Maximum Likelihood Optimization of Adaptive Asynchronous Interference Mitigation Beamformer

Alexandr M. Kuzminskiy, Yuri I. Abramovich, Pei Xiao, Rahim Tafazolli, Jinliang Huang

Abstract—In asynchronous (intermittent) interference scenarios, the content of co-channel interference sources over the data interval may be different from the interferers content over the training interval, typically with extra interference sources presented over the data interval. Under such conditions, conventional adaptive beamformer designed over the training interval may lose its efficiency when applied to the data interval. In this paper, we address the problem by 1) formulating a family of the second order statistics adaptive beamformers regularized by the covariance matrix estimated over the data interval; 2) proposing a maximum likelihood methodology for optimization of the combined (mixed) covariance matrix based on maximization of a product of a likelihood ratio that checks the accuracy of the recovered training signals and a likelihood ratio on equality of the eigenvalues in complementary to the signal subspace defined over the data interval; 3) demonstrating efficiency and robustness of the proposed solution as a linear adaptive beamformer and as an initialization for iterative beamformer with projections to the finite alphabet in different asynchronous interference scenarios comparing with the basic training and data based interference rejection combining receivers.

Index Terms—Asynchronous interference, interference rejection combining, regularization, mixing factor, maximum likelihood.

I. INTRODUCTION

To meet the challenge of ever increasing demands for higher capacity and better user experience in wireless communications industry, a number of promising technologies have been suggested in the literature, e.g. in [1], [2] and references therein. One such technology relies on utilizing more frequency resources via, e.g., carrier aggregation, incorporation of unlicensed spectrum, shared-license spectrum, and millimeter wave spectrum [3], [4]. Another approach is to deploy more network nodes, including both macro BSs and low power nodes such as picos, femtos, small cells, and relays, leading to dense or ultra-dense networks (UDN) [5]. Particularly, coexistence of spectrum sharing UDNs, possibly based on different radio access technologies (RAT) in unlicensed spectrum is considered in [6] as a useful supplement to the conventional 5G cellular licensed band scenarios.

The main limiting factor for spectrum sharing UDNs is the harsh and generally uncontrollable interference environment, which may compromise user experience. In such an environment, interference mitigation becomes a pivotal technology. Interference mitigation by means of antenna arrays (smart antennas) is of special interest because (massive) multiple input multiple output (MIMO) is recognized as one of the key enablers for future networks. Interference rejection combining (IRC) based on the minimum mean square error (MMSE) criterion is a simplest antenna array interference mitigation technology using a linear beamforming at the receiver based on estimated second order statistics [7].

Conventional space-time-frequency equalization and interference rejection techniques in wireless communications exploit known training (pilot) symbols to estimate the weight vector of an antenna array using the estimated propagation channel of the desired signal and space-time-frequency parameters of the interference. The underlying assumption for these kind of techniques is that the training data is reliable since the co-channel interference (CCI) completely overlaps with the training symbols of the desired signal. Normally, this is the case for synchronous CCI, which has the same time-frequency structure as the desired user. Asynchronous cells, packed intermittent transmission, inter-system interference of different time-frequency formats in unlicensed spectrum, and other techniques may lead to more complicated asynchronous or intermittent CCI scenarios, where the interference may partially overlap or not overlap with the training data of the desired signal. The main difficulty with this scenario is that potentially different content of CCI sources may be observed over the training and data intervals. Under these conditions, adaptive beamformers designed over the training interval, occupied by the known (often orthogonal) training sequences, becomes inefficient for interference mitigation over the data interval due to a possibility of the presence of interference signals absent over the training interval. Directly using the estimate of the covariance matrix averaged over the data interval does not result in an efficient useful signal extraction due to the the useful signal presence in the data used for covariance matrix estimation [8].

The asynchronous CCI scenario is addressed in time division multiple access (TDMA) networks [9] - [13], wireless local area networks (WLAN) [14], [15], long term evolution (LTE) networks on the downlink [16] and on the uplink [17]. Significant performance degradation for the conventional training based IRC in the asynchronous CCI scenario is reported in these papers.

The asynchronous CCI scenario becomes especially important for coexisting spectrum sharing technologies in unlicensed spectrum and applications such as Internet of Things (IoT). A typical example leading to the asynchronous CCI scenario is IoT based on the IEEE802.11ah standard [21], which assumes thousands of stations per a single access point with highly

A. M. Kuzminskiy, P. Xiao, R. Tafazolli are with the 5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, UK, (a.kuzminskiy,p.xiao,r.tafazolli)@surrey.ac.uk. Y. I. Abramovich is with WR Systems, USA, yabramovich@wrsystems.com. J. Huang is with Huawei Technologies Sweden AB, jinliang.huang@huawei.com

irregular transmissions in unlicensed band.

Currently, the MMSE-IRC is considered as a basic receiver for the 3GPP New Radio (NR) standard [18]. Although, it is known that a pilot-based estimation of the interference plus noise covariance matrix required for IRC significantly outperforms the data based estimation in the basic synchronous CCI scenario[1] [16], [19], both the pilot and data based versions are indicated in [18] as the ones that need investigation mainly because "IRC receiver performance may depend on network synchronization assumptions."

A possibility to improve performance in such scenarios using constant modulus (CM) algorithms is pointed out in [10], [13]. Similar improvements can be achieved using algorithms with projections to the finite alphabet (FA) [22]. An important feature of these criteria in the considered problem is that they belong not only to the training data of the desired signal, but to the whole receive data block, which always overlaps with the asynchronous CCI. This means that interference mitigation techniques based on such nonlinear criteria are naturally resistant against asynchronous CCI. The main difficulty with such higher order statistics based solutions is the necessity for a reliable initialization, which may be difficult to achieve with the conventional beamformers in the asynchronous CCI scenarios.

As alternative to the conventional pilot and data based IRC, in [14], [23] we have proposed a semi-blind IRC in the asynchronous WLAN scenario, where the data based estimate of the covariance interference plus noise matrix averaged over the whole transmitted medium access control (MAC) protocol data unit (MPDU) of the desired signal is used as a regularization to the conventional pilot-based IRC receiver. In [23], we have developed a second order statistics based non-asymptotic benchmark for such scenarios and demonstrated that performance of the regularized receiver may be close to the benchmark for properly selected regularization parameter. Although, the problem of adaptive selection of such regularization parameter was not addressed in [14], [23].

Selection of regularization parameters generally has no closed form solution because it involves unknown information regarding signal/interference scenarios. Different empirical techniques are proposed in the literature as summarized in [24]. One such technique is referred as decision added a posteriori least square in [24], which selects the regularization parameters by means of minimization of mean square error between the regularized soft and hard (projections to FA) estimates of the desired signals. A version of this technique based on the CM minimization is applied in [25], [26] for selection of the conventional diagonal loading in robust beamforming. Unfortunately, this solution may be computationally very expensive. Indeed, to find one single parameter the weight coefficients, output signals, projections to FA and decision metric need to be recalculated at least several times for each data slot. A likelihood second order statistics based uniform solution for adaptive selection of diagonal loading parameter

for the conventional regularized training based IRC is given in [27].

Considering that maximum likelihood (ML) optimization of adaptive asynchronous interference mitigation beamformers remains open, in this paper we address a generalized asynchronous CCI scenario, and 1) formulate a family of the second order statistics adaptive beamformers regularized by the covariance matrix estimated over the data interval; 2) propose a ML methodology for optimization of the combined (mixed) covariance matrix based on maximization of a product of a likelihood ratio (LR) that checks the accuracy of the recovered training signals and LR on equality of the eigenvalues in complementary to the signal subspace defined over the data interval; 3) demonstrate efficiency and robustness of the proposed solution as a linear adaptive beamformer and as an initialization for iterative beamformer with projections to the finite alphabet in different asynchronous interference scenarios comparing with the basic training and data based MMSE-IRC beamformers.

Typically, the IRC related works study the effects of time-frequency channel variations and limited sampling support for estimation of the propagation channels of the useful signals and interference plus noise covariance matrix. Our emphasis in this paper is on the CCI overlapping effects with the training and data intervals in the asynchronous CCI scenario. Therefore, we use a flat fading assumption for our basic scenario with relatively high sampling support for both training and data intervals for the given number of antennas and users. Then, we demonstrate robustness of the proposed solution in terms of 1) different channel models leading to time-frequency channel variations, and 2) limited sampling support using an additional conventional diagonal loading regularization.

The rest of the paper is organized as follows. In Section II, we present the system model, introduce our regularized interference mitigation solution, and formulate the problem. In Section III, we propose a maximum likelihood methodology for optimization of the regularized beamformer. In Section IV, we develop an algorithm for adaptive selection of the mixing factor. In Section V, we present the simulation results that illustrate sensitivity of the interference mitigation performance on selection of the mixing factor, compare efficiency of the developed algorithm with the optimal (non-implementable) selection of the regularization parameter, and evaluate its performance as a linear adaptive IRC and as an initialization for adaptive beamformer with the FA projections in different asynchronous CCI scenarios. Finally, Section VI concludes the paper.

The key symbols and notations used in the paper are summarized in Table 1.

---

[1]In [20], we have shown that the conventional training-based linear solution practically cannot be improved in the class of second-order semi-blind techniques under the synchronous interference scenario.

