Semi-Blind Sparse Channel Learning in Cell-Free Massive MIMO - a CRB Analysis

Zilu Zhao, Dirk Slock
EURECOM, Sophia Antipolis, France
Email: {zilu.zhao,dirk.slock}@eurecom.fr

***Abstract*—In this paper we consider cell-free (CF) massive MIMO (MaMIMO) systems, which comprise a very large number of geographically distributed access points (APs) serving a much smaller number of users. We exploit channel sparsity to tackle pilot contamination, which originates from the reuse of pilot sequences. Specifically, we consider semi-blind methods for channel estimation in the presence of unknown Gaussian i.i.d. data to resolve the pilot contamination. This task is further aided by exploiting prior channel information in a Bayesian formulation. We develop Bayesian Maximum a Posteriori (MAP) channel estimators and we also provide various Cramer-Rao Bounds to characterize performance limits. The main contribution is the derivation of an original type of Bayesian CRB for the semi-blind problem at hand, in which a certain expectation operation is facilitated by the asymptotics of the large system dimensions considered here. Whereas Bayesian CRBs lead to fairly useless lose bounds, corresponding to unrealistic genie-aided scenarios, the proposed variation turns out to be quite tight as illustrated by performance comparisons with various estimation algorithms.**

I. Introduction

Recently, cell-free (CF) massive MIMO (MaMIMO) systems are attracting extensive research interests as an effective and promising approach for next generation wireless systems thanks to their potential to reap the benefit of both MaMIMO and distributed antenna systems (DAS). CF MaMIMO systems consist of a massive number of access points (APs) which serve a much smaller number of single-antenna users and are geographically distributed over a large coverage area. All the APs are connected through a back-haul network to a central processing unit (CPU). The massive number of antennas improves spectral efficiency [1] whereas energy efficiency [2], [3] and macro-diversity gain result from the distributed topology and ultra-densification. Additionally, since each user is surrounded by a large number of serving APs, with high probability all the users enjoy good channel conditions [4]. Therefore, CF MaMIMO systems are expected to provide significant improvements in terms of spectral/energy efficiency and coverage probability. In [1], [5], the performance of CF MaMIMO and small-cell systems were compared under the assumption of employing maximum ratio (MR) processing. In [6]–[10], the authors advocated the use of more effective processing than MR processing in CF MaMIMO to guarantee superior performance of CF MaMIMO systems compared to small-cell systems. The performance of CF MaMIMO systems is critically affected by the so-called *pilot contamination*. This impairment originates from the reuse of training sequences or pilots utilized in channel estimation, which prevents the possibility of obtaining an adequate estimate of the channel state information (CSI). The detrimental effects of pilot contamination were highlighted in [11] for centralized MaMIMO systems. Specific features of centralized MaMIMO channels such as channel hardening and favorable propagation or limited angular spread could be exploited to "separate" user channels in power domain [12], angular domain [13], [14], or jointly in power and angular domain [15] and thus, mitigate or annihilate pilot contamination. However, these appealing properties of channels in centralized MaMIMO systems are destroyed in a distributed setting and pilot contamination is still an open and challenging problem in CF MaMIMO systems. Several pilot assignment (PA) methods for mitigating pilot contamination in CF MaMIMO systems were proposed recently in [1], [16]–[18]. In [1], a greedy pilot assignment (GPA) based on knowledge of large-scale fading channel coefficients was proposed. In [16], a location-based greedy (LBG) pilot assignment scheme utilized the location information in a GPA algorithm. The structured PA approach proposed in [17] maximized the minimum geographical distance between users sharing the same pilot sequences. An additional PA method based on graph coloring was proposed in [18]. All these techniques address the pilot contamination problem via a careful assignment of pilots and do not exploit the inherent structure of channels and data in CF MaMIMO systems in contrast to blind or semi-blind estimation and detection techniques. A blind pilot decontamination approach was proposed first in [12] for centralized MaMIMO systems and utilized asymptotic orthogonality of user channels to remove undesired interference including pilot contamination from the received signal. The same property was also exploited for semi-blind channel estimation, e.g., [15], in centralized MaMIMO but it does not hold in CF MaMIMO systems [8], [9]. Blind and semi-blind channel estimation have been thoroughly investigated in general settings, see, e.g., [19]–[22] and references therein. In this context, the concept of identifiability was very relevant since it guarantees the non-singularity of the Fisher information matrix (FIM) and thus, the existence of the Cramer-Rao bound (CRB). The corresponding conditions provide fundamental insights into the feasibility of reliable communications in the analyzed system. Conditions under which channel and data signals are blindly and semi-blindly identifiable have been thoroughly studied in various settings for centralized systems, see, e.g., [23], [24]. A semi-blind pilot decontamination approach was proposed in [25] exploiting the inherent sparse structure of channels and correlation structure of data in CF massive MIMO systems. In [26], we proposed a new variable level expectation propagation (VL-EP) algorithm to iteratively conduct the semi-blind method for channel estimation in the presence of Gaussian i.i.d. data to mitigate the pilot contamination problem originating from the reuse of training sequences. The proposed VL-EP algorithm provides an approximate minimum mean square error (MMSE) channel estimator which can not be found analytically. The main constributions in this paper are as follows:

- Whereas we proposed a basic Maximum a Posteriori (MAP) algorithm for the joint estimation of channels and unknown data in [25], we present here further MAP estimation algorithms for the channels separately, with the unknown data symbols being eliminated from the inference problem. The performance evaluations lead us to observe that MAP estimation of a reduced parameter subset leads to estimates that are apparently closer to MMSE estimates, compared to the case in which a joint MAP estimation of all parameters is performed.
- CRBs are obtained as the result of an expectation step in the Fisher Information Matrix, followed by an inversion. In the Bayesian case there is a further expectation step over prior information. Applied to the (semi-)blind estimation of both channels and data, the result is a bound in which the coupling in the estimation between channels and data disappears, and each gets estimated as if the other were known, leading to a very lose bound. We propose an original Bayesian CRB in which the expectations are based on approximate large system asymptotics, and an analytical solution is obtained by interpolating the manageable low and high SNR regimes results. Given that the best algorithms approach the resulting bound quite closely, we conclude that the propose bound is quite tight.
- We also propose another CRB based on a Gaussian (interference) approximation, which we term the extrinsic information lower bound, which is actually a CRB upper bound due to the worst case assumption of Gaussian interference.

