Interference Alignment Using Variational Mean Field Annealing

Badiu, Mihai Alin; Guillaud, Maxime; Fleury, Bernard Henri

Published in:
Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on

DOI (link to publication from Publisher):
10.1109/WIOPT.2014.6850351

Publication date:
2014

Document Version
Early version, also known as pre-print

The 3rd Workshop on Physics-Inspired Paradigms in Wireless Communications and Networks
May 16th, 2014, Hammamet, Tunisia (in conjunction with WiOpt 2014)

Interference Alignment Using Variational Mean Field Annealing

Mihai-Alin Badiu*, Maxime Guillaud† and Bernard Henri Fleury*
* Aalborg University, Denmark
† Vienna University of Technology, Austria

***Abstract*—We study the problem of interference alignment in the multiple-input multiple-output interference channel. Aiming at minimizing the interference leakage power relative to the receiver noise level, we use the deterministic annealing approach to solve the optimization problem. In the corresponding probabilistic formulation, the precoders and the orthonormal bases of the desired signal subspaces are variables distributed on the complex Stiefel manifold. To enable analytically tractable computations, we resort to the variational mean field approximation and thus obtain a novel iterative algorithm for interference alignment. We also show that the iterative leakage minimization algorithm by Gomadam *et al.* and the alternating minimization algorithm by Peters and Heath, Jr. are instances of our method. Finally, we assess the performance of the proposed algorithm through computer simulations.**

I. Introduction

Interference alignment (IA), which was introduced in [1] for the MIMO interference channel, has received a lot of attention in recent years since it is a key ingredient in achieving the full degrees-of-freedom of the channel. It consists in making sure that the interference received from multiple interferers *aligns* at each receiver in a linear subspace of limited dimension; the remaining dimensions can be used for interference-free communication.

The existence of an IA solution given the channel dimensions and the rank of the transmitted signals has been studied in [2], and more recently in [3], [4]. The number of such solutions has been studied in [5], for the single-stream case. Despite its deceptively simple mathematical formulation, no general closed-form solution to the IA equations is available (although it exists for certain dimensions, see e.g. [6]). Numerical algorithms based on alternating optimization were introduced in [7], [8]. A distributed version based on message-passing over a graph is proposed in [9]. In [10], the authors introduce a distributed algorithm which achieves a smooth trade-off between the interference-limited regime (where IA is optimal) and the noise-limited regime (where a selfish approach based on the direct channel only is preferable).

In this paper, we formulate the IA problem as an optimization problem whose cost function is the total *weighted* interference leakage and employ the deterministic annealing (DA) approach. DA is an optimization heuristic inspired by the annealing process (hence its name) in physical chemistry which aims at driving a physical system (e.g., a glass or metal) in its lowest energy state by keeping it at thermal equilibrium while slowly decreasing the temperature. Constructed within a probabilistic framework, DA introduces controlled randomness of the solution (quantified by Shannon entropy), which is gradually reduced. DA was shown to be successful in avoiding poor local minima and initialization issues in optimization problems, such as clustering, classification, regression and others [11], [12]. Note that DA differs from the stochastic optimization method of simulated annealing. Although both have the same underlying principle, the latter algorithm relies on sequentially sampling the solution space at random, and the decision to accept a new possible solution is randomly taken based on the cost reduction relative to the current solution; this sampling procedure makes it slower than deterministic techniques. In contrast, DA analytically estimates expected values of the system variables.

To enable analytical computations, we use DA in combination with the variational mean field method [13] and obtain a distributed iterative algorithm that has the algorithms [7], [8] as special cases. The proposed algorithm is numerically evaluated in terms of convergence and achieved sum rate.

Notation: Boldface lowercase and uppercase letters are used to represent vectors and matrices, respectively; the $n \times n$ identity matrix is written as $\mathbf{I}_n$; superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transposition and Hermitian transposition, respectively. The trace of a matrix is denoted by $\mathrm{tr}(\cdot)$; the scalar-valued function of matrix argument $\mathrm{etr}(\cdot)$ stands for $\exp(\mathrm{tr}(\cdot))$. The expectation of a random variables is denoted by $\langle\cdot\rangle$. We denote by $\mathcal{CV}_{k,n}$, $k \leq n$, the complex Stiefel manifold represented by the set $\mathcal{CV}_{k,n} = \left\{\mathbf{X} \in \mathbb{C}^{n\times k} \mid \mathbf{X}^{\mathrm{H}}\mathbf{X} = \mathbf{I}_k\right\}$.

II. Problem statement

We consider the communication over the K-user MIMO interference channel using linear precoding at the transmitters. Transmitter i, with $i \in \mathcal{K} = \{1, \ldots, K\}$, is equipped with M_i antennas and uses the precoding matrix $\mathbf{V}_i \in \mathcal{CV}_{d_i,M_i}$ to encode d_i data streams. The data symbols in $\mathbf{x}_i \in \mathbb{C}^{d_i}$ are i.i.d. zero mean circularly symmetric complex Gaussian random variables, with $\langle \mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}\rangle = \rho_i\,\mathbf{I}_{d_i}$, where ρ_i is the transmit power per stream. The signal $\mathbf{y}_i \in \mathbb{C}^{N_i}$ acquired by the N_i antennas of the ith receiver reads

$$\mathbf{y}_i = \mathbf{H}_{ii}\mathbf{V}_i\mathbf{x}_i + \sum_{j\in\mathcal{K}\setminus i} \mathbf{H}_{ij}\mathbf{V}_j\mathbf{x}_j + \mathbf{z}_i. \tag{1}$$

In (1), $\mathbf{H}_{ij} \in \mathbb{C}^{N_i\times M_j}$ is the matrix corresponding to the static, flat-fading channel between transmitter j and receiver i, while the noise vector $\mathbf{z}_i \in \mathbb{C}^{N_i}$ is zero mean complex

Gaussian with covariance $\langle \mathbf{z}_i \mathbf{z}_i^{\mathrm{H}} \rangle = \gamma_i^{-1} \mathbf{I}_{N_i}$, where γ_i is the precision (inverse variance).