Table 1. Summary of key symbols / notations

Symbol	Definition
K	Number antenna array elements
M	Number of useful signals
M_{CCI}	Number of interference sources
$\mathbf{H}, \hat{\mathbf{H}}$	$K \times M$ actual and estimated channel matrices
N_t, N_d	Numbers of training and data samples
$\mathbf{X}_t$	$K \times N_t$ received signal at the training interval
$\mathbf{X}_d$	$K \times N_d$ received signal at the data interval
$\hat{\mathbf{R}}_t, \hat{\mathbf{R}}_d$	$K \times K$ training and data covariance matrices
δ	Mixing factor
$\hat{\mathbf{R}}_r(\delta)$	$K \times K$ regularized covariance matrix
$\hat{\mathbf{W}}_d(\delta)$	$K \times M$ beamforming weight matrix
$\mathrm{LR}_t(\delta), \mathrm{LR}_d(\delta)$	Training, and data based likelihood ratios
$\mathrm{LR}(\delta)$	Joint likelihood ratio

II. DATA MODEL AND PROBLEM FORMULATION

We consider the following narrowband data model of the signal received by an antenna array of K elements:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) + \sum_{m=1}^{M_{CCI}} \mathbf{g}_m u_m(n) + \mathbf{z}(n), \quad (1)$$

where $n = 1 \ldots N$; N is the total number of samples in data block; $\mathbf{x}(n) \in \mathcal{C}^{K \times 1}$ is the vector of observed outputs of an antenna array; $\mathbf{s}(n) \in \mathcal{C}^{M \times 1}$ is the vector of M desired signals from user equipment (UE), $u_m(n)$, $m = 1 \ldots M_{CCI}$ are the $M_{CCI} < K - M$ components of CCI:

$$\mathrm{E}\{u_m(n_1)u_m^*(n_2)\} = \begin{cases} p_m, & n_1 = n_2 \in \mathcal{N}_m \\ 0, & n_1 = n_2 \notin \mathcal{N}_m \\ 0, & n_1 \neq n_2 \end{cases}, \quad (2)$$

$\mathcal{N}_m$ is the appearance interval for the m-th interference component; $\mathbf{z}(n) \in \mathcal{C}^{K \times 1}$ is the vector of noise, $\mathrm{E}\{\mathbf{z}(n)\mathbf{z}^*(n)\} = \sigma^2 \mathbf{I}_K$, $\mathrm{E}\{\mathbf{z}(n_1)\mathbf{z}^*(n_2)\} = 0$, $n_1 \neq n_2$, where $\mathrm{E}\{\cdot\}$ and $(\cdot)^*$ denote expectation and transposition complex conjugate correspondingly; $\mathbf{H} \in \mathcal{C}^{K \times M}$ is the matrix of the propagation channels of the desired signals and $\mathbf{g}_m \in \mathcal{C}^{K \times 1}$, $m = 1 \ldots M_{CCI}$ are the vectors of propagation channels of CCI. All propagation channels are assumed to be stationary over the whole data slot and independent for different antenna elements and slots. The desired signals are assumed to be deterministic signals. The training data of N_t samples, where $K < N_t < N$, and their positions inside the data slot are known at the receiver: $\mathbf{S}_t = [\mathbf{s}_t(1), ..., \mathbf{s}_t(N_t)]$, $\mathbf{X}_t = [\mathbf{x}_t(1), ..., \mathbf{x}_t(N_t)]$, $(N_t M)^{-1} \mathrm{tr}(\mathbf{S}_t \mathbf{S}_t^*) = 1$, where $\mathrm{tr}(\mathbf{A})$ is the trace of matrix $\mathbf{A}$, leading to the signal to noise ratio SNR$=\sigma^{-2}$. We assume SNR$\gg 1$. The working data interval of $N_d = N - N_t$ samples is defined as the rest of the slot: $\mathbf{S}_d = [\mathbf{s}_d(1), ..., \mathbf{s}_d(N_t)]$, $\mathbf{X}_d = [\mathbf{x}_d(1), ..., \mathbf{x}_d(N_d)]$. All interference components and noise are assumed to be independent circular Gaussian processes. The CCI propagation channels, powers, and appearance intervals are not known at the receiver. Regarding the UEs propagation channels, we consider two cases with $\mathbf{H}$ 1) being known a priori, 2) being estimated using the same training data. The former represents the case with multiple data slots that can be used for accurate estimation of the stationary slowly varying propagation channels of the desired signals.

Both data and training intervals may contain a number of data blocks, e.g., sub-carriers in frequency domain. An example of the asynchronous CCI scenario with the training interval positions similar to the ones in the LTE resource block (RB) on the uplink is shown in Fig. 1. This example illustrates the main difficulty with the asynchronous interference scenario: some of the interference components may partially overlap or not overlap at all with the training interval of the desired signal.
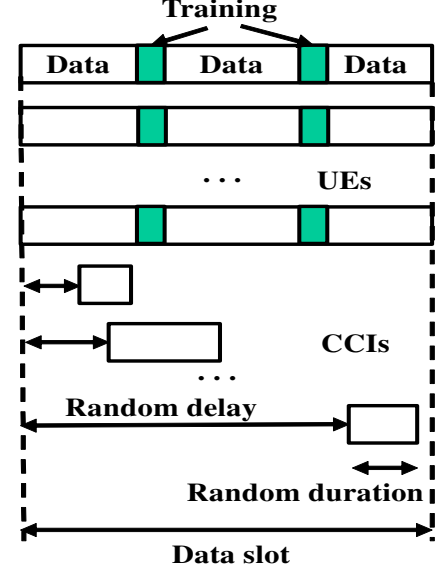

Fig. 1. Data model

Under the introduced model, the interference plus noise covariance matrices that are averaged over the training and working data intervals of the received signal can be expressed as

$$\mathbf{R}_t = \sum_{m=1}^{M_{CCI}} \bar{p}_m^t \mathbf{g}_m \mathbf{g}_m^* + \sigma^2 \mathbf{I}_K, \quad (3)$$

$$\mathbf{R}_d = \sum_{m=1}^{M_{CCI}} \bar{p}_m^d \mathbf{g}_m \mathbf{g}_m^* + \sigma^2 \mathbf{I}_K, \quad (4)$$

where $\bar{p}_m^t \geq 0$ and $\bar{p}_m^d > 0$ are the power coefficients averaged over the corresponding training and data intervals depending on the appearance interval $\mathcal{N}_m$ and actual power p_m of each interference source.

It is important to note that we consider the case, where the number of antenna elements exceeds the total number of useful signals and CCI components

$$K > M + M_{CCI}. \quad (5)$$

In turn, this means that the optimal linear beamformer $\mathbf{W}_d \in \mathcal{C}^{K \times M}$:

$$\mathbf{W}_d = \mathbf{R}_d^{-1} \mathbf{H} (\mathbf{H}^* \mathbf{R}_d^{-1} \mathbf{H})^{-1} \quad (6)$$

should provide efficient interference mitigation both in the data and training intervals, though over the training interval, better extraction of the (known a priori) training signals could be performed by another beamformer $\mathbf{W}_t \in \mathcal{C}^{K \times M}$:

$$\mathbf{W}_t = \mathbf{R}_t^{-1} \mathbf{H} (\mathbf{H}^* \mathbf{R}_t^{-1} \mathbf{H})^{-1}. \quad (7)$$

These properties of the optimal beamformers are fully exploited in the sequel [2].

Clearly, covariance matrices $\mathbf{R}_t$ and $\mathbf{R}_d$ are not known a priori and they have to be adaptively estimated. For the training data with the known $\mathbf{H}$, we have

$$\hat{\mathbf{R}}_t = N_t^{-1}(\mathbf{X}_t - \mathbf{H}\mathbf{S}_t)(\mathbf{X}_t - \mathbf{H}\mathbf{S}_t)^* \tag{8}$$

which is an ML estimate of $\mathbf{R}_t$. Since the training data volume N_t could be small (compared to the data set N_d) for the given number of antennas K, and the training interval may not contain any interferers, one should apply diagonally loaded estimate of $\mathbf{R}_t$ [28]: $\hat{\mathbf{R}}_t^{(DL)} = \hat{\mathbf{R}}_t + \alpha\mathbf{I}_K$, where $\alpha = (0.1 \div 1)\sigma^2$ is the regularization parameter[3].

In the case when $\mathbf{H}$ is unknown, we use its least square estimate

$$\hat{\mathbf{H}} = \mathbf{X}_t\mathbf{S}_t^*(\mathbf{S}_t\mathbf{S}_t^*)^{-1}, \tag{9}$$

leading to the following matrix estimate

$$\hat{\mathbf{R}}_t = (N_t - M)^{-1}\mathbf{X}_t[\mathbf{I}_{N_t} - \mathbf{S}_t^*(\mathbf{S}_t\mathbf{S}_t^*)^{-1}\mathbf{S}_t]\mathbf{X}_t^*. \tag{10}$$

The ML estimate of the data covariance matrix $\mathbf{R}_d$ is

$$\hat{\mathbf{R}}_d = N_d^{-1}\mathbf{X}_d\mathbf{X}_d^*. \tag{11}$$

Note, that

$$\mathrm{E}\{\hat{\mathbf{R}}_d\} = \mathbf{R}_d + N_d^{-1}\mathbf{H}\mathbf{S}_d\mathbf{S}_d^*\mathbf{H}^* \tag{12}$$

due to presence of inseparable (unknown) deterministic desired signals and CCI on the data interval.

Direct application of estimate (8) to the beamformer (7) gives the basic training based MMSE-IRC receiver, which fully exploits information at the training interval for useful signals recovery and CCI mitigation presented at the training interval, but it cannot reject interference appearing at the data interval. On the other hand, direct application of estimate (11) to (6) gives the basic data based MMSE-IRC receiver that allows interference mitigation at the data interval, but leads to significant performance degradation due to presence of the desired signals in $\mathbf{R}_d$ estimate in (11) [8].

Thus, both the conventional MMSE-IRC solutions are inefficient in the considered asynchronous interference scenario. In order to exploit the merits of both basic receivers and at the same time, avoid drawbacks in the asynchronous CCI scenario, similarly to our development in [14], [23], instead of directly using $\hat{\mathbf{R}}_t$ or $\hat{\mathbf{R}}_d$ defined in (8) and (11), we consider a regularized estimate of the corresponding matrix defined as

$$\hat{\mathbf{R}}_r(\delta) = (1 - \delta)\hat{\mathbf{R}}_t + \delta\hat{\mathbf{R}}_d, \tag{13}$$

where $0 \leq \delta \leq 1$ is the mixing (regularization) parameter. Then, the corresponding adaptive beamformer for data signals recovery can be defined as

$$\hat{\mathbf{W}}_d(\delta) = \hat{\mathbf{R}}_r^{-1}(\delta)\mathbf{H}\left[\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}(\delta)\mathbf{H}\right]^{-1} \tag{14}$$

---

[2]We assume interference limited scenarios with high SNR. Thus, in (6), (7) we use the following version of the basic MMSE-IRC formulation $\mathbf{W} = (\mathbf{R} + \mathbf{H}\mathbf{H}^*)^{-1}\mathbf{H} = \mathbf{R}^{-1}\mathbf{H}(\mathbf{I}_M + \mathbf{H}^*\mathbf{R}^{-1}\mathbf{H})^{-1} \approx \mathbf{R}^{-1}\mathbf{H}(\mathbf{H}^*\mathbf{R}^{-1}\mathbf{H})^{-1}$ for $\mathbf{H}^*\mathbf{R}^{-1}\mathbf{H} \gg 1$.

