Circuits, Systems & Signal Processing

Performance and Complexity Analysis of Blind FIR Channel Identification Algorithms Based on Deterministic Maximum Likelihood in SIMO Systems

--Manuscript Draft--

Manuscript Number:	
Full Title:	Performance and Complexity Analysis of Blind FIR Channel Identification Algorithms Based on Deterministic Maximum Likelihood in SIMO Systems
Article Type:	Original Research
Keywords:	Blind Channel Estimation; Deterministic Maximum Likelihood; Performance analysis; DIQML; PQML
Corresponding Author:	Samir Omar Eurecom Sophia Antipolis, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Eurecom
Corresponding Author's Secondary Institution:	
First Author:	Elisabeth de Carvalho
First Author Secondary Information:	
Order of Authors:	Elisabeth de Carvalho
	Samir Omar
	Dirk Slock
Order of Authors Secondary Information:	
Abstract:	We analyze two algorithms that have been introduced previously for Deterministic Maximum Likelihood (DML) blind estimation of multiple FIR channels. The first one is a modification of the Iterative Quadratic ML (IQML) algo- rithm. IQML gives biased estimates of the channel and performs poorly at low SNR due to noise induced bias. The IQML cost function can be "denoised" by eliminating the noise contribution: the resulting algorithm, Denoised IQML (DIQML), gives consistent estimates and outperforms IQML. Furthermore, DIQML is asymptotically globally convergent and hence insensitive to the initialization. Its asymp- totic performance does not reach the DML performance though. The second strategy, called Pseudo-Quadratic ML (PQML), is naturally denoised. The denoising in PQML is furthermore more efficient than in DIQML: PQML yields the same asymptotic performance as DML, as opposed to DIQML, but requires a consistent initialization. We furthermore compare DIQML and PQML to the strategy of alternating minimization w.r.t. symbols and channel for solving DML (AQML). An asymptotic performance analysis, a complexity evaluation and simulation results are also presented. The proposed DIQML and PQML algorithms can immediately be applied also to other subspace problems such as frequency estimation of sinusoids in noise or direction of arrival estimation with uniform linear arrays.

Circuits, Systems, and Signal Processing manuscript No.
(will be inserted by the editor)

Performance and Complexity Analysis of Blind FIR Channel Identification Algorithms Based on Deterministic Maximum Likelihood in SIMO Systems

Elisabeth de Carvalho, Samir-Mohamad Omar, and Dirk T.M. Slock

the date of receipt and acceptance should be inserted later

Abstract We analyze two algorithms that have been introduced previously for Deterministic Maximum Likelihood (DML) blind estimation of multiple FIR channels. The first one is a modification of the Iterative Quadratic ML (IQML) algorithm. IQML gives biased estimates of the channel and performs poorly at low SNR due to noise induced bias. The IQML cost function can be "denoised" by eliminating the noise contribution: the resulting algorithm, Denoised IQML (DIQML), gives consistent estimates and outperforms IQML. Furthermore, DIQML is asymptotically globally convergent and hence insensitive to the initialization. Its asymptotic performance does not reach the DML performance though. The second strategy, called Pseudo-Quadratic ML (PQML), is naturally denoised. The denoising in PQML is furthermore more efficient than in DIQML: PQML yields the same asymptotic performance as DML, as opposed to DIQML, but requires a consistent initialization. We furthermore compare DIQML and PQML to the strategy of alternating minimization w.r.t. symbols and channel for solving DML (AQML). An asymptotic performance analysis, a complexity evaluation and simulation results are also presented. The proposed DIQML and PQML algorithms can immediately be applied also to other subspace problems such as frequency estimation of sinusoids in noise or direction of arrival estimation with uniform linear arrays.

Keywords Blind Channel Estimation · Deterministic Maximum Likelihood · Performance analysis · DIQML · PQML

EURECOM's research is partially supported by its industrial members: BMW Group Research Technology, Bouygues Telecom, Cisco, Hitachi, ORANGE, SFR, Sharp, STMicroelectronics, Swisscom, Thales. The research reported herein was also partially supported by the French ANR project SESAME and by the EU FET project CROWN and Strep WHERE2.

Samir-Mohamad Omar, Dirk T.M. Slock are with Mobile Communications Dept., EURECOM, BP 193, 06904 Sophia Antipolis Cdx, France. email: {omar, slock}@eurecom.fr

Elisabeth de Carvalho is with Aalborg University, Niels Jernes Vej 12, A6-212, 9220 Aalborg, Denmark. email: edc@es.aau.dk

Address(es) of author(s) should be given

1 Introduction

Recently, cognitive radio has become one of the hot topics that attracts a lot of researchers from communication and signal processing societies. In some of its proposed scenarios, the blind channel estimation constitutes an essential block [45], [32], [33]. Hence, engineers exploit from the extensive research that has been done during the last two decades in this field. Unfortunately, the performance of many blind channel identification algorithms that were introduced in that era was assessed only by simulation. This constitutes an obstacle that may hinder the implementation of those algorithms in some applications due to an uncertainty in their behavior. Encouraged by the importance of performing the analytical analysis and by the recent interest in blind channel estimation, we present and analyze in this paper previously introduced algorithms to solve the Deterministic Maximum Likelihood (DML) criterion for FIR multichannel estimation. Two solutions are discussed that appear as a significant improvement of the popular Iterative Quadratic ML (IQML) approach [8]. The algorithms considered here are generic and could also be applied to frequency estimation for sinusoids in noise [16] or direction of arrival (DOA) estimation for plane waves impinging on a uniform linear antenna array [28].

In this paper, however, we shall concentrate on the application to the blind identification of FIR multichannels [40].

DML considers the input symbols as deterministic unknown quantities. It does not use any a priori statistical information on the input symbols, unlike other methods which for example consider the input symbols as i.i.d. random variables, such as covariance matching [20], linear prediction [35, 1], or optimally weighted subspace fitting [26]. Still other approaches model the input symbols as Gaussian random variables such as Gaussian ML in [12], or as belonging to a discrete alphabet such as ILSP [38]. These methods exploit more information and hence may lead to better performance. However, compared to these methods, the deterministic methods [15] offer the advantage of giving (asymptotically) convex cost functions, thus avoiding local minima and often allowing closed-form (one-shot) solutions. The other methods may theoretically be more efficient statistically, but are often computationally more complex with typically local minima issues. In the category of deterministic methods, we also find the Subchannel Response Matching (SRM) method which will be elaborated below, methods based on the singular part of linear prediction [34], (unweighted) subspace fitting [29], and LS smoothing [41] or two-sided linear prediction [6]. Within the limitations of the information exploited by a given approach, ML methods normally provide asymptotically the best performance attainable. The asymptotic ML performance is usually the Cramer-Rao lower bound (CRB). However, the CRB cannot be attained in deterministic blind channel estimation due to the inconsistency of the symbol estimates [12]. So the benchmark performance here is the DML performance, rather than the CRB. The iterative algorithms considered here for optimizing DML require solving quadratic problems at each iteration and converge in 1 or 2 iterations. Furthermore, deterministic methods have the property of allowing the exact solution for a finite amount of data in the absence of noise (consistency in SNR). Methods that use the statistics of the input symbols often allow identifiability of a larger part of the channel than what deterministic methods can identify, but require an infinite amount of data to give the exact solution for these supplementary parameters (such as e.g. the channel magnitude). This "consistency in SNR" property of deterministic methods will be exploited in the paper.

At high SNR, IQML performs very well and gives the DML estimate. At low SNR however, it gives biased estimates due to the presence of noise and performs poorly [37]. The two iterative methods we consider here are in fact the only methods solving DML with reasonable computational cost, as they involve structured quadratic criteria at each iteration, and provide very good estimation performance even at low SNR.

The first method, Denoised IQML (DIQML), subtracts the asymptotic noise contribution in the DML criterion: it gives consistent estimates and outperforms IQML. This method is proved to be asymptotically, in the number of data (or in SNR), insensitive to the initialization and globally convergent. The second iterative method, Pseudo-Quadratic ML (PQML) which has been used recently to estimate the time delay in Ultra-Wideband (UWB) ranging applications [47] attempts to null the actual gradient of DML in each iteration. PQML appears as a modification of IQML from which the noise contribution is removed, but in a more efficient way than in DIQML. An asymptotic performance study of the two algorithms is presented. PQML is proven to give better performance than DIQML and gives the same asymptotic performance as DML provided the initialization is consistent. A complexity study is also provided: DIQML and PQML are computationally attractive solutions with a complexity linear in the number of data samples.

To summarize our contributions, we introduced (in [3]) independently from [26],[17] the denoising operation in IQML. Compared to [26], we introduced a more judicious choice of the denoising parameter that leaves the Hessian of the problem positive semidefinite. In [17] , a different problem formulation leads to the same algorithmic result. We then applied our constraining approach for the Hessian to PQML, which was first introduced in [30] in the context of sinusoids in noise estimation and then applied to DML channel estimation in [22]. In contrast to these references which only used simulations, we provide an asymptotic performance analysis which shows that both DIQML and PQML are useful since only DIQML allows global convergence while only PQML allows to reach the DML performance. We also emphasize the role played by various noise subspace parameterizations and provide a computational complexity analysis. These elements will be elaborated below.

2 Problem Formulation and Notation

We first introduce some notation and acronyms:

$(.)^*$, $(.)^T$, $(.)^H$	conjugate, transpose, conjugate transpose
$(.)^+$	Moore–Penrose pseudo–inverse
$\mathrm{tr}(A)$, $\det(A)$	trace and determinant of matrix A
$\hat{\theta}$, θ^o	estimate of vector θ, true value of θ
$\mathrm{Re}(.)$, $\mathrm{Im}(.)$	real and imaginary part
I	Identity matrix with adequate dimension
w.r.t.	with respect to

We consider here linear modulation (nonlinear modulations such as GMSK can be linearized with good approximation [42]) transmitted over a linear channel. The received signal, after a linear receiver filter, is then the convolution of the transmitted symbols with an overall channel impulse response. In wireless communications terminology, we consider here the single-user case (as opposed to the multi-user case in which the received signal contains a mixture of multiple users). The multichannel model results from the oversampling of the received signal and/or from reception by multiple antennas. It can also come from the separation of the real (in phase) and imaginary (in quadrature) part of the demodulated received signal if the symbol constellation is real [27], [44].

Consider a sequence of symbols $a(k)$ received through m channels of length N and coefficients $\boldsymbol{h}(i)$:

$$\boldsymbol{y}(k) = \sum_{i=0}^{N-1} \boldsymbol{h}(i) a(k-i) + \boldsymbol{v}(k), \tag{1}$$

$\boldsymbol{v}(k)$ is an additive independent white Gaussian noise with $r_{\boldsymbol{vv}}(k-i) = \mathrm{E}\,\boldsymbol{v}(k)\,\boldsymbol{v}(i)^H = \sigma_v^2 I_m\, \delta_{ki}$ and when $\boldsymbol{v}(k)$ is complex $\mathrm{E}\,\boldsymbol{v}(k)\boldsymbol{v}^T(i) = 0$ (circular noise).

Assume we receive M samples, concatenated in the vector $\boldsymbol{Y}_M(k)$:

$$\boldsymbol{Y}_M(k) = \mathcal{T}_M(h)\, A_{M+N-1}(k) + \boldsymbol{V}_M(k)\,. \tag{2}$$

$\boldsymbol{Y}_M(k) = [\boldsymbol{y}^T(k) \cdots \boldsymbol{y}^T(k-M+1)]^T$, similarly for $\boldsymbol{V}_M(k)$, and $A_M(k) = [a(k) \cdots a(k-M-N+2)]^T$. $\mathcal{T}_M(h)$ is a block Toeplitz matrix with M block rows and $\left[\boldsymbol{H}\ \ 0_{m\times(M-1)}\right]$ as first block row, where:

$$\boldsymbol{H} = [\boldsymbol{h}(0)\ \cdots\ \boldsymbol{h}(N-1)] \ \text{ and } h = \left[\boldsymbol{h}^T(0) \cdots \boldsymbol{h}^T(N-1)\right]^T . \tag{3}$$

We furthermore denote $\mathbf{H}(z) = \sum_{i=0}^{N-1} \boldsymbol{h}(i) z^{-i} = [\mathrm{H}_1(z) \cdots \mathrm{H}_m(z)]^T$ the SIMO channel transfer function. We shall simplify the notation in (2) with $k = M-1$ to:

$$\boldsymbol{Y} \;=\; \mathcal{T}(h)\, A + \boldsymbol{V}\ . \tag{4}$$

Note that due to the commutativity of convolution: $\mathcal{T}(h)A = \mathcal{A}h$, where $\mathcal{A} = \mathcal{A}^{'} \otimes I_m$ and the Hankel matrix $\mathcal{A}^{'}$ is filled with the elements of A.

