to derive expressions for the variances and this is not treated here any further.

Let us now return and see how Proposition 1 applies to the problem at hand. For the moment, we assume that the multipath components are adequately separated (i.e., the minimum time delay difference between two components is greater than T). The case of closely-spaced delays is treated at the end of this subsection. From (11) and (19) we get

$$C(\tau) = X(\tau) + V(\tau) + U(\tau), \tag{26}$$

where

$$X(\tau) = G_1^{\mathrm{T}}(\tau) S_{ML}^{\mathrm{H}} S_{ML} G_1(\tau) + G_2^{\mathrm{T}}(\tau) S_{ML}^{\mathrm{H}} S_{ML} G_2(\tau),$$

$$V(\tau) = G_1^{\mathrm{T}}(\tau) Y_2^{\mathrm{H}} Y_2 G_1(\tau) - G_1^{\mathrm{T}}(\tau) Y_2^{\mathrm{H}} Y_1 G_2(\tau),$$

$$U(\tau) = G_2^{\mathrm{T}}(\tau) Y_1^{\mathrm{H}} Y_1 G_2(\tau) - G_2^{\mathrm{T}}(\tau) Y_1^{\mathrm{H}} Y_2 G_1(\tau).$$

Proposition 1 can directly be applied to the two terms of $X(\tau)$. Indeed, due to the i.i.d. property of the input symbols, their sampled autocorrelation matrix $S_{ML}^{\mathrm{H}} S_{ML}$ is diagonally dominant (it actually tends to a diagonal matrix for large M). Moreover, the columns of $G_i(\tau)$ for $i = 1, 2$ contain shifted versions of a raised cosine pulse shaping filter. The inner product of two columns of $G_i(\tau)$ approximates the value of the autocorrelation function of the raised cosine pulse for a lag equal to the difference of the corresponding time delays. The raised cosine autocorrelation function is given in Appendix D and is plotted in Fig. 3 for a roll-off factor equal to 0.3 and $T = 1$. We observe from Fig. 3 that even for a difference as small as one symbol period the value of the inner product is one order of magnitude smaller than its maximum value, which corresponds to a zero time difference. Note also from Fig. 4 that this property is preserved independently of the roll-off factor value. Thus, matrices $G_i(\tau)$, properly normalized, have a structure which is practically close to a unitary matrix. By taking advantage of Proposition 1, we deduce that $X(\tau)$ has a much larger degree of diagonal dominance than $S_{ML}^{\mathrm{H}} S_{ML}$. Note that the parameters m and n of Proposition 1 take the values p and $L + 1$, respectively, and therefore in relations (24) and (25) the factor m/n is very small and even m^2/n has a value smaller than unity for the channels we consider here.

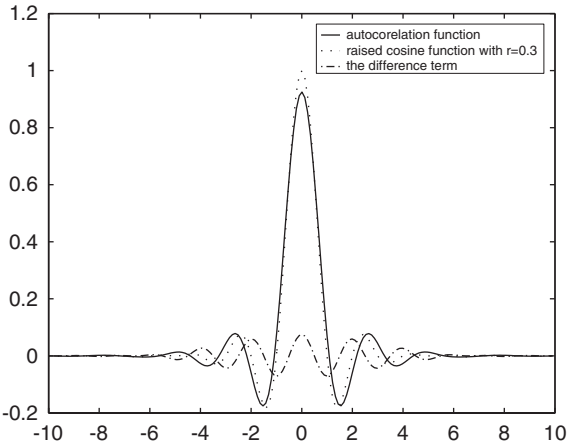

Fig. 3. Raised cosine autocorrelation function.

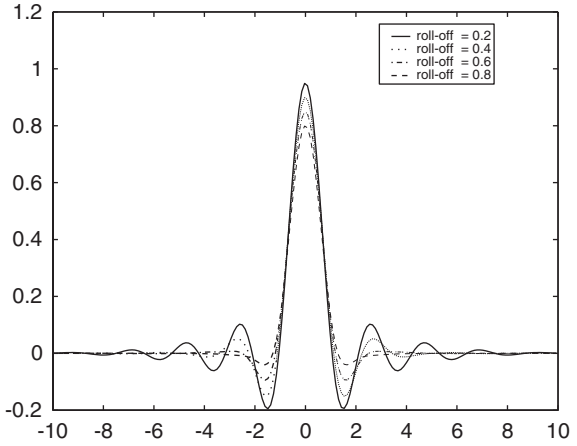

Fig. 4. Raised cosine autocorrelation function for different roll-off factors.

As shown in Appendix B, the contribution of the terms $V(\boldsymbol{\tau})$ and $U(\boldsymbol{\tau})$ in the RHS of (26) is such that the diagonally dominant structure of $C(\boldsymbol{\tau})$, which is due to $X(\boldsymbol{\tau})$, is preserved. More specifically, we concentrate on the matrix $V(\boldsymbol{\tau})$ and show that the magnitude of its elements is much lower compared to the magnitude of the diagonal elements of $X(\boldsymbol{\tau})$. Note that a similar approach can also be applied for the final term $U(\boldsymbol{\tau})$.

From the previous analysis, it turns out that, $C(\boldsymbol{\tau})$ tends to a matrix not simply diagonally dominant but, in fact, with a very 'heavy' diagonal compared to the rest part of the matrix. Due to its special structure, matrix $C(\boldsymbol{\tau})$ can be written as $\varrho(I+\tilde{C}(\boldsymbol{\tau}))$ where $\tilde{C}(\boldsymbol{\tau})$ is a matrix with a norm much less than unity and ϱ a complex number. Thus, it is straightforward to show using Taylor expansion of $C(\boldsymbol{\tau})$, that the inverse of this matrix, i.e., $C^{-1}(\boldsymbol{\tau})$, possesses the same near-diagonal structure. Furthermore, from (19) and the form of $G(\boldsymbol{\tau})$, it is easily shown that each diagonal element of $C(\boldsymbol{\tau})$ is a function of a single delay parameter only. Obviously, the same holds for the diagonal elements of $\tilde{C}(\boldsymbol{\tau})$. Thus, from the first-order Taylor approximation of $C^{-1}(\boldsymbol{\tau})$ i.e., $1/\varrho(I-\tilde{C}(\boldsymbol{\tau}))$, we see that the diagonal elements of $C^{-1}(\boldsymbol{\tau})$ can also be considered dependent on single delay parameters. As a result, cost function $F(\boldsymbol{\tau})$ can be approximated by a summation of p positive terms, each of which is a function of one delay parameter only. Therefore, the optimization search can be performed separately for each τ_i, $i=0,1,\ldots,p-1$ and independently of the other delay parameters. In other words, instead of a p-dimensional search of exponential complexity, p linear searches are adequate to obtain a solution sufficiently close to the optimum one.