[3]Selection of α is a well studied problem, e.g. [27], [28], [29]. We illustrate using a conventional diagonal loading additionally to our data based regularization in Section V.A.

for the known channel matrix and

$$\hat{\mathbf{W}}_d(\delta) = \hat{\mathbf{R}}_r^{-1}(\delta)\hat{\mathbf{H}}\left[\hat{\mathbf{H}}^*\hat{\mathbf{R}}_r(\delta)^{-1}\hat{\mathbf{H}}\right]^{-1} \tag{15}$$

for the estimated channels correspondingly.

The regularized estimate defined in (13) is heuristic, but could be formally introduced as a Bayesian covariance matrix estimate with the Wishart distribution of one of the matrix as a prior. Example of such introduction is given, e.g. in [28].

The problem now is to adaptively estimate the mixing factor $\hat{\delta}_{ML}$ in (14), (15) in order to achieve the best performance of the linear beamformer

$$\hat{\mathbf{S}}_d(\hat{\delta}_{ML}) = \hat{\mathbf{W}}_d^*(\hat{\delta}_{ML})\mathbf{X}_d \tag{16}$$

for the given beamformer structure.

The complexity of the semi-blind solutions (14), (15) is similar to that of the conventional data-based MMSE-IRC.

III. Maximum likelihood methodology for optimization of the regularized beamformer

In this study, we limit our consideration by using only second order statistics of the input data, both on the training and data intervals. This means that in addition to the training signals and known a priori or estimated with the help of the training signals channel propagation matrix $\mathbf{H}$, we can estimate the signal-free interference plus noise covariance matrix $\hat{\mathbf{R}}_t$ on the training interval and the covariance matrix $\hat{\mathbf{R}}_d$ of the mixture of interference and signal on the data interval. Thus, if we confine our adaptive search of the optimal mixing factor δ in (13) only by the second order statistics, no much more measurable parameters are left in our disposal. The main reason is that in terms of their correlation properties, the useful signals on the data interval are indistinguishable from CCI, which is often created by very similar signals. Therefore, a direct analysis of the second order properties of the filtered data covariance matrix

$$\hat{\mathbf{R}}(\delta) = \hat{\mathbf{S}}_d(\delta)\hat{\mathbf{S}}_d^*(\delta), \tag{17}$$

$$\hat{\mathbf{S}}_d(\delta) = \left[\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}(\delta)\mathbf{H}\right]^{-1}\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}(\delta)\mathbf{X}_d \tag{18}$$

within the signal subspace spanned by the manifold matrix $\mathbf{H}$ can say a little about the appropriateness of the selected δ. This is the main reason why higher order statistics properties such as CM and FA have been applied for adaptive regularization, e.g. selection of diagonal loading parameter in [25], [26].

Our central idea is that under the specified conditions, the beamformer $\hat{\mathbf{W}}_d(\hat{\delta}_{ML})$ should efficiently reject interference in both data and training intervals and extract the (known) training signals over the training interval and the data signals over the data interval. For the training interval with the known training signals the corresponding likelihood ratio may be directly evaluated comparing the extracted and the true training signals. At the data interval, the transmitted signals are unknown. Thus, we need to estimate interference "whitening" properties on the data interval to be able to derive the corresponding likelihood ratio. Then, we may expect that $\hat{\delta}_{ML}$ that delivers the maximum of the joint (product) likelihood ratio of the extracted with $\hat{\mathbf{R}}_r(\delta)$ training signals and the likelihood

ratio of the covariance of the "whitened" with $\hat{\mathbf{R}}_r(\delta)$ data signals, is the one we are looking for.

In essence, we use the traditional ML criterion not only for estimation of $\hat{\mathbf{R}}_t$ and $\hat{\mathbf{R}}_d$, but also for finding the right mixing factor $\hat{\delta}_{ML}$ in $\hat{\mathbf{R}}_r(\delta)$.

To estimate "whitening" properties on the data interval, we need to separate estimation of interference properties from the impact of the useful signal components. One way to do that is to consider that the useful signals do not reside in the subspace complementary to the signal subspace spanned by $\mathbf{H}$. Indeed, only remains of the interference and white noise should reside in the subspace spanned by the projector

$$\mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta)\mathbf{H}\right] = \qquad (19)$$

$$= \mathbf{I}_K - \hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta)\mathbf{H}[\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}(\delta)\mathbf{H}]^{-1}\mathbf{H}^*\hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta).$$

Therefore, the properties of the projected by (19) "whitened" by $\mathbf{R}_r^{-\frac{1}{2}}$ input signal on the data interval

$$\hat{\mathbf{Z}}_d = \mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta)\mathbf{H}\right]\hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta)\mathbf{X}_d, \qquad (20)$$

are defined only by the remains of the interference and noise. Then, matrix $\hat{\mathbf{Z}}_d\hat{\mathbf{Z}}_d^*$ can be presented as

$$\hat{\mathbf{Z}}_d\hat{\mathbf{Z}}_d^* = \mathbf{U}_{K-M}\hat{\mathbf{Q}}_{K-M}\mathbf{U}_{K-M}^*, \qquad (21)$$

$$\hat{\mathbf{Q}}_{K-M}(\delta) = \mathbf{U}_{K-M}^*\hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta)\mathbf{X}_d\mathbf{X}_d^*\hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta)\mathbf{U}_{K-M}, \qquad (22)$$

with the following decomposition of projector (19)

$$\mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}(\delta)\mathbf{H}\right] = \mathbf{U}_{K-M}\mathbf{U}_{K-M}^*, \qquad (23)$$

where $\mathbf{U}_{K-M}^*\mathbf{U}_{K-M} = \mathbf{I}_{K-M}$.

Now, analyzing distribution of the eigenvalues of matrix $\hat{\mathbf{Q}}_{K-M}(\delta)$ we may conclude on the interference "whitening" properties in the complementary to the signal subspace on the data interval. The remaining in the complementary to signal subspace interference power should be a function of overlap between the signal and interference subspaces. Considering that the interference subspace is spanned by the similar to the channel vectors, one may expect that for the growing signal subspace dimension M, less interference power should remain in the complementary to signal subspace and properties of matrix $\hat{\mathbf{Q}}_{K-M}(\delta)$ could be directly used for indication of the interference mitigation efficiency. Indeed, for large enough M (compared to the antenna dimension K) the "whitener" with the optimum δ in $\hat{\mathbf{R}}_r(\delta)$ may significantly remove CCI from the complementary subspace and matrix $\hat{\mathbf{Q}}_{K-M}(\delta)$ in (22) may indeed be close to the sample matrix of white noise. Yet, for small $M \ll K$, significant remaining interference power in $\hat{\mathbf{Q}}_{K-M}(\delta)$ may be observed, even if efficient interference mitigation in the (unavailable) signal subspace may be achieved.

Let us illustrate this property in a flat fading scenario shown in Fig. 1 with stationary propagation channels assuming that data slot consists of N_{RB} resource blocks (RBs) of 14 symbols and 12 sub-carriers each with 2 symbols containing pilots in all sub-carriers. CCI sources have a similar structure with random delay within data slot, random duration of 2 or 3 symbols and signal to interference ratio SIR=0 dB. We use 16-QAM and QPSK signaling for the desired signals

and CCI correspondingly. Pilots are generated by means of MATLAB routine "ltePUSCHDRS". Propagation channels of the desired signals and CCI components are generated as i.i.d. unit power Gaussian vectors. In Fig. 2, we show the cumulative distribution functions (CDF) of the ratio between the total CCI power and the CCI power in the complementary to the signal subspace, calculated as the sum of M_{CCI} largest eigenvalues of matrix $\hat{\mathbf{Q}}_{K-M}(\delta_0)$, estimated over 1000 scenario realizations for $K = 16$, $N_{RB} = 2$, SNR=15 dB, $M = 2, 4, 8$, and $M_{CCI} = 1, 2, 3$. The optimal mixing factor δ_0 is estimated by means of the direct bit error rate (BER) minimization for each scenario realization.
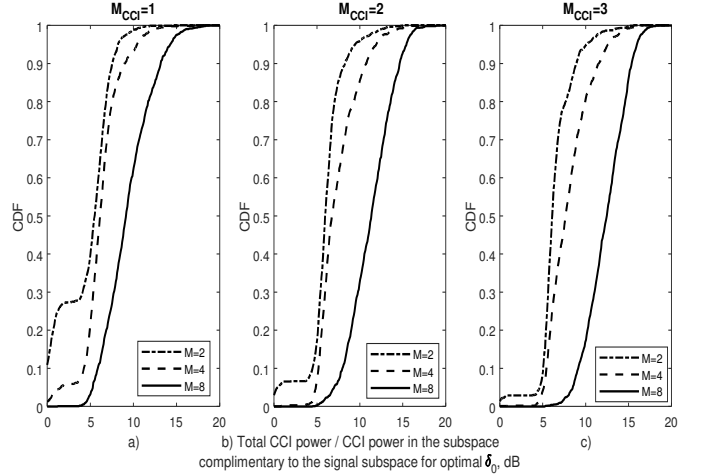

Fig. 2. CDF of the ratio between the total CCI power and the CCI power in the complementary to the signal subspace: $K = 16$, $N_{RB} = 2$, SNR=15 dB, SIR=0 dB, $M = 2, 4, 8$, and $M_{CCI} = 1$ for (a), $M_{CCI} = 2$ for (b), $M_{CCI} = 3$ for (c)

One can see in Fig. 2b for M_{CCI} that, as expected, for $M = 2 \ll K = 16$, the remaining in $\hat{\mathbf{Q}}_{K-M}(\delta_0)$ interference power with probability 0.5 is only 5 dB below the total CCI power, but for $M = 8 = K/2$ it is already 12 dB weaker. For SNR=15 dB and SIR=0 dB one may expect that in most cases the interference eigenvalues in $\hat{\mathbf{Q}}_{K-M}(\delta_0)$ are relatively close to the noise ones for M comparable with K. This dependence on M remains for different numbers of CCI sources as one can see in Figs. 2a and 2c for $M_{CCI} = 1$ and $M_{CCI} = 3$ correspondingly. Indeed, while for a single interferer in Fig. 2a, the remaining in the complementary subspace CCI power is somewhat higher for $M = 8$, for $M_{CCI} = 2$ and $M_{CCI} = 3$ the CDFs in Figs. 2b and 2c are barely distinguishable.