3 Preliminary Facts

3.1 Ambiguity and Constraints

It is well known that in the context of the SIMO systems, the channel can only be estimated blindly up to a scale factor. In the complex signals case, a proper approach requires to convert the problem with Nm complex parameters h into a problem with $2Nm$ real parameters $h_R = \left[\mathrm{Re}(h)^T\ \ \mathrm{Im}(h)^T\right]^T$ and in that case a complex scale factor corresponds to two real parameters (amplitude or norm, and phase). In the real case, $h_R = h$. To make the estimation problem well posed, constraints on the channel need to be introduced to fix its unidentifiable components. We computed the asymptotic performance of DIQML and PQML under the following constraints:

(1) a quadratic constraint:
$$h^H h = h^{oH} h^o \tag{5}$$
which allows to adjust the norm of the channel.

(2) in the complex case, an additional constraint is necessary to adjust the phase factor:
$$h_S^{o\,T} h_R = h_S^{o\,T} h_R^o = 0 \tag{6}$$
where $h_S = \left[-\mathrm{Im}(h)^T \;\; \mathrm{Re}(h)^T\right]^T \; (= \, (jh)_R)$.

The first constraint is a commonly used constraint in blind channel estimation and we have already introduced it earlier. The particular constraints above were chosen to characterize the asymptotic performance of blind DML in [12,9] because they yield the minimal MSE, $\mathrm{E}\|\hat{h}-h^o\|^2$, for a minimal number γ ($\gamma = 2$ in the complex case, $\gamma = 1$ in the real case) of independent constraints. This property holds also for PQML and DIQML. In both real and complex cases, these constraints leave a sign ambiguity on the channel which does not lead to singularities in the matrices involved in the performance derivation (see (53)) or in the CRB, as is the case for any discrete valued parameter ambiguity. For practical MSE evaluation, the ambiguity can be resolved by requiring $h^{o\,H} h > 0$ (constraint (6) only forces $h^{o\,H} h$ to be real).

The constraints (5), (6) are of the form $\mathcal{K}(h_R) = 0$, $\mathcal{K} : \mathbb{R}^{\gamma N m} \to \mathbb{R}^{\gamma}$. We denote by $\mathcal{M}_{h_R^o}$ the tangent subspace to the constraint set at the point h_R^o:
$$\mathcal{M}_{h_R^o} = \left\{ Z \in \mathbb{R}^{\gamma N m} \; ; \; \left(\frac{\partial \mathcal{K}^T(h_R^o)}{\partial h_R}\right)^T \; Z = 0 \right\} . \tag{7}$$
For the constraints (5), (6), we get $\dfrac{\partial \mathcal{K}^T(h_R^o)}{\partial h_R} = [h_R^o \;\; h_S^o]$ in the complex case (or just the first column in the real case). For the asymptotic performance, any constraint set that leads to the same tangent subspace $\mathcal{M}_{h_R^o}$ is equivalent. For instance, we can consider also the following linear constraint for the complex case: $h^{o\,H} h = h^{o\,H} h^o$ which for the corresponding real parameters translates to $[h_R^o \;\; h_S^o]^T h_R = [h_R^o \;\; h_S^o] h_R^o = [h_R^{o\,T} h_R^o \;\; 0]$. One may remark that this linear constraint set leaves no sign ambiguity, and has the same tangent subspace as the constraints in (5), (6).

3.2 Linear Parameterization of the Noise Subspace

As we shall observe in the sequel, the DML criterion (10) is highly nonlinear and its direct optimization would require cumbersome optimization techniques. The key to a computationally attractive solution of the DML problem is a linear parameterization of the noise subspace. We consider here a linear parameterization of the noise subspace in terms of channel coefficients (a parameterization in terms of prediction quantities was also presented in [36]).

Let $\mathbf{H}^{\perp}(z)$ $(p \times m)$, $p \geq m-1$, be such a parameterization: it verifies $\mathbf{H}^{\perp}(z)\mathbf{H}(z) = 0$ and $\mathcal{T}(h^{\perp})\mathcal{T}(h) = 0$; $\mathcal{T}(h^{\perp})$ is the convolution matrix corresponding to the filter $\mathbf{H}^{\perp}(z)$ and the columns of $\mathcal{T}^H(h^{\perp})$ span the noise subspace. For more details about this issue, we refer the reader to [3].

4 Various Deterministic Algorithms

4.1 Blind Deterministic ML

The Deterministic Maximum Likelihood (DML) method was introduced for blind channel estimation in [35,24]. In DML, both channel coefficients and input symbols are considered as deterministic quantities, and are jointly estimated through the criterion:

$$\max_{A,h} f(\boldsymbol{Y}|h) \quad \Leftrightarrow \quad \min_{A,h} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 . \tag{8}$$

$f(\boldsymbol{Y}|h)$ is the complex probability density function (which exists as $\boldsymbol{V}$ is circular). The derivations will be done in the complex case, the real case is similar. We assume here that the blind deterministic identifiability conditions, which ensure $\mathcal{T}(h)$ to have full column rank, are verified. Sufficient conditions are for the channel to be irreducible ($\mathbf{H}(z)$ has no zeros), the number of input symbol excitation modes to exceed $N{-}1{+}2\underline{M}$ and the burst length to exceed $N{-}1{+}\underline{M}$ [13]. $\underline{M}$ is the minimal M for which $\mathcal{T}(h)$ has full column rank. In general $\left\lceil \frac{N-1}{m-1} \right\rceil \leq \underline{M} \leq N{-}1$ where for a random channel the lower bound is attained w.p. 1. The channel is then uniquely identifiable up to a scale factor. We impose the non-triviality constraint $\|h\| = 1$. This constraint is in fact not sufficient to completely identify the channel as it leaves a phase ambiguity: a phase constraint will be imposed later, in the performance study of the proposed algorithms. As for any deterministic method, the channel length N has to be assumed known. If N is unknown, all values should be tried and the best value should be determined on the basis of a criterion, see e.g. [41],[5]. Alternatively, robustness to channel length overestimation may be obtained by anchoring the first channel coefficient (similar to what occurs in linear prediction). This can be done by replacing the constraint $\|h\| = 1$ with $\|\boldsymbol{h}(0)\| = 1$ [4]. We shall not pursue this approach here. The effect of channel length underestimation in DML has been analyzed in [11].

Optimizing (8) first w.r.t. A, we get:

$$A = \left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{-1} \mathcal{T}^H(h)\boldsymbol{Y} \tag{9}$$

which is the output of the (burst–mode) MMSE-ZF equalizer [14]. Substituting (9) in (8), we get the following DML criterion for h (in which the nuisance parameter A is now removed):

$$\min_{\|h\|=1} \boldsymbol{Y}^H P^{\perp}_{\mathcal{T}(h)} \boldsymbol{Y} . \tag{10}$$

$P^{\perp}_{\mathcal{T}(h)} = I - \mathcal{T}(h)\left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{-1} \mathcal{T}^H(h)$ is the orthogonal projection on the noise subspace. The signal subspace is defined as the column space of $\mathcal{T}(h)$ and the noise subspace is its orthogonal complement.

4.2 Subchannel Response Matching (SRM)

The Subchannel Response Matching (SRM) algorithm, which was (re)invented four times in [21,7,46,24], is based on a linear parameterization of the noise subspace in terms of the channel coefficients. Using the commutativity of convolution and the linearity of $h^\perp$ in h, we can write $\mathcal{T}(h^\perp)\boldsymbol{Y}$ as:

$$\mathcal{T}(h^\perp)\boldsymbol{Y} = \mathcal{Y}h \tag{11}$$

where $\mathcal{Y}$ is a matrix filled with the elements of the observation vector $\boldsymbol{Y}$.

In the noiseless case, $\boldsymbol{Y} = \boldsymbol{X} = \mathcal{T}(h)A$ and we have $\mathcal{T}(h^\perp)\boldsymbol{X} = \mathcal{X}h = 0$: from this relation, the channel can be uniquely determined up to a scale factor [46, 24], as the unique right singular vector of $\mathcal{X}$ corresponding to the singular value zero. As for DML, SRM requires the channel to be irreducible; it has to be noted that the burst length requirements are higher than for DML [15]: $M \geq N$ for 2 subchannels and $M \geq 2(N-1)$ for more than 2 subchannels. When noise is present, $\mathcal{Y}h \neq 0$ and the SRM criterion is solved in the least–squares sense under the constraint $\|h\| = 1$:

$$\min_{h:\,\|h\|=1} h^H\, \mathcal{Y}^H\mathcal{Y}\, h\ . \tag{12}$$

The solution is $h = V_{min}(\mathcal{Y}^H\mathcal{Y})$, the eigenvector of $\mathcal{Y}^H\mathcal{Y}$ corresponding to its smallest eigenvalue. Different choices for the linear parameterization of the noise subspace give different channel estimates. Note that $\mathrm{E}\, h^H\, \mathcal{Y}^H\mathcal{Y}\, h = h^H\, \mathcal{X}^H\mathcal{X}\, h + \sigma_v^2 \mathrm{tr}\left\{\mathcal{T}(h^\perp)\mathcal{T}^H(h^\perp)\right\}$. Hence a balanced $h^\perp$ yields asymptotically unbiased and consistent channel estimates whereas unbalanced $h^\perp$ yield biased and inconsistent estimates.

SRM may be viewed as a non-weighted version of the Iterative Quadratic ML (IQML) algorithm described below, and was used in [24] to initialize IQML. We will use it to initialize our algorithms also.

4.3 Iterative Quadratic ML (IQML)

Since $P^\perp_{\mathcal{T}(h)} = P_{\mathcal{T}^H(h^\perp)}$ (only asymptotically true in case of $h^\perp_{min}$), the DML problem (10) can be written as:

$$\min_{h:\,\|h\|=1} \boldsymbol{Y}^H\mathcal{T}^H(h^\perp)\left(\mathcal{T}(h^\perp)\mathcal{T}^H(h^\perp)\right)^+\mathcal{T}(h^\perp)\boldsymbol{Y} \tag{13}$$

where the Moore-Penrose pseudo–inverse needs to be introduced since $\mathcal{T}(h^\perp)$ $\mathcal{T}^H(h^\perp)$ is singular for $m > 2$ for any choice of $h^\perp$ different from $h^\perp_{min}$ (unless $\mathcal{T}^\perp$ is used).

The Iterative Quadratic ML (IQML) algorithm [8] solves (13) iteratively in such a way that at each iteration a quadratic problem appears. Let $\mathcal{R}(h) \triangleq \mathcal{T}(h^\perp)\mathcal{T}^H(h^\perp)$, then (13) becomes:

$$\min_{h:\,\|h\|=1} \boldsymbol{Y}^H\mathcal{T}^H(h^\perp)\ \mathcal{R}^+(h)\ \mathcal{T}(h^\perp)\boldsymbol{Y}\ . \tag{14}$$

In iteration (i) of IQML, the “denominator” $\mathcal{R}(h)$ is computed based on the estimate from the previous iteration/initialization $\hat{h}^{(i-1)}$ and is considered as constant for the current iteration. Hence, $\mathcal{T}(h^\perp)$ being linear in h, the criterion (14) becomes quadratic. Denoting the constant denominator $\mathcal{R}(h) = \mathcal{R}$, the IQML criterion can be rewritten as:

$$\min_{h:\,\|h\|=1} h^H\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}\, h\ . \tag{15}$$

Under constraint $\|h\| = 1$, we get $h = V_{min}(\mathcal{Y}^H \mathcal{R}^+ \mathcal{Y})$.

In an alternative interpretation (which holds even if the columnspace of $\mathcal{T}^H(h^\perp)$ is smaller than the noise subspace), IQML can be viewed as the optimally weighted least-squares version of the SRM least-squares problem: the covariance matrix of the noise contribution in $\mathcal{T}(h^\perp)\boldsymbol{Y} = \mathcal{Y}h$ is indeed $\sigma_v^2 \mathcal{R}(h)$ and we use for h in $\mathcal{R}(h)$ the best estimate available.

In the noisefree case, the IQML algorithm behaves very well: the IQML criterion becomes indeed equivalent to:

$$\min_{h:\,\|h\|=1} \boldsymbol{X}^H \mathcal{T}^H(h^\perp) \mathcal{R}^+ \mathcal{T}(h^\perp) \boldsymbol{X} \tag{16}$$

where $\boldsymbol{X} = \mathcal{T}(h^o)A$ is the noise-free received signal. As $\mathcal{T}(h^{o\perp})\boldsymbol{X} = \mathcal{X}h^o = 0$, h^o nulls exactly the criterion, regardless of the initialization. At high SNR, a first iteration of IQML gives a consistent estimate of the channel whatever the initialization of $\mathcal{R}(h)$ (provided that $Null(\mathcal{R}^+) \cap Range(\mathcal{X}) = 0$, which is guaranteed in general).And it can be proven [24] (see Appendix A also) that a second iteration gives the exact DML estimate.