Our goal is to design the precoding matrices $\mathbf{V}_i$, $i \in \mathcal{K}$, so that interference alignment is achieved. That is, for each i, the interfering signals at receiver i should lie in a subspace of $\mathbb{C}^{N_i}$ whose dimension is at most $N_i - d_i$, such that the d_i-dimensional orthogonal complement is interference-free. The conditions for interference alignment can be written as

$$\operatorname{rank}\left([\{\mathbf{H}_{ij}\mathbf{V}_j\}_{j\neq i}]\right) \leq N_i - d_i, \quad \forall i \in \mathcal{K}, \tag{2}$$

where the N_i-by-$\sum_{j\neq i} d_j$ matrix $[\{\mathbf{H}_{ij}\mathbf{V}_j\}_{j\neq i}]$ is obtained through horizontal concatenation.

III. Design Objective

Similar to [7], [8], our general design principle is to minimize the power of the interference leaking in the desired signal subspace at each receiver. Unlike those works, in the cost function that we define the interference leakage at each receiver is weighted by the receiver noise precision.

Let us define the matrix $\mathbf{W}_i \in \mathcal{CV}_{d_i,N_i}$ whose columns form an orthonormal basis for the subspace where receiver i expects its desired signal to lie. Note that $\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}}$ is the projector onto the desired signal subspace. Based on (1), the leaked interference signal at receiver i is

$$\mathbf{l}_i \triangleq \mathbf{W}_i\mathbf{W}_i^{\mathrm{H}} \sum_{j\in\mathcal{K}\setminus i} \mathbf{H}_{ij}\mathbf{V}_j\mathbf{x}_j.$$

Using the statistical assumptions about the data symbols, we obtain the average power of the leaked interference

$$\begin{aligned} L_i(\mathbf{V}_{\sim i}, \mathbf{W}_i) &= \langle \operatorname{tr}\left(\mathbf{l}_i\mathbf{l}_i^{\mathrm{H}}\right)\rangle \\ &= \sum_{j\in\mathcal{K}\setminus i} \rho_j \operatorname{tr}\left(\mathbf{V}_j^{\mathrm{H}}\mathbf{H}_{ij}^{\mathrm{H}}\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}}\mathbf{H}_{ij}\mathbf{V}_j\right) \end{aligned}$$

where $\mathbf{V}_{\sim i}$ represents all precoders other than $\mathbf{V}_i$. The power of the noise that lies in the signal subspace at receiver i is $\langle \operatorname{tr}\left(\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}}\mathbf{z}_i\mathbf{z}_i^{\mathrm{H}}\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}}\right)\rangle = d_i\gamma_i^{-1}$.

We define the cost function depending on the configuration $\mathbf{\Omega} \triangleq \{\mathbf{V}_1, \ldots, \mathbf{V}_K, \mathbf{W}_1, \ldots, \mathbf{W}_K\}$ as the total *weighted* interference leakage

$$\begin{aligned} E(\mathbf{\Omega}) &= \sum_{i=1}^{K} \kappa_i\, L_i(\mathbf{V}_{\sim i}, \mathbf{W}_i) \\ &= \sum_{i\in\mathcal{K}} \kappa_i \sum_{j\in\mathcal{K}\setminus i} \rho_j \operatorname{tr}\left(\mathbf{V}_j^{\mathrm{H}}\mathbf{H}_{ij}^{\mathrm{H}}\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}}\mathbf{H}_{ij}\mathbf{V}_j\right). \end{aligned} \tag{3}$$

In (3), the normalized weights κ_i, $i \in \mathcal{K}$, are proportional to the precisions of the noise lying in the desired signal subspaces at the corresponding receivers, i.e.,

$$\kappa_i = \frac{\gamma_i/d_i}{\sum_{j\in\mathcal{K}} \gamma_j/d_j}.$$

The motivation behind employing such weighting is to capture in the cost function (3) the fact that interference leakages of the same magnitude at receivers with different noise powers have different impacts on their individual performance.

We aim at finding the configuration $\mathbf{\Omega}_*$ that minimizes (3):

$$\mathbf{\Omega}_* = \arg\min_{\Omega} E(\mathbf{\Omega}) \tag{4}$$

Being a power, $E(\mathbf{\Omega}) \geq 0$; when $E(\mathbf{\Omega}_*) = 0$, the obtained precoders satisfy the K conditions (2), meaning that IA is feasible. Note that the solution to (4) is not unique.

IV. Proposed Method

In this section, we use the deterministic annealing approach to solve the optimization problem (4). To obtain tractable computations, we employ the variational mean field method and thus obtain a distributed, iterative algorithm for computing IA solutions. We also show how the iterative leakage minimization algorithm [7] and the alternating minimization algorithm [8] can be instantiated from our approach.