Therefore, this quite stable behavior of the CCI power within the complementary to the signal subspace, may be used for adaptive mixing factor selection. Ideally, if the remaining interference subspace eigenvalues were somehow predictable, our criterion for appropriate interference mitigation on the data interval could be proximity of the sample eigenvalues in $\hat{\mathbf{Q}}_{K-M}(\delta)$ to the predicted ones for δ_0. Yet, even if such a prediction was possible, a number of additional parameters, such as the number of "new" CCI sources appeared on the data interval, their SIR, etc., would be required. In what follows, we suggest a simple robust approach to provide sufficiently close to the optimum values of mixing factor. Indeed, for M

comparable with K, we observe reduction of the interference power in the complementary to the signal subspace to the level close to the noise power. Therefore, for such relatively large M we may apply as a criterion the likelihood ratio $\mathrm{LR}_0(\delta)$ that tests hypothesis on equal eigenvalues of the true matrix $\mathbf{Q}_{K-M}(\delta)$, given eigenvalues of the sample matrix $\hat{\mathbf{Q}}_{K-M}(\delta)$. Considering that for $M \ll K$ lower interference mitigation in the complementary subspace may be required for the optimal mixing factor compared to the case of comparable values of M and K, we suggest to compensate the likelihood ration degradation by the simple correction factor

$$\mathrm{LR}_d(\delta) = [\mathrm{LR}_0(\delta)]^{\frac{M}{K}}. \tag{24}$$

In Section V.A, we demonstrate robustness of this ad-hoc correction for relatively low dimensions of the signal subspace.

IV. ADAPTIVE SELECTION OF THE MIXING FACTOR

In this section, we specify the likelihood ratio $\mathrm{LR}_t(\delta)$ for testing proximity of the actual training sequence $\mathbf{S}_t$ to the reconstructed one using $\hat{\mathbf{W}}_d(\delta)$ in (14), (15), and the likelihood ratio test $\mathrm{LR}_d(\delta)$ on equality of all eigenvalues of the true matrix $\mathbf{Q}_{K-M}(\delta)$, given the sample ones. Each of these ratios could be constructed using a methodology introduced in [30]. Then, we formulate a joint likelihood ratio for the given volumes of samples on the training N_t and data N_d intervals.

Let us transform the input training data $\mathbf{X}_t \in \mathcal{C}^{K \times N_t}$ into the sufficient statistics data $\hat{\mathbf{Z}}_t \in \mathcal{C}^{M \times N_t}$, which is the output signal of the estimated beamformer in (7). In the case of the known $\mathbf{H}$, we define

$$\hat{\mathbf{Z}}_t = (\mathbf{H}^*\hat{\mathbf{R}}_t^{-1}\mathbf{H})^{-1}\mathbf{H}^*\hat{\mathbf{R}}_t^{-1}\mathbf{X}_t = \tilde{\mathbf{Z}}_t + \mathbf{S}_t, \tag{25}$$

where

$$\tilde{\mathbf{Z}}_t = (\mathbf{H}^*\hat{\mathbf{R}}_t^{-1}\mathbf{H})^{-1}\mathbf{H}^*\hat{\mathbf{R}}_t^{-1}(\mathbf{X}_t - \mathbf{H}\mathbf{S}_t) \tag{26}$$

leading to the ML estimate of the covariance matrix

$$\hat{\mathbf{R}}_z = N_t^{-1}\tilde{\mathbf{Z}}_t\tilde{\mathbf{Z}}_t^* = (\mathbf{H}^*\hat{\mathbf{R}}_t^{-1}\mathbf{H})^{-1}. \tag{27}$$

Therefore, the likelihood ratio for the estimated with beamformer (14) training signal

$$\hat{\mathbf{S}}_t(\delta) = (\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}(\delta)\mathbf{H})^{-1}\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}(\delta)\mathbf{X}_t \tag{28}$$

may be introduced as

$$\mathrm{LR}_t(\delta) = \frac{\mathrm{L}[\hat{\mathbf{S}}_t(\delta), \hat{\mathbf{R}}_z]}{\max_{\hat{\mathbf{S}}} \mathrm{L}[\hat{\mathbf{S}}, \hat{\mathbf{R}}_z]}, \tag{29}$$

where

$$\mathrm{L}[\hat{\mathbf{S}}, \hat{\mathbf{R}}_z] = \frac{1}{\pi^{N_t}\det(\hat{\mathbf{R}}_z)}\exp\left\{-(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}})\hat{\mathbf{R}}_z^{-1}(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}})^*\right\} \tag{30}$$

is the likelihood function, where $\det(\mathbf{A})$ is the determinant of matrix $\mathbf{A}$. Taking into account that

$$\max_{\hat{\mathbf{S}}} \mathrm{L}[\hat{\mathbf{S}}, \hat{\mathbf{R}}_z] = \mathrm{L}[\hat{\mathbf{S}}_t(0), \hat{\mathbf{R}}_z] = \tag{31}$$

$$= \mathrm{L}[\hat{\mathbf{Z}}_t, \hat{\mathbf{R}}_z] = \frac{1}{\pi^{N_t}\det(\hat{\mathbf{R}}_z)},$$

we finally get for the known $\mathbf{H}$

$$\mathrm{LR}_t(\delta) = \tag{32}$$

$$= \exp\left\{-\mathrm{tr}\left[\frac{1}{N_t}\mathbf{H}^*\hat{\mathbf{R}}_t^{-1}\mathbf{H}(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}}_t(\delta))(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}}_t(\delta))^*\right]\right\}^{N_t}.$$

For the channel matrix estimated according to (9), there are only $N_t - M$ independent samples for estimation of $\hat{\mathbf{R}}_t$ in (10). Therefore, $\mathrm{LR}_t(\delta)$ is defined in a slightly different way in this case

$$\mathrm{LR}_t(\delta) = \tag{33}$$

$$= \exp\left\{-\mathrm{tr}\left[\frac{1}{N_t - M}\hat{\mathbf{H}}^*\hat{\mathbf{R}}_t^{-1}\hat{\mathbf{H}}(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}}_t(\delta))(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}}_t(\delta))^*\right]\right\}^{N_t - M},$$

where $\hat{\mathbf{H}}$ is used in (25) and (28) instead of $\mathbf{H}$.

On the data interval, the test on equality of eigenvalues of $\hat{\mathbf{Q}}_{K-M}(\delta)$ defined in (22) is the well-known sphericity test [30]:

$$\mathrm{LR}_d(\delta) = \left[\frac{\prod_{j=1}^{K-M}\lambda_j[\hat{\mathbf{Q}}_{K-M}(\delta)]}{\left(\frac{1}{K-M}\sum_{j=1}^{K-M}\lambda_j[\hat{\mathbf{Q}}_{K-M}(\delta)]\right)^{K-M}}\right]^{\beta N_d}, \tag{34}$$

where $\lambda_j(\hat{\mathbf{Q}}_{K-M})$, $j = 1, ..., (K-M)$ are the eigenvalues of matrix $\hat{\mathbf{Q}}_{K-M}$, and β is the correction factor introduced in (24), i.e. we use $\beta = 1$ for relatively large M and $\beta = M/K$ for $M \ll K$.

Note, that $\mathrm{LR}_d(1) = 1$, while $\mathrm{LR}_t(0) = 1$. We are looking for the mixing factor that maximizes the product of the likelihood ratios in (32) and (34) or some monotonic function of this product, e.g. the N_dth root:

$$\hat{\delta}_{ML} = \arg\max_\delta[\tilde{\mathrm{LR}}(\delta)], \tag{35}$$

$$\tilde{\mathrm{LR}}(\delta) = [\mathrm{LR}_d(\delta)\mathrm{LR}_t(\delta)]^{\frac{1}{N_d}} = \tag{36}$$

$$= \left[\frac{\prod_{j=1}^{K-M}\lambda_j[\hat{\mathbf{Q}}_{K-M}(\delta)]}{\left(\frac{1}{K-M}\sum_{j=1}^{K-M}\lambda_j[\hat{\mathbf{Q}}_{K-M}(\delta)]\right)^{K-M}}\right]^\beta \times$$

$$\times\exp\left\{-\mathrm{tr}\left[\frac{1}{N_t}\mathbf{H}^*\hat{\mathbf{R}}_t^{-1}\mathbf{H}(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}}_t(\delta))(\hat{\mathbf{Z}}_t - \hat{\mathbf{S}}_t(\delta))^*\right]\right\}^{\frac{N_t}{N_d}}.$$

For the channel matrix $\hat{\mathbf{H}}$ estimated over the training interval , we have a similar to (36) expression using $\hat{\mathbf{H}}$ instead of $\mathbf{H}$ and $N_t - M$ instead of N_t as in (33).

The complexity of algorithm (35) is mainly due to the $K \times K$ matrix inversion, eigendecomposition, and square root calculation that require a number of complex multiplications proportional to K^3 for each value of δ to be tested for finding maximum in (35). Assuming that for typical scenarios, a range for reasonable values of δ may be known, this number could be kept at a minimum, leading to a reasonably low computational complexity of the algorithm.

V. Numerical results

A. Basic scenario

We consider the same flat fading scenario as in Section III with random SIR uniformly distributed in the range of [-10,5] dB independent for CCI components. As a performance metric, we use BER averaged over 1000 independent scenario realizations. We begin with the case of relatively high dimension of the signal subspace of $M = K/2$ and use $\beta = 1$ in (36). To avoid effects of mixing different types of regularization, initially in our simulations we consider a relatively high number of training samples of $N_t = 3K$ and do not use any diagonal loading, i.e. $\alpha = 0$. Then, we illustrate importance of using diagonal loading additionally to the data based regularization for lower values of N_t especially for the estimated channel.