At low SNR however, the IQML estimate is biased. Indeed, consider the asymptotic situation in which the number of data M grows to infinity. By the law of large numbers, the IQML criterion becomes essentially equivalent to its expected value, viz.

$$\begin{aligned}\tfrac{1}{M}\,\boldsymbol{Y}^H \mathcal{T}^H(h^\perp)\ \mathcal{R}^+(h)\ \mathcal{T}(h^\perp)\boldsymbol{Y} \ &= \ \tfrac{1}{M}\,\mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)E(\boldsymbol{Y}\boldsymbol{Y}^H)\} + \mathcal{O}_p(\tfrac{1}{\sqrt{M}}) \\ = \ \tfrac{1}{M}\ \big[\,\mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\boldsymbol{X}\boldsymbol{X}^H\} &+ \sigma_v^2\ \mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\}\big] + \mathcal{O}_p(\tfrac{1}{\sqrt{M}})\end{aligned} \tag{17}$$

since $E(\boldsymbol{Y}\boldsymbol{Y}^H) = \boldsymbol{X}\boldsymbol{X}^H + \sigma_v^2 I$. Hence

$$h^1 \ = \ \arg\min_{h:\|h\|=1} \left\{ \boldsymbol{Y}^H \mathcal{T}^H(h^\perp)\ \mathcal{R}^+(h)\ \mathcal{T}(h^\perp)\boldsymbol{Y} \right\} \ = \ h^2 + \mathcal{O}_p(\frac{1}{\sqrt{M}}) \tag{18}$$

where

$$h^2 \ = \ \arg\min_{h:\|h\|=1} \left\{ \mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\boldsymbol{X}\boldsymbol{X}^H\} + \sigma_v^2\ \mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\} \right\}\ . \tag{19}$$

The true IQML minimizer h^1 differs from h^2 by an asymptotically vanishing estimation error. Now, considering the deterministic minimization problem in (19), h^o nulls exactly the first term, but is not in general the minimizer of the second term, even if $\mathcal{R} = \mathcal{R}(h^o)$. More explicitely, in general

$$\min_{h:\|h\|=1} \left\{ \mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+(h^o)\mathcal{T}(h^\perp)\} \right\} < \mathrm{tr}\{P_{\mathcal{T}^H(h^{o\perp})}\} = mM - (M+N-1)\ . \tag{20}$$

Hence, the first term in (19) gets minimized by h^o but the second term gets minimized by $h \neq h^o$, so the sum of the two terms gets minimized by $h^2 \neq h^o$ in general. See [37] for the corresponding issues in the DOA problem. Hence, due to the presence of noise, h^o is not asymptotically near a stationary point of the algorithm and IQML performs poorly even if initialized by a consistent channel estimate.

We propose here a method to “denoise” the IQML criterion: this denoised criterion, solved in the IQML style, will correct the IQML bias and provide a consistent channel estimate.

4.4 Denoised Iterative Quadratic ML (DIQML)

4.5 Asymptotic Amount of Data

The asymptotic noise contribution to the DML criterion is $\sigma_v^2 \ \mathrm{tr}\left\{P_{\mathcal{T}^H(h^\perp)}\right\}$ (see (17)). The denoising strategy simply consists in removing this asymptotic noise term, or more exactly an estimate of it, $\widehat{\sigma_v^2} \ \mathrm{tr} P_{\mathcal{T}^H(h^\perp)}$, from the DML criterion which becomes:

$$\begin{gathered} \min_{\|h\|=1} \mathrm{tr}\left\{P_{\mathcal{T}^H(h^\perp)}\left(\boldsymbol{Y}\boldsymbol{Y}^H - \widehat{\sigma_v^2}\, I\right)\right\} \ \Leftrightarrow \\ \min_{\|h\|=1} \left\{h^H \mathcal{Y}^H \mathcal{R}^+(h)\mathcal{Y}h - \widehat{\sigma_v^2} \ \mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+(h)\mathcal{T}(h^\perp)\}\right\} . \end{gathered} \tag{21}$$

Note that this operation does not change the optimizer of the DML criterion as $\widehat{\sigma_v^2} \ \mathrm{tr}\{P_{\mathcal{T}^H(h^\perp)}\} = \widehat{\sigma_v^2}\,(M(m{-}1){-}N{+}1)$ is constant w.r.t. h. We take $\widehat{\sigma_v^2}$ to be a consistent estimate of the noise variance.

The denoised DML criterion (21) is now solved in the IQML way: considering $\mathcal{R}(h) = \mathcal{R}$ as constant, the optimization problem becomes again quadratic in h:

$$\min_{\|h\|=1} h^H \left\{\mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \widehat{\sigma_v^2}\, \mathcal{D}\right\} h \tag{22}$$

where the matrix $\mathcal{D}(h)$ is such that $h^{``H}\mathcal{D}h^{'} = \mathrm{tr}\{\mathcal{T}^H(h^{``\perp})\mathcal{R}^+(h)\mathcal{T}(h^{'\perp})\}$.

Asymptotically in the number of data, DIQML is globally convergent. Indeed, asymptotically it is essentially equivalent to the denoised criterion:

$$\frac{1}{M}\, h^H \left\{\mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \widehat{\sigma_v^2}\, \mathcal{D}\right\} h \ = \ \frac{1}{M}\, \boldsymbol{X}^H \mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\boldsymbol{X} \ + \ \mathcal{O}_p(\frac{1}{\sqrt{M}}) \tag{23}$$

if $\widehat{\sigma_v^2} - \sigma_v^2 = \mathcal{O}_p(\frac{1}{\sqrt{M}})$. The denoised criterion (the first term of the RHS of (23)) corresponds to the IQML criterion in the noiseless case and hence leads to $h = \alpha\, h^o$ for some scaling factor α, under the identifiability conditions of SRM. One iteration of DIQML hence yields an estimate $h = \alpha\, h^o + \mathcal{O}_p(\frac{1}{\sqrt{M}})$. So the DIQML algorithm behaves asymptotically at any SNR like the IQML algorithm behaves at high SNR:

- the first iteration gives a consistent estimate of the channel,
- this behavior holds whatever the initialization.

The second iteration gives asymptotically the global minimizer of DIQML. Unlike in the high SNR IQML case though, this global minimizer at an arbitrary SNR is not the DML minimizer, as can be seen in Appendix A. As the SNR increases, the difference between DIQML and IQML disappears and we have global convergence, to the DML solution.

4.6 Finite Amount of Data

The choice of $\widehat{\sigma_v^2}$ turns out to be crucial. In practice, with large but finite amount of data M, and the true noise variance value, the central matrix $\mathcal{Q} = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \sigma_v^2\, \mathcal{D}$ in (22) is indefinite, and the minimization problem is no longer well posed. The solution in this case would be $V_{min}(Q)$ corresponding to the smallest eigenvalue

$\lambda_{min}(Q)$, which is negative. Simulations have shown that performance does not improve upon IQML in this case. The central matrix $\mathcal{Q}$ should be constrained to be positive semi-definite.

For the consistent estimate of σ_v^2, we choose here a certain λ that renders $\mathcal{Q} = \mathcal{Q}(\mathcal{R}^+) = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{D}$ exactly positive semi-definite with one singularity. The DIQML criterion becomes:

$$\min_{\|h\|=1,\lambda} h^H\left\{\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{D}\right\}h \tag{24}$$

with the constraint that $\mathcal{Q}$ be positive semi-definite. The solution is $\lambda = \lambda_{min}(\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}, \mathcal{D})$, the minimal generalized eigenvalue of $\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}$ and $\mathcal{D}$, and $h = V_{min}(\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}, \mathcal{D})$, the corresponding generalized eigenvector. Asymptotically, the DIQML criterion (24) becomes

$$\begin{aligned}\frac{1}{M}\,h^H\left\{\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\,\mathcal{D}\right\}h \;&=\; \frac{1}{M}\,\boldsymbol{X}^H\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\boldsymbol{X} \;+\\ &\quad\;\frac{1}{M}\,(\sigma_v^2 - \lambda)\,h^H\mathcal{D}h \;+\; \mathcal{O}_p(\frac{1}{\sqrt{M}})\end{aligned} \tag{25}$$

Optimization w.r.t. λ, subject to the non-negativity constraint, leads to $\lambda = \sigma_v^2 + \mathcal{O}_p(\frac{1}{\sqrt{M}})$, regardless of channel initialization (in $\mathcal{R}$ and $\mathcal{D}$), and the criterion (24) in h and λ becomes equivalent to the criterion (23) in h. Hence, asymptotic global convergence applies for h and for λ (to σ_v^2), with the same properties as mentioned earlier (independently from the initialization).

Other attempts have been undertaken to denoise the IQML strategy. Kristensson [26] independently applied the same strategy in the DOA context: as estimate of the noise variance, he chooses the one which in the context of blind channel estimation would correspond to the minimum value of the SRM criterion: it can indeed be verified that asymptotically

$$\lambda_{min}(\mathcal{Y}^H\mathcal{Y}) = \hat{h}_{SRM}^H\left(\mathcal{Y}^H\mathcal{Y}\right)\hat{h}_{SRM} = \sigma_v^2\,\alpha \tag{26}$$

with $\|\hat{h}_{SRM}\|^2 = 1$ ($\hat{h}_{SRM} = V_{min}(\mathcal{Y}^H\mathcal{Y})$), for balanced $h^\perp$. For a finite amount of data, the noise variance estimate given by SRM underestimates the true σ_v^2 on the average: indeed, as $\hat{h}_{SRM}$ minimizes the SRM criterion, $\hat{h}_{SRM}^H\left(\mathcal{Y}^H\mathcal{Y}\right)\hat{h}_{SRM} \leq h^{o\,H}\left(\mathcal{Y}^H\mathcal{Y}\right)h^o/\|h^o\|^2$, taking the expected value on both sides, we get $\mathrm{E}\,\widehat{\sigma_v^2}_{SRM} \leq \sigma_v^2$. The quadratic cost function of denoised SRM (in case a balanced $h^\perp$ is used) corresponds to $Q(I) = \mathcal{Y}^H\mathcal{Y} - \lambda\,\alpha\,I$, the unweighted version of DIQML. Choosing $\lambda\,\alpha = \lambda_{min}(\mathcal{Y}^H\mathcal{Y})$, the (ordinary) minimal eigenvalue of the first term in Q, allows to guarantee $Q(I) \geq 0$ and λ is a consistent estimate of σ_v^2 as is clear from (26). However, with the weighting matrix $\mathcal{R}^+$ introduced in DIQML, the second term, $\mathcal{D}$, in the Hessian $Q(\mathcal{R}^+)$ is no longer a multiple of identity, and λ needs to be chosen as a generalized eigenvalue of the two matrix terms in order to guarantee $\mathcal{Q}(\mathcal{R}^+) \geq 0$. Both $\lambda_{min}(\mathcal{Y}^H\mathcal{Y}, \alpha I) = \frac{1}{\alpha}\lambda_{min}(\mathcal{Y}^H\mathcal{Y})$ and $\lambda_{min}(\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}, \mathcal{D})$ are consistent estimators of σ_v^2 which tend to underestimate σ_v^2 for finite amount of data due to the non-negativity constraint on the respective Q. However, both are different random quantities and hence $\lambda = \frac{1}{\alpha}\lambda_{min}(\mathcal{Y}^H\mathcal{Y})$ can very well exceed $\lambda = \lambda_{min}(\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}, \mathcal{D})$ in some realizations, making the $Q(\mathcal{R}^+)$ of DIQML indefinite, preventing the corresponding DIQML version to improve upon IQML.

4.7 Pseudo-Quadratic ML (PQML)

The principle of PQML has been introduced in the context of sinusoids in noise estimation [30] and then applied to DML channel estimation in [22]. The gradient of the DML cost function may be arranged as $\mathcal{P}(h)\,h$, where $\mathcal{P}(h)$ is ideally a positive semi-definite matrix with a one-dimensional nullspace. The DML estimate satisfies

$$\mathcal{P}(h)\;h = 0, \tag{27}$$

which is solved for h under the constraint $\|h\| = 1$. The DML gradient in (27) is the same as the gradient of the (pseudo-)quadratic cost function $h^H\,\mathcal{P}(\hat{h})\,h$ evaluated at $\hat{h} = h$. The PQML strategy is now the following. At iteration (i), $\mathcal{P}(\hat{h}^{(i-1)}) \geq 0$ is fixed. The problem $\min_{h:\,\|h\|=1}\; h^H\,\mathcal{P}(\hat{h}^{(i-1)})\,h$ is quadratic and its solution is $\hat{h}^{(i)} = V_{min}(\mathcal{P}(\hat{h}^{(i-1)}))$. This solution is used to reevaluate $\mathcal{P}(h)$ and further iterations may be performed.