3.2.1. Closely-spaced delays

In case there exist two closely-spaced delays, the near-to-unitary structure of matrices $G_i(\boldsymbol{\tau})$, $i=1,2$ is violated and Proposition 1 does not apply as accurately as before. However, by properly extending Proposition 1 (see Appendix C), it can be shown that the decoupling of the delay parameters depends directly on the raised cosine autocorrelation function and even for delay spacings smaller than T, delay decoupling may be preserved. In any case, the closely spaced delays remain decoupled with respect to the other delays and the proposed algorithm still applies by properly adjusting the attenuation parameters. This observation is also verified in the simulations of Section 4.

3.3. Solution refinement using Gauss–Newton

The solution obtained from the procedure described above, say $\boldsymbol{\tau}^{(0)}$, can be further improved by applying a few steps of a G–N search method. More specifically, starting from the cost function in (17) the following update formula can be used for the delay vector

$$\boldsymbol{\tau}^{(k+1)}=\boldsymbol{\tau}^{(k)}-\mu R^{-1}(\boldsymbol{\tau}^{(k)})F'(\boldsymbol{\tau}^{(k)}), \tag{27}$$

where the superscript (k) indicates the kth iteration step of the G–N procedure, for $k=0,1,\ldots$. Parameter μ is the step size of the G–N search and $F'(\cdot)$, $R(\cdot)$ stand for the first differential of $F(\cdot)$ and its Hessian matrix, respectively. It can be shown following a procedure similar to [20] that

$$[F'(\boldsymbol{\tau})]_i=-E^{\mathrm{H}}(\mathbf{a},\boldsymbol{\tau})Y_S\frac{\mathrm{d}\mathbf{g}(\tau_{i-1})}{\mathrm{d}\tau_{i-1}}\alpha_{i-1}, \tag{28}$$

$$[R(\boldsymbol{\tau})]_{ij} = \alpha_{i-1}^{*}\left(\frac{\mathrm{d}\mathbf{g}(\tau_{i-1})}{\mathrm{d}\tau_{i-1}}\right)^{\mathrm{T}} Y_S^{\mathrm{H}}(I-\Phi(\boldsymbol{\tau})) \times(\Phi^{\mathrm{H}}(\boldsymbol{\tau})\Phi(\boldsymbol{\tau}))^{-1}\Phi^{\mathrm{H}}(\boldsymbol{\tau})Y_S\frac{\mathrm{d}\mathbf{g}(\tau_{j-1})}{\mathrm{d}\tau_{j-1}}\alpha_{j-1}, \tag{29}$$

where $[\cdot]_i$ and $[\cdot]_{i,j}$ denote the ith and (i,j)th elements of $F'(\boldsymbol{\tau})$ and $R(\boldsymbol{\tau})$, respectively, and the error vector $E(\mathbf{a},\boldsymbol{\tau})$ is given as follows

$$E(\mathbf{a},\boldsymbol{\tau}) = \boldsymbol{z} - \Phi(\boldsymbol{\tau})\mathbf{a}. \tag{30}$$

The expressions of the derivatives in (28) and (29) are provided in Appendix D, when a raised cosine function is assumed for the pulse shaping filters.

The vector of attenuation parameters is recomputed at each iteration of the G–N search as $\mathbf{a}^{(k)} = \Phi^{\dagger}(\boldsymbol{\tau}^{(k)})\boldsymbol{z}$. The G–N search converges to the global optimum point, provided that $\boldsymbol{\tau}^{(0)}$ is in the neighborhood of this point. As a result, the estimation accuracy is improved after running a few steps of the G–N search.

3.4. Summary of the new semi-blind parametric method

In order to apply the new estimation method, a p-dimensional grid must be initially defined with a linear step size, say δ. Then, $F(\boldsymbol{\tau})$ is evaluated in certain points of this grid according to the procedure described in Section 3.2. The basic steps of the proposed method, called hereafter decoupled estimation of channel parameters (DECP) method, are summarized in Table 1.

Note that the channel length L and the number of multipath channel components p may be overestimated, an issue which is further discussed in Sections 5 and 6. The selection of parameter δ controls the estimation accuracy of the DECP method. Apparently, as the size of δ decreases, the estimation accuracy of the method is improved.

Table 1
Summary of the DECP method

(1)	Set overestimated values $\hat{L}$, $\hat{p}$ for L, p, respectively.
(2)	Initialize τ_i, $i=0,1,\ldots,\hat{p}-1$ with distinct random values in the interval $[0, LT]$.
(3)	Choose a linear search step size, δ and set $i=0$.
(4)	Maximize $F(\boldsymbol{\tau})$ with respect to τ_i. Find $\tau_{i,\mathrm{opt}}$ by evaluating the function at $\tau_i = j\delta$, $j=0,1,\ldots,\frac{\hat{L}T}{\delta}$
(5)	Set $\tau_i = \tau_{i,\mathrm{opt}}$, $i=i+1$ and repeat from step 3 until $i=\hat{p}$.
(6)	Obtain the attenuation parameters from (15).
(7)	Run a G–N search in the neighborhood of $\boldsymbol{\tau}_{\mathrm{opt}}$ to improve the estimation accuracy for parameters $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$.

4. An alternative version of the DECP method

In this section it is shown that by further exploiting the form of $F(\boldsymbol{\tau})$, an alternative equivalent procedure can be developed for estimating in an even more efficient manner the delay parameters. Let us first define the following quantities:

$$\boldsymbol{\phi}_i = Y_S\boldsymbol{g}(\tau_i), \quad \Phi_{i:j} = [\boldsymbol{\phi}_i \ \ \boldsymbol{\phi}_{i+1} \ \ \cdots \ \ \boldsymbol{\phi}_j], \quad i,j = 0,1,\ldots,p-1, \ \ j \geqslant i. \tag{31}$$

That is, $\boldsymbol{\phi}_i$ is the $(i+1)$th column of matrix $\Phi(\boldsymbol{\tau})$ and $\Phi_{i:j}$ stands for its sub-matrix formed by columns $i+1$ through $j+1$.

