Separability and Sum-Rate Maximization in MIMO Interfering Networks

Hadi Ghauch, Student Member, IEEE, Taejoon Kim, Member, IEEE, Mats Bengtsson, Senior Member, IEEE, and Mikael Skoglund, Senior Member, IEEE

Abstract—In this work we highlight the inherent connection between sum-rate maximization problems and separability metrics (arising in the context of linear discriminant analysis), by establishing that maximizing the separability between the signal and interference-plus-noise subspaces, results in optimizing a lower bound on the sum-rate of the network. We propose a distributed algorithm for that purpose, whereby the transmit/receive filters are updated to maximize the latter separability metric, establish the fact that it is a generalization of max-SINR, and argue its superior performance. We also advocate the use of yet another separability metric - a so-called DLT bound (i.e., a difference of log and trace). We show that it is a lower bound on the sum-rate shed light on its tightness, and underline a major advantage of using such a bound: it leads to separable convex subproblems that decouple at both the transmitters and receivers. Moreover, we derive the solution to the latter subproblem, that we dub non-homogeneous waterfilling (a variation on the MIMO waterfilling solution), and underline an inherent desirable feature: its ability to turn-off streams exhibiting low-SINR, thereby greatly speeding up the convergence of the proposed algorithm. We then show the convergence of the resulting algorithm, max-DLT, to a stationary point of the DLT bound (a lower bound on the sum-rate). Finally, we rely on extensive simulation of various network configurations, to establish the superior performance of our proposed schemes, with respect to other state-of-the-art methods.

Index Terms—Distributed optimization, Difference of Log and Trace (DLT), Non-homogeneous Waterfilling, max-DLT, Alternating Iterative Maximal Separation (AIMS), Generalized Multi-dimensional Rayleigh Quotient (GMRQ), Trace Quotient (TQ)

I. INTRODUCTION

The exponentially increasing demand for wireless data rate has been identified as a major challenge that is to be addressed by 5G communication. Thus, future cellular networks have to be orders of magnitude more spectrally efficient, than current ones. Furthermore, as has been already established by the research community, inter-/intra-cell interference puts a hard ceiling on the achievable data rates in wireless networks, devising ways of combating interference become an active area of research. Coordinated Multipoint (CoMP) - namely its Coordinated Beamforming variant, was one of the later attempts at turning interference from foe to friend [2]. Moreover, the advent of the so-called Interference Alignment (IA) concept [2], [3] (where interfering signals are aligned in a predetermined subspace at each receiver) spurred a widespread interest in the research community. On a more fundamental level, problems such as CoMP or IA are precoding problems: the aim is to (jointly) design the transmit and the receive filters to optimize a selected metric.

Over the years, earlier work focused on tackling those problems in an iterative distributed manner, where only local CSI is needed. Such schemes employ Forward-Backward (F-B) iterations (also known as ping-pong iterations), to iteratively optimize the transmit and receive filters (detailed in Definition [1]). Most of the earlier distributed algorithms (e.g., Distributed IA / Leakage Minimization and max-SINR [4], (weighted) Minimum MSE [5], [6], sum-rate maximization [7]), were originally developed within the context of the MIMO Interference Channel (MIMO IFC) - although earlier approaches employing F-B training can be traced back to [8]. Note that the plethora of such algorithms that have been developed so far, all use the same underlying structure of F-B training, to accomplish the optimization in a distributed fashion. The inherent communication overhead is thus a primary concern, since such algorithms typically require hundreds/thousands of over-the-air iterations. Despite the critical nature of this issue, this line of work remains largely unexplored. In line with recent work, [9]–[11], investigating algorithms that operate in the low-overhead regime (where just a few F-B iterations are performed), will be one of the main aims of this work.

With the exception of [12], [13], fewer work focused on the MIMO Interfering Broadcast Channels (MIMO IBC) - a natural generalization of the MIMO IFC to multi-cellular multi-user settings. Moreover, to the best of our knowledge, no well-known scheme has been designed for the MIMO Interfering Multiple-Access Channels (MIMO IMAC) - another extension of the MIMO IFC. As a matter of fact, with the exception of [14], little is known about the optimal degrees-of-freedom and the feasibility of IA, in such scenarios. Shedding light on the MIMO IMAC case will be one of the main goals of this work.

In particular, we address the problem of sum-rate maximization in MIMO Interfering Multiple-Access Channels (MIMO IMAC), by formulating lower bounds on the latter. In a first part, we highlight the inherent connection between this formulation, and separability metrics. We propose a distributed algorithm dubbed Alternating Iterative Maximal Separation (AIMS), that optimizes the latter, and establish the fact that the algorithm is a generalization of max-SINR [4]. Furthermore, we argue (and later verify via simulations) that this generaliza-
tion offers superior performance over max-SINR. In addition, we advocate the use of a lower bound on the sum-rate, that we refer to as a difference of log and trace (DLT) expression. We highlight the main advantages of using such an expression, namely that it yields optimization problems that decouple in both the transmit and receive filters, and subproblems that are convex, at each transmitter / receiver. We derive the solution to each of the subproblems - that we dub non-homogeneous waterfilling, and underline its ability for stream control (by turning off streams that have low SINR). We then propose a corresponding distributed algorithm, max-DLT, and establish its convergence to a stationary point of the DLT expression - a lower bound on the sum-rate. Finally, we argue the crucial role of the stream control mechanism, in significantly speeding up the convergence of the algorithm.

Though the paper addresses the problem at hand for a MIMO IMAC, it can be easily verified that the latter framework and methods are exactly applicable to the network-dual problem, the MIMO IBC, with no modifications. Needless to say, it also applies to all the ensuing special cases, such as the MIMO IFC, and the MIMO Multiple-Access Channel.

Notation: we use bold upper-case to denote matrices, and bold lower-case denote vectors. Furthermore, for a given matrix \( \mathbf{A} \), we define \( \text{tr}(\mathbf{A}) \) as its trace, \( \| \mathbf{A} \|_F^2 \) as its Frobenius norm, \( |\mathbf{A}| \) as its determinant, \( \mathbf{A}^\dagger \) as its conjugate transpose, and \( \mathbf{A}^{-1} \) as \( (\mathbf{A}^\dagger)^{-1} \). In addition, \( \mathbf{A}(i,j) \) denotes its \( i \)th column, \( \mathbf{A}_{i,j} \) element \((i,j)\) in \( \mathbf{A} \), \( \lambda_i[\mathbf{A}] \) the \( i \)th eigenvalue of a Hermitian matrix \( \mathbf{A} \) (assuming the eigenvalues are sorted in decreasing order), and \( v_{i,d}[\mathbf{A}] \) denotes the \( d \) dominant eigenvectors of \( \mathbf{A} \). \( \mathbf{S}_{n,n} \) (resp. \( \mathbf{S}_{n,n}^- \)) is the set of complex \( n \times n \) positive (resp. positive definite) matrices. Furthermore, \( \mathbf{A} \succ \mathbf{0} \) (resp. \( \mathbf{A} \succeq \mathbf{0} \)) implies that \( \mathbf{A} \) is positive definite (resp. positive semi-definite), and \( \mathbf{A} \succ \mathbf{0} \) (resp. \( \mathbf{A} \succeq \mathbf{0} \)) implies that \( \mathbf{A} - \mathbf{B} \succ \mathbf{0} \) (resp. \( \mathbf{A} - \mathbf{B} \succeq \mathbf{0} \)). Finally, \( \mathbf{I}_n \) denotes the \( n \times n \) identity matrix, \( \{n\} = \{1, \ldots, n\} \), and \( x^+ = \max\{0, x\} \).

II. System Model and Problem Formulation

Consider a system with \( K \) cells (each having one BS), where each cell is serving \( K \) users (Fig. [1]). Let \( L \) be the set of BSs, \( K_L \) the set of users served by BSs \( l \in L \), and \( K_j \) the index of user \( j \in K_j \), at BS \( l \in L \). We denote by \( \mathcal{T} \) the total set of users, i.e., \( \mathcal{T} = \{k_j \mid l \in L, j \in K_l\} \). The received signal at Base Station (BS) \( l \in L \) is given by,

\[
y_{l} = \sum_{i \in L} \sum_{k \in K_l} h_{l,i,k} V_{i,k} s_{i,k} + n_{l}, \quad l \in L
\]

After linear processing with the receive filter, the recovered signal vector of user \( l_j \in \mathcal{I} \), is given by,

\[
\hat{s}_{l_j} = U_{l_j}^\dagger H_{l,j} V_{l_j} s_{l_j} + \sum_{i \in L} \sum_{k \in K_l} \sum_{j \in K_l} U_{l_j}^\dagger H_{l,i,k} V_{i,k} s_{i,k} + U_{l_j}^\dagger n_{l}, \quad \forall l_j \in \mathcal{I}
\]

where the first terms represents the desired signal, the second both the intra and inter-cell interference. In the above, \( s_{i,k} \) represents the \( d \)-dimensional vector of independently encoded symbols for user \( i_k \in \mathcal{I} \), with covariance matrix \( \mathbb{E}[s_{i,k} s_{i,k}^\dagger] = I_d \). In addition, \( V_{i,k} \) denotes the \( M \times d \) transmit filter of user \( i_k \in \mathcal{I} \), \( U_{l_j} \) the \( N \times d \) receive filter of user \( l_j \in \mathcal{I} \), and \( H_{l,i,k} \) the \( N \times M \) matrix of channel coefficients from user \( i_k \in \mathcal{I} \), to BS \( l \). \( n_l \) represents the \( N \)-dimensional AWGN noise at BS \( l \), such that \( \mathbb{E}[n_n n_l^\dagger] = \sigma^2 I_N \).