The difficulty consists in defining the right $\mathcal{P}(h)$ in the DML gradient, especially with the positive semi-definiteness constraint. In general, and in particular for the DML problem at hand, the choice for $\mathcal{P}(h)$ is indeed not unique. Denoting $\mathcal{T}\left(\dfrac{\partial h^\perp}{\partial h_i}\right) = \Delta\mathcal{T}_i^\perp$, the gradient of the DML cost function consists of two terms (here we write the gradient w.r.t. h_i, which is also component i of the gradient w.r.t. h):

$$\begin{aligned}(\mathcal{P}(h)\underline{\boldsymbol{h}})\,(i) &= \boldsymbol{Y}^H\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h)\mathcal{T}(\underline{\boldsymbol{h}}^\perp)\boldsymbol{Y}\\ -\;& \boldsymbol{Y}^H\mathcal{T}^H(h^\perp)\mathcal{R}^+(h)\left[\mathcal{T}(\underline{\boldsymbol{h}}^\perp)\Delta\mathcal{T}_i^{\perp\,H}\right]\mathcal{R}^+(h)\mathcal{T}(h^\perp)\boldsymbol{Y}\,.\end{aligned} \tag{28}$$

Here, we consider that h is complex and complex derivation w.r.t. h^* is applied; for a real h, the results are similar. We assume here also that the pseudo–inverse (if $\mathcal{R}$ is singular) is computed by regularization so that we simply need to derive w.r.t. a regular inverse. For a small regularization constant $\delta > 0$ we get indeed

$$(\mathcal{R} + \delta\, I)^{-1} \;=\; \mathcal{R}^+ - \delta\,(\mathcal{R}^+)^2 + \frac{1}{\delta}\,P_{\mathcal{R}}^\perp + \mathcal{O}(\delta^2)\;. \tag{29}$$

Hence $(\mathcal{R}+\delta\, I)^{-1}\mathcal{T}(h^\perp) \;=\; \mathcal{R}^+\mathcal{T}(h^\perp)+\mathcal{O}(\delta)$. So we get the correct DML gradient by regularization as the regularization factor $\delta \to 0$.

In each iteration, $\mathcal{P}(h)$ will be considered as constant. The question now is which factors h should be considered as variable and which instances of h are considered as part of $\mathcal{P}(h)$. $\underline{\boldsymbol{h}}$ in (28) designates those instances of h that we consider as variable (on which minimization will be done) and h designates those instances of h that are considered as part of the constant $\mathcal{P}(h)$. The first term of $\mathcal{P}(h)\underline{\boldsymbol{h}}$ is $\mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y}\underline{\boldsymbol{h}}$, which is the IQML gradient, and the second term is $\mathcal{B}^H(h)\mathcal{B}(h)\underline{\boldsymbol{h}}$, with $\boldsymbol{Y}^H\mathcal{T}^H(h^\perp)\mathcal{R}^+(h)\mathcal{T}(\underline{\boldsymbol{h}}^\perp) = \underline{\boldsymbol{h}}^T\mathcal{B}^T(h)$ (note that $\boldsymbol{Y}^H P_{\mathcal{T}^H(h^\perp)}\boldsymbol{Y} = \left(\boldsymbol{Y}^H P_{\mathcal{T}^H(h^\perp)}\boldsymbol{Y}\right)^*$). Then $\mathcal{P}(h)$ has the following form:

$$\mathcal{P}(h) = \mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y} - \mathcal{B}^H(h)\mathcal{B}(h)\,. \tag{30}$$

The second term of $\frac{1}{M}\mathcal{P}(h)$ asymptotically tends to its expected value by the law of large numbers. In Appendix A, we prove that $\mathrm{E}\left(\mathcal{B}^H(h)\mathcal{B}(h)\right)$ has a noise component equal to $\sigma_v^2\mathcal{D}$, the asymptotic noise component of the IQML Hessian, but it also has a non–zero signal component when $\mathcal{P}(h)$ is evaluated at $h \neq h^o$. This prevents PQML from being asymptotically insensitive to the initialization, unlike DIQML. However, when $\mathcal{P}(h)$ is evaluated at a consistent h, the previously mentioned signal component becomes negligible. PQML gives furthermore better performance than DIQML, and in fact offers the same performance as DML.

The matrix $\mathcal{P}(h)$ is indefinite for a finite data length M, and applying the PQML strategy directly will not work. In [22], h is chosen as the eigenvector corresponding to the smallest eigenvalue magnitude of $\mathcal{P}(h)$; it gives poor performance except at high SNR.

PQML is closely related to DIQML as the first term of (22) and (30) are the same and $\mathrm{E}(\mathcal{B}^H(h^o)\mathcal{B}(h^o)) = \sigma_v^2\mathcal{D}(h^o)$. By analogy with DIQML for which $\mathcal{Q}$ was also indefinite for finite M if an arbitrary $\widehat{\sigma_v^2}$ were to be used, we introduce a variable λ such that $\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\, \mathcal{B}^H\mathcal{B}$ is exactly positive semi-definite. PQML then becomes the following minimization problem:

$$\min_{\|h\|=1,\lambda} h^H \left\{\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\, \mathcal{B}^H\mathcal{B}\right\} h \tag{31}$$

with a semi-definite positivity constraint on the central matrix. The solution is again
$h = V_{min}(\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}, \mathcal{B}^H\mathcal{B})$ corresponding to $\lambda = \lambda_{min}(\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}, \mathcal{B}^H\mathcal{B})$. Asymptotically for a consistent initialization, there is global convergence for h, as described previously, as well as for λ ($\to 1$). However, for a finite amount of data, and for an arbitrary h,

$$\lambda = \lambda_{min}(\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}, \mathcal{B}^H\mathcal{B}) = \min_{\hat{h}} \frac{\hat{h}^H\mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y}\hat{h}}{\hat{h}^H\mathcal{B}^H(h)\mathcal{B}(h)\hat{h}} \leq \frac{h^H\mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y}h}{h^H\mathcal{B}^H(h)\mathcal{B}(h)h} = 1 \tag{32}$$

which means that using $\lambda = 1$, as in the original PQML algorithm, systematically leads to an indefinite $\mathcal{P}(h)$.

The identifiability conditions for both DIQML and PQML are the same as for SRM: the channel has to be irreducible and the burst length should be sufficiently large (see section 4.2).

4.8 Alternating Quadratic ML (AQML)

In addition to comparing the performance of DIQML and PQML to the optimal DML performance, we will compare them to an algorithm we call Alternating Quadratic ML (AQML), which was also introduced in [19,31,2]. AQML corresponds in fact to the ILSP algorithm [43,38] in which the exploitation of the finite alphabet gets dropped.

AQML proceeds by alternating minimizations w.r.t. A and w.r.t h of the DML criterion:

$$\min_{h,A} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 \tag{33}$$

(1) Initialization: $\hat{h}^{(0)}$.
(2) Iteration $(i+1)$:

– Minimization w.r.t. A, $h = \hat{h}^{(i)}$: $\min_A \|\boldsymbol{Y} - \mathcal{T}(\hat{h}^{(i)})A\|^2$

$$\widehat{A}^{(i+1)} = \left(\mathcal{T}^H(\hat{h}^{(i)})\mathcal{T}(\hat{h}^{(i)})\right)^{-1}\mathcal{T}^H(\hat{h}^{(i)})\boldsymbol{Y} \tag{34}$$

– Minimization w.r.t. h, $A = \widehat{A}^{(i+1)}$: $\min_h \|\boldsymbol{Y} - \mathcal{T}(h)\widehat{A}^{(i+1)}\|^2 = \min_h \|\boldsymbol{Y} - \widehat{\mathcal{A}}^{i+1}h\|^2$

$$\hat{h}^{(i+1)} = \left(\widehat{\mathcal{A}}^{(i+1)\,H}\widehat{\mathcal{A}}^{(i+1)}\right)^{-1}\widehat{\mathcal{A}}^{(i+1)\,H}\boldsymbol{Y} \tag{35}$$

(3) Repeat (2) until $(\widehat{A}^{(i+1)}, \hat{h}^{(i+1)}) \approx (\widehat{A}^{(i)}, \hat{h}^{(i)})$.

At any iteration $(i+1)$, we assume that the algorithm gives a unique solution: $\mathcal{T}(\hat{h}^{(i)})$ has full-column rank (*i.e.* $\hat{\boldsymbol{H}}(z)$ is irreducible), as well as $\hat{\mathcal{A}}^{(i+1)}$ (at least N excitation modes), otherwise as suggested in [39], we take the minimum-norm solution (*i.e.* the regular inverse is replaced by the pseudo-inverse). That case is unlikely though if h^o and A^o are well-conditioned.

5 Computational Complexity Analysis

5.1 Complexity of IQML

A complexity analysis of the IQML algorithm for blind multichannel estimation has been provided in [24]. However, in [24] $\mathcal{T}(h^\perp)$ has been constructed as a Toeplitz block matrix instead of the block Toeplitz structure we use. As a result, the banded structure of (our) $\mathcal{T}(h^\perp)$ is overlooked and the complexity results provided in [24] are not particularly attractive. By taking into account the bandedness of $\mathcal{T}(h^\perp)$, the complexity of IQML becomes of the same nature as for the case of sinusoids in noise [23], and in particular is linear in the burst length M. So $\mathcal{T}(h^\perp)$ is banded block Toeplitz with $p \times m$ blocks when $\mathbf{H}^\perp(z)$ is $p \times m$ (p is the number of subchannel pairs used in the construction of $\mathbf{H}^\perp(z)$). The matrix $\mathcal{R}$ is banded and block Toeplitz with $2N-1$ block diagonals containing blocks of size $p \times p$.

For $h^\perp$ other than $h^\perp_{min}$ (or hence $p > m-1$) or for $\mathcal{T}(h^\perp)$ other than $\mathcal{T}^\perp$, $\mathcal{R}$ is singular. The main implementation choice then is between computing the pseudo-inverse $\mathcal{R}^+$ or the inverse of a regularized version such as in (29). Although a vanishing regularization has no effect on the DML gradient, the use of regularization in the iterative quadratic strategies slows down convergence for small regularization factors δ (large $\frac{1}{\delta}$): the projection matrix in (29) constrains the next iterate to be close to the previous one (all the more so as δ is smaller). Hence, using the exact pseudo-inverse is preferable. Note that on the other hand, $\mathcal{R} + \delta I$ has the same banded block Toeplitz structure as $\mathcal{R}$. In the complexity considerations below, we assume that M is much larger than N.

Computation of $\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}$. The complexity of the Lower-Diagonal-Upper (LDU) triangular factorization of $\mathcal{R}\,(+\delta I) = LDL^H$ is of order $\mathcal{O}(p^3MN^2)$ using the fact that $\mathcal{R}$ is banded. Now, using the block Toeplitz structure of $\mathcal{R}$ and the Schur algorithm [25], the complexity can be lowered to $\mathcal{O}(p^2MN)$. D is diagonal and L is a unit diagonal banded lower triangular block matrix with $p \times p$ blocks and N non-zero block diagonals. In the regular(ized) case, $\mathcal{R}^{-1} = L^{-H}D^{-1}L^{-1}$. Otherwise, the singularity of $\mathcal{R}$ manifests itself in D in which the $p \times p$ diagonal blocks contain only $m-1$ non-zero (positive) elements after an initial transient. Let D_1 be a diagonal matrix that contains only the non-zero diagonal elements of D and let L_1 contain the corresponding columns of L: $\mathcal{R} = L_1 D_1 L_1^H$. L_1 has the same structure as L except that the size of the blocks changes from initially $p \times p$ to eventually $p \times (m-1)$. The pseudo-inverse is then $\mathcal{R}^+ = L_1 (L_1^H L_1)^{-1} D_1^{-1} (L_1^H L_1)^{-1} L_1^H = L_1 (L_1^H \mathcal{R} L_1)^{-1} L_1^H$. To compute $\mathcal{R}^+$, one computes the LDU decomposition of $L_1^H L_1$ or $L_1^H \mathcal{R} L_1$ (again banded) and proceeds as in the case when $\mathcal{R}$ is regular.

Computing the product $\mathcal{Z} = L^{-1}\mathcal{Y}$ is equivalent to solving $\mathcal{Y} = L\mathcal{Z}$. Because L is lower triangular, each column of $\mathcal{Z}$ can be computed by backsubstitution (and hence the complexity for multiplying with L^{-1} or L is the same). The complexity for solving this system is $\mathcal{O}(p^2 m M N^2)$. The complexity for computing $\mathcal{Z}^H\, D^{-1}\, \mathcal{Z} = \mathcal{Y}^H \mathcal{R}^{-1} \mathcal{Y}$ is $\mathcal{O}(pm^2 M N^2)$. So the total complexity to compute $\mathcal{Y}^H \mathcal{R}^{-1} \mathcal{Y}$ is $\mathcal{O}(p^3 M N^2)$, which is $\mathcal{O}(m^3 M N^2)$ if $p = \mathcal{O}(m)$, or $\mathcal{O}(m^6 M N^2)$ if $p = \mathcal{O}(m^2)$. In the case of exact pseudo-inverse, the complexity is $\mathcal{O}(m^3 M N^2)$ (but with a higher coefficient than in the non-singular $\mathcal{R}$ case).

The term $\mathcal{Y}^H \mathcal{R}^{-1} \mathcal{Y}$ could also be computed directly using the Schur algorithm applied to partial Cholesky or LDU factorization [18] of the matrix $\begin{bmatrix} \mathcal{R} & \mathcal{Y} \\ \mathcal{Y}^H & 0 \end{bmatrix}$ which has displacement structure, leading to overall complexity of $\mathcal{O}(p^2 M N)$. The displacement rank only increases marginally when $\mathcal{T}^{\perp}$ is used instead of $\mathcal{T}(h^{\perp})$.