A. Principles of Deterministic Annealing

In DA, the admissible (candidate) solutions are governed by a pdf $p(\mathbf{\Omega})$ of the configurations. Noting that $E(\mathbf{\Omega})$ represents the cost of operating over the interference channel with the precoders and bases in $\mathbf{\Omega}$, we define the expected cost

$$U(p) = \langle E(\mathbf{\Omega})\rangle_p$$

The expected cost is minimized with respect to $p(\mathbf{\Omega})$ subject to a constraint on the level of randomness of the admissible solutions, quantified by the Shannon entropy $S(p) = -\langle \log p(\mathbf{\Omega})\rangle_p$. Therefore, the original optimization problem (4) is restated as that of minimizing the extended objective function

$$F(p) = U(p) - T\,S(p), \tag{5}$$

in which entropy acts as a penalty and the positive parameter T controls the tradeoff between minimizing the expected cost and entropy maximization. When T is very large, we basically maximize the entropy, while as T approaches zero we fall back to solving the original problem (4) to obtain a hard (deterministic) solution.[1] For a fixed value of T, minimizing F with respect to p gives the pdf

$$p_{\mathrm{G}}(\mathbf{\Omega}) = \frac{1}{Z}\exp\left(-\frac{E(\mathbf{\Omega})}{T}\right) \tag{6}$$

where Z is the normalizing constant.

An analogy to statistical physics is in order. Our $2K$ matrix variables, namely the precoders $\mathbf{V}_i$ and orthonormal bases $\mathbf{W}_i$, $i \in \mathcal{K}$, characterize a physical system in state $\mathbf{\Omega}$. The internal energy of the system is our expected cost U, while its entropy is given by S. The objective function (5) is merely the Helmholtz free energy and the parameter T is therefore the system temperature controlling the level of randomness. According to the fundamental principle of minimal free energy, the system achieves the minimum of F at thermal equilibrium, at which point it is governed by the Gibbs pdf p_{G} in (6).

The Gibbs pdf (6) is parameterized by the temperature. High values of T smooth the pdf – at very high values it

[1] Indeed, when $T = 0$, direct minimization of $U(p)$ gives the Dirac δ function $\delta(\mathbf{\Omega} - \mathbf{\Omega}_*)$ with $\mathbf{\Omega}_*$ in (4).

approaches the uniform pdf, in which case any configuration $\mathbf{\Omega}$ is equally good. As T is lowered, more and more "structure" (complexity) of $E(\mathbf{\Omega})$ is revealed. When T approaches zero, the pdf becomes highly peaked about the solution $\mathbf{\Omega}_*$.

For a given temperature, the main idea of DA is to analytically approximate the expectations of the optimization variables at thermal equilibrium. Starting at high temperature, DA tracks the evolution of the expectations by gradually reducing the temperature. However, in most cases it is not possible to analytically compute expectations with respect to the Gibbs pdf p_{G}. To enable analytical computations, the extended objective function (5) is minimized over a restricted class of pdfs (one that allows for tractable computations).

B. Distributed Iterative Algorithm for Interference Alignment

We employ the mean field approximation to obtain analytically tractable computations. That is, for a given T, we consider the minimization of the objective function (5) over the class of fully-factorized pdfs of the form

$$q(\mathbf{\Omega}) = \prod_{i=1}^{K} q(\mathbf{W}_i) \prod_{j=1}^{K} q(\mathbf{V}_j) \tag{7}$$

where each factor—called the belief of the respective variable—is defined on the corresponding complex Stiefel manifold. It can be shown that minimizing the free energy over the restricted class is equivalent to finding that member of the class with minimum Kullback-Leibler divergence from the Gibbs pdf p_{G}. The minimization is performed with respect to each belief in turn while keeping the other beliefs fixed. The optimal belief of a certain variable $v \in \mathbf{\Omega}$ is [13]

$$q(v) \propto \exp\left(-\frac{1}{T}\langle E(\mathbf{\Omega})\rangle_{q(\sim v)}\right) \tag{8}$$

where $\langle\cdot\rangle_{q(\sim v)}$ denotes the expectation with respect to the product of the beliefs of all variables in $\mathbf{\Omega}$ other than v. The beliefs are updated sequentially and the resulting iterations are guaranteed to converge. Note that such iterative computations are performed for each temperature value. In this way, the annealing process becomes the outer loop of the overall optimization algorithm.

In the following we compute the expressions of the beliefs of the precoders and orthonormal bases. To fix some notations, we denote the second moments of these matrices by $\mathbf{Q}_j = \langle \mathbf{V}_j \mathbf{V}_j^{\mathrm{H}} \rangle_{q(\mathbf{V}_j)}$ and $\mathbf{R}_i \triangleq \langle \mathbf{W}_i \mathbf{W}_i^{\mathrm{H}} \rangle_{q(\mathbf{W}_i)}$, $i, j \in \mathcal{K}$. Plugging (3) in (8), we obtain

$$\begin{aligned} q(\mathbf{V}_j) &\propto \operatorname{etr}\left(-\frac{\rho_j}{T}\mathbf{V}_j^{\mathrm{H}} \sum_{i\neq j} \kappa_i \mathbf{H}_{ij}^{\mathrm{H}} \langle \mathbf{W}_i \mathbf{W}_i^{\mathrm{H}} \rangle_{q(\mathbf{W}_i)} \mathbf{H}_{ij} \mathbf{V}_j\right) \\ &= \frac{1}{c(\mathbf{S}_j)} \operatorname{etr}\left[\mathbf{V}_j^{\mathrm{H}} \mathbf{S}_j \mathbf{V}_j\right] \end{aligned} \tag{9}$$