Based on these definitions, matrix $C(\boldsymbol{\tau})$ given in (19) can be written as

$$C(\boldsymbol{\tau}) = \begin{bmatrix} \Phi_{0:p-2}^{\mathrm{H}}\Phi_{0:p-2} & \Phi_{0:p-2}^{\mathrm{H}}\boldsymbol{\phi}_{p-1} \\ \boldsymbol{\phi}_{p-1}^{\mathrm{H}}\Phi_{0:p-2} & \boldsymbol{\phi}_{p-1}^{\mathrm{H}}\boldsymbol{\phi}_{p-1} \end{bmatrix}. \tag{32}$$

It is then straightforward to show, using the matrix inversion lemma for partitioned matrices, that the inverse of $C(\boldsymbol{\tau})$ is expressed as

$$C^{-1}(\boldsymbol{\tau}) = \begin{bmatrix} (\Phi_{0:p-2}^{\mathrm{H}}\Phi_{0:p-2})^{-1} + \dfrac{\Phi_{0:p-2}^{\dagger}\boldsymbol{\phi}_{p-1}\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(\Phi_{0:p-2}^{\dagger})^{\mathrm{H}}}{\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(I-\Phi_{0:p-2}\Phi_{0:p-2}^{\dagger})\boldsymbol{\phi}_{p-1}} & \dfrac{-\Phi_{0:p-2}^{\dagger}\boldsymbol{\phi}_{p-1}}{\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(I-\Phi_{0:p-2}\Phi_{0:p-2}^{\dagger})\boldsymbol{\phi}_{p-1}} \\ \dfrac{-\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(\Phi_{0:p-2}^{\dagger})^{\mathrm{H}}}{\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(I-\Phi_{0:p-2}\Phi_{0:p-2}^{\dagger})\boldsymbol{\phi}_{p-1}} & \dfrac{1}{\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(I-\Phi_{0:p-2}\Phi_{0:p-2}^{\dagger})\boldsymbol{\phi}_{p-1}} \end{bmatrix}. \tag{33}$$

If we now substitute (33) in (18) and take into consideration that vector $\boldsymbol{v}(\boldsymbol{\tau})$ can be partitioned as

$$\boldsymbol{v}^{\mathrm{H}}(\boldsymbol{\tau}) = \boldsymbol{z}^{\mathrm{H}}[\Phi_{0:p-2} \quad \boldsymbol{\phi}_{p-1}] \tag{34}$$

then after some algebra we end up with the following expression for the cost function $F(\boldsymbol{\tau})$

$$F(\boldsymbol{\tau}) = \boldsymbol{z}^{\mathrm{H}}\Phi_{0:p-2}(\Phi_{0:p-2}^{\mathrm{H}}\Phi_{0:p-2})^{-1}\Phi_{0:p-2}^{\mathrm{H}}\boldsymbol{z} + \frac{\|\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(I-\Phi_{0:p-2}\Phi_{0:p-2}^{\dagger})\boldsymbol{z}\|^2}{\|\boldsymbol{\phi}_{p-1}^{\mathrm{H}}(I-\Phi_{0:p-2}\Phi_{0:p-2}^{\dagger})\|^2}. \quad (35)$$

We notice that the first term in the right-hand side (RHS) of (35) is the cost function of order $p-1$, i.e. the cost function which arises by assuming that the CIR consists of $p-1$ components. By further decomposing this term using the same procedure, $F(\boldsymbol{\tau})$ can be written in the final form

$$F(\boldsymbol{\tau}) = \frac{\|\boldsymbol{\phi}_0^{\mathrm{H}}\boldsymbol{z}\|^2}{\|\boldsymbol{\phi}_0\|^2} + \sum_{i=1}^{p-1}\frac{\|\boldsymbol{\phi}_i^{\mathrm{H}}(I-\Phi_{0:i-1}\Phi_{0:i-1}^{\dagger})\boldsymbol{z}\|^2}{\|\boldsymbol{\phi}_i^{\mathrm{H}}(I-\Phi_{0:i-1}\Phi_{0:i-1}^{\dagger})\|^2}. \quad (36)$$

We observe from the last equation that $F(\boldsymbol{\tau})$ consists of p non-negative terms. The first term is a function of τ_0 only and the ith term depends on $\tau_0,\ldots,\tau_{i-1}$. Note that due to the non-negative definiteness of these terms, their optimum points would coincide with the corresponding components of the global optimum point of $F(\boldsymbol{\tau})$, if they were independent. Although, this is not true, in general, the decoupling of the delay parameters discussed in Section 3.2 allows for their efficient estimation based on the form of $F(\boldsymbol{\tau})$ given in (36). As a result, $F(\boldsymbol{\tau})$ can be efficiently maximized following the alternative procedure summarized in Table 2 (so-called DECPa).

Table 2
Summary of the alternative DECP (DECPa) method

(1) Set overestimated values $\hat{L}$, $\hat{p}$ for L, p, respectively.
(2) Maximize the first term of $F(\boldsymbol{\tau})$ in (36) with respect to τ_0. Let $\tau_{0,\mathrm{opt}}$ be the optimum point.
(3) Substitute τ_0 with $\tau_{0,\mathrm{opt}}$ in the second term of $F(\boldsymbol{\tau})$ and maximize this term with respect to τ_1, which gives say $\tau_{1,\mathrm{opt}}$ as the optimum point.
(4) Repeat the same procedure (i.e., step 3) for τ_i, $i=2,\ldots,\hat{p}-1$ by substituting $\tau_0,\ldots,\tau_{i-1}$, with the optimum values obtained in the previous steps and maximizing the $i+1$th term of $F(\boldsymbol{\tau})$ with respect to τ_i.
(5) Obtain the attenuation parameters from (15).
(6) Run a G–N search in the neighborhood of $\boldsymbol{\tau}_{\mathrm{opt}}$ to improve the estimation accuracy for parameters $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$.

5. Performance issues

5.1. Computational complexity

It can be easily shown that the two versions of the new method presented in Tables 1 and 2 (i.e., DECP and DECPa, respectively) are completely equivalent, in terms of performance, by assuming a perfect decoupling of the delay parameters. However, even in practical cases, where decoupling is not perfect, the two algorithms perform in a very similar manner. This has been verified by extensive simulations, some of which are presented in Section 6.

To compare the two algorithms in terms of complexity, it should be taken into account that the DECPa algorithm consists of p maximization steps of orders $1,2,\ldots,p$, corresponding to the increasing orders of the terms in the RHS of (36). On the other hand, the original DECP approach, which evaluates $F(\boldsymbol{\tau})$ according to (17), requires the realization of p line searches of order p each. More specifically, the number of elementary operations (multiplications and additions) per iteration of the line search is $4(L+1)^2+2p(L+1)+p^3+p^2+2p$ for the DECP algorithm and $4(L+1)^2+p(L+1)+p^2/2+p$ for the DECPa algorithm. Thus, the DECPa algorithm is less computationally demanding compared to the DECP algorithm, with no substantial difference in performance. In general, as far as computational complexity is concerned, there exists a trade-off between the number of evaluation points K in the line searches, which is a function of parameter δ (i.e., $K=(LT/\delta)p$), and the number of iterations of the G–N method. However, it has been observed that by doubling the size of δ (thus reducing the number of evaluation points by half), the number of iterations of the G–N search needs to be increased only slightly to preserve the same performance. Therefore, it is preferable, in most cases, to use a relatively large value for the step parameter δ (e.g. $\delta=0.2T$) in order to achieve a low overall computational complexity.