Notation: In the following, superscripts ^{T}, *, and ^{H} stand for transpose, conjugate, and conjugate transpose, respectively. The expectation operator is indicated by $\mathbb{E}\{.\}$ and $\mathbf{I}_P$ is the $P \times P$ identity matrix. Here, $\|\cdot\|$ and $\text{diag}(.)$ denote the Euclidean norm operator and the squared diagonal matrix consisting of the diagonal elements of matrix argument, respectively. $\text{vec}(.)$ denotes $\text{vec}(\mathbf{A}) = \left[\mathbf{A}_{:,1}^T \;\; \mathbf{A}_{:,2}^T \cdots \mathbf{A}_{:,n}^T\right]^T$, where $\mathbf{A}_{:,j}$ is the j-th column of matrix $\mathbf{A}$ and $\text{tr}\{.\}$ is the trace operator. The Kronecker operator is denoted by $\otimes$. Finally, $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{CN}(\mu, \sigma^2)$ denote a real and a complex Gaussian distribution with mean μ and variance σ^2, respectively.

II. System Model

We consider a CF massive MIMO system in uplink in which M APs serve K users in the same time-frequency resource. All APs and users equipped with a single antenna are randomly distributed over a $D \times D$ square area. Furthermore, all APs are connected to a CPU via a back-haul network. The channel is assumed to remain constant over L consecutive symbol intervals, i.e., a block. In the uplink transmission, each user k sends P pilot sequences known by the CPU followed by $L-P$ unknown data symbols. The pilot sequences are assumed to be ortho-normal, i.e., orthogonal with unit norm. The received signal $\mathbf{Y} \in \mathbb{C}^{M\times L}$ at the M APs over the block interval is given by

$$\mathbf{Y} = \sqrt{\rho}\,\mathbf{H}\,\mathbf{X} + \mathbf{V}, \tag{1}$$

where ρ denotes the transmit power at each user terminal normalized by the noise variance, $\mathbf{X}^T = [\mathbf{x}_1 \ldots \mathbf{x}_K] \in \mathbb{C}^{L\times K}$ is a matrix of the transmitted symbols and $\mathbf{x}_k \in \mathbb{C}^{L\times 1}$ is the signal vector sent by user k. The channel vector between user k and M APs is denoted by $\mathbf{h}_k = [h_{1k} \ldots h_{Mk}]^T \in \mathbb{C}^{M\times 1}$, then the channel matrix between the APs and users is given by $\mathbf{H} = [\mathbf{h}_1 \ldots \mathbf{h}_K] \in \mathbb{C}^{M\times K}$. The matrix $\mathbf{V} \in \mathbb{C}^{M\times L}$ represents the additive white Gaussian noise (AWGN) with i.i.d. components having zero mean and unit variance. The channel coefficient h_{mk} between AP m and user k is modeled as follows

$$h_{mk} = \sqrt{\beta_{mk}}\, g_{mk}, \tag{2}$$

where β_{mk} represents the large-scale fading coefficient which accounts for path loss and shadowing effects and g_{mk} represents the small-scale fading. We assume that $g_{mk},\ m = 1, \cdots M,\ k = 1, \cdots K$, are independent and identically distributed (i.i.d.) complex normal random variables, i.e., $g_{mk} \sim \mathcal{CN}(0, 1)$. Furthermore, we assume perfect knowledge of the large-scale fading coefficients β_{mk}, $m = 1, \cdots M$, $k = 1, \cdots K$ at the CPU.

Let the matrices $\mathbf{X}_p \in \mathbb{C}^{K\times P}$ and $\mathbf{X}_d \in \mathbb{C}^{K\times(L-P)}$ denote the pilot sequences and data symbols, respectively. Then, $\mathbf{X} = [\mathbf{X}_p \;\; \mathbf{X}_d]$ and $\mathbf{x}_k = [\mathbf{x}_{p,k}^T \;\; \mathbf{x}_{d,k}^T]^T$. Similarly, $\mathbf{Y} = [\mathbf{Y}_p \;\; \mathbf{Y}_d]$ where $\mathbf{Y}_p \in \mathbb{C}^{M\times P}$ and $\mathbf{Y}_d \in \mathbb{C}^{M\times(L-P)}$ represent the matrices of received training and data signals, respectively

III. CRB for Deterministic Semi-Blind Joint Channel Estimation and Data Detection

To analyze the performance of the semi-blind channel estimation, we first derive the CRB in a deterministic framework. In the deterministic framework, both data signal $\mathbf{X}_d$ and relevant channel coefficients $\mathbf{H}_I$ are modeled as unknown deterministic quantities. In the deterministic framework, we are forced to split the channel as $\mathbf{H} = \mathbf{H}_I + \mathbf{H}_0$ in which $\mathbf{H}_I$ contains significant channel coefficients that need to be estimated, and $\mathbf{H}_0$ contains channel coefficients that are considered sufficiently small and that will not be estimated [25] (in the Bayesian framework below, this split will be unnecessary). Thus, we have

$$\mathbf{y} \sim \mathcal{CN}\big(m_{\mathbf{y}}(\boldsymbol{\theta}), \mathbf{C}_{\mathbf{yy}}\big) \tag{3}$$

where $\mathbf{y} = \text{vec}(\mathbf{Y})$ and $\boldsymbol{\theta} = [\mathbf{h}_I^H \;\; \text{vec}^H(\mathbf{X}_d)]^H$ is the complex unknown parameter vector to be estimated. Here, $\mathbf{h}_I$ is a vector deduced from the non-zero elements of the matrix $\mathbf{H}_I$, whose support is known. Mean and covariance of received signal $\mathbf{y}$ are given by $m_{\mathbf{y}}(\boldsymbol{\theta}) = \sqrt{\rho}\,\text{vec}(\mathbf{H}_I\mathbf{X})$ and $\mathbf{C}_{\mathbf{yy}} = \mathbf{I}_L \otimes \mathbf{C}_{\mathbf{YY}}$, respectively, with $\mathbf{C_{YY}} = \mathbf{I}_M + \rho\,\mathbf{C}_0$ and covariance matrix $\mathbf{C}_0$ specified in the following:

$$\mathbf{C}_0 = \mathbb{E}\Big\{\mathbf{H}_0\mathbf{H}_0^H\Big\} = \mathrm{diag}\Big(\sum_{(1,k)\in\mathcal{K}_0}\beta_{1k},\cdots,\sum_{(M,k)\in\mathcal{K}_0}\beta_{Mk}\Big).$$