where $\mathbf{S}_j = -\frac{\rho_j}{T} \sum_{i\neq j} \kappa_i \mathbf{H}_{ij}^{\mathrm{H}} \mathbf{R}_i \mathbf{H}_{ij}$ and $c(\mathbf{S}_j)$ is the normalizing constant determined by $\mathbf{S}_j$. Similarly, we obtain

$$q(\mathbf{W}_i) = \frac{1}{c(\mathbf{T}_i)} \operatorname{etr}\left(\mathbf{W}_i^{\mathrm{H}} \mathbf{T}_i \mathbf{W}_i\right) \tag{10}$$

where $\mathbf{T}_i = -\frac{\kappa_i}{T} \sum_{j\neq i} \rho_j \mathbf{H}_{ij} \mathbf{Q}_j \mathbf{H}_{ij}^{\mathrm{H}}$ and $c(\mathbf{T}_i)$ is the normalizing constant.

Observe that (9) and (10) are pdfs of a distribution on the complex Stiefel manifold. The form of the pdfs resembles that of the matrix Bingham distribution on the real Stiefel manifold [14]. However, to the best of our knowledge, we are not aware of any work that extends the matrix Bingham distribution to the complex Stiefel manifold. Therefore, we refer to the distribution as the complex matrix Bingham distribution and compute the normalizing constant and second moment of its pdf in the Appendix.

According to our results in the Appendix, the normalizing constant (15) of the complex matrix Bingham pdf is determined by the eigenvalues of its matrix parameter, while its second order moment (19) is given by both the eigenvalues and the eigenvectors of the parameter. Thus, the second moments $\mathbf{Q}_j$ and $\mathbf{R}_i$ are computed based on the eigenvalues and eigenvectors of the matrix parameters $\mathbf{S}_j$ and $\mathbf{T}_i$, respectively.

Algorithm 1 outlines the main steps of the proposed deterministic annealing algorithm. The total number of iterations is the number of times the inner loop is executed. The annealing loop can be stopped when T drops below a minimum value or when a specific number of iterations is reached. The inner loop is repeated until the relative reduction of the average cost from one iteration to the next becomes smaller than a threshold. Note that at the end of the iterative process the pdfs $q(\mathbf{V}_j)$ and $q(\mathbf{W}_i)$ will be highly peaked about their mode (practically they are Dirac delta functions).

Algorithm 1 Outline of the deterministic annealing algorithm

```
Set the initial temperature T ← T_0    (e.g., T_0 = 100)
Initialize Q_j = I_M, for all j ∈ K
repeat
    repeat
        Compute T_i and R_i, for all i ∈ K
        Compute S_j and Q_j, for all j ∈ K
    until convergence
    T ← η T    (e.g., η = 0.9)
until convergence
V_j ← the d_j most dominant eigenvectors of S_j, ∀j ∈ K
W_i ← the d_i most dominant eigenvectors of T_i, ∀i ∈ K
```

C. Special Instances

To obtain the iterative leakage minimization algorithm [7], in the following we keep the temperature fixed to $T = 1$ (i.e., no annealing) and constrain the beliefs in (7) to be Dirac delta pdfs: $\hat{q}(\mathbf{V}_j) = \delta\left(\mathbf{V}_j - \hat{\mathbf{V}}_j\right)$ and $\hat{q}(\mathbf{W}_i) = \delta\left(\mathbf{W}_i - \hat{\mathbf{W}}_i\right)$. Consequently, the second moments are $\mathbf{Q}_j = \hat{\mathbf{V}}_j \hat{\mathbf{V}}_j^{\mathrm{H}}$ and $\mathbf{R}_i = \hat{\mathbf{W}}_i \hat{\mathbf{W}}_i^{\mathrm{H}}$. According to [15], the point estimates $\hat{\mathbf{V}}_j$ and $\hat{\mathbf{W}}_i$ are the maximizers of expressions (9) and (10), respectively. This means that, for all $j \in \mathcal{K}$, the columns of $\hat{\mathbf{V}}_j$ are the d_j least dominant eigenvectors of the matrix $\sum_{i\neq j} \kappa_i \mathbf{H}_{ij}^{\mathrm{H}} \hat{\mathbf{W}}_i \hat{\mathbf{W}}_i^{\mathrm{H}} \mathbf{H}_{ij}$. Similarly, the columns

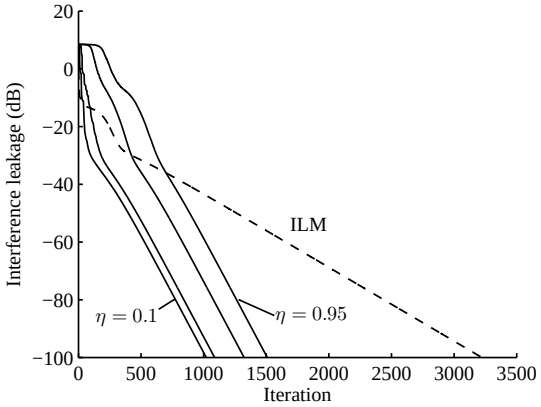

Fig. 1. Convergence of the interference leakage for one random realization of the channel. The continuous lines correspond to the DA algorithm with initial temperature $T_0 = 100$ and $\eta \in \{0.95, 0.9, 0.5, 0.1\}$.