5.2. Overmodeling

It is well known from the relevant literature (e.g., [9]) that the second-order statistics (SOS) blind channel identification algorithms suffer from the channel overmodeling problem, i.e., their performance dramatically deteriorates in case the channel order is overestimated. Such a performance degradation also arises in the blind SRM method, which forms the basis of the approach developed in this paper. However, two important issues distinguish the cost function of DECP (and DECPa) from the original SRM cost function. First, the proposed cost-function is a semi-blind one, thus incorporating properly the few input samples which are assumed to be known at the receiver. Even though the size of the "training part" of the algorithm is small compared to the size of the blind part, it is sufficient to set constraints on the possible solutions of the initial LS problem. Second, the parametric modeling of the channel leads to sound channel estimates, even if the channel order is overdetermined. Indeed, it has been shown in [9] that the blind subspace method [13], combined with a parametric channel modeling, gives consistent channel estimates when only an upper bound of the channel order is known. The ideas presented in [9] can be directly applied to the blind SRM method, from which the new methods originate. The immunity of the new semi-blind parametric technique in case of channel overmodeling is further highlighted in the following section, where a number of simulation results are presented.

6. Simulation results

In this section, we investigate the performance of the new parametric estimation method under various conditions. The two versions of the new method (i.e., DECP and DECPa) are compared with the semi-blind SRM-based method presented in [3], in which the CIR is estimated in a non-parametric manner, using a linear combination of a blind and a non-blind cost function. As a comparison measure we have used the root-mean-square-error between actual and estimated CIRs, i.e.

$$\mathrm{RMSE}=\sqrt{(1/P)\sum_{k=1}^{P}\sum_{i=1}^{2L}(h_{\mathrm{act}}(i)-h_{\mathrm{est}}^{k}(i))^{2}},$$

where P is the number of independent experiments, $h_{\mathrm{act}}(i)$ is the ith element of the actual $T/2$-spaced CIR and $h_{\mathrm{est}}^{k}(i)$ is the ith element of the estimated CIR during the kth simulation run. In our experiments, $P=100$ and both CIR sequences $\{h_{\mathrm{act}}\}$ and $\{h_{\mathrm{est}}^{k}\}$ are normalized to have unity norm. The input sequence was taken from a QAM-16 alphabet and the experiments have been conducted for SNR values between 16 and 36.

In the first experiment, the CIR to be estimated has the parameter vectors $\boldsymbol{\tau}=[0\ 6.11\ 18.17\ 23.34]$ and $\mathbf{a}=[0.9+0.1j\ \ -0.31-0.26j\ \ -0.4-0.3j\ \ 0.3-\ 0.15j]$. The pulse shaping filter is a raised cosine filter with a roll-off factor equal to 0.3 and $L_g=3$, i.e. the total channel span is $L=29$, assuming $T=1$. The number of output and training samples used for semi-blind estimation are 200 and 50, respectively (i.e. $N=200$ and $M=20$).The linear step size δ is set to 0.2. The step size of the G–N step is also initialized to 0.2 and is reduced during execution according to an appropriate strategy, while the total number of G–N iterations is preset to 20. In order to verify the validity of our approach, we first plot in Fig. 5 the mean and variance of the amplitude of $C(\boldsymbol{\tau})$, where averaging is considered over all delay parameter values that are used in the search. We observe that $C(\boldsymbol{\tau})$ retains a diagonally dominant structure during the search procedure in accordance with the theoretical analysis. In Fig. 6 the RMSE of the two versions of the new parametric semi-blind channel estimation method (i.e., DECP and DECPa) is compared with the RMSE of the algorithm proposed in [3] for different signal to noise ratios (SNRs).[3] To verify the need for solution refinement we have tested the new methods with and without the corresponding

[3]Performance comparisons have been also carried out with the semi-blind subspace (SBS) algorithm of [4], which is representative of a different class of semi-blind techniques. However, the RMSE of the SBS algorithm was higher even compared with the RMSE of the semi-blind SRM method [3] and we do not include this method in our simulations.

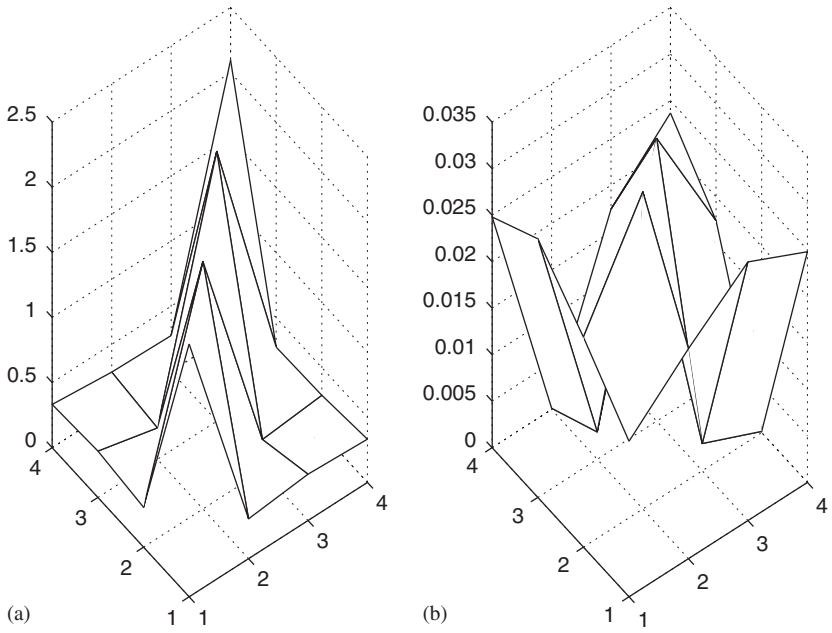

Fig. 5. (a) Mean and (b) variance of the amplitude of the elements of $C(\tau)$.

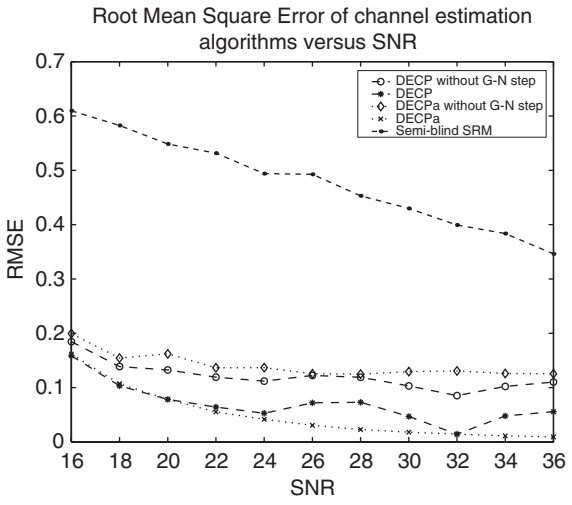

Fig. 6. Comparison of the two versions of the new method with semi-blind SRM.