First, we demonstrate the impact of the proper selection of the mixing parameter δ on the performance in the considered scenario. The BER results for two configuration settings are shown in Fig. 3
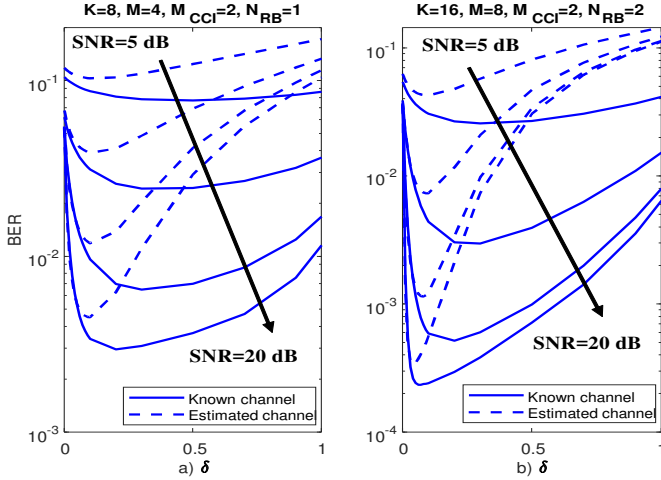
Fig. 3. BER results depending on δ: a) $K = 8$, $M = 4$, $M_{CCI} = 2$, $N_{RB} = 1$, $N_t = 24$, $N_d = 144$, and b) $K = 16$, $M = 8$, $M_{CCI} = 2$, $N_{RB} = 2$, $N_t = 48$, $N_d = 288$

One can observe from Fig. 3 that:

- Proper selection of the mixing factor is critically important and it gives much better results compared to both conventional boundary cases of the training based for $\delta = 0$ and data based for $\delta = 1$ adaptive beamforming.
- The optimal mixing factor that minimizes the BER performance depends on scenario and network configuration.
- BER dependence on δ is similar for both known and estimated channels with slightly better results for the known channel and significantly different optimal values of the mixing parameter.

CDFs for the optimal and the optimized according to Section IV mixing factors for each scenario realization and averaged BER results for the known and estimated channels are presented in Figs. 4 and 5 with the same configuration settings as in Fig. 3 The optimal values of the mixing factor are obtained by means of (non-implementable) direct BER

minimization for each scenario realization for both cases of the known and estimated channels of the desired signals. The BER results for the conventional training ($\delta = 0$) and data ($\delta = 1$) based beamformers are shown in Fig. 5 for comparison.
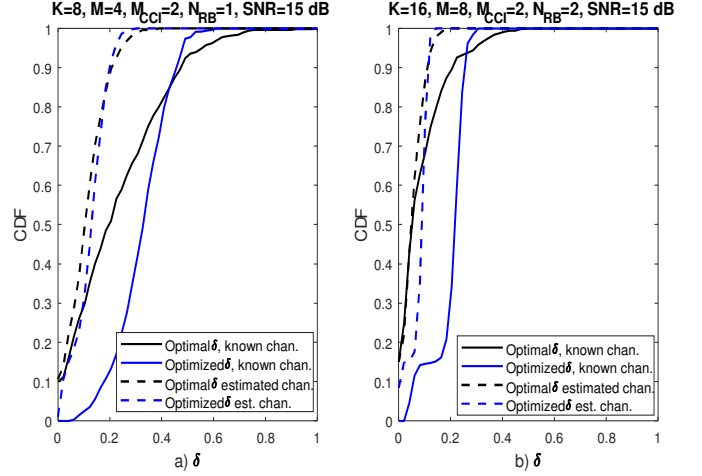

Fig. 4. Distributions of the optimal and optimized values of δ in the same scenarios as in Fig. 3 for SNR=15 dB
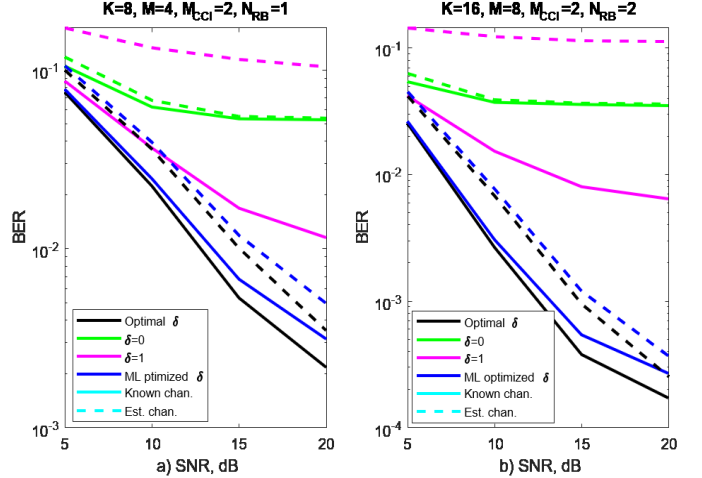

Fig. 5. BER results for different values of δ in the same scenarios as in Fig. 3

The presented simulation results show the following:

- The proposed adaptive selection of the mixing factor gives its optimized values that are close to the optimal ones.
- The BER performance for the optimized mixing factor is close to the optimal one in the considered scenarios.
- The BER results for the proposed optimized mixing factor are much better compared to both conventional boundary cases of the training based for $\delta = 0$ and data based for $\delta = 1$ MMSE-IRC receivers.

Selection of the mixing factor is further illustrated in Fig. 6, which shows the optimization statistics $\tilde{LR}$ in (36) and BER results depending on δ for a typical realization of the extreme scenario, where all CCI components do not overlap with pilots.
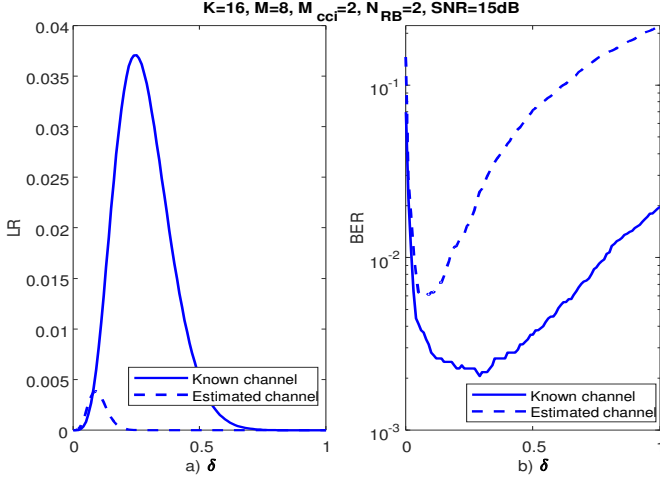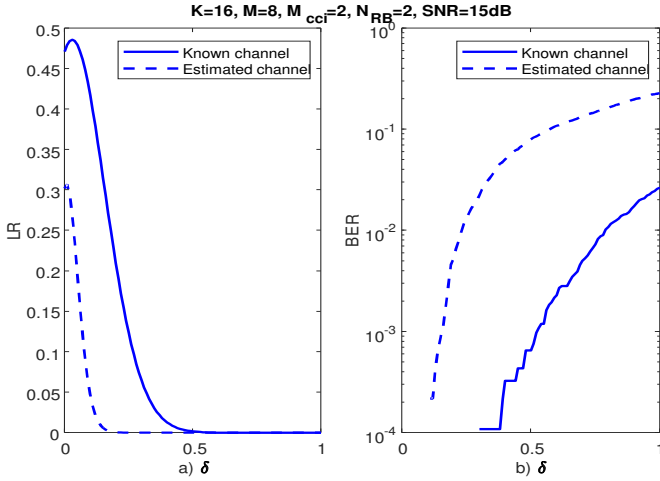
Fig. 6. Illustration of selection of the mixing factor when CCI do not overlap with pilots in the same scenarios as in Fig. 3b for SNR=15 dB

It is important to note that although $\mathbf{R}_t$ estimate is based only on noise samples in this case, we still use the training interval (estimated pilots $\hat{\mathbf{S}}_t(\delta)$ in (32), (33)) to control distortion of the recovered desired signals that allows us to balance it with the "whitening" of interference on the data interval in (34) leading to selection of the effective mixing factors corresponding to the maximum of the decision statistics as one can see in Fig. 6a.

An example of another extreme scenario with all interference components overlapping with the pilots is presented in Fig. 7. One could see another illustration of efficiency of our

Fig. 7. Illustration of selection of the mixing factor when CCI overlaps with pilots in the same scenarios as in Fig. 3b for SNR=15 dB

solution, where very small values of δ are selected, which corresponds to the optimal in this case conventional training-based beamformer.

While the presented above results demonstrate significant performance improvement by the optimized adaptive semi-blind beamformer compared to the training ($\delta = 0$) and data ($\delta = 1$) based adaptive beamformers, it is also important to demonstrate limitations of this approach based exclusively on

second order statistics. To do this, we ones again consider scenarios as per Figs. 3a and 3b with the known $\mathbf{H}$. This time we analyze effects of improvement in interference plus noise covariance matrix estimation on the data interval achieved by means of subtraction ("cleaning") of the different estimates of the useful signals from the received signal on data interval. Figs. 8 and 9 present the corresponding results for SNR=15 dB in 10000 scenario realizations for the following estimates of $\mathbf{R}_d$:

1) the basic one as in (13) without any "cleaning" marked as "Initial";
2) the one with "cleaning" of the data interval with the initially estimated desired signal

$$\hat{\mathbf{R}}_d^{soft} = N_d^{-1} \left[\mathbf{X}_d - \mathbf{H}\hat{\mathbf{S}}_d(\hat{\delta}_{ML}) \right] \left[\mathbf{X}_d - \mathbf{H}\hat{\mathbf{S}}_d(\hat{\delta}_{ML}) \right]^* \tag{37}$$

marked as "Soft cleaning";
3) the one with "cleaning" of the data interval with the initially estimated desired signal projected to the FA $\tilde{\mathbf{S}} = \mathrm{P}_{FA}\{\hat{\mathbf{S}}\}$:

$$\hat{\mathbf{R}}_d^{hard} = N_d^{-1} \left[\mathbf{X}_d - \mathbf{H}\tilde{\mathbf{S}}_d(\hat{\delta}_{ML}) \right] \left[\mathbf{X}_d - \mathbf{H}\tilde{\mathbf{S}}_d(\hat{\delta}_{ML}) \right]^* \tag{38}$$

marked as "Hard cleaning";
4) the one with the ideal (non-implementable) "cleaning"

$$\hat{\mathbf{R}}_d^{ideal} = N_d^{-1} \left[\mathbf{X}_d - \mathbf{H}\mathbf{S}_d \right] \left[\mathbf{X}_d - \mathbf{H}\mathbf{S}_d \right]^* \tag{39}$$

marked as "Ideal cleaning".