The computation of h as the eigenvector associated with the minimal eigenvalue of a $mN \times mN$ matrix requires $\mathcal{O}(m^3 N^3)$ computations.

5.2 Complexity of DIQML and PQML

For DIQML and PQML, the same computations as for IQML (and hence the factorization of $\mathcal{R}$) need to be performed. The computational complexity for a generalized eigenvector is similar to that for an ordinary eigenvector. We furthermore get the following additional computations.

Computation of $\mathcal{D}$. DIQML requires the computation of $\mathcal{D}$ of size $Nm \times Nm$, each entry of which is the sum of elements of $\mathcal{R}^{-1}$. The entries in $\mathcal{R}^{-1}$ that are needed for $\mathcal{D}$ are limited to the main band, of the same size as the band in $\mathcal{R}$. This band can again be computed by partial Cholesky or LDU factorization of the matrix $\begin{bmatrix} \mathcal{R} & I \\ I & 0 \end{bmatrix}$. By limiting the computations in the $(2,2)$ block to the main band, the complexity becomes $\mathcal{O}(p^3 M N^2)$, which can be reduced to $\mathcal{O}(p^2 M N)$ by exploiting the displacement structure. In the case of an exact pseudo-inverse, one can work with $\begin{bmatrix} L_1^H \mathcal{R} L_1 & L_1^H \\ L_1 & 0 \end{bmatrix}$ which again leads to a complexity of $\mathcal{O}(m^3 M N^2)$ with a higher coefficient.

$\mathcal{D}$ is block Toeplitz with $m \times m$ blocks. The complexity for generating $\mathcal{D}$ corresponds to $\mathcal{O}(4\frac{p^2}{m} M N)$ entries of $\mathcal{R}^+$ that are involved in additions.

Computation of $\mathcal{B}^H \mathcal{B}$. For PQML, the matrix $\mathcal{B}^*$ is filled with elements of the vector $\mathcal{R}^+ \mathcal{T}(h^{\perp}) \boldsymbol{Y}$. In the regular case, this becomes the computation of the vector $L^{-H}D^{-1}L^{-1}\mathcal{T}(h^{\perp})\boldsymbol{Y}$ which gets performed by executing the consecutive matrix-vector multiplications from the right. The main complexity term is $\mathcal{O}(p(m{+}2p)MN)$. In the singular case with exact pseudo-inverse, the complexity becomes $\mathcal{O}(m^2MN)$ with a higher coefficient. $\mathcal{B}$ is pre- and postwindowed block Toeplitz with $M \times N$ blocks of size $m \times m$. Hence $\mathcal{B}^H\mathcal{B}$ is block Toeplitz with blocks of size $m \times m$ and its computation from $\mathcal{B}$ requires $\mathcal{O}(m^3MN)$ operations.

6 Comparison of the Performance of DIQML and PQML

In Appendix A, we compute the asymptotic performance of DIQML and PQML under constraints with the above tangent subspace, for the complex case. We find for the channel estimation error covariance matrix:

$$C^{\mathrm{DIQML}}_{\Delta h\Delta h} = \mathrm{CRB} + \mathrm{CRB}\,\mathcal{D}^{\frac{1}{2}}P^{\perp}_{\mathcal{D}^{\frac{H}{2}}h^o}\mathcal{D}^{\frac{H}{2}}\,\mathrm{CRB} \;\geq\; \mathrm{CRB} \tag{36}$$

$$C^{\mathrm{PQML}}_{\Delta h\Delta h} = \mathrm{CRB} + \mathrm{CRB}\,\mathcal{D}^{''}\,\mathrm{CRB} \;\geq\; \mathrm{CRB} \tag{37}$$

$$= C^{\mathrm{DIQML}}_{\Delta h\Delta h} - \mathrm{CRB}\left[I - \frac{\mathcal{D}h^o h^{o\,H}}{h^{o\,H}\mathcal{D}h^o}\right]\mathcal{D}^{'}\left[I - \frac{h^o h^{o\,H}\mathcal{D}}{h^{o\,H}\mathcal{D}h^o}\right]\mathrm{CRB} \;\leq\; C^{\mathrm{DIQML}}_{\Delta h\Delta h} \tag{38}$$

where $\mathrm{CRB} = \sigma_v^2\,(\mathcal{A}^H P^{\perp}_{\mathcal{T}(h^o)}\mathcal{A})^{+} = \sigma_v^2\,(\mathcal{X}^H\mathcal{R}^{+}\mathcal{X})^{+}$, $\mathcal{D}^{'}_{i,j} = \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^{+}\Delta\mathcal{T}_j^{\perp} P_{\mathcal{T}^H(h^{o\,\perp})}\right\}$, $\mathcal{D}^{''}_{i,j} \;=\; \mathrm{tr}\left\{\Delta\mathcal{T}_i^H P^{\perp}_{\mathcal{T}(h^o)}\Delta\mathcal{T}_j(\mathcal{T}^H(h^o)\mathcal{T}(h^o))^{-1}\right\}$, $\Delta\mathcal{T}_i \;=\; \mathcal{T}\left(\dfrac{\partial h}{\partial h_i}\right)$, $\mathcal{D}^{'} \;\geq\; 0$, $\mathcal{D}^{''} \;\geq\; 0$, and $\mathcal{D} \;=\; \mathcal{D}^{'} + \mathcal{D}^{''}$. CRB and $\mathcal{D}^{''}$, and hence $C^{\mathrm{PQML}}_{\Delta h\Delta h}$, are independent of the choice of $h^{\perp}$, whereas $\mathcal{D}$ and hence $C^{\mathrm{DIQML}}_{\Delta h\Delta h}$ do depend on the choice of $h^{\perp}$. In the case of $h^{\perp}_{max}$, $\mathcal{D}\,h^o \sim h^o$ and $C^{\mathrm{DIQML}}_{\Delta h\Delta h} = \mathrm{CRB}{+}\mathrm{CRB}\;\mathcal{D}\;\mathrm{CRB}$.

The following conclusions can be drawn from this analysis:

- PQML has better performance than DIQML, and both exceed the CRB (in terms of $\mathrm{MSE} = \mathrm{tr}\,C_{\Delta h\Delta h}$). For both, $C_{\Delta h\Delta h}{-}\mathrm{CRB} \sim \sigma_v^4$ whereas $\mathrm{CRB} \sim \sigma_v^2$.
- PQML has the same asymptotic performance as DML. The PQML global minimizer is different however from the DML global minimizer.
- Asymptotically the performance of PQML remains unchanged when λ gets forced to $\lambda = 1$.
- At high SNR, DIQML, PQML and DML exhibit the same performance and all attain the CRB. It was shown in [24] that the same is true for IQML.

The analysis suggests to use the various algorithms in the following sequence

- An initial estimate can be obtained with SRM, which is non-iterative. A balanced $h^{\perp}$ should be used to avoid bias.
- The SRM estimate can be used to initialize the asymptotically globally convergent DIQML. Since the SRM estimate is consistent, asymptotically only one iteration of DIQML is required to achieve the performance attainable by DIQML. This performance is better than that of SRM since DIQML can be viewed as an optimally weighted version of (the noise part in) SRM. The use of $\mathcal{T}^{\perp}$ is recommended since it avoids singularities in $\mathcal{R}$ and hence leads to lower complexity. The use of a more elaborate $h^{\perp}$ may lead to better performance though.

– The DIQML estimate is finally used to initialize PQML which is the only algorithmic version discussed so far that attains DML performance asymptotically. Asymptotically, only one iteration is required to attain this performance since the initialization is consistent. Since the asymptotic performance of PQML is insensitive to the choice of $h^{\perp}$ as long as the full noise subspace gets spanned, the use of $\mathcal{T}^{\perp}$ is recommended.

7 Simulation Results

For the first set of simulations, we consider an irreducible channel $\boldsymbol{H}$ of length $N=4$ with $m=2$ subchannels, complex and randomly generated:

$$\boldsymbol{H}=\begin{bmatrix} -0.8285-0.1753i & 0.0557-0.2706i & 0.3411-1.2932i & 0.5545-0.7925i \\ -0.0681-0.3266i & 0.0594+0.2082i & -1.6307-0.1314i & -0.2047+0.7507i \end{bmatrix}. \tag{39}$$

The input symbols are drawn from an i.i.d. QPSK symbol sequence. The initialization of the DIQML/PQML algorithms is done by SRM.

In Figure 1, we plot the Normalized channel estimation MSE (NMSE): NMSE= $\|h^o-\hat{h}\|^2/\|h^o\|^2$ (computed under constraints (5) and (6)), the DML cost function (10), the generalized eigenvalue for PQML and the ratio between the generalized eigenvalue and σ_v^2 for DIQML, averaged over 500 Monte-Carlo runs of the noise. The burst length is $M=100$. The SNR, defined as $\dfrac{\sigma_a^2\,\|h\|^2}{m\sigma_v^2}$ (average SNR per subchannel), is 10dB. We notice that the averaged minimal generalized eigenvalue of DIQML tends to the noise variance σ_v^2 and that of PQML to 1, while remaining smaller than these values in both cases. We note however a better convergence for PQML than for DIQML. After 1 or 2 iterations, DIQML and PQML reach their steady state.

In Figures 1 and 2, the NMSEs are shown for a burst length of 100 and 200 and SNR values of 10dB and 20 dB. They are compared to the theoretical performance (derived in Appendix A) of DIQML and PQML, the last one being also the DML performance. The deterministic Cramér–Rao bound (CRB) computed under constraints (5) and (6) [3,9] is also shown; we recall that DML does not reach the CRB asymptotically in the number of data, except at high SNR. An improvement w.r.t. to SRM initialization can be observed for both algorithms, especially for PQML which outperforms DIQML. Performance can be seen to be close to the theoretical performance.

In Figure 3, we compare PQML and DIQML to AQML, to illustrate the slow convergence of AQML.

In the next two simulations, NMSE averaging is performed over independent channel realizations with i.i.d. channel coefficients. In Figure 4, we compare for $m=2$, $N=4$ and $M=100$ the NMSE as a function of SNR, for SRM, IQML, DIQML with Kristensson's denoising factor, our DIQML, the original PQML ($\lambda=1$), and our PQML. One can observe the bad performance of IQML, the improved performance of PQML w.r.t. DIQML on one hand, and due to the improved denoising factor on the other hand. In Figure 5, we compare for $m=4$, $N=4$ and $M=200$ the NMSE as a function of SNR, for SRM and various versions of PQML: using $h^{\perp}_{min}$, or $h^{\perp}_{bal,min}$ with λ being forced to 1 or not, and

$\mathcal{R}$ regularized or not. At burst length $M = 200$, the suboptimality of using $h^{\perp}_{min}$ is apparently still quite substantial. We also see that the regularized versions of PQML fail to improve upon SRM, and that at burst length $M = 200$ the proper choice of the denoising factor is still important at low SNR. Finally, we refer to [6] for a simulation comparison between smoothing or two-sided linear prediction and PQML. For the particular simulations shown there, the former algorithms are able to attain performance close to the DML performance, but no performance analysis to confirm such a trend is available.

8 Concluding Remarks

We have presented two methods, DIQML and PQML, to solve DML. These two methods correct the IQML flaw which consists of giving biased estimates at low SNR. DIQML is asymptotically globally convergent but does not reach the DML performance. PQML reaches asymptotically the DML performance with a consistent initialization, which can be provided by SRM or DIQML. Semi–blind extensions of PQML were presented in [3] and are shown to give better performance than their blind counterparts. A (blind and semi–blind) extension of PQML has also been proposed in a multiuser context (Spatial Division Multiple Access (SDMA)) in [10].

Appendix A Asymptotic Performance Study of DIQML and PQML

A.1 Asymptotic behavior of PQML ($M \to \infty$)

We prove here that PQML needs a consistent initialization in order to give a consistent estimate of the channel.