G–N step. As it can be readily seen, the new methods outperform significantly the non-parametric semi-blind SRM method for all SNRs considered, while they exhibit almost similar performance. We observe that the situation is improved for both versions of the proposed algorithm, when the G–N step is activated even for a small number of iterations. Taking into account the significantly lower complexity of the second version (i.e., DECPa), we will use this method only in the rest comparisons.

In Fig. 7 the performance of the DECPa method is studied in case of closely-spaced time delays. More specifically, the delay of the third component of the CIR considered above is set to 22.77, resulting in a time difference equal to 0.57 symbol periods with respect to the fourth time delay parameter. Comparing Figs. 6 and 7 we notice that the estimation error of the parametric method remains unaffected.

The performance of the two methods in case of overestimation in the channel order L and the number of multipath components p is depicted in Fig. 8. In this experiment, the SNR is set to 25 dB and a number of 56 training symbols is used. Note that the number of training symbols has been slightly increased, in order to test the performance of the algorithms in a wide enough range of overmodeling values. We observe that the new

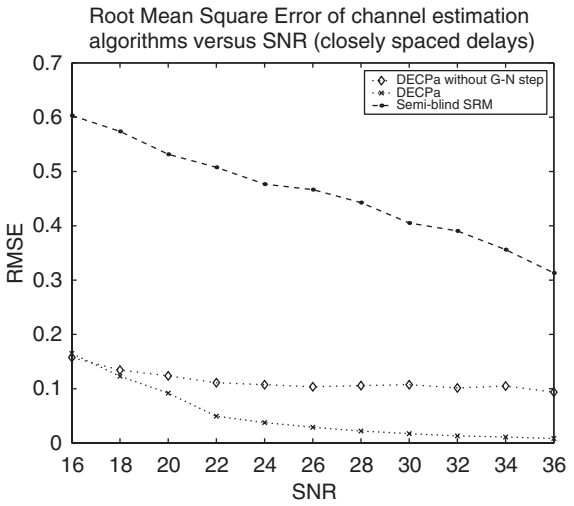

Fig. 7. Comparison of the new method with semi-blind SRM, in case of closely spaced delays.

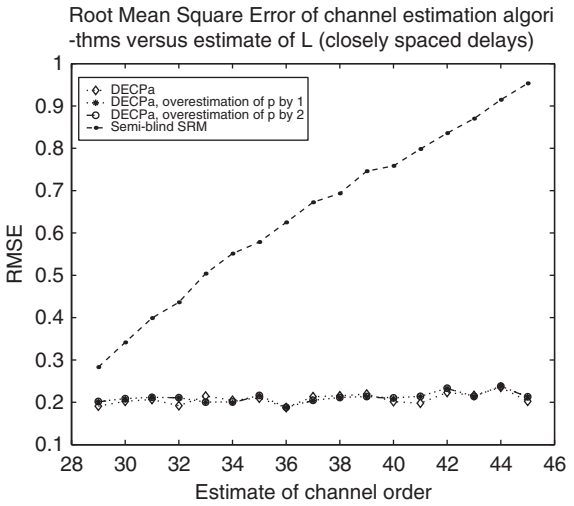

Fig. 8. Comparison of the new method with semi-blind SRM, in case of overmodeling.

method is only slightly affected in case of overmodeling by one or two with respect to the number of paths p. Furthermore, it is quite robust to channel order over-estimation, in contrast to the non-parametric semi-blind SRM, whose performance deteriorates significantly as the degree of overmodeling increases.

7. Conclusion

In this paper, we have proposed a new semi-blind estimation algorithm that exploits the specular channel structure and a-priori knowledge of the pulse shaping filter. The main contribution of this work is the observation that the optimized cost function is almost decoupled with respect to the time delays, a fact which allows for efficient estimation of the delay parameters. The new method is very simple to implement, it works well even for closely-spaced multipath components and is quite robust to overmodeling. It provides very good channel estimates with a very small number of training symbols, thus achieving a significant saving in bandwidth compared to training-based methods. The algorithm can be used as the first stage of an equalization module (e.g. interacting with a decision feedback equalizer) to obtain the transmitted data symbols [21]. The idea developed in this paper can also be extended to multicarrier (e.g. OFDM) systems, assuming a parametric channel modeling [14]. These are topics under current investigation.

Appendix A. Proof of Proposition 1

From the definition of B in (21) it can be shown that the (i,j)th element of B can be written as

$$b_{i,j} = \delta_{i,j} + e_{i,j}, \tag{37}$$

where

$$\delta_{i,j} = \sum_{k=1}^{n} a_{k,k} q_{i;k}^{*} q_{j;k} \tag{38}$$

and

$$e_{i,j} = \sum_{\substack{l=1 \\ l \neq r}}^{n} \sum_{r=1}^{n} a_{l,r} q_{i;l}^{*} q_{j;r}, \tag{39}$$

where $q_{i;j}$ is the jth element of the ith column of matrix Q, which, in turn, is denoted as $\mathbf{q}_i$. Exploiting the unitarian structure of Q and the fact that quantity $\delta_{i,j}$ contains the diagonal

elements of A, we obtain that

$$\delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{40}$$

Quantity $e_{i,j}$ is, by definition, a sum of weighted uncorrelated gaussian random variables (i.e., $a_{l,r}$ weighted by $q^*_{i;l}q_{j;r}$). It can be readily shown that $e_{i,j}$ is a zero mean complex gaussian random variable with variance equal to

$$\sigma_e^2 = \sigma_a^2 \sum_{\substack{l=1 \\ l\neq r}}^{n} \sum_{r=1}^{n} |q^*_{i;l}|^2 |q_{j;r}|^2. \tag{41}$$

By adding in the double sum of (41) the terms for which $l = r$ we get the inequality,

$$\sigma_e^2 \leqslant \sigma_a^2 \sum_{l=1}^{n} \sum_{r=1}^{n} |q^*_{i;l}|^2 |q_{j;r}|^2. \tag{42}$$

To proceed further let us define matrix P as

$$P = \boldsymbol{q}_j \boldsymbol{q}_i^{\mathrm{H}}. \tag{43}$$

The Frobenious norm of matrix P equals

$$\begin{aligned} \|P\|_{\mathrm{F}} &= \left(\sum_{l=1}^{n} \sum_{r=1}^{n} |p_{l,r}|^2 \right)^{1/2} \\ &= \left(\sum_{l=1}^{n} \sum_{r=1}^{n} |q^*_{i;l}|^2 |q_{j;r}|^2 \right)^{1/2}. \end{aligned} \tag{44}$$