Figs. 8a and 9a show histograms of the likelihood ratio values in (36) obtained after maximization with the indicated estimates of $\mathbf{R}_d$. Figs. 8b and 9b present CDFs of the optimized values of the mixing factor. Figs. 8c and 9c show CDFs of the corresponding BER results.
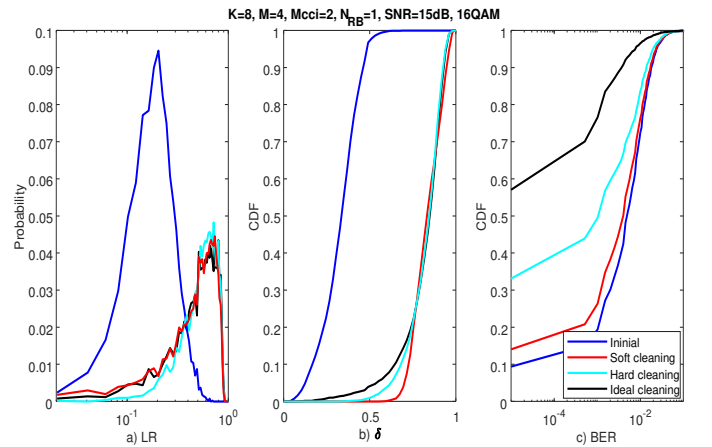

Fig. 8. Comparison of the initial and "cleaned" beamforming versions in the same scenario as in Fig. 3a for SNR=15 dB

Additionally, in Fig. 10, we illustrate the decision metric (36) and N_dth roots of its training and data components in the particular scenario realization.

The presented simulation results show the following:

- The likelihood ratio values for the soft, hard and ideally "cleaned" estimates of the interference plus noise covariance matrix on the data interval are practically the same
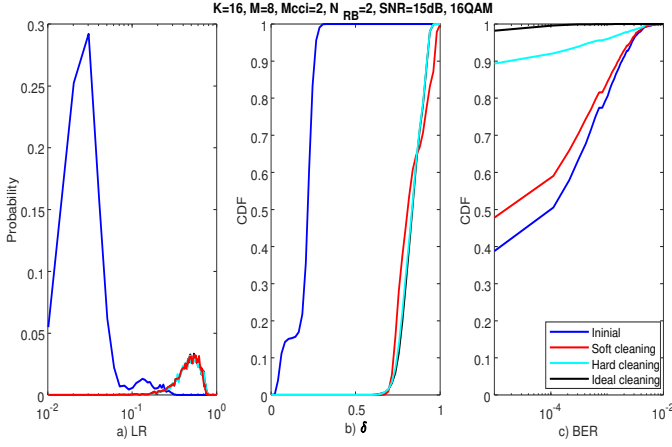
Fig. 9. Comparison of the initial and "cleaned" beamforming versions in the same scenario as in Fig. 3b for SNR=15 dB
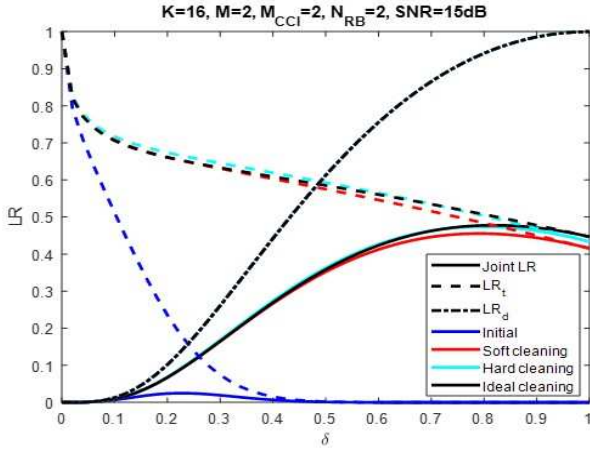

Fig. 10. Illustration of the joint decision metric and its training and data components in the same scenario as in Fig. 3b for SNR=15 dB

as one can see in Figs. 8a and 9a. This means that better in the likelihood sense than the soft "cleaning" estimates cannot be obtained by means of second order statistics. The main reason for that is that $\mathrm{LR}_d(\delta)$ in (34) does not depend on "cleaning" of the data based covariance matrix. The analytical demonstration of this remarkable property is given in Appendix and it is further illustrated in Fig. 10, where one can see the same $\mathrm{LR}_d(\delta)$ values for the initial and all the introduced "cleaned" estimates of $\mathbf{R}_d$. On the other hand, the joint likelihood ratio is very different for the initial and "cleaned" estimates of $\mathbf{R}_d$. The reason is that the training component LR_t in (32) is very different for these estimates as one can see in Fig. 10. At the same time, no significant difference in LR_t behavior between different types of "cleaning" in (37) - (39) could be observed. This is because for high enough values of δ in (13), the presence of the useful signals in $\hat{\mathbf{R}}_d$ in (11) leads to rejection of the training data, but even the soft "cleaning" in (37) reduces the signal component in the covariance matrix estimate to the levels that do not significantly affect the extracted training signal accuracy.

- In Figs. 8b and 9b, our selection algorithm (36) gives much higher and close values of the optimized mixing factor for all "cleaned " versions, which is natural for better estimates of the interference plus noise covariance matrix compared with the initial one due to accurate extraction of the desired signals. This situation demonstrates ability and efficiency of our proposed mixing factor selection algorithm.

- In Figs. 8c and 9c, the BER performance for the soft "cleaned" estimates is only slightly better compared to the initial one illustrating limitations of using only second order statistics solution, while the ideally "cleaned" results are much better indicating a room for possible performance improvements using higher order statistics techniques. One way for such improvements could be "cleaning" with projections to FA leading to significant improvement of the BER results as one can see in Figs. 8c and 9c, although the likelihood "quality" of such estimate remains the same as for the soft "cleaning" as one can see in Figs. 8a and 9a. More results on the iterative "cleaning" with projections to the FA with initialization from the adaptive linear beamformer with selection of the mixing factor defined in (36) are given in Section V.B.

Next, we demonstrate robustness of the proposed adaptive solution for lower dimensions of the signal subspace. In Fig. 11, we present the BER results in the same scenario as in Fig. 5b, but for $M = 2, 4$ using $\beta = M/K$ in (36).
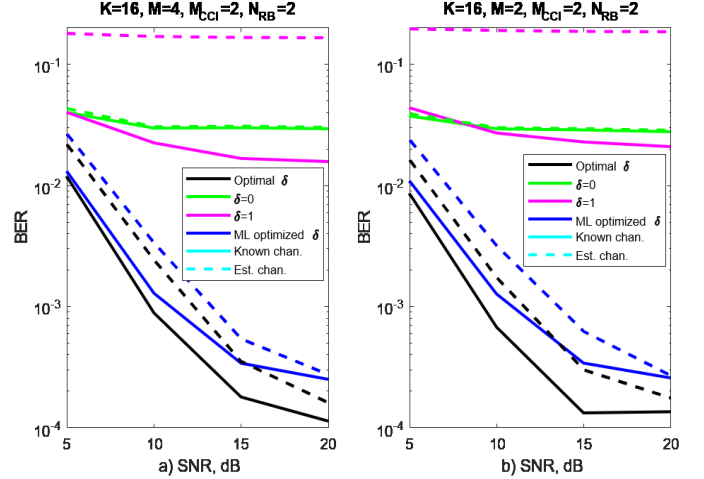

Fig. 11. BER results for different values of δ in the same scenario as in Fig. 5b for $M = 4$ (a) and $M = 2$ (b)

One can see in Fig. 11 that our optimized beamformer significantly outperforms both conventional boundary cases of the training based for $\delta = 0$ and data based for $\delta = 1$ MMSE-IRC solutions, and its BER results are still close to the optimal ones demonstrating robustness of the proposed algorithm.

In Fig. 12, we show the simulation results in the same scenario as in Fig. 5b, but for $N_{RB} = 1$. Now, for $K = 16$ antennas and $M = 8$ users we have only $N_t = 24$ samples at the training interval. In Fig. 12a, we still do not use any diagonal loading, but in Fig. 12b we apply a typical conventional regularization of $\alpha = 0.5\sigma^2$. One can see mach

better results for the optimized mixing factor for the estimated channel in Fig. 12b, which illustrates importance of using both types of regularization in the asynchronous CCI scenario with low sampling support.
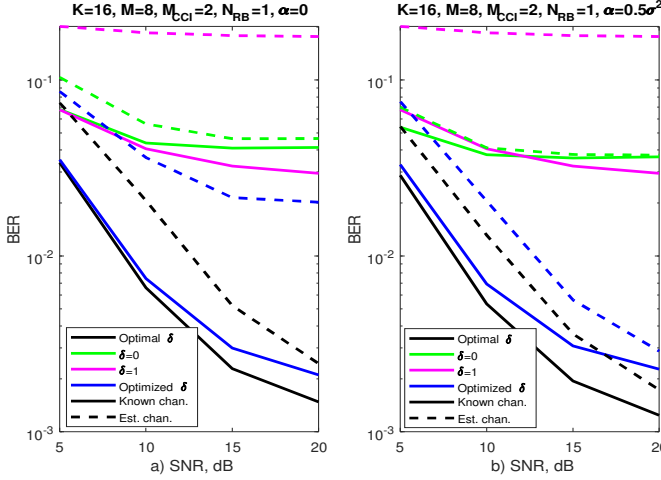

Fig. 12. BER results for different values of δ in the same scenario as in Fig. 5b for $N_{RB} = 1$ with no diagonal loading $\alpha = 0$ (a) and with diagonal loading $\alpha = 0.5\sigma^2$ (b)

B. OFDM scenario

We simulate an OFDM network with the 5G-NR numerology on the uplink [31] for the similar interference scenario as in Section V.A using the 16QAM signaling for both the signals and CCI. We assume 3.4 GHz carrier frequency, 10 MHz bandwidth, 15 kHz carrier spacing and 3 km/h Doppler speed for both the desired signals and CCIs that occupy the whole band. The difference compared to the interference scenario in Section IV.A is that CCI symbols can partially overlap with the desired signal symbols in the OFDM case. We consider the case of $M = K/2$ and use $\beta = 1$ in (36). The BER results estimated over 200 scenario realizations (10000 RBs) are presented in Figs. 13 and 14 for the Extended Pedestrian A (EPA) channel model [32] with 410 ns maximum delay spread. for the same scenarios and algorithms as in Fig. 5 for the estimated fixed channel in time and frequency for all resource elements for $N_{RB} = 1$ in Fig. 13, and $N_{RB} = 2$ in Fig. 14. Additionally, the BER results for the hard "cleaning" as in (38) for 1 and 2 iterations are plotted in Figs. 13 and 14 with initialization from all the considered algorithms using $\delta = 1$ for both iterations taking into account that the optimized mixing parameter approaches to one for the "cleaned" estimates of the interference plus noise covariance matrix on the data interval as illustrated in Figs. 8b and 9b.