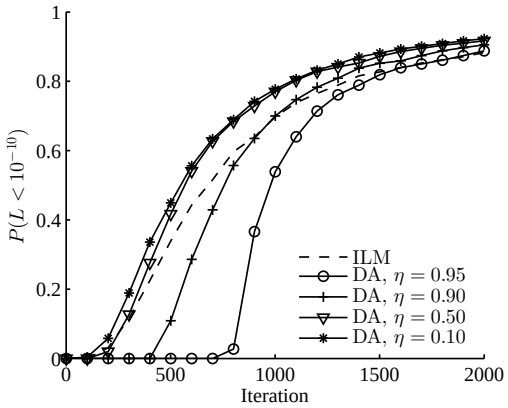

Fig. 2. Empirical probability of $L < 10^{-10}$ at a given iteration.

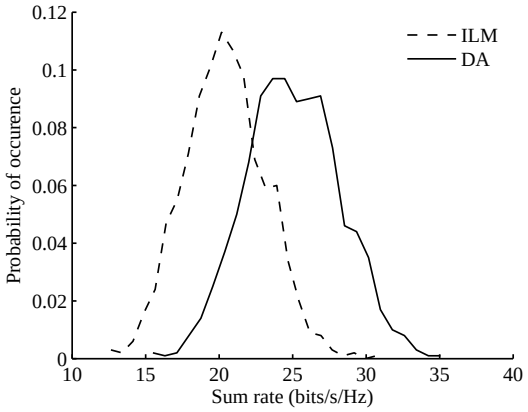

Fig. 3. Empirical distribution of the sum rate at the 500th iteration based on 1000 channel realizations for the asymmetric scenario. For the DA algorithm, $T_0 = 100$ and $\eta = 0.9$.

of $\hat{\mathbf{W}}_i$ are the d_i least dominant eigenvectors of the matrix $\sum_{j \neq i} \rho_j \mathbf{H}_{ij} \hat{\mathbf{V}}_j \hat{\mathbf{V}}_j^{\mathrm{H}} \mathbf{H}_{ij}^{\mathrm{H}}$, for all $i \in \mathcal{K}$. Note that by setting the weights $\kappa_i = 1$, for all $i \in \mathcal{K}$, and iteratively updating $\hat{\mathbf{V}}_j$ and $\hat{\mathbf{W}}_i$ we obtain the distributed algorithm [7].

It can be shown in a similar way that the alternating minimization algorithm [8] is an instance of our algorithm. For this, we have to re-parameterize the problem so that $\mathbf{W}_i$ is an $N_i \times (N_i - d_i)$ matrix whose columns are an orthonormal basis for the *interference* subspace of receiver i. At the same time, we need to set $\kappa_i = \rho_j = 1$, for all $i, j \in \mathcal{K}$.

V. Simulation Results

We use computer simulations to evaluate the proposed deterministic annealing (DA) algorithm and compare it against the iterative leakage minimization (ILM) algorithm [7].

In the first experiment, we analyze the convergence of the algorithm for the following system parameters making IA feasible: $K = 3$, $M = N = 4$, $d = 2$. The entries of the channel matrices are independent and have a complex Gaussian distribution with unit variance. The scenario is symmetric, i.e., the transmitters and receivers have same powers and noise levels, respectively. The total interference leakage is $L = \sum_{i=1}^{K} L_i(\mathbf{V}_{\sim i}, \mathbf{W}_i)$. For a random channel realization, Fig. 1 shows that L of DA converges faster than ILM and that, to some extent, using lower values of the annealing factor η tends to speed up convergence. To further analyze convergence, we illustrate in Fig. 2 the empirical probability of $L < 10^{-10}$ (basically, of achieving IA) at a given iteration computed from 1000 realizations of the channel. It can be noticed that DA with higher values of η require more iterations to converge, while for $\eta = 0.5$ and $\eta = 0.1$ DA converges faster than ILM.

In the second experiment, we consider a system setting for which IA is not feasible. In particular, we set $K = 4$, $M = N = 4$, $d = 2$ and assume an asymmetric scenario, i.e., the receiver noise precisions are different: $\gamma_1 = 100$, $\gamma_2 = 40$, $\gamma_3 = 10$, $\gamma_4 = 5$. This could correspond, for example, to a situation when the users experience different levels of uncoordinated interference. We notice in Fig. 3, which shows the probability of occurrence of the sum rate, that DA ensures significantly higher sum rates in this scenario. This is also supported by Fig. 4 which displays the average sum rate and the individual average rates of the users with highest and lowest noise precisions. Due to the weighting in the cost function (3), DA tries to align the interference mostly at user 1, while ILM attempts to achieve IA at each user, as it can also be seen in Fig. 5.

VI. Conclusion

We formulated the interference alignment as an optimization problem whose cost function weights the interference leakages at the receivers according to their inverse noise power. By employing the deterministic annealing method in conjunction with the mean field approximation, we obtained a novel iterative algorithm that includes some existing methods as special instances. For some settings of the annealing scheme, the DA algorithm showed better convergence performance than the ILM method. Moreover, in an asymmetric scenario where

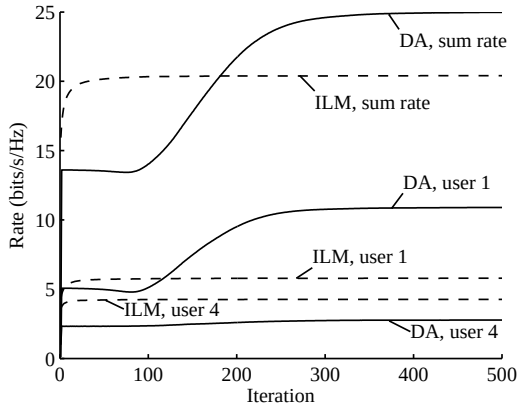

Fig. 4. Convergence of the rate averaged over 1000 channel realizations for the asymmetric scenario.