The probability density function[1] (pdf) of the observations $\mathbf{Y}$ in the parameter $\boldsymbol{\theta}$ is given by

$$p(\mathbf{Y}|\boldsymbol{\theta}) = \frac{1}{\pi^{ML}(\det(\mathbf{C_{YY}}))^L}\exp\Big(-\|\mathbf{C_{yy}}^{-1/2}(\mathbf{y}-m_{\mathbf{y}})\|^2\Big).$$

Computing the Jacobian of $m_{\mathbf{y}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, the deterministic complex Fisher information matrix (FIM) denoted as $\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}$ on the basis of the data $\mathbf{Y}$ is given by

$$\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}} = \Big(\frac{\partial m_{\mathbf{y}}^H}{\partial\boldsymbol{\theta}^*}\Big)\mathbf{C_{yy}}^{-1}\Big(\frac{\partial m_{\mathbf{y}}^H}{\partial\boldsymbol{\theta}^*}\Big)^H = \rho\,\big[\mathbf{Q}^{'}\ \mathbf{R}^{'}\big]^H\big[\mathbf{Q}^{'}\ \ \mathbf{R}^{'}\big] \quad (4)$$

where $\mathbf{Q}^{'} = \mathbf{C_{yy}}^{-1/2}\mathbf{Q}$, $\mathbf{Q} = \frac{1}{\sqrt{\rho}}\frac{\partial m_{\mathbf{y}}}{\partial\mathbf{h}_I^T}$ and $\mathbf{R}^{'} = \mathbf{C_{yy}}^{-1/2}\mathbf{R}$, $\mathbf{R} = \frac{1}{\sqrt{\rho}}\frac{\partial m_{\mathbf{y}}}{\partial\mathrm{vec}^T(\mathbf{X}_d)}$. Note that $\frac{1}{\sqrt{\rho}}m_{\mathbf{y}} = \mathrm{vec}(\mathbf{H}_I\,\mathbf{X}) = \mathbf{Q}\,\mathbf{h}_I = \mathrm{vec}(\mathbf{H}_I\,[\mathbf{X}_p\ \mathbf{0}]) + \mathbf{R}\,\mathrm{vec}(\mathbf{X}_d)$. The FIM $\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}$ is a 2×2 block matrix. The deterministic CRB^d is obtained as the inverse of the Fisher information matrix $\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}$

$$\mathrm{CRB}^d = \big(\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}\big)^{-1}. \quad (5)$$

The blocks $(1,1)$ and $(2,2)$ of the CRB^d in (5) relative to the estimation of the channel coefficients $\mathbf{h}_I$ and data symbols $\mathrm{vec}(\mathbf{X}_d)$, respectively are given as follows

$$\mathrm{CRB}^d_{\mathbf{h}_I} = \frac{1}{\rho}\big(\mathbf{Q}^{'H}\,P^{\perp}_{\mathbf{R}^{'}}\,\mathbf{Q}^{'}\big)^{-1} \quad (6)$$

$$\mathrm{CRB}^d_{\mathrm{vec}(\mathbf{X}_d)} = \frac{1}{\rho}\big(\mathbf{R}^{'H}\,P^{\perp}_{\mathbf{Q}^{'}}\,\mathbf{R}^{'}\big)^{-1} \quad (7)$$

where $P_{\mathbf{A}} = \mathbf{A}\big(\mathbf{A}^H\mathbf{A}\big)^{-1}\mathbf{A}^H$ and $P^{\perp}_{\mathbf{A}} = \mathbf{I} - P_{\mathbf{A}}$ denote the projection matrices on the column space of matrix $\mathbf{A}$ and its orthogonal complement, respectively. In the deterministic identifiability analysis that follows, we shall ignore $\mathbf{C}_0$ ($\mathbf{C}_0 = \mathbf{0}$) and hence $\mathbf{C_{YY}} = \mathbf{I}_M$, $\mathbf{C_{yy}} = \mathbf{I}_{ML}$.

IV. Pilot Based Bayesian Performance Bounds

From the moment that prior channel information is exploited, we consider estimating the whole channel $\mathbf{H}$ (limiting to $\mathbf{H}_I$ will not affect estimation performance much since $\mathbf{H}_0$ is small). We consider that pilots and data have the same power $\sigma_x^2 = 1$. Observe that $\mathrm{vec}(\mathbf{H}\,\mathbf{X}_p) = (\mathbf{X}_p^T\otimes\mathbf{I}_M)\,\mathbf{h}$. As a result the pilot portion leads to the following

$$\mathrm{FIM}_p = \rho\,(\mathbf{X}_p^*\mathbf{X}_p^T)\otimes\mathbf{I}_M \quad (8)$$

for $\mathbf{h}$, which is of course singular, due to pilot reuse. We can get a first idealized pilot only based CRB, by assuming that the pilots would somehow be orthogonal

$$\mathrm{CRB}_{p,o} = \mathrm{FIM}_{p,o}^{-1},\ \mathrm{FIM}_{p,o} = \rho\,\mathrm{diag}(\mathbf{X}_p^*\mathbf{X}_p^T)\otimes\mathbf{I}_M = \rho\,P\,I_{MK}\,. \quad (9)$$

[1]For the sake of compactness, we adopt an identical notation $p(\mathbf{Y}|\boldsymbol{\theta})$ to indicate the pdf of random variable (r.v.) $\mathbf{Y}$ in vector parameter $\boldsymbol{\theta}$ or conditioned to r.v. $\boldsymbol{\theta}$ when $\boldsymbol{\theta}$ is assumed to be a deterministic unknown vector parameter or a r.v., respectively.

The pilot contamination can be alleviated by prior channel information, leading to the Bayesian pilot based CRB

$$\mathrm{CRB}_{p,B} = (\mathrm{FIM}_p + \mathbf{C}_o^{-1})^{-1}\,. \quad (10)$$

For the semi-blind approaches considered here, another idealized MSE lower bound can be considered, considering to the genie-aided scenario in which the data $\mathbf{X}_d$ would be detected exactly, hence becoming also pilots for the channel estimation, leading to the genie-aided semi-blind Bayesian CRB

$$\mathrm{CRB}_{p+d,B} = (\rho\,(\mathbf{X}^*\mathbf{X}^T)\otimes\mathbf{I}_M + \mathbf{C}_o^{-1})^{-1} \quad (11)$$

where $\mathbf{X} = [\mathbf{X}_p\,\mathbf{X}_d]$. For any of the CRBs considered, we get a corresponding Normalized MSE (NMSE) bound in the form of $\mathrm{NMSE} = \mathrm{tr}\{\mathrm{CRB}\}/\mathrm{tr}\{\mathbf{C}_o\}$, where asymptotically $\|\mathbf{h}\|^2 = \mathrm{E}\,\|\mathbf{h}\|^2 = \mathrm{tr}\{\mathbf{C}_o\}$. Note that we get for all Bayesian approaches $\mathrm{NMSE} < 1$.