One can see in Figs. 13 and 14 that relatively low variations of the channels in time and frequency for the EPA channel and low Doppler speed allow effective averaging of the second order statistics over the considered number of RBs leading to the situation, where for both scenarios the BER performance for the proposed optimized mixing factor is close to the optimal (but intractable and non-implementable) mixing factor results and it significantly outperforms both conventional training
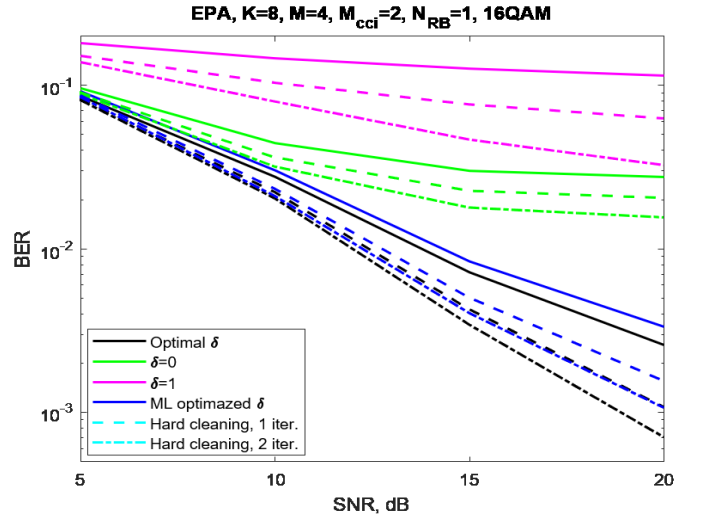

Fig. 13. BER results for different values of δ in OFDM scenario with estimated channels for EPA channel model: $K = 8$, $M = 4$, $M_{CCI} = 2$, $N_{RB} = 1$, $N_t = 24$, $N_d = 144$

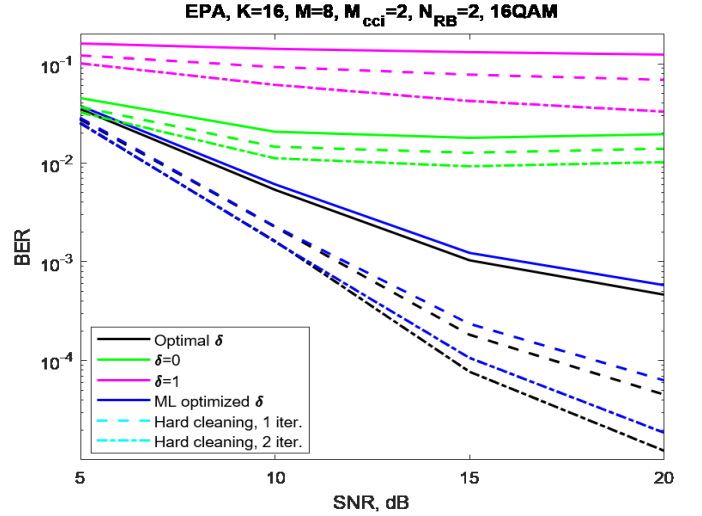Fig. 14. BER results for different values of δ in OFDM scenario with estimated channels for EPA channel model: $K = 16$, $M = 8$, $M_{CCI} = 2$, $N_{RB} = 2$, $N_t = 48$, $N_d = 288$

based ($\delta = 0$) and data based ($\delta = 1$) solutions. As expected, the algorithms with projections to the FA demonstrate significant improvements for the proposed initialization, but negligible effects on the one with the conventional training and data based beamformers.

In Fig. 15, we show the BER results in the same scenario as in Fig 14, but for the Extended Vehicular A (EVA) channel model [32] with 2510 ns maximum delay spread. Now, significant channel variations in the frequency domain do not allow effective averaging over two RBs as in Fig. 14. Thus, we use $N_{RB} = 1$ in Fig.15 and additional diagonal loading similarly to the results in Fig. 12.

One can see an overall performance degradation in Fig. 15 compared to Fig. 14, but the BER performance for the proposed optimized mixing factor is still close to the (non-
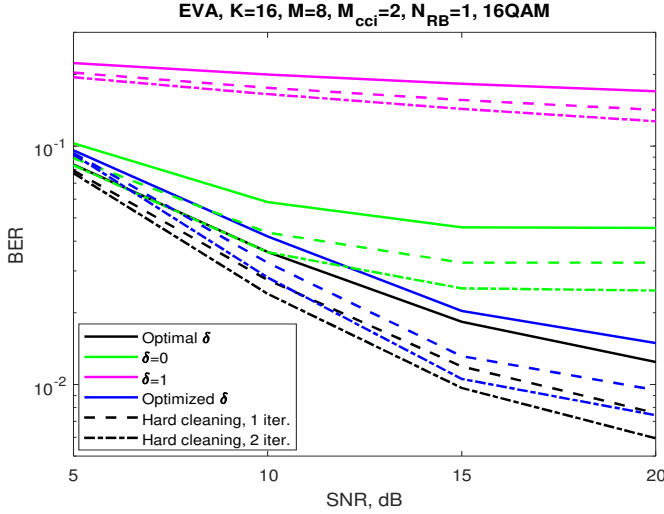
Fig. 15. BER results for different values of δ in OFDM scenario with estimated channels for EVA channel model: $K = 16$, $M = 8$, $M_{CCI} = 2$, $N_{RB} = 1$, $N_t = 24$, $N_d = 144$, $\alpha = 0.5\sigma^2$

implementable) optimal mixing factor results and it significantly outperforms both conventional MMSE-IRC solutions.

VI. CONCLUSIONS

In this paper, we have introduced a family of the second order statistics adaptive beamformers regularized by the covariance matrix estimate over the data interval and proposed a novel maximum likelihood methodology for their optimization. We have developed the algorithm for adaptive selection of the regularization (mixing) factor and demonstrated its efficiency as a linear adaptive beamformer and as an initialization for adaptive beamformer with the FA projections in different asynchronous CCI scenarios. Our regularization solution is widely applicable to systems with different configuration settings. It demonstrates a near-optimal performance and significantly outperforms the conventional training and data based schemes. The proposed approach constitutes a promising and viable solution to mitigate interference in practical asynchronous transmission systems.

ACKNOWLEDGMENT

APPENDIX

In Appendix, we provide an analytical evidence on the independence of the sphericity test (34) with respect to the "cleaning" operation. We, therefore, have to demonstrate that the eigenspectrum of the $(K - M)$ dimension matrix $\hat{\mathbf{Q}}_{K-M}$ (22) is the same for the initial covariance matrix estimate $\hat{\mathbf{R}}_d$ (11) and any of its "cleaned" versions (37) - (39).

Let $\hat{\mathbf{R}}_r$ is the regularized matrix (13) with any of $\mathbf{R}_d$ estimates (11) or (37) - (39). Then, matrix $\hat{\mathbf{Q}}_{K-M}$ (22) can be represented as

$$\hat{\mathbf{Q}}_{K-M} = \mathbf{U}^*_{K-M}\mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}\right]\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{X}_d\times \qquad (40)$$

$$\times\mathbf{X}_d^*\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}\right]\mathbf{U}_{K-M},$$

where projector $\mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}\right]$ is defined in (19) with the corresponding decomposition based on $\mathbf{U}_{K-M}$ in (23).

Now, note that

$$\mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}\right]\hat{\mathbf{R}}_r^{-\frac{1}{2}} = \hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r), \qquad (41)$$

where

$$\mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r) = \mathbf{I}_K - \mathbf{H}\left[\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}. \qquad (42)$$

Similarly, we get

$$\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathrm{P}_\perp\left[\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}\right] = \mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H})\hat{\mathbf{R}}_r^{-\frac{1}{2}}, \qquad (43)$$

where

$$\mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H}) = \mathbf{I}_K - \hat{\mathbf{R}}_r^{-1}\mathbf{H}\left[\mathbf{H}^*\hat{\mathbf{R}}_r^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^* \qquad (44)$$

with the properties

$$\begin{cases} \mathbf{H}^*\mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H}) = 0 \\ \mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r)\mathbf{H} = 0 \end{cases} \qquad (45)$$

Then, we get

$$\hat{\mathbf{Q}}_{K-M} = \mathbf{U}^*_{K-M}\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r)\mathbf{X}_d\times \qquad (46)$$

$$\times\mathbf{X}_d^*\mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H})\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{U}_{K-M}.$$

From (45), it follows that $\hat{\mathbf{Q}}_{K-M}$ in (46) remains the same if we replace $\mathbf{X}_d$ with its any "cleaned" version $\mathbf{X}_d - \mathbf{H}\hat{\mathbf{S}}$ irrespective of $\hat{\mathbf{R}}_r$.