A.1.1 Inconsistent Initialization

The element (i, j) of the "Hessian" $\mathcal{P}(h)$ of the PQML cost function with $\lambda = 1$ (introducing the generalized eigenvalue does not change the following arguments much, see the next subsection also) can be written as:

$$\mathcal{P}(h)(i,j) = \underbrace{\frac{1}{M}\boldsymbol{Y}^H \Delta\mathcal{T}_i^{\perp\,H} \mathcal{R}^+(h) \Delta\mathcal{T}_j^{\perp} \boldsymbol{Y}}_{\mathcal{P}_1(h)(i,j)} \\ \underbrace{-\frac{1}{M}\boldsymbol{Y}^H \mathcal{T}^H(h^\perp)\mathcal{R}^+(h)[\Delta\mathcal{T}_j^{\perp}\Delta\mathcal{T}_i^{\perp\,H}]\mathcal{R}^+(h)\mathcal{T}(h^\perp)\boldsymbol{Y}}_{\mathcal{P}_2(h)(i,j)} \tag{40}$$

Recall that $\mathrm{E}\,\boldsymbol{Y}\boldsymbol{Y}^H = \boldsymbol{X}\boldsymbol{X}^H + \sigma_v^2 I = \mathcal{T}(h^o)AA^H\mathcal{T}^H(h^o) + \sigma_v^2 I$. Asymptotically both terms $\mathcal{P}_1(h)$ and $\mathcal{P}_2(h)$ differ from their expected value by $\mathcal{O}_p(\frac{1}{\sqrt{M}})$, and

$$M\,\mathrm{E}\,\mathcal{P}_1(h)(i,j) = \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h)\Delta\mathcal{T}_j^{\perp}\boldsymbol{X}\boldsymbol{X}^H\right\} \\ +\sigma_v^2 \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h)\Delta\mathcal{T}_j^{\perp}\right\} \tag{41}$$

$$\begin{aligned}
&M\,\mathrm{E}\,\mathcal{P}_2(h)(i,j) = \\
&\mathrm{tr}\left\{\mathcal{T}^H(h^\perp)\mathcal{R}^+(h)\left[\Delta\mathcal{T}_j^\perp \Delta\mathcal{T}_i^{\perp\,H}\right]\mathcal{R}^+(h)\mathcal{T}(h^\perp)\boldsymbol{X}\boldsymbol{X}^H\right\} \\
&\qquad\qquad +\sigma_v^2\,\mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h)\Delta\mathcal{T}_j^\perp\right\}\,. \qquad (42)
\end{aligned}$$

Let $\mathrm{E}\,\mathcal{P}_i(h) = \mathrm{E}\,\mathcal{P}_{i1}(h) + \mathrm{E}\,\mathcal{P}_{i2}(h)$, $i = 1, 2$, be a decomposition in signal and noise terms. Note that $\mathrm{E}\,\mathcal{P}_{12}(h) = \mathrm{E}\,\mathcal{P}_{22}(h)$ so that we have cancellation of the noise terms in $\mathrm{E}\,\mathcal{P}(h)$. For $h \neq \alpha h^o$, for any $\alpha \in \mathbb{C}$, $\mathrm{E}\mathcal{P}(h) \neq \mathrm{E}\mathcal{P}_{11}(h)$ (*i.e.* the noise–free IQML Hessian) because of $\mathrm{E}\mathcal{P}_{21}(h)$, the signal contribution in $\mathrm{E}\mathcal{P}_2(h)$. So, since $\mathrm{E}\mathcal{P}_{21}(h)\,h^o = \mathcal{O}(1)$ if $h - \alpha h^o = \mathcal{O}(1)$ for any $\alpha \in \mathbb{C}$, an iteration of PQML yields asymptotically an inconsistent estimate for an inconsistent initialization.

A.1.2 Consistent Initialization

Assume h is a consistent estimate of h^o, *i.e.* $h = h^o + \Delta h$, where typically $\Delta h = \mathcal{O}_p(\frac{1}{\sqrt{M}})$. We get

$$\begin{aligned}
&\mathrm{E}\mathcal{P}_{21}(h^o + \Delta h)(i,j) = \\
&\tfrac{1}{M}\,\mathrm{tr}[A^H\mathcal{T}^H(h^o)\mathcal{T}^H(\Delta h^\perp)\mathcal{R}^+(h^o)\Delta\mathcal{T}_j^\perp\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h^o) \\
&\qquad \mathcal{T}(\Delta h^\perp)\mathcal{T}(h^o)A] = O(\|\Delta h\|^2) \,=\, \mathcal{O}_p(\tfrac{1}{M}) \qquad (43)
\end{aligned}$$

whereas the other term in $\mathrm{E}\mathcal{P}(h)$, $\mathrm{E}\mathcal{P}_{11}(h)$, can be verified to be of order 1. So $\mathcal{P}_{21}(h)$ is asymptotically negligible: with a consistent initialization, the role of $\mathcal{P}_2$ is to remove the noise contribution in $\mathcal{P}_1$. Apart from terms in $\mathcal{O}_p(\frac{1}{\sqrt{M}})$, $\mathcal{P}$ becomes asymptotically equivalent to the noise–free IQML Hessian, so the estimation of h is consistent. So an iteration of PQML yields asymptotically a consistent estimate for a consistent initialization.

A.2 Performance of DIQML and PQML

We consider the following general generalized eigenvalue problem for blind channel estimation:

$$\min_{h,\lambda}\; h^H\left\{\widehat{F}(\boldsymbol{Y}, h^c) - \lambda\,\widehat{G}(\boldsymbol{Y}, h^c)\right\}h \qquad (44)$$

subject to $\widehat{F}(\boldsymbol{Y}, h^c) - \lambda\,\widehat{G}(\boldsymbol{Y}, h^c) \geq 0$ and constraints on h. h^c is a consistent estimate of h. $\widehat{F}(\boldsymbol{Y}, h^c) = \frac{1}{M}\mathcal{Y}^H\mathcal{R}^+(h^c)\mathcal{Y} = \mathcal{P}_1(h^c)$ for DIQML and PQML, $\widehat{G}(\boldsymbol{Y}, h^c) = \frac{1}{M}\mathcal{D}(h^c)$ for DIQML and $\widehat{G}(\boldsymbol{Y}, h^c) = \frac{1}{M}\mathcal{B}^H(h^c)\mathcal{B}(h^c) = \mathcal{P}_2(h^c)$ for PQML. It can be shown that the channel estimation performance given by (44) is asymptotically unchanged when one replaces $\widehat{F}(\boldsymbol{Y}, h^c)$ and $\widehat{G}(\boldsymbol{Y}, h^c)$ by $\widehat{F}(\boldsymbol{Y}) = \widehat{F}(\boldsymbol{Y}, h^o)$ and $\widehat{G}(\boldsymbol{Y}) = \widehat{G}(\boldsymbol{Y}, h^o)$ respectively (since $\mathcal{O}(||\Delta h^c||^2) = \mathcal{O}_p(\frac{1}{M})$). Asymptotically, we also have:

$$\begin{cases}\widehat{F}(\boldsymbol{Y}) = F^o + \mathcal{O}_p(\frac{1}{\sqrt{M}}) \\ \widehat{G}(\boldsymbol{Y}) = G^o + \mathcal{O}_p(\frac{1}{\sqrt{M}})\end{cases} \qquad (45)$$

where $F^o(h^c) = \mathrm{E}\,\widehat{F}(\boldsymbol{Y}, h^c)$, $G^o(h^c) = \mathrm{E}\,\widehat{G}(\boldsymbol{Y}, h^c)$ and $F^o = F^o(h^o)$, $G^o = G^o(h^o)$. Although we will not need this, one may also remark that $F^o(h^o) = \lim\limits_{M\to\infty} F^o(h^o) + \mathcal{O}(\dfrac{1}{M})$ and similarly for G^o.

A.2.1 Asymptotic Expression for $\Delta\lambda$

The solution of (44) for λ and h is the minimal generalized eigenvalue and corresponding eigenvector of $\widehat{F}(\boldsymbol{Y})$ and $\widehat{G}(\boldsymbol{Y})$.

$$\widehat{F}(\boldsymbol{Y})\,\hat{h} - \hat{\lambda}\,\widehat{G}(\boldsymbol{Y})\,\hat{h} = 0 \;\;\Rightarrow\;\; \hat{\lambda} = \frac{\hat{h}^H\,\widehat{F}(\boldsymbol{Y})\,\hat{h}}{\hat{h}^H\,\widehat{G}(\boldsymbol{Y})\,\hat{h}}\,. \tag{46}$$

We denote $\hat{h} = h^o + \Delta h$, and $\hat{\lambda} = \lambda^o + \Delta\lambda$, where $\Delta h \overset{M\to\infty}{\longrightarrow} 0$, $\Delta\lambda \overset{M\to\infty}{\longrightarrow} 0$. We have $\lambda^o = \dfrac{h^{o\,H}\,F^o\,h^o}{h^{o\,H}\,G^o\,h^o}$ and, performing a series expansion, we get:

$$\Delta\lambda = \frac{h^{o\,H}\,[\widehat{F}(\boldsymbol{Y}) - \lambda^o\widehat{G}(\boldsymbol{Y})]\,h^o}{h^{o\,H}\,G^o\,h^o} \;+\; \mathcal{O}_p(\frac{1}{M})\,. \tag{47}$$

A.2.2 Asymptotic Expressions for Δh and $C_{\Delta h\Delta h} = E(\hat{h}-h^o)(\hat{h}-h^o)^H$

After substitution of the solution for λ, the estimation problem for h becomes:

$$\min_h \;\left\{h^H\left\{\widehat{F}(\boldsymbol{Y}) - \hat{\lambda}(\boldsymbol{Y})\widehat{G}(\boldsymbol{Y})\right\}h = \mathcal{F}(h)\right\}\,. \tag{48}$$

The estimation of h is performed under constraints $\mathcal{K}(h_R) = 0$ with tangent subspace $\mathcal{M}_{h_R^o}$ at $h_R = h_R^o$. Let $\mathcal{V}_R^o$ be a matrix whose columns form an orthonormal basis of $\mathcal{M}_{h_R^o}$. Then locally we can write $\Delta h_R = \mathcal{V}_R^o\,\theta$ where θ are the unconstrained parameter variations. A Taylor series expansion of $\mathcal{F}(h)$ at h^o in terms of θ gives

$$\begin{aligned} \mathcal{F}(h) &= \mathcal{F}(h^o) + \theta^T\mathcal{V}_R^{o\,T}\frac{\partial\mathcal{F}(h^o)}{\partial h_R} \\ &+\tfrac{1}{2}\theta^T\mathcal{V}_R^{o\,T}\frac{\partial^2\mathcal{F}(h^o)}{\partial h_R\partial h_R^T}\mathcal{V}_R^o\theta + \mathcal{O}(||\theta||^3) \end{aligned} \tag{49}$$

Optimization of (49) up to second order w.r.t. θ gives for $\Delta h_R = \mathcal{V}_R^o\,\theta$

$$\Delta h_R = \mathcal{V}_R^o\left(\mathcal{V}_R^{o\,T}\frac{\partial^2\mathcal{F}(h^o)}{\partial h_R\partial h_R^T}\mathcal{V}_R^o\right)^{-1}\mathcal{V}_R^{o\,T}\;\frac{\partial\mathcal{F}(h^o)}{\partial h_R} \tag{50}$$

assuming that the matrix inverse exists (which will be the case here). The expression becomes easier to work with when expressed in terms of complex quantities (see [12]):

$$\Delta h = \mathcal{V}^o\left(\mathcal{V}^{o\,H}\frac{\partial}{\partial h^*}\left(\frac{\partial\mathcal{F}(h^o)}{\partial h^*}\right)^H\mathcal{V}^o\right)^{-1}\mathcal{V}^{o\,H}\frac{\partial\mathcal{F}(h^o)}{\partial h^*}\;\;. \tag{51}$$

For the constraints (5), (6) or equivalent, the columns of $\mathcal{V}^o$ form a basis for the orthogonal complement of h^o. We shall also require

$$\begin{cases} J_{hh}^{(1)} = \mathrm{E}\left(\dfrac{\partial\mathcal{F}(h^o)}{\partial h^*}\right)\left(\dfrac{\partial\mathcal{F}(h^o)}{\partial h^*}\right)^H \\ J_{hh}^{(2)} = \mathrm{E}\,\dfrac{\partial}{\partial h^*}\left(\dfrac{\partial\mathcal{F}(h^o)}{\partial h^*}\right)^H\,. \end{cases} \tag{52}$$

Note that if $\mathcal{F}$ would have been the log likelihood function, then $J_{hh}^{(1)} = -J_{hh}^{(2)}$, but this equality does not hold here. We now obtain

$$\begin{aligned}
&\Delta h = \mathcal{V}^o \left(\mathcal{V}^{o\,H} J_{hh}^{(2)} \mathcal{V}^o\right)^{-1} \mathcal{V}^{o\,H} \frac{\partial \mathcal{F}(h^o)}{\partial h^*} + \mathcal{O}_p(\frac{1}{M}) \\
&C_{\Delta h \Delta h} = \\
&\mathcal{V}^o \left(\mathcal{V}^{o\,H} J_{hh}^{(2)} \mathcal{V}^o\right)^{-1} \mathcal{V}^{o\,H} J_{hh}^{(1)} \mathcal{V}^o \left(\mathcal{V}^{o\,H} J_{hh}^{(2)} \mathcal{V}^o\right)^{-1} \mathcal{V}^{o\,H} \\
&\qquad + o(\tfrac{1}{M})
\end{aligned} \tag{53}$$

For the quadratic problem in (48), we have (using (47) and the fact that ΔF and ΔG have zero mean):

$$J_{hh}^{(2)} = \mathrm{E}\left(\widehat{F}(\boldsymbol{Y}) - \hat{\lambda}(\boldsymbol{Y})\widehat{G}(\boldsymbol{Y})\right) = F^o - \lambda^o G^o + \mathcal{O}(\frac{1}{M}) \tag{54}$$

where we shall neglect the last term.