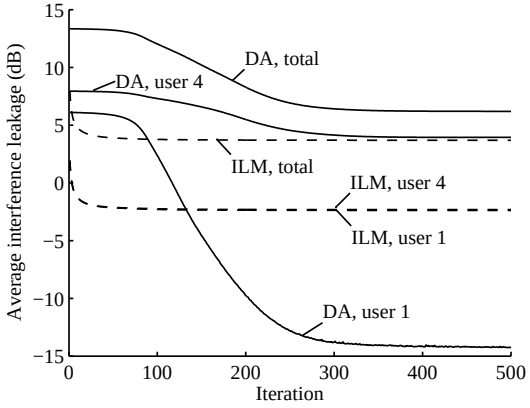

Fig. 5. Convergence of the interference leakage averaged over 1000 channel realizations for the asymmetric scenario.

IA is not feasible, the proposed weighting enables significantly higher sum rates than ILM.

The paper opens several interesting directions. To further improve convergence speed, it could be relevant to study other approximations than mean field (e.g., the Bethe approximation). At the same time, it would be pertinent to extend the method to include channel uncertainty or to study other cost functions that are more focused on sum rate optimization.

ACKNOWLEDGMENT

This work was supported in part by the EC FP7 Network of Excellence NEWCOM# (Grant agreement no. 318306). M. Guillaud also acknowledges the funding of the FP7 project HIATUS (grant 265578) of the European Commission (EC) and of the Austrian Science Fund (FWF) through grant NFN SISE (S106).

APPENDIX

The Bingham distribution of a random vector on the real or complex unit sphere and that of a random matrix on the real Stiefel manifold are well established [14], [16]. In this appendix, we extend the matrix Bingham distribution to the complex-variate case. The complex Stiefel manifold $\mathcal{CV}_{k,n}$ is represented by the set $\mathcal{CV}_{k,n} = \left\{\mathbf{X} \in \mathbb{C}^{n\times k} \mid \mathbf{X}^{\mathrm{H}}\mathbf{X} = \mathbf{I}_k\right\}$. For $k = 1$, $\mathcal{CV}_{k,n}$ is the complex sphere, while for $k = n$ it is the unitary group $\mathcal{U}(n)$. By analogy with the real matrix Bingham pdf, we consider that the complex matrix Bingham distribution has the pdf

$$f(\mathbf{X};\mathbf{A}) = \frac{1}{c(\mathbf{A})}\operatorname{etr}\left(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}\right) \tag{11}$$

with respect to the invariant measure on $\mathcal{CV}_{k,n}$, where the $n\times k$ matrix $\mathbf{A}$ is Hermitian and $c(\mathbf{A})$ is the normalizing constant. In the following we determine the normalizing constant and second-order moment of the pdf (11).

A. The normalizing constant

The normalizing constant is given by

$$c(\mathbf{A}) \triangleq \int_{\mathcal{CV}_{k,n}} \operatorname{etr}\left(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}\right)\left(\mathbf{X}^{\mathrm{H}}\,\mathrm{d}\mathbf{X}\right) \tag{12}$$

where the differential form $(\mathbf{X}^{\mathrm{H}}\,\mathrm{d}\mathbf{X})$ is the unnormalized invariant measure on $\mathcal{CV}_{k,n}$. It is important to notice that the normalizing constant (12) is actually determined only by the eigenvalues of the Hermitian matrix $\mathbf{A}$ expressed as $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}}$, where $\mathbf{U} \in \mathcal{U}(n)$ and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1,\ldots,\lambda_n) \in \mathbb{R}^{n\times n}$. Indeed, since $(\mathbf{X}^{\mathrm{H}}\,\mathrm{d}\mathbf{X})$ is invariant, we make the transformation $\mathbf{X} \to \mathbf{U}\mathbf{X}$ in (12) and obtain that $c(\mathbf{A}) = c(\mathbf{\Lambda})$.

We start by evaluating the integral

$$I(\mathbf{A},\mathbf{B}) \triangleq \int_{\mathcal{U}(n)} \operatorname{etr}\left(\mathbf{A}\mathbf{Z}\mathbf{B}\mathbf{Z}^{\mathrm{H}}\right)\left(\mathbf{Z}^{\mathrm{H}}\,\mathrm{d}\mathbf{Z}\right) \tag{13}$$

where $\mathbf{B} = \operatorname{diag}(b_1,\ldots,b_n)$, $b_i \in \mathbb{R}$ for all $i = 1,\ldots,n$. The differential form $(\mathbf{Z}^{\mathrm{H}}\,\mathrm{d}\mathbf{Z})$ is the unnormalized invariant measure on $\mathcal{U}(n)$ [17]. The normalized invariant measure, i.e., the uniform probability measure on $\mathcal{U}(n)$, is $(\mathrm{d}\mathbf{Z}) \triangleq \frac{1}{\operatorname{vol}(\mathcal{U}(n))}(\mathbf{Z}^{\mathrm{H}}\,\mathrm{d}\mathbf{Z})$, where $\operatorname{vol}(\mathcal{U}(n))$ is the volume of the unitary group. The integral (13) evaluates to

$$I(\mathbf{A},\mathbf{B}) = \operatorname{vol}(\mathcal{U}(n))\frac{\prod_{i=1}^{n-1} i! \times \det\left(\left[e^{\lambda_i b_j}\right]_{1\le i,j\le n}\right)}{\prod_{i<j}^{n}(\lambda_j-\lambda_i)\prod_{i<j}^{n}(b_j-b_i)} \tag{14}$$

where we used the Harish-Chandra-Itzykson-Zuber integral formula [18].