V. Semi-Blind with Gaussian i.i.d. Inputs

In this section we consider approaches in which the unknown $\mathbf{X}_d$ are (still) modeled as i.i.d. Gaussian and hence can be eliminated, leading to the Gaussian distribution $p(\mathbf{Y}|\mathbf{h})$. Some (unofficial) motivations:

- Overall joint estimation approaches, with EP or other, are not very simple.
- The joint estimation of $\mathbf{X}_d$ also may slow down the overall convergence.
- It looks like the introduction of an i.i.d. Gaussian prior on $\mathbf{X}_d$ does not improve much the identifiability (w.r.t. deterministic unknown $\mathbf{X}_d$), though still the exploitation of the blind part $\mathbf{Y}_d$ appears to be important. The identifiability probably relies mostly on the prior information on $\mathbf{h}$ (leading in particular to the sparsity structure in $\mathbf{H}_I$).
- Still the joint estimation of $\mathbf{h}_I$ and $\mathbf{X}_d$ is important and is expected to play a larger role when more prior information on $\mathbf{X}_d$ would be exploited.

So, eliminating the i.i.d. Gaussian $\mathbf{X}_d$, and approximating $\mathbf{Y}' = \mathbf{H}_0\,\mathbf{X} + \mathbf{V}$ as Gaussian noise (assuming pilots and data have the same per symbol power) we get

$$\begin{aligned}&\ln p(\mathbf{Y}|\mathbf{h}_I)\\&= -\mathrm{tr}\{(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{H}_I\mathbf{X}_p)^H\mathbf{C}_{\mathbf{Y}'\mathbf{Y}'}^{-1}(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{H}_I\mathbf{X}_p)\}\\&-(L-P)\ln\det(\mathbf{H}_I\mathbf{H}_I^H + \mathbf{C}_{\mathbf{Y}'\mathbf{Y}'})\\&-\mathrm{tr}\{(\mathbf{H}_I\mathbf{H}_I^H + \mathbf{C}_{\mathbf{Y}'\mathbf{Y}'})^{-1}\,\mathbf{Y}_d\mathbf{Y}_d^H\} + c^t\end{aligned} \quad (12)$$

where $\mathbf{y}' = \mathrm{vec}(\mathbf{Y}')$. So, the per channel use covariance matrix in the blind data part is $\mathbf{C_{YY}} = \mathbf{H}_I\mathbf{H}_I^H + \mathbf{C}_{\mathbf{Y}'\mathbf{Y}'}$. The non-quadratic appearance of $\mathbf{h}_I$ in the last two terms in (12) complicates the obtention of the posterior $p(\mathbf{h}_I|\mathbf{Y})$. The MAP estimator does not require the posterior, and can be obtained by maximizing $p(\mathbf{Y}|\mathbf{h}_I)\,p(\mathbf{h}_I)$.

VI. Gaussian Inputs Semi-Blind Bayesian CRB

From now on we drop the split in $\mathbf{H} = \mathbf{H}_I + \mathbf{H}_0$ and estimate all of $\mathbf{H}$. Eliminating the i.i.d. Gaussian $\mathbf{X}_d$, we get

$$\begin{aligned}&\ln p(\mathbf{Y}|\mathbf{h})\\&= -\mathrm{tr}\{(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{H}\mathbf{X}_p)^H\mathbf{C}_{\mathbf{V}_p\mathbf{V}_p}^{-1}(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{H}\mathbf{X}_p)\}\\&-(L-P)\ln\det(\rho\,\mathbf{H}\mathbf{H}^H + \mathbf{C}_{\mathbf{V}_d\mathbf{V}_d})\\&-\mathrm{tr}\{(\rho\,\mathbf{H}\mathbf{H}^H + \mathbf{C}_{\mathbf{V}_d\mathbf{V}_d})^{-1}\,\mathbf{Y}_d\mathbf{Y}_d^H\} + c^t\end{aligned} \quad (13)$$

The per channel use noise covariance matrices are $\mathbf{C}_{\mathbf{V}_p\mathbf{V}_p} = \mathbf{C}_{\mathbf{V}_d\mathbf{V}_d} = \mathbf{I}_M$. The blind FIM per channel use can be shown to be

$$\text{FIM}_b = \rho^2\, (\mathbf{H}^H\mathbf{C}^{-1}\mathbf{H})^* \otimes \mathbf{C}^{-1}\,,\; \mathbf{C} = \mathbf{I}_M + \rho\,\mathbf{H}\mathbf{H}^H \quad (14)$$

which results in the deterministic semi-blind CRB

$$\text{CRB}_{sb,d} = (\text{FIM}_p + (L{-}P)\,\text{FIM}_b)^{-1} \quad (15)$$

which depends on the true channel. $\text{CRB}_{sb,d}$ could be compared to its genie-aided version $\text{CRB}_{p+d,d} = \frac{1}{\rho}\,(\mathbf{X}^*\mathbf{X}^T)^{-1} \otimes \mathbf{I}_M$. The corresponding Bayesian semi-blind CRB

$$\begin{aligned}\text{CRB}_{sb,B} &= (\text{FIM}_p + (L{-}P)\,\text{E}_{\mathbf{h}}\{\text{FIM}_b\} + \mathbf{C}_o^{-1})^{-1}\\ &= (\text{CRB}_{p,B}^{-1} + (L{-}P)\,\text{E}_{\mathbf{h}}\{\text{FIM}_b\})^{-1}\end{aligned} \quad (16)$$

is difficult to compute analytically (except at low/high SNR) and might be loose. In any case, assuming that $\mathbf{H}$ is tall ($M > K$), at high SNR we get the dominating term

$$\text{FIM}_b^{hSNR} = \rho\,\mathbf{I}_K \otimes \mathbf{P}_{\mathbf{H}}^{\perp} \quad (17)$$

where $\mathbf{P}_{\mathbf{H}}^{\perp}$ denotes the projection on the orthogonal complement of the column space of $\mathbf{H}$. Then we get approximately

$$\text{E}_{\mathbf{h}}\,\text{FIM}_b^{hSNR} \approx \rho\,(1 - \frac{K}{M})\,\mathbf{I}_{KM} \quad (18)$$