A.2.3 Application to DIQML and PQML

Specializing to DIQML and PQML, we get first of all $F^o - \lambda^o G^o = \frac{1}{M}\,\mathcal{X}^H \mathcal{R}^+ \mathcal{X}$. To show the relation of this expression to the CRB, consider for any h, $h^{'}$:

$$\begin{aligned}
&h^H \mathcal{X}^H \mathcal{R}^+ \mathcal{X} h^{'} = \\
&\boldsymbol{X}^H \mathcal{T}^H(h^{\perp}) \mathcal{R}^+ \mathcal{T}(h^{'\,\perp}) \boldsymbol{X} \\
&= A^H \mathcal{T}^H(h^o) \mathcal{T}^H(h^{\perp}) \mathcal{R}^+ \mathcal{T}(h^{'\,\perp}) \mathcal{T}(h^o) A \\
&= A^H \mathcal{T}^H(h) \mathcal{T}^H(h^{o\,\perp}) \mathcal{R}^+ \mathcal{T}(h^{o\,\perp}) \mathcal{T}(h^{'}) A \\
&= h^H \mathcal{A}^H P^{\perp}_{\mathcal{T}(h^o)} \mathcal{A} h^{'}
\end{aligned} \tag{55}$$

where $\mathcal{T}(h)A = \mathcal{A}h$. Hence $\sigma_v^{-2} \mathcal{X}^H \mathcal{R}^+ \mathcal{X} = \sigma_v^{-2} \mathcal{A}^H P^{\perp}_{\mathcal{T}(h^o)} \mathcal{A}$ which is the Fisher information matrix for deterministic ML. As $F^o - \lambda^o G^o$ admits h^o as unique eigenvector corresponding to the eigenvalue zero, and $\mathcal{V}^o$ spans the orthogonal complement of h^o,

$$\mathcal{V}^o \left(\mathcal{V}^{oH} J_{hh}^{(2)} \mathcal{V}^o\right)^{-1} \mathcal{V}^{o\,H} = (F^o - \lambda^o G^o)^+ \ , \tag{56}$$

the Moore-Penrose pseudo-inverse of $F^o - \lambda^o G^o$. Hence

$$\Delta h = (F^o - \lambda^o G^o)^+ \ \frac{\partial \mathcal{F}(h^o)}{\partial h^*} \tag{57}$$

neglecting $\mathcal{O}_p(\frac{1}{M})$ terms. Now, using (47), we also get

$$\begin{aligned}
\frac{\partial \mathcal{F}(h^o)}{\partial h^*} &= \left(\widehat{F}(\boldsymbol{Y}) - \hat{\lambda}(\boldsymbol{Y})\,\widehat{G}(\boldsymbol{Y})\right) h^o \\
&= \left(\widehat{F}(\boldsymbol{Y}) - \lambda^o\,\widehat{G}(\boldsymbol{Y})\right) h^o - G^o\,h^o\,\Delta\lambda(\boldsymbol{Y}) \\
&= \left[I - \frac{G^o\,h^o\,h^{o\,H}}{h^{o\,H}\,G^o\,h^o}\right] \left(\widehat{F} - \lambda^o\,\widehat{G}\right) h^o
\end{aligned} \tag{58}$$

which leads to

$$\Delta h = M \left(\mathcal{X}^H \mathcal{R}^+ \mathcal{X}\right)^+ \left[I - \frac{G^o\, h^o\, h^{o\,H}}{h^{o\,H}\, G^o\, h^o}\right] \left(\widehat{F} - \lambda^o\, \widehat{G}\right) h^o \tag{59}$$

neglecting $\mathcal{O}_p(\frac{1}{M})$ terms. For DML, the same kind of analysis gives [12]:

$$\Delta h_{DML} = M \left(\mathcal{X}^H \mathcal{R}^+ \mathcal{X}\right)^+ \left(\widehat{F} - \lambda^o\, \widehat{G}\right) h^o \tag{60}$$

where $\widehat{F}(\boldsymbol{Y})$ and $\widehat{G}(\boldsymbol{Y})$ are the same as in the PQML case. So the estimate $\hat{h}$ given by DIQML and PQML is different from the DML estimate (though the difference with PQML is only $\mathcal{O}_p(\frac{1}{M})$). From (59), we see that the channel estimation performance depends on the matrix

$$\mathcal{W} = \mathrm{E}\left\{\left(\widehat{F} - \lambda^o \widehat{G}\right) h^o h^{o^H} \left(\widehat{F} - \lambda^o \widehat{G}\right)^H\right\} . \tag{61}$$

Recall that for both DIQML and PQML, $\widehat{F}(\boldsymbol{Y}) = \frac{1}{M}\mathcal{Y}^H\mathcal{R}^+(h^o)\mathcal{Y}$, $F^o = \frac{1}{M}\mathcal{X}^H\mathcal{R}^+(h^o)\mathcal{X} + \frac{\sigma_v^2}{M}\mathcal{D}(h^o)$.

Performance of DIQML

For DIQML, $\widehat{G}(\boldsymbol{Y}) = \frac{1}{M}\mathcal{D}(h^o) = G^o$, $\lambda^o = \sigma_v^2$ and hence $F^o - \lambda^o G^o = \frac{1}{M}\mathcal{X}^H\mathcal{R}^+(h^o)\mathcal{X}$. We have:

$$\mathcal{W}^{\mathrm{DIQML}} = \frac{1}{M^2}\left[\sigma_v^2\, \mathcal{X}^H \mathcal{R}^+ \mathcal{X} + \sigma_v^4\, \mathcal{D}\right] \tag{62}$$

which leads to (36).

Performance of PQML

Now $\widehat{G}(\boldsymbol{Y}) = \frac{1}{M}\mathcal{B}^H(h^o)\mathcal{B}(h^o)$, $G^o = \frac{\sigma_v^2}{M}\mathcal{D}(h^o)$, $\lambda^o = 1$ and $F^o - \lambda^o G^o = \frac{1}{M}\mathcal{X}^H\mathcal{R}^+(h^o)\mathcal{X}$. We get:

$$\begin{aligned} &\mathcal{W}_{i,j}^{\mathrm{PQML}} = \\ &\tfrac{\sigma_v^2}{M^2}\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)_{i,j} + \tfrac{\sigma_v^4}{M^2}\mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+\Delta\mathcal{T}_j^{\perp}P_{\mathcal{T}(h^o)}\right\} \\ &\qquad = \mathcal{W}_{i,j}^{\mathrm{DIQML}} - \tfrac{\sigma_v^4}{M^2}\mathcal{D}'_{i,j} \end{aligned} \tag{63}$$

where $\mathcal{D}'$ is defined below (38). Note that $\mathcal{D}' h^o = \mathcal{D}\, h^o$ and for any h', $h'^{\,H}\mathcal{W}^{\mathrm{PQML}}h^o = \frac{\sigma_v^2}{M^2}h'^{\,H}\,\mathcal{X}^H\mathcal{R}^+\mathcal{X}\,h^o + \frac{\sigma_v^4}{M^2}\mathrm{tr}\left\{\mathcal{T}^H(h'^{\perp})\mathcal{R}^+\mathcal{T}(h^{o\perp})P_{\mathcal{T}(h^o)}\right\} = 0 + 0 = 0$: $\mathcal{W}^{\mathrm{PQML}}$

has a null space spanned by h^o. Now, for any h, $h^{'}$, we have

$$
\begin{aligned}
h^H(\mathcal{D}-\mathcal{D}^{'})h^{'} &= \mathrm{tr}\left\{\mathcal{T}^H(h^{\perp})\mathcal{R}^{+}(h^o)\mathcal{T}(h^{'\perp})\right\} - \\
&\quad \mathrm{tr}\left\{\mathcal{T}^H(h^{\perp})\mathcal{R}^{+}(h^o)\mathcal{T}(h^{'\perp})P_{\mathcal{T}^H(h^{o\perp})}\right\} \\
&= \mathrm{tr}\left\{\mathcal{T}^H(h^{\perp})\mathcal{R}^{+}(h^o)\mathcal{T}(h^{'\perp})P^{\perp}_{\mathcal{T}^H(h^{o\perp})}\right\} \\
&= \mathrm{tr}\left\{\mathcal{T}^H(h^{\perp})\mathcal{R}^{+}(h^o)\mathcal{T}(h^{'\perp})P_{\mathcal{T}(h^o)}\right\} \\
&= \mathrm{tr}\left\{\mathcal{T}^H(h^o)\mathcal{T}^H(h^{\perp})\mathcal{R}^{+}(h^o)\mathcal{T}(h^{'\perp})\right. \\
&\qquad\qquad \left.\mathcal{T}(h^o)(\mathcal{T}^H(h^o)\mathcal{T}^(h^o))^{-1}\right\} \\
&= \mathrm{tr}\left\{\mathcal{T}^H(h)\mathcal{T}^H(h^{o\perp})\mathcal{R}^{+}(h^o)\mathcal{T}(h^{o\perp})\right. \\
&\qquad\qquad \left.\mathcal{T}(h^{'})(\mathcal{T}^H(h^o)\mathcal{T}(h^o))^{-1}\right\} \\
&= \mathrm{tr}\left\{\mathcal{T}^H(h)P_{\mathcal{T}^H(h^{o\perp})}\mathcal{T}(h^{'})(\mathcal{T}^H(h^o)\mathcal{T}(h^o))^{-1}\right\} \\
&= \mathrm{tr}\left\{\mathcal{T}^H(h)P^{\perp}_{\mathcal{T}(h^o)}\mathcal{T}(h^{'})(\mathcal{T}^H(h^o)\mathcal{T}(h^o))^{-1}\right\} = h^H\,\mathcal{D}^{''}\,h^{'}
\end{aligned}
\tag{64}
$$

or hence $\mathcal{D}^{''} = \mathcal{D}-\mathcal{D}^{'}$, where $\mathcal{D}^{''}$ is defined below (38) and we used the commutativity of convolution, leading to $\mathcal{T}(h^{\perp})\mathcal{T}(h^o) = \mathcal{T}(h^{o\perp})\mathcal{T}(h)$. (37), (38) now follow. Note that the factor $\left[I - \frac{G^o\,h^o\,h^{o\,H}}{h^{o\,H}\,G^o\,h^o}\right]$ in (59), which is due to $\Delta\lambda$, has asymptotically no effect on $C^{\mathrm{PQML}}_{\Delta h\Delta h}$. So asymptotically $C^{\mathrm{PQML}}_{\Delta h\Delta h} = C^{\mathrm{DML}}_{\Delta h\Delta h}$ [12]. In fact, for PQML, $\Delta\lambda = \mathcal{O}_p(\frac{1}{M})$, whereas $\left(\widehat{F} - \lambda^o\,\widehat{G}\right)h^o = \mathcal{O}_p(\frac{1}{\sqrt{M}})$. Hence also, forcing $\lambda = 1$ in PQML does not influence the performance asymptotically.

Recall that the various substitutions above of $P_{\mathcal{T}^H(h^{\perp})}$ by $P^{\perp}_{\mathcal{T}(h)}$ are correct for all versions of $h^{\perp}$ with $p \geq m$ that include $h^{\perp}_{bal,min}$, or for $\mathcal{T}^{\perp}$, but for $h^{\perp}_{min}$ ($p = m-1$) constitute an approximation that becomes justifiable only asymptotically (in M).