Note that our model (and the results presented thereafter) can easily be extended to cases where \( M, N, d \) are different across users and BSs.

If we assume that joint encoding/decoding of each user’s streams is performed at the users and BSs, respectively, and treating interference as noise, the achievable rate of user \( l_j \in \mathcal{I} \) is given by,

\[
r_{l_j} = \log_2 |I_d + (U_{l_j}^\dagger R_{l_j} U_{l_j}) (U_{l_j}^\dagger Q_{l,j} U_{l_j})^{-1}|, \quad l_j \in \mathcal{I}
\]

where \( R_{l_j} \) and \( Q_{l,j} \) are the desired signal and interference-plus-noise covariance matrices for user \( j \), at BS \( l \), respectively, and are given by,

\[
R_{l_j} = H_{l,j} V_{l_j} V_{l_j}^\dagger H_{l,j}^\dagger l_j \in \mathcal{I}
\]

\[
Q_{l,j} = \sum_{i=1}^L \sum_{k=1}^K H_{l,i,k} V_{i,k} V_{i,k}^\dagger H_{l,i,k}^\dagger + \sigma^2 I_N - R_{l_j}, l_j \in \mathcal{I}
\]

Moreover, we define

\[
R_{i_k} = \sum_{i=1}^L \sum_{j=1}^K H_{i,j} V_{i,j} V_{i,j}^\dagger H_{i,j}^\dagger i_k \in \mathcal{I}
\]

\[
Q_{i_k} = \sum_{j=1}^L \sum_{i=1}^K H_{i,j} V_{i,j} V_{i,j}^\dagger H_{i,j}^\dagger + \sigma^2 I_N - R_{i_k}, i_k \in \mathcal{I}
\]

as the signal and interference-plus-noise covariance matrices of user \( i_k \), in the reverse network (where \( \sigma^2 \) is the noise variance at user \( i_k \)). Finally, we henceforth denote \( L_{i_k} L_{i_k}^\dagger \) as the Cholesky Decomposition of \( Q_{i_k} \), and \( K_{i_k} K_{i_k}^\dagger \) as that of \( Q_{i_k} \). We formulate the (unweighted) sum-rate maximization problem as follows,

\[
\max \sum_{l_j \in \mathcal{I}} \{r_{l_j}\} = \max \sum_{l_j \in \mathcal{I}} \{r_{l_j}\}
\]

In the next section we generalize the well-known max-SINR algorithm from a stream-by-stream optimization algorithm, to an algorithm that optimizes the whole transmit/receive filter.
For that purpose, we show that the latter can be formulated using separability metrics. We next highlight the central assumptions/definitions of this work.

### A. Preliminaries

The schemes we consider in the present work fall under the category of Forward-Backward training, recapped in the definition below.

**Definition 1** (F-B Training). Schemes employing Forward-Backward (F-B) iterations (also known as ping-pong iterations, or bi-directional training), consist of optimizing the receive filters in the forward training phase, then the transmit filters in the reverse training phase. They exploit channel reciprocity in Time-Division Duplex (TDD) systems, and result in fully distributed algorithms. The basic iteration structure is show in Fig. 3.

**Definition 2** (Separability). Given two sets of points with covariance matrices $R$ and $Q$, separability is a measure of the distance between the sets, after projecting on a subspace $U$. Separability metrics - the building blocks of areas such as linear discriminant analysis [13], Chap. 4.1, include the *Generalized Multi-dimensional Rayleigh Quotient*

$$\text{GMRQ} = \frac{|U^\dagger RU|}{|U^\dagger RU|}$$

*Trace Quotient* $\text{TQ} = \text{tr}(U^\dagger RU)/\text{tr}(U^\dagger RU)$, and *Quotient Trace* $\text{QT} = \text{tr}\left((U^\dagger RU)/(U^\dagger RU)^{-1}\right)$

The above quantities are inherent to dimensionality reduction problems [16]. In the context of this work, $R$ and $Q$ represent the signal and interference-plus-noise covariance matrices, respectively, and $U$ linear filtering at the receiver.

**Assumption 1** (Local CSI). We assume that the users and BSs have local Channel State Information (CSI), i.e., each user (resp. BS) knows the channels to its desired and interfering BSs (resp. users). Although we underline that methods in [17] are applicable for acquiring such quantities (discussion in Sect. V-A), investigating the CSI acquisition mechanism is not part of this work. Moreover, CSI at each BS and user is assumed to be perfectly known.

**Assumption 2** (Distributed Operation). All schemes are required to use local CSI only, using the framework of Forward-Backward training.

**Assumption 3** (Low-Overhead Regime). We restrict our proposed schemes to operate in the low-overhead regime, where only a small number of F-B iterations is used (in line with recent work such as [9] [11]).

**Assumption 4** (Decoding). Note that a common assumption in uplink communication scenarios such as ours, is that multi-user decoding is performed (e.g. successive interference cancellation). Due to the fact that such receivers are hard to realize in practice, we do not make such assumptions.

### III. SEPARABILITY AND SUM-RATE MAXIMIZATION

#### A. Problem Formulation

We make use of the fact that $\log |X|$ is monotonically increasing on the positive-definite cone, i.e.,

$$\log |X_2| \geq \log |X_1|, \text{ for } X_2 \succeq X_1 \succ 0$$

Applying the above property, we lower bound $r_{l_j}$ in (3) as,

$$r_{l_j} \geq \log_2 |(U_j^\dagger R_{l_j} U_{l_j})(U_j^\dagger Q_{l_j} U_{l_j})^{-1}|$$

$$= \log_2 \left| \frac{U_j^\dagger R_{l_j} U_{l_j}}{U_j^\dagger Q_{l_j} U_{l_j}} \right| \equiv \tilde{r}_{l_j}, \forall l_j \in I$$

Note that $\tilde{r}_{l_j}$ is an approximation of the actual user rate $r_{l_j}$, where the approximation error is $O(\text{tr}(|U_j^\dagger Q_{l_j} U_{l_j}|)(U_j^\dagger R_{l_j} U_{l_j})^{-1}))$ (Refer to Appendix D). Thus, the sum-rate $R_S$ can be bounded below, as follows,

$$R_S > \sum_{l_j \in I} \tilde{r}_{l_j} = \log_2 \prod_{l_j \in I} q_{l_j}, \text{ where } q_{l_j} \equiv \left| \frac{U_j^\dagger R_{l_j} U_{l_j}}{U_j^\dagger Q_{l_j} U_{l_j}} \right|$$

Since the log function is monotonic, the sum-rate maximization problem in (3) is lower bounded as,

$$\text{(SRM)} \left\{ \begin{array}{l} \max \prod_{l_j \in I} q_{l_j} \\ \text{s. t. } \|U_{l_j}\|_F^2 \leq P_r, \|V_{l_j}\|_F^2 \leq P_t, \forall l_j \in I \end{array} \right.$$

### Remark 1 (Power Constraint). In distributed optimization schemes employing Forward-Backward (F-B) training, receivers are active in one of the phases (i.e., by sending data / pilots). Thus, generally, one needs a maximum transmit power constraint for the receiver filter, in addition to the maximum transmit power constraint of the transmitter. In addition, in scenarios involving a multi-cellular downlink communication, each BS employs a sum-power constraint for its users, e.g., [12]. However, the same does not hold in the considered setup (multi-cellular uplink), since it would lead to a sum-power constraint, across all UEs: clearly this is not applicable in practice. Moreover, a per-user power constraint is easier to motivate than a per-BS sum-power constraint, since the latter might result in highly uneven power allocations across users of the same BS (this is often the case in the WMMSE algorithm). For the reasons above, we assume in this work that equal power allocation is done, across the users. This being said, we underline to the fact that, from a mathematical perspective, the cost function in (SRM) renders the presence (or absence) of a receive power constraint, irrelevant.