which would be exact if the elements of $\mathbf{H}$ were i.i.d. On the other hand, at low SNR we get $\mathbf{C} \approx \mathbf{I}_M$ and hence $\text{FIM}_b^{lSNR} \approx \rho^2\,(\mathbf{H}^H\mathbf{H})^* \otimes \mathbf{I}_M$ from which

$$\text{E}_{\mathbf{h}}\,\text{FIM}_b^{lSNR} \approx \rho^2\,\text{tr}_M\{\mathbf{C}_o\} \otimes \mathbf{I}_M \quad (19)$$

where $\text{tr}_M\{\mathbf{A}\}$ is a diagonal matrix obtained by taking the trace of $\mathbf{A}$ over consecutive diagonal element portions of size M. In other words $(\text{tr}_M\{\mathbf{C}_o\})_{k,k} = \text{E}\|\mathbf{h}_k\|^2$. Remains to find an interpolation between low and high SNR. For that purpose, consider the SVD of $\mathbf{H} = \mathbf{U}[\mathbf{\Sigma}\;\mathbf{0}]^T\mathbf{V}^H$ (where $\mathbf{U}$, $\mathbf{V}$ and the diagonal $\mathbf{\Sigma}$ are square) and note that $\mathbf{C}^{-1} = \mathbf{I} - \mathbf{H}(\mathbf{H}^H\mathbf{H} + \frac{1}{\rho}\mathbf{I})^{-1}\mathbf{H}^H$, then we get

$$\begin{aligned}&\text{FIM}_b = \rho\,(\mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H} + \tfrac{1}{\rho}\mathbf{I})^{-1})^* \otimes \mathbf{C}^{-1} =\\ &\rho\,(\mathbf{V}\mathbf{\Sigma}^2(\mathbf{\Sigma}^2{+}\tfrac{1}{\rho}\mathbf{I})^{-1}\mathbf{V}^H)^* \otimes (\mathbf{U}\begin{bmatrix}(\rho\mathbf{\Sigma}^2{+}\mathbf{I})^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{M-K}\end{bmatrix}\mathbf{U}^H)\,.\end{aligned} \quad (20)$$

If $\mathbf{H}$ would have had i.i.d. elements then $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$ would be independent. This incites us to take the expectation of the two factors in the Kronecker product separately. With some futher approximation, we then get

$$\begin{aligned}\text{E}_{\mathbf{h}}\text{FIM}_b \approx \rho\,(1 - \tfrac{K}{M} + \tfrac{1}{M}\,\text{tr}\{(\rho\,\text{tr}_M\{\mathbf{C}_o\} + \mathbf{I}_K)^{-1}\})\\ \text{tr}_M\{\mathbf{C}_o\}(\text{tr}_M\{\mathbf{C}_o\} + \tfrac{1}{\rho}\mathbf{I}_K)^{-1} \otimes \mathbf{I}_M\end{aligned} \quad (21)$$

which is consistent with the high and low SNR limits in (18), (19), and which needs to be plugged into (16).

Should check here also the Modified CRB and other variations on the Bayesian CRB theme.

VII. Gaussian-Gaussian Extrinsic Information Lower Bound

Another performance bound can be obtained by considering $\mathbf{H}_{\overline{k}}\mathbf{X}_{d,\overline{k}}$ as Gaussian. Since for the estimation of the signal of user k, considering the interfering signals to be Gaussian corresponds to a worst case interference for given interference covariance, the resulting mutual information lower bound should lead to an information matrix lower bound and hence to an error covariance (MSE) upper bound. For the signal k in the resulting model, the Gaussian input $\mathbf{x}_{d,k}$ can then be eliminated, and the CRB for $\mathbf{h}_k$ can be computed. In the absence of a prior on $\mathbf{h}_k$, this would correspond to extrinsic information on $\mathbf{h}_k$. Once we exploit prior information on the channel (at least for the interferers), estimation of the whole channel vector $\mathbf{h}_k$ (and not just $\mathbf{h}_{I,k}$) becomes conceivable.

Let $\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{C}_{o,k})$ with $\mathbf{C}_{o,k} = \text{diag}(\beta_{1k}, \ldots, \beta_{Mk})$. Eliminating the Gaussian $\mathbf{H}_{\overline{k}}$, $\mathbf{x}_k$ and $\mathbf{H}_{\overline{k}}\mathbf{X}_{d,\overline{k}}$, we get

$$\begin{aligned}&\ln p(\mathbf{y}|\mathbf{h}_k) = -(\mathbf{y}_p - \sqrt{\rho}\,\mathbf{x}_{p,k} \otimes \mathbf{h}_k)^H\\ &(\mathbf{C}_{\mathbf{v},p} + \rho\textstyle\sum_{i\neq k}\mathbf{x}_{p,i}\mathbf{x}_{p,i}^H \otimes \mathbf{C}_{o,i})^{-1}(\mathbf{y}_p - \sqrt{\rho}\,\mathbf{x}_{p,k} \otimes \mathbf{h}_k)\\ &-\ln\det(\mathbf{C}_{\mathbf{v},d} + \rho\mathbf{I}_{L-P} \otimes (\mathbf{h}_k\mathbf{h}_k^H + \textstyle\sum_{i\neq k}\mathbf{C}_{o,i}))\\ &-\mathbf{y}_d^H(\mathbf{C}_{\mathbf{v},d} + \rho\mathbf{I}_{L-P} \otimes (\mathbf{h}_k\mathbf{h}_k^H + \textstyle\sum_{i\neq k}\mathbf{C}_{o,i}))^{-1}\mathbf{y}_d + c^t.\end{aligned} \quad (22)$$

We assume here normalized white noise $\mathbf{C}_{\mathbf{v}} = \mathbf{I}$. In that case the data portion of (22) simplifies to

$$\begin{aligned}&\ln p(\mathbf{y}_d|\mathbf{h}_k) = -(L-P)\,\ln\det(\mathbf{\Sigma})\\ &-\text{tr}\{\mathbf{Y}_d^H\mathbf{\Sigma}^{-1}\mathbf{Y}_d\} + c^t, \qquad \text{with}\\ &\mathbf{\Sigma} = \mathbf{I}_M + \rho\,(\mathbf{h}_k\mathbf{h}_k^H + \textstyle\sum_{i\neq k}\mathbf{C}_{o,i})\end{aligned} \quad (23)$$

where we used $\det(\mathbf{A} \otimes \mathbf{B}) = (\det(\mathbf{A})^{\dim(\mathbf{B})}(\det(\mathbf{B})^{\dim(\mathbf{A})}$, and $\text{vec}^T(\mathbf{A})\,(\mathbf{D} \otimes \mathbf{B})\,\text{vec}(\mathbf{C}) = \text{tr}\{\mathbf{A}^T\mathbf{B}\,\mathbf{C}\,\mathbf{D}^T\}$. Using the FIM for a circularly complex Gaussian pdf, we get