References

1. Abed Meraim, K., Moulines, E., Loubaton, P.: Prediction Error Method for Second-Order Blind Identification. IEEE Transactions on Signal Processing **45**(3), 694–705 (1997)
2. Alberge, F., Duhamel, P., Nikolova, M.: Adaptive Solution for Blind Identification/Equalization Using Deterministic Maximum Likelihood. IEEE Transactions on Signal Processing **50**(4), 923–936 (2002)
3. Ayadi, J., de Carvalho, E., Slock, D.: Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification. In: Proc. ICASSP 98 Conf. Seattle, USA (1998)
4. Ayadi, J., Deneire, L., Slock, D.: Channel-based Blind Multichannel Identification without Order Overestimation Problems. In: 13^{th} Int'l Conf. on Digital Signal Processing. Santorini, Greece (1997)
5. Ayadi, J., Slock, D.: Blind Channel Estimation and Joint Order Detection by MMSE ZF Equalization. In: IEEE Vehic. Tech. Conf. (VTC). Amsterdam, The Netherlands (1999)
6. Ayadi, J., Slock, D.: Multichannel Estimation by Blind MMSE ZF Equalization. In: IEEE-SP Workshop on Sig. Proc. Advances in Wireless Comm's (SPAWC). Annapolis, Maryland, USA (1999)
7. Baccala, L., Roy, S.: A New Time-Domain Blind Identification Method. IEEE Signal Processing Letters **1**(6), 89–91 (1994)
8. Bresler, Y., Macovski, A.: Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise. IEEE Transactions on Acoustics, Speech, Signal Processing **35**(10), 1081–1089 (1986)

9. de Carvalho, E., Cioffi, J., Slock, D.: Cramer-Rao Bounds for Blind Multichannel Estimation. In: IEEE Globecom Conf. San Fransisco, USA (2000)
10. de Carvalho, E., Deneire, L., Slock, D.: Blind and Semi-Blind Maximum Likelihood Techniques for Multiuser Multichannel Identification. In: European Association for Signal Processing EUSIPCO 98. Island of Rhodes, Greece (1998)
11. de Carvalho, E., Slock, D.: Maximum-Likelihood Blind Equalization of Multiple FIR Channels. In: Proc. ICASSP 96 Conf. Atlanta, USA (1996)
12. de Carvalho, E., Slock, D.: Asymptotic Performance of ML Methods for Semi-Blind Channel Estimation. In: Proc. Asilomar Conference on Signals, Systems & Computers. Pacific Grove, CA (1997)
13. de Carvalho, E., Slock, D.: Identifiability Conditions for Blind and Semi-Blind Multichannel Estimation. In: European Association for Signal Processing EUSIPCO 98. Island of Rhodes, Greece (1998)
14. de Carvalho, E., Slock, D.: Burst Mode Equalization: Optimal Approach and Suboptimal Continuous–Processing Approximation. Signal Processing (EURASIP), Special Issue on Signal Processing Technologies for Short–Burst Wireless Communications (2000)
15. de Carvalho, E., Slock, D.: Blind and Semi–Blind FIR Multichannel Estimation: Identifiability Conditions. IEEE Transactions on Signal Processing **52**, 1053–1064 (2004)
16. Chan, K., So, H., Amin, M.T., Chan, C., Lau, W..: Iterative quadratic maximum likelihood based estimator for a biased sinusoid. EURASIP Signal Processing **90**(6), 20832086 (2010)
17. Chotikakamthorn, N., Chambers, J.: On the IQML Algorithm for Multiple Signal Parameter Estimation. IEE Proceedings-Radar, Sonar and Navigation **5**(144), 237–244 (1997)
18. Chun, J.: Fast array algorithms for structured matrices. Ph.D. thesis, Stanford University (1989)
19. Gesbert, D., Duhamel, P., Mayrargue, S.: Blind Least–Squares Criteria for Joint Data/Channel Estimation. In: Proc. IEEE Digital Signal Processing Workshop (1996)
20. Giannakis, G., Halford, S.: Asymptotically Optimal Blind Fractionally-Spaced Channel Estimation and Performance Analysis. IEEE Transactions on Signal Processing **45**(7), 1815–1830 (1997)
21. Gürelli, M., Nikias, C.: A New Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals. In: Proc. ICASSP, pp. 448–451 (1993)
22. Harikumar, G., Bresler, Y.: Analysis and Comparative Evaluation of Techniques for Multichannel Blind Deconvolution. In: Proc. 8th IEEE Sig. Proc. Workshop Statistical Signal and Array Proc., pp. 332–335. Corfu, Greece (1996)
23. Hua, Y.: The Most Efficient Implementation of the IQML Algorithm. IEEE Transactions on Signal Processing **42**(8), 2203–2204 (1994)
24. Hua, Y.: Fast Maximum Likelihood for Blind Identification of Multiple FIR Channels. IEEE Transactions on Signal Processing **44**(3), 661–672 (1996)
25. Kailath, T., Sayed, A.: Displacement Structure: Theory and Applications. SIAM Review **37**(3), 297–386 (1995)
26. Kristensson, M., Jansson, M., Ottersten, B.: Modified IQML and Weighted Subspace Fitting without Eigendecomposition. Signal Processing **79**(1), 29–44 (1999)
27. Kristensson, M., Ottersten, B., Slock, D.: Blind Subspace Identification of a BPSK Communication Channel. In: Proc. of the 30th Asilomar Conference on Signals, Systems & Computers. Pacific Grove, CA (1996)
28. Mani, V., Rose, R.: Direction of Arrival Estimation and Beamforming of Multiple Coherent UWB Signals. In: Proc. ICC 2010 Conf. Cape town, USA (2010)
29. Moulines, E., Duhamel, P., Cardoso, J., Mayrargue, S.: Subspace Methods for the Blind Identification of Multichannel FIR filters. IEEE Transactions on Signal Processing **43**(2), 516–526 (1995)
30. Osborne, M., Smyth, G.: A Modified Prony Algorithm for Fitting Functions Defined by Difference Equations. SIAM J. Sci. Stat. Comput. **12**(2), 362–382 (1991)
31. Pité, E., Duhamel, P.: Bilinear Methods for Blind Channel Equalization: (No) Local Minimum Issue. In: Proc. ICASSP 98 Conf. Seattle, USA (1998)
32. Ramkumar, B., Bose, T., Radenkovic, M.: Combined Blind Equalization and Automatic Modulation Classification for Cognitive Radios. In: IEEE DSP/SPE 2009 (2009)
33. Ramkumar, B., Bose, T., Radenkovic, M.: Robust Multiuser Automatic Modulation Classifier for Multipath Fading Channels. In: IEEE Symposium on New Frontiers in Dynamic Spectrum (2010)
34. Slock, D., Papadias, C.: Blind Fractionally-Spaced Equalization Based on Cyclostationarity. In: Proc. Vehicular Technology Conf. Stockholm, Sweden (1994)

35. Slock, D.: Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction. In: Proc. ICASSP 94 Conf. Adelaide, Australia (1994)
36. Slock, D.: Blind Joint Equalization of Multiple Synchronous Mobile Users Using Oversampling and/or Multiple Antennas. In: Proc. 28th Asilomar Conference on Signal, Systems & Computers. Pacific Grove, CA (1994)
37. Stoica, P., Li, J., Söderström, T.: On the Inconsistency of IQML. Signal Processing pp. 185–190 (1997)
38. Talwar, S., Viberg, M., Paulraj, A.: Blind Separation of Synchronous Co-Channel Digital Signals Using an Antenna Array. Part I. Algorithms. IEEE Transactions on Signal Processing **44**(5), 1184–1197 (1996)
39. Talwar, S.: Blind space-time algorithms for wireless communication system. Ph.D. thesis, Stanford University (1996)
40. Tong, L., Perreau, S.: Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods. Proceedings of the IEEE **86**(10), 1951–1968 (1998)
41. Tong, L., Zhao, Q.: Joint Order Detection and Blind Channel Estimation by Least Squares Smoothing. IEEE Transactions on Signal Processing **47**(9) (1999)
42. Trigui, H., Slock, D.: Cochannel Interference Cancellation within the Current GSM Standard. In: Proc. Workshop COST254 (1997). Toulouse, France
43. der Veen, A.V., Talwar, S., Paulraj, A.: Blind Estimation of Multiple Digital Signals transmitted over FIR Channels. IEEE Signal Processing Letters **2**(5), 99–102 (1995)
44. van der Veen, A.: Analytical Method for Blind Binary Signal Separation. IEEE Transactions on Signal Processing **45**(4), 1078–1082 (1997)
45. Wu, H., Saquib, M., Yun, Z.: Novel Automatic Modulation Classification Using Cumulant Features for Communications via Multipath Channels. IEEE Transactions on Wireless Communications **7**(8), 3098 – 3105 (2008)
46. Xu, G., Liu, H., Tong, L., Kailath, T.: A Least Squares Approach to Blind Channel Identification. IEEE Transactions on Signal Processing **43**(12), 2982–2993 (1995)
47. Zhan, H., Ayadi, J., Farserotu, J., Boudec, J.Y.L.: Impulse radio ultra-wideband ranging based on maximum likelihood estimation. IEEE Transactions on Wireless Communications **8**(12), 2982–2993 (2009)

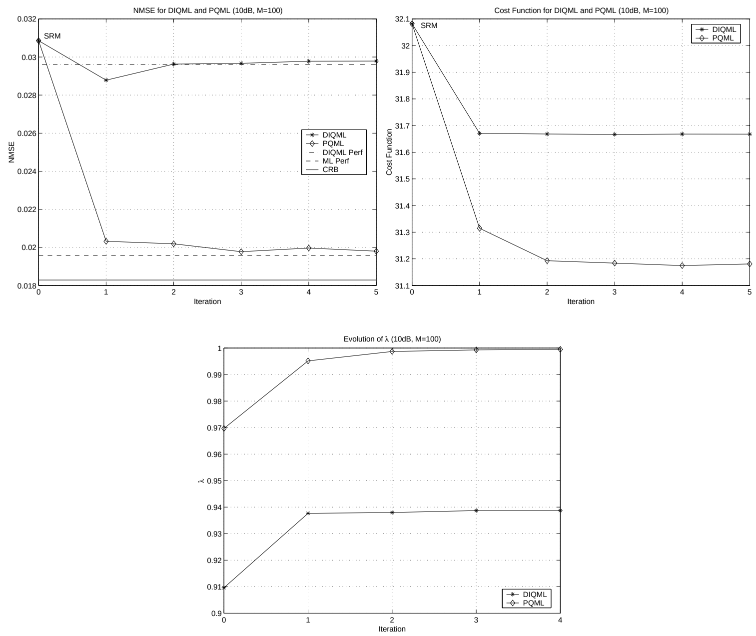

Fig. 1 NMSE, cost function, generalized eigenvalue for DIQML and PQML at 10dB, for a burst length of 100.

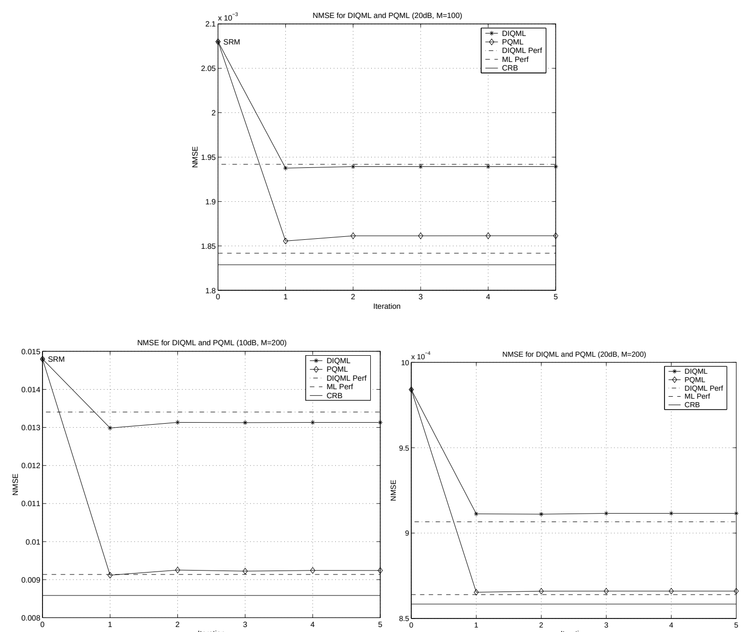

Fig. 2 NMSE for DIQML and PQML at 10dB and 20dB for a burst length of 100 and 200.

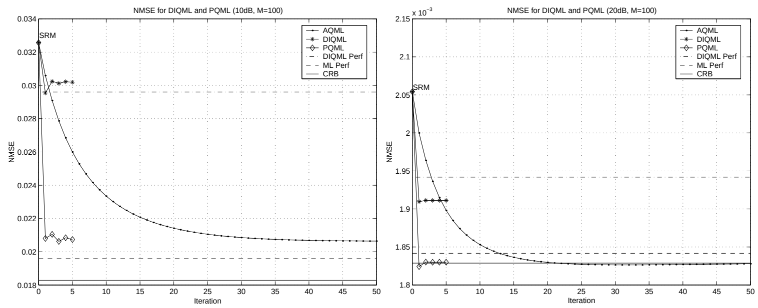

Fig. 3 Comparison between DIQML, PQML and AQML at 10dB and 20 dB for a burst length of 100.

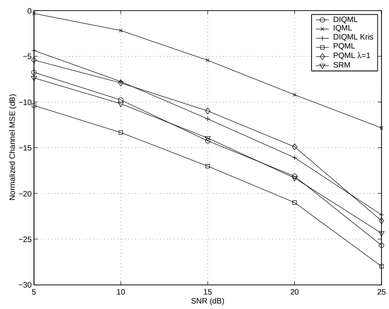

Fig. 4 NMSE for SRM, IQML, DIQML with Kristensson's denoising factor, our DIQML, the original PQML ($\lambda = 1$), and

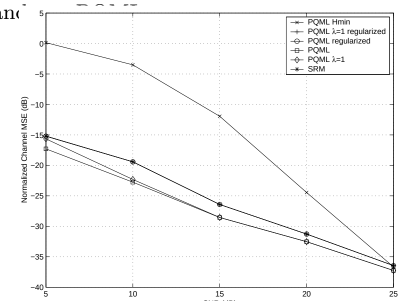

Fig. 5 NMSE for SRM and various versions of PQML: using $h^{\perp}_{min}$, or $h^{\perp}_{bal,min}$ with λ being forced to 1 or not, and $\mathcal{R}$ regularized or not.