Referring to (SRM), $q_{l_j}$ is nothing but the GMQR separability metric (Definition 2). Consequently, given the signal and
interference-plus-noise covariance matrices, $R_{ij}$ and $Q_{ij}$, each receiver chooses its filter such to maximize the separation of between signal and interference-plus-noise subspace.

B. Maximization of Generalized Multi-dimensional Rayleigh Quotient

The main limitation of solving problems such (SRM) is the fact it is not jointly convex in all the optimization variables. Though Block Coordinate Decent (BCD) stands out as a strong candidate, one major obstacle persists: while the problem decouples in the receive filters (as shown in (SRM)), attempting to write a similar expression by factoring out the transmit filters, leads to a coupled problem. Therefore, we propose an alternative (purely heuristic) method: the receive filters are updated as the solution to maximize the sum-rate (assuming fixed transmit filters), while the transmit filters are chosen as the solution of the reverse network sum rate maximization (this same structure is implicitly exploited in max-SINR [4]), i.e.,

$$(SRM_F) \begin{cases} \max \left\{ \prod_{j \in \mathbb{Z}} q_{ij}(U_{ij}) = \frac{|U_{ij}^\dagger R_{ij} U_{ij}|}{|U_{ij}^\dagger Q_{ij} U_{ij}|} \right\} \\
\text{s.t. } |U_{ij}|^2 \leq P_r, \forall j \end{cases}$$

and

$$(SRM_B) \begin{cases} \max \left\{ \prod_{k \in \mathbb{Z}} p_k(V_{ik}) = \frac{|V_{ik}^\dagger R_{ik} V_{ik}|}{|V_{ik}^\dagger Q_{ik} V_{ik}|} \right\} \\
\text{s.t. } |V_{ik}|^2 \leq P_t, \forall i_k \end{cases}$$

In other words, assuming transmit filters as fixed, the receive filters are updated such as to maximize the separability metric in the forward phase. Similarly, the transmit filters are chosen to maximize the separability in the backward training phase. Moreover, as seen from the above problems, the objective in each subproblem is invariant to scaling of the optimal solution. Thus, they can be solved as unconstrained problems, and optimal solutions can be scaled, without loss of optimality.

Before we proceed, we first derive a generic solution to the Generalized Multi-dimensional Rayleigh Quotient (GMRQ) maximization. The solution to the latter problem was earlier proposed in [18]. We provide an alternate and more concise proof to the problem (Appendix B).

**Lemma 1.** Consider the following maximization of the $r$-dimensional Generalized Multi-dimensional Rayleigh Quotient (GMRQ).

$$X^* = \arg \max_{X \in \mathbb{C}^{n \times r}} q(X) = \frac{|X^\dagger RX|}{|X^\dagger QX|}, \quad (10)$$

where $Q \in \mathbb{S}^{n \times n}_+$, $R \in \mathbb{S}^{n \times n}$ and $r < n$. The optimal solution to this non-convex problem is given by

$$X^* = L^{-1} \Psi \hat{V}, \quad (11)$$

where $LL^\dagger = Q$, $L \in \mathbb{C}^{n \times n}$, $\Psi = v_{1:1} L_{1:1}^{-1} L_{1:1}^{-\dagger}$, $\hat{V} \in \mathbb{C}^{r \times r}$ is an arbitrary non-singular square matrix.

*Proof:* Refer to Appendix B.

It is worth mentioning that the above solution is a generalized formulation of the well-known generalized eigenvalues solution: this result was also obtained in [18].

**Corollary 1.** Consider a special case of (11) where $\hat{V} = I_r$. Then, this corresponds to the generalized eigenvalues solution, i.e.,

$$X^* = L^{-1} \Psi \Rightarrow RX^* = QX^* \Lambda_r$$

where $\Lambda_r \in \mathbb{R}^{r \times r}$ be the (diagonal) matrix of eigenvalues for $L^{-1} RL^{-1}$.

*Proof:* Refer to Appendix B.

With this in mind, we can write the optimal transmit and receive filter updates, as follows,

$$U_{ij}^* = L_{1:1}^{-1} \Psi_{ij}, \quad \Psi_{ij} \triangleq v_{1:1} d_i L_{1:1}^{-1} R_{ij} L_{1:1}^{-\dagger}, \forall j$$

$$V_{ik}^* = K_{ik}^{-1} \Theta_{ik}, \quad \Theta_{ik} \triangleq v_{1:1} d_i K_{ik}^{-1} R_{ik} K_{ik}^{-\dagger}, \forall i_k$$

where we used the fact we can set $\hat{V} = I_d$ in the solution of (11). We note that the optimal filter updates for the transmitter are more heuristic than the receiver ones: While the receive filter updates directly maximize a lower bound on the sum-rate - as seen in (SRM), no such claim can be made about the transmit filter updates. The details of our algorithm, Alternating Iterative Maximal Separation (AIMS), are shown in Algorithm 1 (where $T$ denotes the number of F-B iterations).

**Algorithm 1** Alternating Iterative Maximal Separation (AIMS)

for $t = 1, 2, ..., T$ do

// forward network optimization: receive filter update

Estimate $R_{ij}, Q_{ij}$, and compute $L_{ij}$, $\forall j$

$U_{ij} \leftarrow L_{1:1}^{-1} v_{1:1} d_i [L_{1:1}^{-1} R_{ij} L_{1:1}^{-\dagger}], \forall j$

$U_{ij} \leftarrow \sqrt{T_r} U_{ij} / \|U_{ij}\|_F$

// reverse network optimization: transmit filter update

Estimate $R_{ik}, Q_{ik}$, and compute $K_{ik}$, $\forall i_k$

$V_{ik} \leftarrow K_{ik}^{-1} v_{1:1} d_i [K_{ik}^{-1} R_{ik} K_{ik}^{-\dagger}], \forall i_k$

$V_{ik} \leftarrow \sqrt{T_t} V_{ik} / \|V_{ik}\|_F$

end for

The solution that is derived above, allows us to draw interesting insights. Proposition 1 establishes the expected but insightful result that employing unitary filters is not optimal, from the perspective of separability.

**Proposition 1.** Consider the optimal receive filter given in (13), i.e., $U_{ij}^* = L_{ij}^{-1} \Psi_{ij}$, where $\Psi_{ij} = v_{1:1} d_i [L_{ij}^{-1} R_{ij} L_{ij}^{-\dagger}]$: it is not orthonormal, almost surely.

*Proof:* A sketch of the proof is given in Appendix C.

A few comments are in order at this stage, regarding the difference between AIMS and max-SINR. Referring to (SRM) and (SRM$_B$), it is clear that our proposed algorithm reduces to max-SINR, in case of single-stream transmission, i.e., setting $d = 1$. Moreover, an inherent property of the max-SINR solution is that it yields equal power allocation across all the streams (since the individual columns of each transmit/receive filter are normalized to unity). However, as evident from (13), our proposed solution does not normalize...
the individual columns of the receive filter, but rather the whole filter norm (as seen in Algorithm 1). This allows for different power allocation, across columns of the same filter. That being said, the proposed solution is expected to yield better sum-rate performance (w.r.t. max-SINR), especially in the interference-limited regime. This is due to the intuitive fact that much can be gained from allocating low power to streams that suffer from severe interference, and higher power to streams with lesser interference (this will be validated in the numerical results section). We next introduce a rank adaptation mechanism that further enhances the interference suppression capabilities of the algorithm.

C. AIMS with Rank Adaptation

We introduce one additional (heuristic) mechanism to robustify AIMS against severely interference-limited scenarios, by introducing a mechanism of Rank Adaptation (RA); in addition to the transmit / receive filter optimization (Lemma 1), the latter allows the filter rank to be optimized as well. Mathematically speaking, RA addresses the following problem,

$$r^* \triangleq \arg\max_r \left[ X^* \triangleq \arg\max_{X \in \mathbb{R}^{n \times r}} \left[ X^R X^* \right] \right],$$

(14)

Using the same argument as Lemma 1 one can verify that \(X^*\) and \(r^*\) are as follows,

$$X^* = [L^{-1} \Psi]_{1:r^*}, \text{ where } \Psi = v_{1:n}[L^{-1} R L^{-1}]$$

(15)

$$r^* = \arg\max_r [\Lambda_r] = \{i | \lambda_i [L^{-1} R L^{-1}] \geq 1 \},$$

where \(\Lambda_r \in \mathbb{R}^{r \times r}\) is the (diagonal) matrix consisting of the \(r\)-largest eigenvalues of \(L^{-1} R L^{-1}\). Simply put, \(r^*\) is the number of eigenvalues greater than one.