$$\begin{aligned}&\text{FIM}_{\mathbf{h}_k}^{GGei} = \rho^2\,(L-P)\,\mathbf{h}_k^H\mathbf{\Sigma}^{-1}\mathbf{h}_k\,\mathbf{\Sigma}^{-*} +\\ &\rho(\mathbf{x}_{p,k}^H \otimes \mathbf{I}_M)(\mathbf{I} + \rho\textstyle\sum_{i\neq k}\mathbf{x}_{p,i}\mathbf{x}_{p,i}^H \otimes \mathbf{C}_{o,i})^{-1}(\mathbf{x}_{p,k} \otimes \mathbf{I}_M)\end{aligned} \quad (24)$$

where $\mathbf{\Sigma}^{-*} = (\mathbf{\Sigma}^*)^{-1}$. Then we have an extrinsic information CRB upper bound: $\text{CRB}_{\mathbf{h}_k}^{GGei} = (\text{FIM}_{\mathbf{h}_k}^{GGei})^{-1}$.

VIII. Channel MAP

Here we consider the Gaussian i.i.d. $\mathbf{x}_d$ eliminated, as in Section V.

A. Joint Channel MAP for All Users

As in (13) but now estimating $\mathbf{h}$ and not just $\mathbf{h}_I$, we get

$$\begin{aligned}&-\ln p(\mathbf{Y}|\mathbf{h}) = -\ln p(\mathbf{Y}_p|\mathbf{h}) - \ln p(\mathbf{Y}_d|\mathbf{h})\\ &= \text{tr}\{(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{H}\,\mathbf{X}_p)^H(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{H}\,\mathbf{X}_p)\}\\ &+(L{-}P)\ln\det(\rho\,\mathbf{H}\,\mathbf{H}^H + \mathbf{I})\\ &+\text{tr}\{(\rho\,\mathbf{H}\,\mathbf{H}^H + \mathbf{I})^{-1}\,\mathbf{Y}_d\mathbf{Y}_d^H\} + c^t\end{aligned} \quad (25)$$

The pilot part is convex. For the blind part $-\ln p(\mathbf{Y}_d|\mathbf{h})$, we construct a convex majorizer as in [27, Section V.A 4)], [28] which can actually also been derived with an EM approach. The construction of the majorizer is simply based on first-order

Taylor series expansion of concave functions, either w.r.t. $\mathbf{H}$ directly or w.r.t. a covariance type expression (which is then quadratic in $\mathbf{H}$). So let $\mathbf{H}'$ be a current estimate of $\mathbf{H}$. Then

$$\begin{array}{l}\ln\det(\rho\,\mathbf{H}\,\mathbf{H}^H+\mathbf{I})=\ln\det(\rho\,\mathbf{H}^H\,\mathbf{H}+\mathbf{I})\\ \leq \text{tr}\{(\rho\,\mathbf{H}'^H\,\mathbf{H}'+\mathbf{I})^{-1}\rho\,\mathbf{H}^H\mathbf{H}+c^t\,.\end{array} \quad (26)$$

On the other hand, using the matrix inversion lemma and Taylor series expansion,

$$\begin{array}{l}\text{tr}\{(\rho\,\mathbf{H}\,\mathbf{H}^H+\mathbf{I})^{-1}\,\mathbf{Y}_d\mathbf{Y}_d^H\}\\ =\text{tr}\{[\mathbf{I}-\mathbf{H}(\mathbf{H}^H\,\mathbf{H}+\frac{1}{\rho}\,\mathbf{I})^{-1}\mathbf{H}^H]\,\mathbf{Y}_d\mathbf{Y}_d^H\}\\ \leq \text{tr}\{\mathbf{H}\,\mathcal{A}'\,\mathbf{H}^H-\mathcal{B}'\,\mathbf{H}^H-\mathbf{H}\,\mathcal{B}'^H\}+c^t\end{array} \quad (27)$$

where

$$\begin{array}{l}\mathcal{A}'=(\mathbf{H}'^H\,\mathbf{H}'+\frac{1}{\rho}\,\mathbf{I})^{-1}\mathbf{H}'^H\mathbf{Y}_d\mathbf{Y}_d^H\mathbf{H}'(\mathbf{H}'^H\,\mathbf{H}'+\frac{1}{\rho}\,\mathbf{I})^{-1}\\ \mathcal{B}'=\mathbf{Y}_d\mathbf{Y}_d^H\mathbf{H}'(\mathbf{H}'^H\,\mathbf{H}'+\frac{1}{\rho}\,\mathbf{I})^{-1}\,.\end{array} \quad (28)$$

Now (26), (27) lead to the following quadratic majorizer in (25)

$$\begin{array}{l}-\ln p(\mathbf{Y}_d|\mathbf{h})\leq \text{tr}\{(\rho(L-P)(\rho\,\mathbf{H}'^H\,\mathbf{H}'+\mathbf{I})^{-1}+\mathcal{A}')\mathbf{H}^H\mathbf{H}\\ \quad -\mathcal{B}'\,\mathbf{H}^H-\mathbf{H}\,\mathcal{B}'^H\}+c^t\,.\end{array} \quad (29)$$

At this point, note that the quantities in (26), (28) have the following interpretation

$$\begin{array}{l}\widehat{\mathbf{X}}_d=\frac{1}{\sqrt{\rho}}\,(\mathbf{H}'^H\,\mathbf{H}'+\frac{1}{\rho}\,\mathbf{I})^{-1}\mathbf{H}'^H\mathbf{Y}_d\\ \mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}=(L-P)\,(\rho\,\mathbf{H}'^H\,\mathbf{H}'+\mathbf{I})^{-1}\end{array} \quad (30)$$

which are the LMMSE estimate and associated error covariance matrix of $\mathbf{X}_d$ (which is i.i.d. across channel uses). This means that the majorizer in (29) has the following Expectation Maximization (EM) interpretation (which was not observed in [28], in spite of EM being discussed there also):