Equivalently the Frobenious norm of P is given by

$$\begin{aligned} \|P\|_{\mathrm{F}} &= (\mathrm{tr}(P^{\mathrm{H}}P))^{1/2} = (\mathrm{tr}(\boldsymbol{q}_i \boldsymbol{q}_j^{\mathrm{H}} \boldsymbol{q}_j \boldsymbol{q}_i^{\mathrm{H}}))^{1/2} \\ &= (\mathrm{tr}(\boldsymbol{q}_i \boldsymbol{q}_i^{\mathrm{H}}))^{1/2} = 1, \end{aligned} \tag{45}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. From (44) and (45) we get

$$\sum_{l=1}^{n} \sum_{r=1}^{n} |q^*_{i;l}|^2 |q_{j;r}|^2 = 1. \tag{46}$$

Thus using the above equation and (42) we obtain

$$\sigma_e^2 \leqslant \sigma_a^2. \tag{47}$$

Notice that all random variables $e_{i,j}$ have the same statistical properties independently of the indices i,j, that is, they are gaussian random variables with zero mean and variance bounded as in (47).

The aim now is to compare the mean values and variances of random variables r_i^A and r_i^B. Starting from the definition of r_i^B in (23) and using (37) and (40) we get

$$r_i^B = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{m} |e_{i,j}|}{|1 + e_{i,i}|}. \tag{48}$$

Since $e_{i,i}$ in the denominator of the RHS of (48) is a zero mean random variable with a very small variance, we consider this variable as a negligible perturbation to unity. Thus it suffices to compare the numerators of r_i^A and r_i^B. Regarding, first, the numerator of r_i^A in (22) it can be shown, after some mathematical manipulations, that

$$\begin{aligned} \mathrm{mean}\left\{ \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{i,j}| \right\} &= (n-1)E\{|a_{i,j}|\} \\ &= (n-1)\left(\frac{\sigma_a^2 \pi}{4}\right)^{1/2} \end{aligned} \tag{49}$$

and

$$\mathrm{var}\left\{ \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{i,j}| \right\} = (n-1)\sigma_a^2 \left(\frac{4-\pi}{4}\right), \tag{50}$$

where for deriving the latter relation we have used the assumption that $a_{i,j}$ are uncorrelated random variables which implies that $E\{|a_{i,k}||a_{i,l}|\} = E\{|a_{i,k}|\}E\{|a_{i,l}|\}$.

Similarly the mean of the numerator of r_i^B in (48) is given by

$$\begin{aligned} \mathrm{mean}\left\{ \sum_{\substack{j=1 \\ j\neq i}}^{m} |e_{i,j}| \right\} &= (m-1)E\{|e_{i,j}|\} \\ &= (m-1)\left(\frac{\sigma_e^2 \pi}{4}\right)^{1/2} \\ &\leqslant (m-1)\left(\frac{\sigma_a^2 \pi}{4}\right)^{1/2}. \end{aligned} \tag{51}$$

To derive an expression for the variance of the numerator of r_i^B it should be taken into account that the random variables $|e_{i,j}|$ involved in the sum $s_e \equiv \sum_{\substack{j=1 \\ j\neq i}}^{m} |e_{i,j}|$ are correlated between each other. The mean squared value of s_e equals

$$E\{s_e^2\} = E\left\{\sum_{\substack{j=1 \\ j\neq i}}^{m} |e_{i,j}|^2\right\} + 2E\left\{\sum_{k,\, s\neq i,\, k\neq s} \{|e_{i,k}||e_{i,s}|\}\right\}, \tag{52}$$

where the second sum in the RHS of the above relation consists of $(m-1)(m-2)$ terms. Since the statistical properties of $|e_{i,j}|$ are independent of the values i,j, relation (52) can be written as

$$E\{s_e^2\} = (m-1)E\{|e_{i,j}|^2\} + (m-1)(m-2)E\{|e_{i,k}||e_{i,l}|\}. \tag{53}$$

According to Schwarz' inequality for the covariance of two random variables

$$E\{|e_{i,k}||e_{i,l}|\} \leqslant E\{|e_{i,j}|^2\}. \tag{54}$$

Using (54) in (53) we get

$$E\{s_e^2\} \leqslant (m-1)E\{|e_{i,j}|^2\} + (m-1)(m-2)E\{|e_{i,j}|^2\} = (m-1)^2\sigma_e^2. \tag{55}$$

Therefore the variance of s_e is written as

$$\mathrm{var}\{s_e\} = E\{s_e^2\} - (E\{s_e\})^2 \leqslant (m-1)^2\sigma_e^2 - (m-1)^2\frac{\sigma_e^2\pi}{4} \tag{56}$$

or

$$\mathrm{var}\{s_e\} \leqslant (m-1)^2\sigma_e^2\left(\frac{4-\pi}{4}\right) \leqslant (m-1)^2\sigma_a^2\left(\frac{4-\pi}{4}\right). \tag{57}$$

Based on the expressions (49), (50), (51) and (57), as well as on the approximation made for the denominator of (48), the relations (24) and (25) can be readily obtained, which was to be proved.

Appendix B. Magnitude of the elements of $V(\tau)$

Matrix $V(\tau)$ can be written as follows

$$V(\tau) = G_1^{\mathrm{T}}(\tau)Y_2^{\mathrm{H}}(Y_2G_1(\tau) - Y_1G_2(\tau)). \tag{58}$$

In the noise-free case, the output matrices Y_i, $i = 1,2$ can be expressed as

$$Y_i = SH_i \tag{59}$$

S is a $(N-L)\times(2L+1)$ input data matrix, whose $(j+1)$th row is $[s(k+L+j)\ s(k+L-1+j)\cdots s(k-L+j)]$ for $j = 0,1,\ldots,N-L-1$ and H_i stands for the $(2L+1)\times(L+1)$ convolution matrix formed by the impulse response $\mathbf{h}_i$ of subchannel i, i.e., H_i is a Toeplitz matrix with first row $[h(\frac{(i-1)}{2}T)\ \mathbf{0}_L^{\mathrm{T}}]$ and first column $[\mathbf{h}_i^{\mathrm{T}}\ \mathbf{0}_L^{\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{0}_L$ is a $L\times 1$ zero vector. Substituting (59) into (58) we get

$$V(\tau) = G_1^{\mathrm{T}}(\tau)H_2^{\mathrm{H}}S^{\mathrm{H}}S(H_2G_1(\tau) - H_1G_2(\tau)) \approx cG_1^{\mathrm{T}}(\tau)H_2^{\mathrm{H}}(H_2G_1(\tau) - H_1G_2(\tau)), \tag{60}$$