When RA is incorporated into AIMS, this mechanism will boost the performance of the algorithm (namely in interference-limited settings). However, one still needs to ensure that the filter ranks for each transmit-receive pair are the same, i.e., rank \((U_{ij})\) = rank \((V_{ij})\) ∀ \(i,j\). One quick (heuristic) solution is as follows. For each transmit-receive pair, compute the optimal filter rank for both the transmit and receive filter, and use the minimum. Needless to say, ensuring this condition requires additional signalling overhead. We thus envision RA, as potential “add-on” for AIMS, when one can afford the resulting overhead increase. The additional performance boost provided by RA, is also highlighted in the numerical results section.

D. Maximization of Trace Quotient

In addition to the optimization based on the GMRQ, we also advocate the use of other metrics (described in Definition 2), namely the Trace Quotient (TQ). The generic formulation of the TQ problem is,

$$U_{ij}^* = \begin{cases} \arg\max_{U_{ij}} \text{tr}(U_{ij}^1 R_{ij} U_{ij})/\text{tr}(U_{ij}^1 Q_{ij} U_{ij}) & \text{s. t. } U_{ij} \in \mathcal{U}_{ij} \\ \end{cases}$$

(16)

Although this section is not part of the contributions of the present work, we include it for completeness. From the context of this work, the TQ problem can be interpreted as follows. This metric is an extension of the per-stream SINR, in the sense that inter-stream interference is now part of the useful signal. In the presence of a unitary constraint, i.e.,

$$U_{ij} = \{U_{ij} \in \mathbb{C}^{N \times d} | U_{ij}^* U_{ij} = I_d \},$$

there is work that attempts to characterize the optimal solution of the TQ problem, e.g., [16], [19]. Note that such approaches are iterative, and thus not best suited for our case. Another variant was considered in [20], where a solution was obtained under an interference-whiteness constraint, i.e.,

$$U_{ij} = \{U_{ij} \in \mathbb{C}^{N \times d} | U_{ij}^* Q_{ij} U_{ij} = I_d \},$$

Finally, when [16] is unconstrained, the TQ problem admits a simple (albeit trivial) solution:

$$U_{ij}^* = L_{ij}^1 \Psi_{ij} \Sigma^*, \quad \Sigma^* = \text{diag}(\sigma_1^*, 0, \ldots, 0)$$

where $\sigma_i^*$ is the \(i\)-th largest eigenvalue of \(L_{ij}^{-1} R_{ij} L_{ij}^{-1}\). This result can be derived based on similar arguments as Lemma 1. Stated simply, the above solution corresponds to the first column in \(U_{ij}\) being the largest generalized eigenvector of \(R_{ij}\) and \(Q_{ij}\) (scaled by \(\sigma_i\)), while the rest are zero.

IV. MAXIMIZING BOUNDS ON SUM-RATE

In this section we propose another approach to tackle the sum-rate optimization problem. The central idea behind this approach is to use a lower bound on the sum-rate, that results in separable sub-problems.

A. Problem Formulation

We focus the derivations to the interference-limited case, where the following holds,

$$\lambda_i[|U_{ij}^1 Q_{ij} U_{ij}|] \rightarrow \infty, \quad \forall i \in \{d\},$$

$$\Leftrightarrow \lambda_i[|U_{ij}^1 Q_{ij} U_{ij}|^{-1}] \rightarrow 0, \quad \forall i \in \{d\}$$

(17)

**Proposition 2.** When \(\sum_{i=1}^d |U_{ij}^1 Q_{ij} U_{ij}| \geq d\), the user-rate \(r_{ij}\) in (5) is lower bounded by,

$$r_{ij} \geq \log_2 |I_d + U_{ij}^1 R_{ij} U_{ij}|- \log_2 |U_{ij}^1 Q_{ij} U_{ij}|, \quad \text{ (b.1)}$$

$$\geq \log_2 |I_d + U_{ij}^1 R_{ij} U_{ij}|- \text{tr}(U_{ij}^1 Q_{ij} U_{ij}) \triangleq r_{ij}^{(LB)} \text{,}$$

(18)

where \(r_{ij}^{(LB)}\) is such that,

$$\Delta_{ij} \triangleq r_{ij} - r_{ij}^{(LB)}$$

$$= \text{tr}(U_{ij}^1 Q_{ij} U_{ij}) - \log_2 |U_{ij}^1 Q_{ij} U_{ij}|$$

$$+ \mathcal{O}(\text{tr}([U_{ij}^1 Q_{ij} U_{ij}](U_{ij}^1 R_{ij} U_{ij}^{-1}))), \quad \forall i \in \mathcal{I}$$

(19)

**Proof:** Refer to Appendix D.

We refer to expressions such as \(r_{ij}^{(LB)}\), as a Difference of Log-Trace (DLT) expressions. They shall be used as basis for the optimization algorithm. With that in mind, the sum-rate\(R_{\Sigma}\), can be lower bounded by \(R_{\Sigma}^{(LB)}\).

$$R_{\Sigma}^{(LB)} = \sum_{i \in \mathcal{I}} \log_2 |I_d + U_{ij}^1 R_{ij} U_{ij}|- \text{tr}(U_{ij}^1 Q_{ij} U_{ij})$$

(20)
\[
\sum_{i_k \in \mathcal{I}} \log_2 |I_d + V_{i_k}^\dagger R_{i_k} V_{i_k}| - \text{tr}(V_{i_k}^\dagger Q_{i_k} V_{i_k}) \tag{21}
\]

where the last equality is due to \( \log |I + AB| = \log |I + BA| \), and the linearity of \( \text{tr}(\cdot) \). Then, the sum-rate optimization problem in (4) can be bounded below by solving the following,

\[
\begin{align*}
\max_{\{V_{l_j}^{	ext{(LB)}}\}} & R_{\Sigma}^{\text{(LB)}} \\
\text{s. t.} & \|U_{l_j}\|_F^2 \leq P_r, \|V_{l_j}\|_F^2 \leq P_t, \forall l_j \in \mathcal{I}
\end{align*}
\tag{22}
\]

Note that the above problem is not jointly convex in all the optimization variables, mainly due to the coupling between the transmit and receive filters.

B. Proposed Algorithm

The formulation in (22) is ideal for a Block Coordinate Descent (BCD) approach. We use the superscript \( (n) \) to denote the iteration number: at the \( n \)th iteration, the transmit filters, \( \{V_{l_j}^{(n)}\} \), are fixed, and the update for the receive filters, \( \{U_{l_j}^{(n+1)}\} \), is the one that maximizes the objective. The same is done for the transmit filter update. In each of the two stages, BCD decomposes the original coupled problem (22) into a set of parallel subproblems, that can solved in distributed fashion. This is formalized in (23), and each of the resulting subproblems are detailed below.

When the transmit filters are fixed, the problem decouples in the receive filters \( \{U_{l_j}\} \) (as seen from (22)), and the resulting subproblems are given by,

\[
(\text{j1}) \begin{align*}
\min_{U_{l_j}} & \text{tr}(U_{l_j}^\dagger Q_{l_j} U_{l_j}) - \log_2 |I_d + U_{l_j}^\dagger R_{l_j} U_{l_j}| \\
\text{s. t.} & \|U_{l_j}\|_F^2 \leq P_r
\end{align*}
\tag{24}
\]

By recalling that (22) can rewritten as (21), we see that the latter objective decouples in the transmit filters, i.e.,

\[
(\text{j2}) \begin{align*}
\min_{V_{i_k}} & \text{tr}(V_{i_k}^\dagger Q_{i_k} V_{i_k}) - \log_2 |I_d + V_{i_k}^\dagger R_{i_k} V_{i_k}| \\
\text{s. t.} & \|V_{i_k}\|_F^2 \leq P_t
\end{align*}
\tag{25}
\]

Thus, choosing DLT expressions is rather advantageous, since they lead to subproblems that are convex, and decouple in both \( \{U_{l_j}\} \) and \( \{V_{i_k}\} \). The solution to each of the subproblems is given by the following result.