Using singular value decomposition of $\mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r)$ and $\mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H})$

$$\mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r) = \tilde{\mathbf{U}}_{K-M}\mathbf{\Lambda}_{K-M}\tilde{\mathbf{V}}^*_{K-M}, \qquad (47)$$

$$\mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H}) = \tilde{\mathbf{V}}_{K-M}\mathbf{\Lambda}_{K-M}\tilde{\mathbf{U}}^*_{K-M}, \qquad (48)$$

where $\mathbf{\Lambda}_{K-M}$ is the diagonal matrix of $K - M$ non-zero singular values, and

$$\begin{cases} \mathbf{\Lambda}_{K-M}\tilde{\mathbf{V}}^*_{K-M}\tilde{\mathbf{U}}_{K-M} = \mathbf{I}_{K-M} \\ \mathbf{\Lambda}_{K-M}\tilde{\mathbf{U}}^*_{K-M}\tilde{\mathbf{V}}_{K-M} = \mathbf{I}_{K-M} \end{cases} \qquad (49)$$

since $\mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r)$ and $\mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H})$ are idempotent matrices, we get from (46)

$$\hat{\mathbf{Q}}_{K-M} = \mathbf{T}^*\left[\tilde{\mathbf{V}}^*_{K-M}\mathbf{X}_d\mathbf{X}_d^*\tilde{\mathbf{V}}_{K-M}\right]\mathbf{T}, \qquad (50)$$

where

$$\mathbf{T} = \mathbf{\Lambda}_{K-M}\tilde{\mathbf{U}}^*_{K-M}\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{U}_{K-M}. \qquad (51)$$

Now, let us prove that $\mathbf{T}^*\mathbf{T}$ is orthogonal to $\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}$ irrespective of $\hat{\mathbf{R}}_r^{-\frac{1}{2}}$. Indeed,

$$\mathbf{T}^*\mathbf{T} = \mathbf{U}^*_{K-M}\mathbf{Z}\mathbf{U}_{K-M}. \qquad (52)$$

where

$$\mathbf{Z} = \hat{\mathbf{R}}_r^{-\frac{1}{2}}\tilde{\mathbf{U}}_{K-M}\mathbf{\Lambda}^2_{K-M}\tilde{\mathbf{U}}^*_{K-M}\hat{\mathbf{R}}_r^{-\frac{1}{2}} = \qquad (53)$$

$$\hat{\mathbf{R}}_r^{-\frac{1}{2}}\tilde{\mathbf{U}}_{K-M}\mathbf{\Lambda}_{K-M}\tilde{\mathbf{V}}^*_{K-M}\tilde{\mathbf{V}}_{K-M}\mathbf{\Lambda}_{K-M}\tilde{\mathbf{U}}^*_{K-M}\hat{\mathbf{R}}_r^{-\frac{1}{2}} =$$

$$\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathrm{P}_\perp(\mathbf{H}, \hat{\mathbf{R}}_r)\mathrm{P}_\perp(\hat{\mathbf{R}}_r, \mathbf{H})\hat{\mathbf{R}}_r^{-\frac{1}{2}},$$

taking into account decomposition in (47) and (48).

Then, using (42) and (44) in (53), it becomes clear that

$$\begin{cases} \mathbf{H}\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{Z} = 0 \\ \mathbf{Z}\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H} = 0 \end{cases} \qquad (54)$$

irrespective of $\hat{\mathbf{R}}_r^{-\frac{1}{2}}$.

Therefore, matrix $\mathbf{T}^*\mathbf{T}$ is free of any components in the "whitened" signal subspace spanned by $\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}$ irrespective of $\hat{\mathbf{R}}_r^{-\frac{1}{2}}$. Similarly, matrix $\tilde{\mathbf{V}}_{K-M}^*\mathbf{X}_d\mathbf{X}_d^*\tilde{\mathbf{V}}_{K-M}$ in (50) does not contain any components spanned by $\mathbf{H}$ because $\tilde{\mathbf{V}}_{K-M}^*\mathbf{H} = 0$, which follows from (45) and (47). For this reason, we may conclude that within the subspace orthogonal to the "whitened" signal subspace, no components spanned by $\hat{\mathbf{R}}_r^{-\frac{1}{2}}\mathbf{H}$ or $\mathbf{H}$ are present. Therefore, the transformation of the input data ("cleaning") that modifies only signal subspace of the input data, should not affect properties of the components that belong to the complementary to the "whitened" signal subspace, irrespective to the "whitening" matrix $\hat{\mathbf{R}}_r^{-\frac{1}{2}}$. Particularly, the eigenspectrum of matrix $\hat{\mathbf{Q}}_{K-M}$ should not depend on such transformation of the input signal.

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, J. C. Zhang, "What will 5G be?" IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1065-1082, 2014.
[2] X. You, et al. "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," Science China. Information Sciences, vol. 64, Jan. 2021.
[3] "The future role of spectrum sharing for mobile and wireless data services: Licensed sharing, Wi-Fi, and dynamic spectrum access," Ofcom, Statement Report, Apr. 2014.
[4] "Spectrum sharing in wireless networks: Fairness, efficiency, and security," Eds. J. D. Matyjas, S. Kumar, F. Hu, CRS Press, 2016.
[5] M. Kamel, W. Hamouda, A. Youssef, "Ultra-dense networks: A survey," IEEE Commun. Surveys and Tutorials, vol. 18, no. 4, pp. 2522-2542, May 2016.
[6] 5G White Paper, Next Generation Mobile Networks (NGMN) Alliance, 17 Feb. 2015. Available: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf
[7] D. Astely, B. Öttersten, Spatiotemporal interference rejection combining," in Smart Antennas: State of the Art. Ed. By T. Kaiser, et al., Hindawi Publishing Corp., 2005.
[8] R. A. Monzingo, T. W. Miller, "Introduction to adaptive arrays," Wiley Inter-science, 1980.
[9] E. Lindskog, "Combatting co-channel Interferers in a TDMA system using interference estimates from adjacent frames," in Proc. Asilomar Conf. Signals, Systems and Computers, Nov. 1995.
[10] A. K. Wang, J. Leary, "SMI based beamforming algorithms for TDMA Signals," in Proc. Asilomar Conf. Signals, Systems and Computers, Nov. 1997.
[11] E. Villier, L. Lopes, S. Aftelak, "On the application of uplink optimum combining to base station reception," in Proc. VTC, May 1998.
[12] F. Siddiqui, V. Sreng, F. Danilo-Lemoine, D. Falconer, "Antenna array training and adaptation techniques in an unpredictable and uncontrolled interference environment, in Proc. VTC Fall, Oct. 2003.
[13] M-H. Yeon, J. J. Shynk, "A stochastic analysis of an iterative semi-blind beamformer for TDMA systems," Signal Processing, vol. 89, pp. 1347-1358, 2009.
[14] A. M. Kuzminskiy, H. R. Karimi, "Multiple-antenna interference cancellation for WLAN with MAC interference avoidance in open access networks," EURASIP J. Wireless Communications and Networking, ID 51358, pp. 1-11, 2007.
[15] P. Mary, J. M. Gorce, G. Villemaud, M. Dohler, M. Arndt, "Reduced complexity MUD-MLSE receiver for partially-overlapping WLAN-like interference," in Proc. VTC Spring, Apr. 2007.
[16] Y. Ohwatari, N. Miki, T. Abe, H. Taoka, "Investigation on advanced receiver employing interference rejection combining in asynchronous network for LTE-advanced downlink," in Proc. VTC Spring, May 2012.
[17] Y. Lost, M. Abdi, R. Richter, M. Jeschke, "Interference rejection combining in LTE networks," Bell Labs Technical Journal, vol. 17, no.1, pp. 25-50, 2012.
[18] RP-201118, "Views on Rel-17 NR Demodulation requirements," Intel Corporation, 3GPP TSG RAN #88e, July 2020.
[19] P. Karunakaran, T. Wagner, A. Scherb, W. Gerstacker, "On interference rejection combining for LTE-A systems: Analysis of covariance estimators and an iterative algorithm for frequency-selective channels," in Proc. VTC Spring, May 2015.
[20] Y. I. Abramovich, A. M. Kuzminskiy, "On correspondence between training-based and semiblind second-order adaptive techniques for mitigation of synchronous CCI," IEEE Trans. Signal Processing, vol. 54, no. 6, pp. 2347 - 2351, June 2006.
[21] 802.11ah-2016 - IEEE Standard for Information technologyTelecommunications and information exchange between systems-Local and metropolitan area networksSpecific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Sub 1 GHz License Exempt Operation.
[22] M. C. Wells, Increasing the capacity of GSM cellular radio using adaptive antennas," IEE Proc. Communications, vol. 143, n.5, pp. 304-310, 1996.
[23] A. M. Kuzminskiy, Y. I. Abramovich, "Second-order asynchronous interference cancellation: Regularized semi-blind technique and non-asymptotic maximum likelihood benchmark," Signal Processing, vol. 86, pp. 3849-3863, 2006.
[24] J. Tong, P. J. Schreier, "A unified framework for regularized linear estimation in communication systems," Signal Processing, vol. 93, pp. 2671-2686, 2013.
[25] W. Liu, S. Ding, "An efficient method to determine the diagonal loading factor using the constant modulus feature," IEEE Trans. Signal Processing, vol. 56, no. 12, pp. 6102-6106, Dec. 2008.
[26] X. Gou, Z. Liu, Y. Xu, "Fully automatic robust adaptive beamforming using the constant modulus feature," IET Signal Processing, vol. 8, no. 8, pp. 823-830, 2014.
[27] A. M. Kuzminskiy, Y. I. Abramovich, P. Xiao, R. Tafazolli, "Uniform expected likelihood solution for interference rejection combining regularization," in Proc. ICASSP, Mar. 2016.
[28] Y. I. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," Radio Engineering and Electronic Physics, vol. 26, no. 3, pp. 87-95, 1981.
[29] H. Cox, R. M. Zeskind, M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust. Speech Signal Processing, vol. ASSP-35, pp. 1365-1376, Oct. 1987.
[30] R. J. Muirhead, "Aspects of multivariate statistical theory," Wiley Inter-science, 2005.
[31] 5G; NR; Physical channels and modulation, 3GPP TS 38.211 version 15.3.0 Release 15, 2018.
[32] LTE. Evolved Universal Terrestrial Radio Access (E-UTRA). Relay radio transmission and reception. 3GPP TS 36.115 version 11.4.0 Release 11, 2015.