where we set $S^{\mathrm{H}}S \approx cI$ and c is of the order of $N-L$. Note that with proper normalization of the initial cost function, the diagonal elements of $S_{ML}^{\mathrm{H}}S_{ML}$ in the expression of $X(\tau)$, as well as c can be set to 1. Thus, after ignoring c and denoting by $\mathbf{g}_i(\tau_j)$, $j = 0,1,\ldots,p-1$, the $(j+1)$th column of $G_i(\tau)$, $i = 1,2$, the $(k+1,j+1)$th element of $V(\tau)$ is written as

$$[V(\tau)]_{k+1,j+1} = \mathbf{g}_1^{\mathrm{T}}(\tau_k)H_2^{\mathrm{H}}(H_2\mathbf{g}_1(\tau_j) - H_1\mathbf{g}_2(\tau_j)). \tag{61}$$

In the last equation, vector $H_2\mathbf{g}_1(\tau_j)$ coincides with the convolution of vectors $\mathbf{h}_2$ and $\mathbf{g}_1(\tau_j)$. Similarly, $H_1\mathbf{g}_2(\tau_j)$ is the convolution of $\mathbf{h}_1$ and $\mathbf{g}_2(\tau_j)$. Due to the form of $\mathbf{h}_1$, $\mathbf{h}_2$ expressed as

$$\mathbf{h}_1 = G_1(\tau)\mathbf{a} = \sum_{i=0}^{p-1}\alpha_i\mathbf{g}_1(\tau_i),$$

$$\mathbf{h}_2 = G_2(\tau)\mathbf{a} = \sum_{i=0}^{p-1}\alpha_i\mathbf{g}_2(\tau_i)$$

after some algebraic manipulations (61) can be written as follows

$$[V(\boldsymbol{\tau})]_{k+1,j+1} = \mathbf{g}_1^{\mathrm{T}}(\tau_k) H_2^{\mathrm{H}} \sum_{\substack{i=0 \\ i\neq j}}^{p-1} \alpha_i [\mathbf{g}_2(\tau_i) \star \mathbf{g}_1(\tau_j) - \mathbf{g}_1(\tau_i) \star \mathbf{g}_2(\tau_j)]. \quad (62)$$

It is not difficult to verify that both terms $\mathbf{g}_2(\tau_i)\star\mathbf{g}_1(\tau_j)$ and $\mathbf{g}_1(\tau_i)\star\mathbf{g}_2(\tau_j)$ in the RHS of (62) approximate the raised cosine autocorrelation function sampled at the same time instants, i.e., at $(\tau_i + \tau_j - T/2) \pm nT$, where n integer. Therefore, the difference of these terms for every i is a vector with a norm which tends to zero. More specifically, we can say that this norm is $\mathrm{O}(\varepsilon^l)$, where ε is a small positive real number with $0<\varepsilon<1$ and l an integer with $l>1$. By taking Euclidean norms on both sides of (62) and applying the Cauchy–Schwarz inequality we get

$$|[V(\boldsymbol{\tau})]_{k+1,j+1}| \leqslant \|\mathbf{g}_1^{\mathrm{T}}(\tau_k) H_2^{\mathrm{H}}\| \cdot \|E_j\| \cdot \|\mathbf{a}_j\|, \quad (63)$$

where E_j is a matrix whose columns are equal to $\mathbf{g}_2(\tau_i)\star\mathbf{g}_1(\tau_j) - \mathbf{g}_1(\tau_i)\star\mathbf{g}_2(\tau_j)$, for $i = 0, 1, \ldots, p-1$, $i\neq j$ and $\mathbf{a}_j$ comes from the attenuations' vector $\mathbf{a}$ by excluding element α_j. By assuming that $\|\mathbf{a}\|$ ($\|\mathbf{a}_j\|$) is O(1) and taking into consideration the form of $\mathbf{g}_1(\tau_k)$ and H_2, it is easily shown that $\|\mathbf{g}_1^{\mathrm{T}}(\tau_k) H_2^{\mathrm{H}}\|$ is also O(1). Thus, from (63), the size of $|[V(\boldsymbol{\tau})]_{k+1,j+1}|$ is upper bounded by $\|E_j\|$, which is $(p-1)\mathrm{O}(\varepsilon^l)$. Therefore, we conclude that the elements of $V(\boldsymbol{\tau})$ are much smaller than the diagonal elements of $X(\tau)$, which with proper normalization are O(1).

Appendix C. Closely-spaced delays

Let us assume that there exist two closely spaced delays.[4] In such a case, the assumption $Q^{\mathrm{H}}Q = I$ in Proposition 1, has to be replaced by the following expression

$$Q^{\mathrm{H}}Q = I + E, \quad (64)$$

where E is a symmetric matrix with only two non-zero elements equal to, say, ε_{ik}. With regard to the problem at hand, ε_{ik} corresponds approximately to the value of the raised cosine autocorrelation function with a lag equal to the time delays' difference. It is easily verified that after following the derivation of Appendix A with only slight modifications, Eq. (51) is rewritten as follows

$$\mathrm{mean}\{r_i^B\} \approx (m-2)\left(\frac{\sigma_e^2\pi}{4}\right)^{1/2} + \mathrm{mean}\{|e_{ik} + \varepsilon_{ik}|\}. \quad (65)$$

Since e_{ik} is Gaussian distributed with zero mean and variance σ_e^2, the random variable $|e_{ik} + \varepsilon_{ik}|$ follows a Ricean distribution. Note, however, that, for closely spaced delays, the ratio $\varepsilon_{ik}^2/\sigma_e^2$ is expected to be large. In such a case, the Ricean distribution tends to a Gaussian distribution with mean $|\varepsilon_{ik}|$ and variance σ_e^2. As a result, (65) takes the final form

$$\mathrm{mean}\{r_i^B\} \approx (m-2)\left(\frac{\sigma_e^2\pi}{4}\right)^{1/2} + |\varepsilon_{ik}|. \quad (66)$$

From the above approximation we see that, the diagonal dominance of matrix B depends mainly on $|\varepsilon_{ik}|$, which is the term that dominates in the RHS of (66). Translating this back to the problem under consideration, it turns out that the delay decoupling depends directly on the corresponding raised cosine autocorrelation value. From Fig. 4, we observe that even for a delay spacing as small as $0.5T$, delay decoupling may be preserved and is only slightly affected by the value of the roll-off factor.