**Lemma 2. Non-homogeneous Waterfilling.**

Consider the following convex problem,

\[
\begin{align*}
\min_{X \in \mathbb{C}^{n \times r}} & f(X) \triangleq \text{tr}(X^\dagger Q X) - \log_2 |I_d + X^\dagger R X| \\
\text{s. t.} & \|X\|_F^2 \leq \zeta
\end{align*}
\tag{26}
\]

where \( Q \succ 0 \) and \( R \succeq 0 \), \( r < n \). Let \( Q \triangleq LL^\dagger \) be the Cholesky factorization of \( Q \), and \( M \triangleq L^{-1} R L^{-1} \), \( M \succeq 0 \), and define the following, \( \{\alpha_i \triangleq \lambda_i |M|\}_{i=1}^r \), \( \Psi \triangleq v_{1:r} |M| \), \( \beta_i \triangleq \Psi_{i}^\dagger (L^\dagger L)^{-1} \Psi_{i} \) (for \( t \)). Then the globally optimal solution for the above problem is given by,

\[
X^* = L^{-1} \Psi \Sigma^*,
\tag{27}
\]

where \( \Sigma^* \) (diagonal) is the optimal power allocation

\[
\Sigma_{i, i}^* = \sqrt{\frac{1}{1 + \mu^* \beta_i - 1/\alpha_i}}, \forall_i
\tag{28}
\]

\( \mu^* \) is the unique root to,

\[
g(\mu) \triangleq \sum_{i=1} L_i (1/(1 + \mu^* \beta_i) - 1/\alpha_i) + \zeta,
\]

on the interval \([-1/(\max \beta_i), \infty]\), and \( g(\mu) \) is monotonically decreasing on that latter.

**Proof:** Refer to Appendix E

With this in mind, we can write the optimal transmit and receive filter updates, as follows,

\[
U_{l_j}^* = L_{l_j}^\dagger \Psi_{l_j} \Sigma_{l_j}^*, \quad \Psi_{l_j} \triangleq v_{1:d} [L_{l_j}^{-1} R_{l_j} L_{l_j}^{-1}]^{\dagger}, \forall l_j,
\]

\[
V_{i_k}^* = K_{i_k}^\dagger \Theta_{i_k} \Lambda_{i_k}^*, \quad \Theta_{i_k} \triangleq v_{1:d} [K_{i_k}^{-1} R_{i_k} K_{i_k}^{-1}]^{\dagger}, \forall i_k,
\tag{29}
\]

where \( \Sigma_{l_j}^* \) and \( \Lambda_{i_k}^* \) are the optimal power allocation, given in Lemma 2. The resulting algorithm, max-DLT, is detailed in Algorithm 2 (where \( T \) is the number of F-B iterations). Moreover, due to the monotone nature of \( g(\mu) \), \( \mu^* \) can be found using simple 1D search methods, such as bisection.

**Remark 2.** We note that \( \Sigma_{l_j}^* \) and \( \Lambda_{i_k}^* \) are both required to ensure the monotonically increasing nature of the updates. And despite the fact that the actual user rate in (4) is invariant to \( \Sigma_{l_j} \), the latter is indeed heavily dependent on the choice of \( \Lambda_{i_k} \).

**Algorithm 2** Maximal DLT (max-DLT)

\[
\begin{array}{l}
\text{for } t = 1, 2, \ldots, T \text{ do} \\
\hspace{1cm} \text{// forward network optimization: receive filter update} \\
\hspace{2cm} \text{Estimate } R_{l_j}, Q_{l_j}, \text{ and compute } L_{l_j}, \forall l_j\, \\
\hspace{1cm} \text{// reverse network optimization: transmit filter update} \\
\hspace{2cm} \text{Estimate } R_{i_k}, Q_{i_k}, \text{ and compute } K_{i_k}, \forall i_k\, \\
\end{array}
\]

end for

C. Connection to other methods

Though the proposed approach seems close to other heuristics such as successive convex programming (SCP) and the convex-concave procedure (CCP), this is misleading. The latter methods start with (non-convex) expressions such as (b.1) (Proposition 3) and approximate \( \log_2 |U_{l_j}^\dagger Q_{l_j} U_{l_j}| \) with a linear function (in the case of CCP (21)), or lower bound it with a quadratic one (in the case of SCP (22)). The approximation is iteratively updated until convergence.

Through our investigation of such methods we found that the associated optimization problems are hard to tackle analytically. Moreover, they result in schemes with higher complexity (due to their iterative nature), require additional information (e.g., Hessian), and do not employ cost functions that decouple at the transmitters/receivers. Thus, despite their widespread effectiveness, such tools are not best suited for the problem at hand.
\[
\{V_{lj}^{(n+1)}\} \triangleq \arg\max_{\{V_{lj}\}} R^{(LB)}_{\Sigma} \left( \{U_{lj}^{(n+1)}\} \right), \quad n = 1, 2, \ldots
\]

\[j_2\]

\[j_1\]

D. Discussions

We provide an intuitive interpretation of the problem in Lemma 3 and its solution. It can be easily verified that \(\{\alpha_i \triangleq \lambda_i [L^{-1} RL^{-1}]\}_{i=1}^{\Omega}\) are also the eigenvalues of \(Q^{-1}R\) (where \(R\) and \(Q\) represent the signal and interference-plus-noise covariance matrix, respectively). Thus, \(\{\alpha_i\}\) acts as a (quasi)-SINR measure, for each of the data streams. Moreover, it can be seen that the optimal power allocated to stream \(i\), \(\Sigma_{(i,i)}\) in (28), tends to zero as \(\alpha_i \to 0\), i.e., no power is allocated to streams that have low-SINR. Moreover, note that \(\{\beta_i\}\) represents the cost of allocating power to each of the streams (the latter becomes clear by rewriting (27) in an equivalent form, as done Appendix E). Thus, the non-homogeneous waterfilling solution in (27) simply allocates power to each of the streams, based on the SINR and cost of each (possibly not allocating power to some streams). In addition, we note that the latter solution reduces to that of the GMRQ problem, (11), when equal power allocation is assumed.

Regarding convergence of the proposed method in (23) for max-DLT, it is established using standard BCD convergence results.

**Proposition 3.** Let \(\psi_n = R^{(LB)}_{\Sigma} (\{U_{lj}^{(n)}\}, \{V_{lj}^{(n)}\})\), \(n = 1, 2, \ldots\) be the sequence of iterates for the objective value. Then, \(\{\psi_n\}\) is non-decreasing in \(n\), and converges to a stationary point of \(R^{(LB)}_{\Sigma} (\{U_{lj}\}, \{V_{lj}\})\)

**Proof:** The proof follows from well-known BCD convergence results such as (23).

V. PRACTICAL ASPECTS

We underline in this section some practical issues that relate to the proposed schemes, such as the mechanism for distributed CSI acquisition, and the resulting communication overhead and computational complexity. Although additional issues such as robustness have to considered as well, such matter are outside the scope of this paper. However, when applying the proposed approach to MIMO IBC and MIMO IFC settings, methods and techniques from [17] are fully applicable.

A. Distributed CSI Acquisition

Evidently, the operation of such schemes is contingent upon each transmitter / receiver being able to estimate the signal and the interference-plus-noise covariance matrices, in a fully distributed manner. From the perspective of this work, this is accomplished via the use of precoded pilots to estimate the effective channels. The methods developed in [17] are fully applicable, and we thus summarize the basic underlying structure (focusing on uplink transmission for a cellular system). First, in the uplink phase, the signal covariance matrix for receiver \(l_j\), \(R_{lj}\), can be computed after estimating the effective signal channel \(H_{lj} V_{lj}\), and the interference-plus-noise covariance matrix is computed after estimating the effective interfering channels, \(\{H_{lj} V_{lj} (i,k)\}_{(i, j) \neq (l, j)}\). The receive filters at the base stations are updated following any of the proposed algorithms (summarized in Fig. 3). Then, in the downlink phase, the same procedure is used to estimate the signal and interference-plus-noise covariance matrices, and update the filters at the receivers. This aforementioned process constitutes one forward-backward (F-B) iteration. Recall that \(T\) is the total number of such iterations that are carried out.

B. Communication Overhead

Thus, for such schemes to be fully distributed (i.e. requiring only local CSI), the required CSI quantities have to be obtained via uplink-downlink pilots and training. As we can see, each F-B iteration has an associated communication overhead, namely the cost of bi-directional transmission of pilots. We adopt a simplistic definition of the communication overhead, as the number of (minimal orthogonal) pilots symbols needed for estimating the required CSI quantities. We note that almost all prior algorithms that have been proposed for that purpose, focus on a regime with a high enough number of F-B iterations (\(T = 100 \sim 1000\)). On the contrary, and in line with earlier attempts such as [9]–[11], we assume that this modus operandi is not feasible in a cellular network (since F-B iterations are carried out over-the-air, and the associated overhead would be higher than the potential gains). We focus on a regime where \(T = 2 \sim 5\). In addition, we assume that the minimal number of orthogonal pilots is used, i.e., \(d\) orthogonal pilot slots for each uplink/downlink effective channel. Moreover, the pilots are orthogonal across users and cells, resulting in a total of \(KLd\) orthogonal pilots for each uplink/downlink training phase. Consequently, the total overhead of both algorithms is

\[\Omega = T (KLd + Kd) = 2T Kd, \text{ channel uses} \quad (30)\]

As mentioned earlier, the non-homogeneous waterfilling solution clearly shows that streams that have low SINR are turned-off, and power is only allocated to the ones that exhibit

3 Although the optimal power allocation to stream \(i\) is zero for some streams, i.e., \(\Sigma_{(i,i)} = 0\), in the actual implementation of the algorithm, \(\Sigma_{(i,i)} = \delta\) where \(\delta \ll 1\).
relatively high SINR. This greatly speeds up the convergence of max-DLT, and allows it to achieve its required performance, with that limited number of F-B iterations.