$$\begin{array}{l}-\ln p(\mathbf{Y}_d|\mathbf{H})=-\ln \text{E}_{\mathbf{X}_d}\,p(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{H})\\ =-\ln \text{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{H}'}\,(p(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{H})\,p(\mathbf{X}_d)/p(\mathbf{X}_d|\mathbf{Y}_d,\mathbf{H}'))\\ =-\ln \text{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{H}'}\,p(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{H})+c^t\\ \leq \text{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{H}'}\{-\ln p(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{H})\}+c^t\\ =\text{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{H}'}\|\mathbf{Y}_d-\sqrt{\rho}\,\mathbf{H}\,\mathbf{X}_d\|^2+c^t\\ =\|\mathbf{Y}_d-\sqrt{\rho}\,\mathbf{H}\,\widehat{\mathbf{X}}_d\|^2+\rho\,\mathbf{H}\,\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}\mathbf{H}^H+c^t\end{array} \quad (31)$$

where the inequality follows from Jensen's inequality and the convexity of $-\ln(.)$, and c^t denotes (various) terms that are constant w.r.t. $\mathbf{H}$. The MMSE estimate $\widehat{\mathbf{X}}_d$ and error covariance matrix $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}$ are defined in (30) and correspond to LMMSE estimation due to the joint Gaussianity of $p(\mathbf{Y}_d,\mathbf{X}_d|\mathbf{H}')$. At this point we have the following quadratic majorizer

$$\begin{array}{l}-\ln p(\mathbf{Y}|\mathbf{h})\leq f'(\mathbf{h}|\mathbf{h}')\\ =\rho\,\text{tr}\{\mathbf{H}^H\mathbf{H}(\mathbf{X}_p\mathbf{X}_p^H+\widehat{\mathbf{X}}_d\widehat{\mathbf{X}}_d^H+\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d})\}\\ -2\sqrt{\rho}\,\Re\,\text{tr}\{\mathbf{H}(\mathbf{X}_p\mathbf{Y}_p^H+\widehat{\mathbf{X}}_d\mathbf{Y}_d^H)\}+c^t\end{array} \quad (32)$$

which is separable between the channel use dimension and the Rx antenna dimension. When we add the channel prior, which contains different channel covariance matrices for different users, we need to switch from $\mathbf{H}$ to $\mathbf{h}$ and we get with $-\ln p(\mathbf{h})=\mathbf{h}^H\mathbf{C}_o^{-1}\mathbf{h}+c^t$ $(\mathbf{C}_o=\mathbf{C}_{\mathbf{hh}})$,

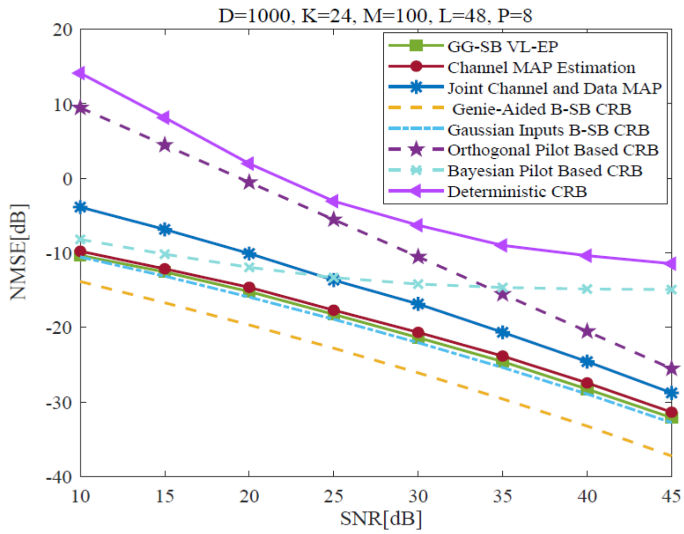

Fig. 1. NMSE [dB] versus SNR [dB]

$$\begin{array}{l}-\ln p(\mathbf{Y}|\mathbf{h})-\ln p(\mathbf{h})\leq f(\mathbf{h}|\mathbf{h}')\\ =\mathbf{h}^H(\mathbf{C}_o^{-1}+\rho\,((\mathbf{X}_p\mathbf{X}_p^H+\widehat{\mathbf{X}}_d\widehat{\mathbf{X}}_d^H+\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d})^T\otimes\mathbf{I}_M))\,\mathbf{h}\\ -2\sqrt{\rho}\,\Re\,\{\mathbf{h}^H\,\text{vec}(\mathbf{Y}_p\mathbf{X}_p^H+\mathbf{Y}_d\widehat{\mathbf{X}}_d^H)\}+c^t\end{array} \quad (33)$$

which leads to the following estimate

$$\begin{array}{l}\widehat{\mathbf{h}}=\sqrt{\rho}\,(\mathbf{C}_o^{-1}+\rho\,((\mathbf{X}_p\mathbf{X}_p^H+\widehat{\mathbf{X}}_d\widehat{\mathbf{X}}_d^H+\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d})^T\otimes\mathbf{I}_M))^{-1}\\ \quad\text{vec}(\mathbf{Y}_p\mathbf{X}_p^H+\mathbf{Y}_d\widehat{\mathbf{X}}_d^H)\,.\end{array} \quad (34)$$

This (34) needs to be solved iteratively, with $\widehat{\mathbf{X}}_d$ and $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}$ in (30) computed with the previous channel estimate. The iterative process can be initialized with $\widehat{\mathbf{h}}^{(-1)}=\mathbf{0}$, which leads to a first iterate $\widehat{\mathbf{h}}^{(0)}$ being based only on pilots and prior information. Note that the previous Bayesian Semi-Blind MAP can be obtained from the channel MAP considered here by putting $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}=\mathbf{0}$ in (34).

B. Joint Channel MAP for Users on Same Pilot

Other user signals $\mathbf{z}_i$ modeled as Gaussian noise. Given that the extrinsic information FIM does not appear to correspond to a lot of information, this approach, as well as the next one, is perhaps less interesting, although it would lead to reduced complexity and be potentially more amenable to distributed implementation. The extrinsic information FIM showing reduced information may mean that the modeling of $\mathbf{z}_{\overline{k}}$ as Gaussian interference leads to significant performance loss. For Separate Channel MAP per User: as in the extrinsic information FIM formulation.