Appendix D. Autocorrelation function and derivative of the raised-cosine function

By assuming that a raised cosine is used as a pulse shaping function, it can be shown that its autocorrelation function is given as follows

$$r_{RC}(t_1) = \frac{\sin(\pi t_1)\cos(\pi r t_1)}{\pi t_1(1-(2rt_1)^2)} - \frac{\cos(\pi t_1)\sin(\pi r t_1)}{\pi t_1(4-(2rt_1)^2)},$$

where r stands for the roll-off factor. Note that the first term in the RHS of the above equation coincides with the raised cosine function. Furthermore, the contribution of the second term is such

[4] Generalization to more than two closely spaced delays is straightforward.

that the overall autocorrelation function very closely resembles the raised cosine function itself. This is also verified in Fig. 3 for $r = 0.3$.

Concerning the derivative of the raised cosine function, it is easily shown that the kth element of the gradient under consideration is given by

$$\left(\frac{\mathrm{d}\mathbf{g}(\tau_i)}{\mathrm{d}\tau_i}\right)_k = \frac{(r\sin(\pi(t-\tau_i))\sin(r\pi(t-\tau_i)) - \cos(\pi(t-\tau_i))\cos(r\pi(t-\tau_i)))}{(t-\tau_i)(1-4r^2(t-\tau_i)^2)} - \frac{(12\pi r^2(t-\tau_i)^2-\pi)\sin(\pi(t-\tau_i))\cos(r\pi(t-\tau_i))}{\pi^2(t-\tau_i)^2(1-4r^2(t-\tau_i)^2)^2}, \tag{67}$$

where t is related to k according to the following relation

$$t = \begin{cases} (k-1)T & \text{if } 1 \leqslant k \leqslant L+1, \\ (L+2-k)T + T/2 & \text{if } L+1 \leqslant k \leqslant 2(L+1). \end{cases}$$

Note that due to the restricted support of the raised cosine function, the RHS of (67) is set to zero for $|t-\tau_i| > L_g T$. It is also zero for $t = \tau_i$, while in the singular points $\tau_i = t \pm \frac{1}{2r}$, it can be computed as follows

$$\left(\frac{\mathrm{d}\mathbf{g}(\tau_i)}{\mathrm{d}\tau_i}\right)_k\Bigg|_{\tau_i = t\pm\frac{1}{2r}} = \left(\left(\frac{\mathrm{d}\mathbf{g}(\tau_i)}{\mathrm{d}\tau_i}\right)_k\Bigg|_{\tau_i = t\pm\frac{1}{2r}-\varepsilon} + \left(\frac{\mathrm{d}\mathbf{g}(\tau_i)}{\mathrm{d}\tau_i}\right)_k\Bigg|_{\tau_i = t\pm\frac{1}{2r}+\varepsilon}\right)\Bigg/ 2,$$

where ε is a very small positive number.

References

[1] T.S. Rappaport, Wireless Communications Principles and Practice, Prentice-Hall, Englewood Cliffs, NJ, 1996.

[2] Z. Ding, Multipath channel identification based on partial system information, IEEE Trans. Signal Process. 45 (January 1997) 235–240.

[3] J.L. Babat, S.V. Schell, Channel identification using a combination of blind and nonblind methods, SPIE 2605 (1995) 202–211.

[4] A. Gorokhov, Ph. Loubaton, Semi-blind second order identification of convolutive channels, in: Proceedings of the ICASSP-97, Munich, April; 1997, pp. 3905–3908.

[5] E. de Carvalho, D. Slock, Deterministic quadratic semi-blind FIR multichannel estimation: Algorithms and performance, in: Proceedings of ICASSP, Instanbul, June 2000.

[6] G. Li, Z. Ding, Semi-blind channel identification for individual data bursts in GSM wireless systems, Signal Process. 80 (2000) 2017–2031.

[7] I. Kang, M.P. Fitz, S.B. Gelfand, Blind estimation of multipath channel parameters: A modal analysis approach, IEEE Trans. Comm. 47 (8) (August 1999) 1140–1150.

[8] M.C. Vanderveen, A. Paulraj, Improved blind channel identification using a parametric approach, IEEE Comm. Lett. 2 (8) (August 1998) 226–228.

[9] L. Perros-Meilhac, E. Moulines, K. Abed-Meraim, P. Chevalier, P. Duhamel, Blind identification of multipath channels: a parametric subspace approach, IEEE Trans. Signal Process. (July 2001) 1468–1480.

[10] M.C. Vanderveen, A. van der Veen, A. Paulraj, Estimation of multipath channel parameters in wireless communications, IEEE Trans. Signal Process. 46 (3) (March 1998) 682–690.

[11] J.-T. Chen, A. Paulraj, U. Reddy, Multichannel maximum-likelihood sequence estimation (MLSE) equalizer for GSM using a parametric channel model, IEEE Trans. Comm. 47 (1) (January 1999) 53–63.

[12] S. Özen, M.D. Zoltowski, Time-of-arrival (TOA) estimation based structured sparse channel estimation algorithm with applications to digital TV-receivers, in: Proceedings of ICASSP-03, vol. IV, Hong Kong, April 2003, pp. 481–484.

[13] E. Moulines, P. Duhamel, J.-F. Cardoso, S. Mayrargue, Subspace methods for the blind identification of multichannel FIR filters, IEEE Trans. Signal Process. 43 (February 1995) 516–525.

[14] B. Yang, K.B. Letaief, R.S. Cheng, Z. Cao, Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling, IEEE Trans. Comm. 49 (3) (March 2001) 467–479.

[15] A.A. D'Amico, U. Mengali, M. Morelli, Multipath channel estimation for the uplink of a DS-CDMA system, in: Proceedings of ICC-02, New York, April 2002.

[16] E. Aktas, U. Mitra, Single-user sparse channel acquisition in multiuser DS-CDMA systems, IEEE Trans. Comm. 51 (4) (April 2003) 682–693.

[17] A.A. Rontogiannis, A. Marava, K. Berberidis, J. Palicot, Efficient multipath channel estimation using a semi-blind parametric technique, in: Proceedings of ICASSP-03, vol. IV, Hong Kong, April 2003, pp. 477–480.

[18] G. Xu, H. Liu, L. Tong, T. Kailath, A least-squares approach to blind channel identification, IEEE Trans. Signal Process. 43 (12) (December 1995) 2982–2993.

[19] A. Bjorck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, PA, 1996.
[20] L.S.H. Ngia, Separable nonlinear least-squares methods for efficient off-line and on-line modeling of systems using Kautz and Laguerre filters, IEEE Trans. Circuits Systems II: Analog Digital Signal Process. 48 (6) (June 2001) 562–579.
[21] A.A. Rontogiannis, K. Berberidis, Efficient decision feedback equalization for sparse wireless channels, IEEE Trans. Wireless Comm. 2 (3) (May 2003) 570–581.