C. Complexity

By noticing that operations such as matrix multiplication and bisection search are quite negligible compared to other operations, both AIMS and max-DLT have similar computational complexity: the latter is dominated by the Cholesky Decomposition of the interference covariance matrix, $Q = LL^H \in \mathbb{C}^{N \times N}$, and the Eigenvalue Decomposition of $M = L^{-1}RL^{-1} \in \mathbb{C}^{N \times N}$, both of which have similar complexity of $O(N^3)$. Moreover, we recall that the complexity of max-SINR is dominated by matrix inversion of $Q \geq 0$, whose complexity can be approximated by $O(N^3)$.

D. Comparison

A few remarks are in order at this stage, regarding the similarities and differences between AIMS and max-DLT. Referring to the optimal update equations for each algorithm, i.e., (13) and (29), we clearly see that both span the same subspace, i.e. the generalized eigenspace between the signal and interference-plus-noise covariance matrices. In addition, max-DLT computes the optimal power allocation for each stream. Despite this significant similarity among the two solutions, recall that they are derived from two fundamentally different problems. While (13) is a heuristic (an extension of max-SINR) that greedily maximizes the separability at each BS and user, the updates in (29) maximize a lower bound on the sum-rate capacity (and are shown to converge to a stationary point of the latter bound). That being said, their performance evaluation is done via numerical results.

VI. NUMERICAL RESULTS

A. Simulation Methodology

We use the achievable sum-rate in the network as the performance metric, where the achievable user rate is given by (3). Because the approach here is presented in the context of MIMO IMAC, a significant fraction of the results will be under the latter. As mentioned earlier, we will also investigate the proposed approach in alternate scenarios. We specialize our results to some MIMO IFC scenarios (a special case of the MIMO IMAC by setting $K = 1$), where interference alignment has been shown to be feasible (24). We also investigate the MIMO IBC setup, since the proposed algorithms are equally applicable to that case, with little-to-no modification.

We benchmark our algorithms against widely adopted ones:

- max-SINR (4) in the MIMO IMAC, MIMO IFC and MIMO IBC
- (Weighted)-MMSE (6), (12) in the MIMO IFC and MIMO IBC.
- IWU (11) in the MIMO IFC, due to its fast converging nature.

With the exception of max-SINR, no algorithm that we know of is able to operate in all three scenarios (hence the need for different algorithms for each scenario).

In this work, we assume a block-fading channel model with static users, where channel coefficients are drawn from independent and identically distributed complex Gaussian random variables, with zero mean and unit variance, for the sake of simplicity. We note at this point that we applied our approaches to a much more realistic 5G setup. Since a description of the resulting simulation methodology is rather lengthy, we refer the interested readers to [23][Sect. 3.3.3]. In addition, we limit the number of F-B iterations, $T$, to a small number. We further assume that both the signal and interference-plus-noise covariance matrices are perfectly estimated at the transmitter / receiver, i.e., we do not model channel estimation errors. Finally, we note that all curves are averaged over 500 channel realizations.

B. Results

1) Feasible MIMO IFC: We start with a feasible MIMO IFC scenario, by setting $M = N = 4, d = 2$ and fixing the number of F-B iterations, $T = 4$, for all schemes. We evaluate the sum-rate of both our algorithms against other well-known algorithms such max-SINR (4), MMSE (6), the rank-reducing algorithm (IWU-RR) earlier proposed in [11] (since such algorithms are designed for scenarios where fast convergence is desired). We also included Weighted-MMSE with the corresponding number of F-B iterations ($T = 4$), and a large enough number of iterations (as an upper bound). It is clear from Fig. 4 that while max-DLT has similar performance as W-MMSE (for $T = 4$) in the low-to-medium SNR regime, the gap increases in the high-SNR region (SNR $\geq 20$ dB). Moreover, we note that our proposed schemes, outperform all the benchmarks, across all SNR regimes. In particular, the performance gap between max-DLT and the benchmarks, is quite significant in the medium-to-high SNR region. Moreover, despite the fact that only the rank-reducing scheme and max-DLT are able to achieve some degrees-of-freedom gain, max-DLT offers a 35% gain over the rank reducing
scheme. Though max-DLT and IWU-RR are able to turn off some streams in view of reducing interference, the significant performance gap is due to the fact that max-DLT also optimizes the signal subspace as well. Finally, we note that the high-SNR performance of max-DLT is indicative of the fact it is able to achieve some to degrees-of-freedom, while the others algorithms seem to have a significant amount of residual interference (i.e. no degrees-of-freedom gain).

2) Proper MIMO IMAC: We next evaluate the convergence behavior of our proposed algorithms, for a large setting with \( M = N = 10, L = 3, K = 3, d = 2 \) (fixing \( T = 4 \)), i.e., a MIMO IMAC setting that is known to be proper. Due to lack of algorithms that are specially tailored to such settings, we use max-SINR as a benchmark. While we clearly see that the proposed schemes slightly outperform max-SINR at 10dB (Fig. 3), the performance gap (between both our proposed schemes, and max-SINR) becomes extremely more significant at 20dB. Note that we observe this particular behavior, where the performance gap between both our schemes and max-SINR increases sharply, in all the numerical results. As discussed earlier in Sect. III-B, this is due to the fact that unlike max-SINR, our schemes offer significantly better performance by allocating different powers, to different streams. Similarly to the previous results, max-DLT largely outperforms all the other schemes. Moreover, though both AIMS and max-DLT show quite a fast convergence, in the case of max-DLT it is impressive: 90% of its final performance after 2 iterations. This is due the stream-control feature of non-homogeneous waterfilling (as discussed in Sect. [V]), and the fact that its effectiveness is increased in severely interference-limited scenarios (such as this one).

3) Improper scenarios: Finally, we test the resilience of the proposed algorithms to operating in improper scenarios. It is clear from Fig. 4 that the latter exhibit the same performance characteristics as what has been observed and discussed thus far. In addition, while max-SINR and AIMS seem to suffer from a collapse in their achievable degrees-of-freedom, max-DLT achieves some (suboptimal) degrees-of-freedom gain.

4) A larger, denser setting: We investigate a “dense” system, with \( L = 3, K = 3, M = 5, N = 5 \) (fixing \( T = 4 \)), and evaluate the robustness of our algorithms to operating in an improper system. Therefore, Fig. 5 shows the sum-rate performance of such a system, for \( d = 1 \) (proper system) and \( d = 2 \) (improper system). The benefits of rank adaptation for AIMS\(^4\) can be clearly seen here: AIMS w/ RA significantly outperforms all other approaches, due to its ability to select the transmit/receive filter rank. We note as well that the performance of max-DLT is similar for \( d = 1 \) and \( d = 2 \), thus exhibiting robustness to operating in improper scenarios.

The above observation is due to the fact that the non-homogeneous waterfilling solution in max-DLT can freely allocate different powers to different directions of the subspace spanned by the transmit / receive filter (possibly turning off some directions, when they suffer from significant interference). As a result, it transforms an improper system (\( d = 2 \)),

\(^4\)As described in Sect. III-C rank adaption is performed at each iteration of AIMS, by taking the minimum rank of each pair of transmit/receive filters.
into a (virtually) proper one by turning off one of the streams, for each user (for this particular case). We attempt to capture the latter effect for the same scenario as the one shown in Fig. 7 setting $d = 2$. We simulate the average value of the smallest singular value of the transmit filter (across all users),

$$
\gamma(\text{SNR}) \triangleq \mathbb{E}\left\{ \frac{1}{|I|} \sum_{l_j \in I} \sigma_d^2[V_{l_j}] \right\}
$$

for several SNR values. As we can see from Table I max-DLT is able to arbitrarily reduce the smallest singular value of the transmit filters, especially in the very high SNR (interference-limited) regime: this is a critical, since reducing interference is vital to increasing the sum-rate. Note that this adaptation is clearly not present in the case for AIMS and max-SINR.