On the other hand, Lemma 9.5.3 in [19] enables us to first integrate in (13) over the last $n-k$ columns of $\mathbf{Z}$ with the first k columns being fixed, and then to integrate over the first k columns. Defining $\mathbf{B}_1 = \operatorname{diag}(b_1,\ldots,b_k)$ and $\mathbf{B}_2 = \operatorname{diag}(b_{k+1},\ldots,b_n)$, we obtain

$$\begin{aligned} I(\mathbf{A},\mathbf{B}) = &\int_{\mathbf{X}\in\mathcal{CV}_{k,n}} \operatorname{etr}\left(\mathbf{A}\mathbf{X}\mathbf{B}_1\mathbf{X}^{\mathrm{H}}\right) \\ &\times \int_{\mathbf{K}\in\mathcal{U}(n-k)} \operatorname{etr}\left(\mathbf{A}(\mathbf{G}\mathbf{K})\mathbf{B}_2(\mathbf{G}\mathbf{K})^{\mathrm{H}}\right)(\mathbf{K}^{\mathrm{H}}\,\mathrm{d}\mathbf{K})(\mathbf{X}^{\mathrm{H}}\,\mathrm{d}\mathbf{X}) \end{aligned}$$

where $\mathbf{G} = \mathbf{G}(\mathbf{X})$ is any $n\times(n-k)$ matrix with orthonormal columns that are orthogonal to those of $\mathbf{X}$. Specializing

$I(\mathbf{A},\mathbf{B})$ for $b_1 = \ldots = b_k = 1$ and $b_{k+1} = \ldots = b_n = 0$, we get

$$I(\mathbf{A},\mathbf{B})\Big|_{\substack{\mathbf{B}_1=\mathbf{I}_k \\ \mathbf{B}_2=\mathbf{0}}} = c(\mathbf{A})\,\mathrm{vol}(\mathcal{U}(n-k))$$

Using now the result (14), we obtain

$$c(\mathbf{A}) = \frac{\mathrm{vol}(\mathcal{U}(n))\prod_{i=1}^{n-1} i!}{\mathrm{vol}(\mathcal{U}(n-k))\prod_{i<j}^{n}(\lambda_j-\lambda_i)} \times \lim_{\substack{b_1,\ldots,b_k\to 1 \\ b_{k+1},\ldots,b_n\to 0}} \frac{\det\left(\left[e^{\lambda_i b_j}\right]_{1\le i,j\le n}\right)}{\prod_{i<j}^{n}(b_j-b_i)}$$

Employing [20, Th. 2.9] to evaluate the limit and using the fact that $\mathrm{vol}(\mathcal{U}(p)) = 2^p\pi^{p^2}/\mathcal{C}\Gamma_p(p)$ [17], where $\mathcal{C}\Gamma_p(\cdot)$ is the complex multivariate gamma function, we finally obtain the normalizing constant

$$c(\mathbf{\Lambda}) = \frac{2^k\pi^{kn}}{\mathcal{C}\Gamma_k(k)} \times \frac{(-1)^{k(k-n)}\det\left(\mathbf{M}(\mathbf{\Lambda})\right)}{\prod_{i<j}^{n}(\lambda_j-\lambda_i)} \qquad (15)$$

where the matrix $\mathbf{M}(\mathbf{\Lambda})$ is

$$\begin{bmatrix} e^{\lambda_1} & \lambda_1 e^{\lambda_1} & \cdots & \lambda_1^{k-1}e^{\lambda_1} & 1 & \lambda_1 & \cdots & \lambda_1^{n-k-1} \\ e^{\lambda_2} & \lambda_2 e^{\lambda_2} & \cdots & \lambda_2^{k-1}e^{\lambda_2} & 1 & \lambda_2 & \cdots & \lambda_2^{n-k-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{\lambda_n} & \lambda_n e^{\lambda_n} & \cdots & \lambda_n^{k-1}e^{\lambda_n} & 1 & \lambda_n & \cdots & \lambda_n^{n-k-1} \end{bmatrix}$$

B. The second-order moment

Given that the measure $(\mathbf{X}^{\mathrm{H}}\,\mathrm{d}\mathbf{X})$ is invariant, we make the transformation $\mathbf{X} \to \mathbf{U}\mathbf{X}$ and obtain the second moment

$$\mathbf{\Sigma} \triangleq \langle \mathbf{X}\mathbf{X}^{\mathrm{H}}\rangle_{f(\mathbf{X};\mathbf{A})} = \mathbf{U}\mathbf{\Gamma}\mathbf{U}^{\mathrm{H}} \qquad (16)$$

where $c(\mathbf{A}) = c(\mathbf{\Lambda})$ is given by (15) and $\mathbf{\Gamma} = \langle \mathbf{X}\mathbf{X}^{\mathrm{H}}\rangle_{f(\mathbf{X};\mathbf{\Lambda})}$. It can be shown that $\mathbf{\Gamma}\mathbf{D} = \mathbf{D}\mathbf{\Gamma}$ for any diagonal unitary matrix $\mathbf{D}$ by using again the invariance of $(\mathbf{X}^{\mathrm{H}}\,\mathrm{d}\mathbf{X})$ under $\mathbf{X} \to \mathbf{D}\mathbf{X}$. It follows that $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1,\ldots,\gamma_n)$.