IX. Performance Evaluation

In this section, we provide numerical results verifying the analytical derivations and the performance of the proposed approaches. The $M=100$ APs and $K=24$ users are uniformly distributed at random over a square area of size 1000×1000. The large-scale fading coefficient β_{mk} in (2) models the path loss and shadow fading as follows

$$\beta_{mk}=10^{\frac{\text{PL}_{mk}}{10}}\;10^{\frac{\sigma_{sh}z_{mk}}{10}} \quad (35)$$

where PL_{mk} represents the path loss (expressed in dB), and $10^{\frac{\sigma_{sh} z_{mk}}{10}}$ represents the shadow fading with standard deviation σ_{sh}, and $z_{mk} \sim \mathcal{N}(0,1)$, i.e., we assume uncorrelated shadow fading. The three-slope model in [29] is adopted for the path loss. The performance of the different channel estimators is assessed by the normalized mean square error (NMSE) versus SNR. The NMSE is defined as $\mathrm{NMSE} = \frac{avg\|\mathbf{h}-\widehat{\mathbf{h}}\|^2}{avg\|\mathbf{h}\|^2}$ where avg stands for average. Fig. 1 compares the performance of the channel estimation VL-EP algorithm for GG-SB proposed in [26] and channel MAP estimation discussed above and presents NMSE [dB] versus SNR [dB]. The VL-EP algorithm outperforms the channel MAP estimation and a joint channel and data MAP algorithm, termed Bayesian semi-blind approach in [25]. The joint channel and data MAP alternatingly estimates the channel or data as if the estimate for the other quantity is perfect, whereas the channel MAP estimation takes into account the data error covariance matrix $\mathbf{C}_{\widetilde{\mathbf{X}}_d \widetilde{\mathbf{X}}_d}$. Therefore the channel MAP estimation outperforms the Bayesian semi-blind iterative algorithm. The performance of these three different semi-blind channel estimation algorithms is compared to the different Cramer-Rao bounds (CRBs). For the semi-blind approaches one can consider the genie-aided scenario in which the data $\mathbf{X}_d$ would be detected exactly, hence becoming also pilots for the channel estimation, leading to the genie-aided Bayesian semi-blind (B-SB) CRB. For the VL-EP or channel MAP scenario, we consider Gaussian channels with the Gaussian input symbols eliminated, leading to the Gaussian inputs B-SB CRB. The deterministic CRB curve in the figure corresponds to a deterministic framework in which both data signal and channel coefficients are modeled as unknown deterministic quantities [25]. The performance of the different CRBs is evaluated by $\mathrm{NMSE} = \mathrm{tr}\{\mathrm{CRB}\}/\mathrm{tr}\{\mathbf{C}_{\mathbf{hh}}\}$, where $\mathbf{C}_{\mathbf{hh}} = \mathrm{diag}(\beta_{11}, \ldots, \beta_{M1} \ldots \beta_{1K}, \ldots, \beta_{MK})$. The simulations show that exploiting prior information gives significant performance gains. Compared to a fictitious scenario of just orthogonal pilot based channel estimation (pilots still of length P), deterministic semi-blind does not do as well whereas Bayesian semi-blind still does much better. On the other hand, the Bayesian pilot based CRB shows that just adding channel prior information to the contaminating pilots allows already to significantly improve MSE at low to moderate SNR, but floors at higher SNR. Adding the blind channel information from the data second-order statistics breaks this flooring, and both channel MAP and especially VL-EP allow to get performance close to the corresponding CRB, which behaves with just an SNR offset compared to the genie-aided CRB.

X. Conclusion

In this paper, we considered semi-blind methods for channel estimation in the presence of Gaussian i.i.d. data, exploiting prior channel information to mitigate the pilot contamination which originates from reusing pilot sequences, in CF massive MIMO systems. Whereas Bayesian Cramer-Rao Bounds are often very lose, we propose for the semi-blind channel estimation scenario an original variation on the Bayesian CRB which turns out to be quite tight, as has been demonstrated by numerical comparisons with powerful semi-blind channel estimation algorithms that provide an approximate MMSE channel estimator.

References

[1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE T=WC*, 2017.

[2] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE JSAC*, 2013.

[3] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system for multiple users," *IEEE JSTSP*, 2014.

[4] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, 2019.

[5] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE T-WC*, vol. 16, no. 7, 2017.

[6] Z. Chen and E. Björnson, "Can we rely on channel hardening in cell-free massive MIMO?" in *IEEE Globecom Workshops*, 2017.

[7] ——, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE T-COM*, no. 11, 2018.

[8] R. Gholami, L. Cottatellucci, and D. Slock, "Favorable Propagation and Linear Multiuser Detection for Distributed Antenna Systems," in *IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[9] ——, "Channel Models, Favorable Propagation and MultiStage Linear Detection in Cell-Free massive MIMO," in *IEEE Int'l Symp. Information Theory (ISIT)*, 2020.

[10] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Communications*, vol. 19, no. 1, 2019.

[11] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE T-WC*, Nov. 2010.

[12] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE JSTSP*, 2014.

[13] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE JSAC*, no. 2, 2013.

[14] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE JSTSP*, no. 5,201 4.

[15] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE T-SP*, no. 11, 2016.

[16] Y. Zhang, H. Cao, P. Zhong, C. Qi, and L. Yang, "Location-based greedy pilot assignment for cell-free massive MIMO systems," in *2018 IEEE ICCC*, 2018.

[17] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive MIMO wireless networks," in *2018 IEEE Int'l Conf. Communications (ICC) Workshops*, IEEE, 2018.

[18] H. Liu, J. Zhang, S. Jin, and B. Ai, "Graph coloring based pilot assignment for cell-free massive MIMO systems," *IEEE T-VT*, 2020.

[19] E. De Carvalho and D.T. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *First IEEE SPAWC Workshop*, 1997.

[20] E. de Carvalho and D. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Asilomar Conf. SSC*, 1997.

[21] A. K. Jagannatham and B. D. Rao, "Whitening-rotation-based semi-blind MIMO channel estimation," *IEEE T-SP*, 2006.

[22] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems," *IEEE SPL*, 2008.

[23] E. de Carvalho and D. T. Slock, "Blind and semi-blind FIR multichannel estimation:(global) identifiability conditions," *IEEE T-SP*, 2004.

[24] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE T-SP*, 2003.

[25] R. Gholami, L. Cottatellucci, and D. Slock, "Tackling Pilot Contamination in Cell-Free Massive MIMO by Joint Channel Estimation and Linear Multi-User Detection," in *IEEE Int'l Symp. IT (ISIT)*, 2021.

[26] ——, "Message Passing for a Bayesian Semi-Blind Approach to Cell-Free Massive MIMO," in *IEEE Asilomar Conf. SSC*, 2021.

[27] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning," *IEEE Trans. Signal Proc.*, Feb. 2017.

[28] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-Complexity Algorithms for Low Rank Clutter Parameters Estimation in Radar Systems," *IEEE Trans. Signal Proc.*, Apr. 2016.

[29] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *IEEE Veh. Tech. Conf.*, Spring, 2001.