5) Proper MIMO IBC: As mentioned earlier, our schemes are equally applicable to MIMO IBCs. For that matter, we investigate their performance in a proper MIMO IBC setup with $8 \times 2$ (single-stream) MIMO links, with $L = 3$, $K = 3$ (for several SNR values). We benchmark our results against the well-known Weighted-MMSE (WMMSE) algorithm [12].

Note that for the latter, we keep the sum-power constraint that is employed by WMMSE, and adjust the per-user transmit power constraint $P_t$, for our algorithms, assuming equal power allocation among users. This implies that a more stringent constraint is placed on our schemes. Despite this unfavorable setup, as Fig. 8 shows, both our schemes significantly outperform the benchmark, in the low-overhead regime (i.e., for small $T$). We reiterate the fact that this is the regime of interest in this work. Needless to say, the full-performance that WMMSE is expected to deliver, is reached after more iterations are performed.

### VII. Conclusion

In this work, we attempted to shed light on the sum-rate maximization problem in multi-user multi-cellular settings, and established its inherent connection to separability between each user’s signal and interference-plus-noise subspaces. We thus proposed an algorithm that alternately maximizes the latter metric for both the uplink and downlink, and established the fact that the latter is a generalization of the well-known max-SINR. Moreover, we advocated the use of DLT bounds and highlighted their significant advantage in yielding convex optimization problems that decouple at both the transmitters and receivers. We provided a generic solution for the latter, the so-called non-homogeneous waterfilling (underlining its built-in stream-control feature), and proposed another distributed algorithm, max-DLT, that solves the latter problem in a distributed manner. We later verified through simulations that our proposed algorithms massively outperform other relevant benchmark algorithms (especially in interference-limited multi-user environments), for several communication scenarios.

### VIII. Acknowledgment

The first author wishes to acknowledge Dr. Rasmus Brandt for the many discussions leading up to this work, and later comments on the manuscript.

### Appendix

#### A. Uniqueness of SVD

The proofs in this work rely on the central premise of mapping the problem into a series of equivalent forms (where equivalence is ensured by the uniqueness of each mapping [26]. One of the steps is rewrite the problem by mapping the variable into its SVD form: for example, let $\max_{\mathbf{X}} f(\mathbf{X}) = \text{tr}(\mathbf{X}^* \mathbf{Q} \mathbf{X})$ and let $\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*$ be the SVD of...
Generally speaking, the SVD is unique only up to rotations of the left and right singular vectors, i.e., the actual SVD of $X$ takes this form, $X = (U\Theta)\Sigma(T^V V^T)$ where $\Theta$ is diagonal with phase elements. Due to the quadratic nature of $f(X)$ it is easy to verify that the ambiguity brought by $\Theta$ is lifted, i.e., it is easy to verify that $f(U\Sigma V^T) = f((U\Theta)\Sigma(T^V V^T)).$ Note that the same arguments holds for all the objective functions that we use in this work.

### B. Proof of Lemma 1 and Corollary 1

**Lemma:**
Let $Q = LL^T$ be the Cholesky decomposition of $Q$, where $L$ is a lower triangular matrix and non-singular (since $Q$ is full-rank almost surely), and $Y = L^TY \Rightarrow Y = L^{-1}Y$. Then, (10) can be written as

$$Y = \arg\max_{Y} q(Y) = \frac{|Y^T L^{-1} RL^{-1} Y|}{|Y^T L^{-1} QL^{-1} Y|} = \frac{|Y^T M Y|}{|Y^T Y|}, \quad (31)$$

where we defined $M = L^{-1}RL^{-1}$. Note that the equivalence between (10) and (31) follows from the fact that $L$ is non-singular by construction. Let $Y = TSv^T$ be the SVD of $Y$ ($T \in \mathbb{C}^{n \times r}, S \in \mathbb{R}^{r \times r}$). Plugging this in $q(Y)$ we write,

$$q(T) = \frac{|VS^T M T S V^T|}{|VS^T T S V^T|} = \frac{\|T^*MT\|}{\|T^*T\|}, \quad (32)$$

where we used the fact that the determinant is unitary invariant. Thus, (31) is equivalent to,

$$\arg\max_{T} \|T^*MT\| \Leftrightarrow T^*T = I_d \quad (32)$$

The equivalence of (31) and (32) is established in Appendix A. Since $M \succeq 0$ ($M$ is Hermitian, and has only non-negative eigenvalues), the proof of the above problem is well known determinant maximization problem:

$$v_{1:d}(M) \triangleq \Psi, \quad \text{where } M\Psi = \Psi \Lambda_r \quad (33)$$

As a result, the solution to (10) is given by

$$X^* = L^{-1}\Psi S v^T$$

where $S$ is an arbitrary diagonal matrix, $V$ is an arbitrary unitary matrix.

Moreover, recall that $q(X^*S) = q(X^*)$, for any $S \in \mathbb{C}^{r \times r}$ that is non-singular. Then, $X^* = L^{-1}\Psi S v^T$

### C. Sketch of proof for Proposition 1

We need to show that $(U_i^* U_i^*) = \alpha I_d$ happens with probability zero.

**Note:**
Note that due to the generic i.i.d. nature of the channel matrices, the eigenvalues of $Q_{ij}$ can be assumed to be distinct (and $Q_{ij}$ is full rank), almost surely. Then, it can be verified that the same holds for $L_i^1, L_i^1$ and $(L_i^1 L_i^1)^{-1}$. Then, $(L_i^1 L_i^1)^{-1} \approx \alpha I_N$ happens with probability zero. Recalling that $\Psi_i^T \Psi_i = I_d$, then the following equivalent statements happen with probability zero,

$$\Psi_i^T (L_i^1 L_i^1)^{-1} \Psi_i \approx \Psi_i (\alpha I_N) \Psi_i$$

$$\Leftrightarrow (U_i^* U_i^*) (L_i^1) \Psi_i \approx \alpha I_d$$

$$\Leftrightarrow (U_i^* U_i^*) \approx \alpha I_d$$

### D. Proof of Proposition 1

We start by lower bounding the user rate in (9), as

$$r_{ij} \geq \log_2 |(U_i^T Q_{ij} U_i)^{-1} + (U_i^T R_i U_i) (U_i^T Q_{ij} U_i)^{-1}|$$

$$\geq \log_2 (|I_d + U_i^T R_i U_i| (U_i^T Q_{ij} U_i)^{-1})$$

$$\geq \log_2 (|I_d + U_i^T R_i U_i| - \log_2 (U_i^T Q_{ij} U_i |)$$

$$\geq \log_2 (|I_d + U_i^T R_i U_i| - \text{tr} (U_i^T Q_{ij} U_i) \Delta_i) \triangleq r_{ij}^{(LB)}$$

Note that the last inequality follows from combining (17) and the monotonically increasing nature of log $|X|$ in (9). Moreover the last one follows from using $\log |A| \leq \text{tr} (A)$ for $A \succeq 0$.

We rewrite $r_{ij}$ in (8) as,

$$r_{ij} = \log_2 |(U_i^T R_i U_i)(U_i^T Q_{ij} U_i)^{-1}|$$

$$\geq \log_2 (|I_d + U_i^T R_i U_i| (U_i^T Q_{ij} U_i)^{-1})$$

$$\geq \log_2 (|I_d + U_i^T R_i U_i| - \text{tr} (U_i^T Q_{ij} U_i) \Delta_i) \triangleq r_{ij}^{(LB)}$$

Thus, $r_{ij}$ is approximated by $\log_2 (|U_i^T R_i U_i| - \log_2 (|U_i^T Q_{ij} U_i|), \quad (34)$

(where the error is given in the above equation). Plugging this result in $\Delta_i$ yields,

$$\Delta_i = \log_2 |(U_i^T R_i U_i)(U_i^T Q_{ij} U_i)^{-1}) - \log_2 (|U_i^T Q_{ij} U_i|)$$

$$\geq \log_2 (|I_d + U_i^T R_i U_i| - \text{tr} (U_i^T Q_{ij} U_i) \Delta_i)$$

$$\geq \Delta_i \approx \text{tr} (U_i^T Q_{ij} U_i) (U_i^T R_i U_i)^{-1})$$

Referring to the above, in the interference-limited regime (17), the first and third terms become negligible w.r.t. the second and fourth. Consequently,

$$\Delta_i = \log_2 |(U_i^T Q_{ij} U_i) - \log_2 (|U_i^T Q_{ij} U_i|)$$

$$\geq \log_2 (|I_d + U_i^T R_i U_i| - \text{tr} (U_i^T Q_{ij} U_i) \Delta_i)$$

$$\geq \Delta_i \approx \text{tr} (U_i^T Q_{ij} U_i) (U_i^T R_i U_i)^{-1})$$

### E. Proof of Lemma 2

Despite the convex nature of the problem, solving the problem using standard Lagrangian techniques on (26) yields complicated expressions. Thus, we rewrite the problem into a
series of equivalent forms. Letting $Z = L^T X \Leftrightarrow X = L^{-\dagger} Z$, then (26) is equivalent to,

$$
\begin{align*}
(1) & \quad \min_{Z} f(Z) \triangleq \text{tr}(Z^\dagger Z) - \log_2 |I_d + Z^\dagger M Z| \\
& \quad \text{s. t. } \text{tr}(Z^\dagger A Z) \leq \zeta
\end{align*}
$$

where $A = (L^T L)^{-1}$. Letting $Z = T \Sigma^{\dagger}$ be the SVD of $Z (T \in \mathbb{C}^{n \times r}, \Sigma \in \mathbb{R}^{r \times r})$, we rewrite (1) as an equivalent form (refer to [Appendix A]).