Defining the diagonal matrix $\mathbf{T} = \mathrm{diag}(t_1,\ldots,t_n)$ with $t_i \in \mathbb{R}$, we compute the expected value of the expression $\mathrm{etr}\left(\mathbf{X}^{\mathrm{H}}\mathbf{T}\mathbf{X}\right)$ with respect to $f(\mathbf{X};\mathbf{\Lambda})$ in two different ways. First, direct evaluation gives

$$\langle \mathrm{etr}\left(\mathbf{X}^{\mathrm{H}}\mathbf{T}\mathbf{X}\right)\rangle_{f(\mathbf{X};\mathbf{\Lambda})} = [c(\mathbf{\Lambda})]^{-1}\,c(\mathbf{T}+\mathbf{\Lambda}) \qquad (17)$$

Then, we use the Taylor series expansion of the exponential function. Applying $\langle\cdot\rangle_{f(\mathbf{X};\mathbf{\Lambda})}$ on both sides of

$$\mathrm{etr}\left(\mathbf{X}^{\mathrm{H}}\mathbf{T}\mathbf{X}\right) = 1 + \mathrm{tr}\left(\mathbf{X}^{\mathrm{H}}\mathbf{T}\mathbf{X}\right) + \frac{1}{2!}\left[\mathrm{tr}\left(\mathbf{X}^{\mathrm{H}}\mathbf{T}\mathbf{X}\right)\right]^2 + \ldots$$

we obtain

$$[c(\mathbf{\Lambda})]^{-1}\,c(\mathbf{T}+\mathbf{\Lambda}) = 1 + \mathrm{tr}(\mathbf{T}\mathbf{\Gamma}) + \ldots$$

where the higher-order terms (not displayed) are homogeneous polynomials in $t_1,\ldots,t_n$ of degree higher than two. So, we take the derivative with respect to each t_i on both sides and evaluate the result at $t_1 = \ldots = t_n = 0$. Doing so, we obtain

$$\gamma_i = [c(\mathbf{\Lambda})]^{-1}\,\frac{\partial c(\mathbf{T}+\mathbf{\Lambda})}{\partial t_i}\Big|_{\mathbf{T}=\mathbf{0}} = [c(\mathbf{\Lambda})]^{-1}\,\frac{\partial c(\mathbf{\Lambda})}{\partial \lambda_i} \qquad (18)$$

for all $i = 1,\ldots,n$. Thus, based on (16), the second-order moment is

$$\mathbf{\Sigma} = \mathbf{U}\,\mathrm{diag}\,(\gamma_1,\ldots,\gamma_n)\,\mathbf{U}^{\mathrm{H}} \qquad (19)$$

with γ_i given by (18).

REFERENCES

[1] T. Gou and S. A. Jafar, "Degrees of freedom of the K user $M \times N$ MIMO interference channel," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6040–6057, Dec. 2010.

[2] C. M. Yetis, S. A. Jafar, and A. H. Kayran, "Feasibility conditions for interference alignment," in *Proc. IEEE Globecom*, Honolulu, HI, USA, Nov. 2009.

[3] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, Feb. 2012.

[4] G. Bresler, D. Cartwright, and D. Tse, "Settling the feasibility of interference alignment for the MIMO interference channel: the symmetric square case," Apr. 2011, preprint available from http://arxiv.org/abs/1104.0888.

[5] D. Schmidt, W. Utschick, and M. Honig, "Large system performance of interference alignment in single-beam MIMO networks," in *Proc. IEEE Global Telecommunications (Globecom) Conference*, Miami, FL, USA, Dec. 2010.

[6] R. Tresch, M. Guillaud, and E. Riegler, "On the achievability of interference alignment in the K-user constant MIMO interference channel," in *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, Cardiff, U.K., Sep. 2009.

[7] K. S. 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, no. 6, pp. 3309–3322, 2011.

[8] S. Peters and R. Heath, Jr., "Interference alignment via alternating minimization," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 2445–2448.

[9] M. Guillaud, M. Rezaee, and G. Matz, "Interference alignment via message-passing," in *Proc. IEEE International Conference on Communications (ICC)*, Sydney, Australia, Jun. 2014.

[10] G. Alexandropoulos and C. Papadias, "A reconfigurable distributed algorithm for k-user mimo interference networks," in *IEEE International Conference on Communications (ICC)*, Jun. 2013, pp. 3063–3067.

[11] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proceedings of the IEEE*, vol. 86, no. 11, Nov. 1998.

[12] T. Hofmann and J. Buhmann, "Pairwise data clustering by deterministic annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, pp. 1–14, Jan 1997.

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

[14] Y. Chikuse, *Statistics on special manifolds*. Springer, 2003.

[15] E. Riegler, G. Kirkelund, C. Manchon, M. Badiu, and B. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *Information Theory, IEEE Transactions on*, vol. 59, no. 1, pp. 588–602, Jan 2013.

[16] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 2000.

[17] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Jacobians and hypergeometric functions in complex multivariate analysis," *Canadian Applied Mathematics Quarterly*, vol. 12, no. 2, pp. 213–239, 2004.

[18] C. Itzykson and J. B. Zuber, "The planar approximation. II," *Journal of Mathematical Physics*, vol. 21, no. 3, pp. 411–421, 1980.

[19] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley-Interscience, 2005.

[20] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, New York: Cambridge University Press, 2011.