The latter problem is convex and separable in $T$ and $\Sigma$. Thus fixing $\Sigma$, the optimal $T$ can be easily obtained from Hadamard’s Inequality, applied to the determinant maximization problem, i.e., $T^\star \triangleq v_{1,1} [M] = \Psi$. Then, (2) can be written in the following equivalent form,

$$
\begin{align*}
\min_{\{x\}} & \quad \sum_{i=1}^{r} \left( \sigma_i^2 - \log_2 (1 + \alpha_i \sigma_i^2) \right) \\
& \quad \text{s. t. } \sum_{i=1}^{r} \beta_i \sigma_i^2 \leq \zeta
\end{align*}
$$

Let $x_1 = \sigma_1^2$. It can be verified that the inequality constraint is always satisfied with equality, and we can rewrite the problem as,

$$
\begin{align*}
(3) & \quad \min_{\{x_i\}} \sum_{i=1}^{r} \left( x_i - \log_2 (x_i + \frac{1}{\alpha_i}) \right) \\
& \quad \text{s. t. } \sum_{i=1}^{r} \beta_i x_i = \zeta, \quad x_i \geq 0, \forall i
\end{align*}
$$

(3) is a generalization of the well-known waterfilling problem: in fact, (3) reduces to latter problem, if $\beta_i = 1, \forall i$, and by dropping the first term in the objective. We start by writing the associated KKT conditions.

\begin{align*}
1 - (x_i + \alpha_i^{-1})^{-1} + \mu \beta_i - \lambda_i = 0, \forall i \\
\sum_i \beta_i x_i = \zeta, \quad x_i \geq 0 \\
\lambda_i x_i = 0, \quad \lambda_i \geq 0, \quad \mu \neq 0, \forall i
\end{align*}

Firstly, note that $\lambda_i$ act as slack variables and can thus easily be eliminated. Then, considering two cases, $\lambda_i = 0, \forall i$ or $\lambda_i > 0, \forall i$, the optimal solution can be found in a straightforward manner,

$$
x_i^\star = \begin{cases} 
(1 + \mu \beta_i)^{-1} - \alpha_i^{-1}, & \text{if } \mu < (\alpha_i - 1)/\beta_i \\
0, & \text{if } \mu > (\alpha_i - 1)/\beta_i \\
\frac{1}{1/(1 + \mu \beta_i) - 1/\alpha_i}, & \forall i
\end{cases}
$$

where $\mu^*$ is the unique root to

$$
g(\mu) \triangleq \sum_{i=1}^{r} \beta_i \left( \frac{1}{1 + \mu \beta_i} - 1/\alpha_i \right) - \zeta
$$

Note that $g(\mu)$ is monotonically decreasing, for $\mu > -1/(\max_i \beta_i)$, and $\mu^*$ can be found using standard 1D search methods, such as bisection.

Thus, the optimal solution for (1) is $Z^\star = \Psi \Sigma^\star$, of which $\Sigma^\star = \sqrt{x_i^\star}, \forall i$, and of that (26) is $X^\star = L^{-\dagger} \Psi \Sigma^\star$.

REFERENCES

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” IEEE Journal on Selected Areas in Communications, vol. 28, no. 9, pp. 1380–1408, 2010.

[2] M. Maddah-Ali, A. Motahari, and A. Khandani, “Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis,” IEEE Transactions on Information Theory, vol. 54, no. 7, pp. 3457–3470, Aug. 2008.

[3] V. Cadambe and S. Jafar, “Interference alignment and degrees of freedom of the K-user interference channel,” IEEE Transactions on Information Theory, vol. 54, no. 8, pp. 3425–3441, 2008.

[4] K. Gomadam, V. R. Cadambe, and S. A. Jafar, “A distributed numerical approach to interference alignment and applications to wireless interference networks,” IEEE Transactions on Information Theory, vol. 57, pp. 3309–3322, June 2011.

[5] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, “Minimum mean squared error interference alignment,” in 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pp. 1106–1110, Nov. 2009.

[6] S. W. Peters and R. W. Heath, “Cooperative algorithms for MIMO interference channels,” IEEE Transactions on Vehicular Technology, vol. 60, pp. 206–218, Jan. 2011.

[7] I. Santamaria, O. Gonzalez, R. W. Heath, and S. W. Peters, “Maximum sum-rate interference alignment algorithms for MIMO channels,” in 2010 IEEE Global Communications Conference (GLOBECOM 2010), pp. 1–6, IEEE, Dec. 2010.

[8] M. Bengtsson, “A pragmatic approach to multi-user spatial multiplexing,” in Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002, pp. 130–134, Aug 2002.

[9] P. Komulainen, A. Tölö, and M. Juntti, “Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems,” IEEE Transactions on Signal Processing, vol. 61, pp. 2204–2218, May 2013.

[10] R. Brandt and M. Bengtsson, “Fast-convergent distributed coordinated precoding for TDD multicell MIMO systems,” in Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on, pp. 457–460, Dec 2015.

[11] H. Ghaush, T. Kim, M. Bengtsson, and M. Skoglund, “Distributed low-overhead schemes for multi-stream MIMO interference channels,” IEEE Transactions on Signal Processing, vol. 63, pp. 1737–1749, April 2015.

[12] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” IEEE Transactions on Signal Processing, vol. 59, no. 9, pp. 4331–4340, 2011.

[13] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J. S. Pang, “Decomposition by partial linearization: Parallel optimization of multi-agent systems,” IEEE Transactions on Signal Processing, vol. 62, pp. 641–656, Feb 2014.

[14] T. Kim, D. Love, B. Clerckx, and D. Hwang, “Spatial degrees of freedom of the multicell MIMO multiple access channel,” in Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE, pp. 1–5, Dec 2011.

[15] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[16] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, “Trace ratio vs. ratio trace for dimensionality reduction,” in 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, June 2007.

[17] R. Brandt and M. Bengtsson, “Distributed CSI acquisition and coordinated precoding for TDD multicell MIMO systems,” IEEE Transactions on Vehicular Technology, vol. PP, no. 99, pp. 1–1, 2015.

[18] R. Prato, “A general solution to the maximization of the multidimensional generalized rayleigh quotient used in linear discriminant analysis for signal classification,” in Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP ’03). 2003 IEEE International Conference on, vol. 6, pp. VI–157–60 vol.6, April 2003.

[19] Z. Zhang, W. Yang, and L.-Z. Liao, “A note on the trace quotient problem,” Optimization Letters, vol. 8, no. 5, pp. 1637–1645, 2014.

[20] M. Sadek, A. Tarighat, and A. Sayed, “A leakage-based precoding scheme for downlink multi-user MIMO channels,” IEEE Transactions on Wireless Communications, vol. 6, pp. 1711–1721, May 2007.

[21] L.-H. Zhang, W. Yang, and L.-Z. Liao, “A note on the trace quotient problem,” Optimization Letters, vol. 8, no. 5, pp. 1637–1645, 2014.

[22] L.-H. Zhang, W. Yang, and L.-Z. Liao, “A note on the trace quotient problem,” Optimization Letters, vol. 8, no. 5, pp. 1637–1645, 2014.

[23] K. Schittkowski and C. Zillober, Stochastic Programming: Numerical Techniques and Engineering Applications, ch. Sequential Convex Programming Methods, pp. 123–141. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995.

[24] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization,” SIAM Journal on Optimization, vol. 23, sep 2012.
[24] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, “On feasibility of interference alignment in MIMO interference networks,” IEEE Transactions on Signal Processing, vol. 58, pp. 4771–4782, Sept. 2010.

[25] METIS D3.3, “Final performance results and consolidated view on the most promising multi-node/multi-antenna transmission technologies,” Feb 2015.

[26] S. Boyd and L. Vandenberghe, Convex Optimization. New York, NY, USA: Cambridge University Press, 2004.