

A Riemannian Geometric Approach to Blind Signal Recovery for Grant Free Radio Network Access

Carlos Feres, *Student Member, IEEE*, Zhi Ding, *Fellow, IEEE*

Department of Electrical and Computer Engineering

University of California, Davis, CA, USA

Abstract—We propose a new nonconvex framework for blind multiple signal demixing and recovery. The proposed Riemannian geometric approach extends the well known constant modulus algorithm to facilitate grant-free wireless access. For multiple signal demixing and recovery, we formulate the problem as non-convex problem optimization problem with signal orthogonality constraint in the form of Riemannian Orthogonal CMA (ROCMA). Unlike traditional stochastic gradient solutions that require large data samples, parameter tuning, and careful initialization, we leverage Riemannian geometry and transform the orthogonality requirement of recovered signals into a Riemannian manifold optimization. Our solution demonstrates full recovery of multiple access signals without large data sample size or special initialization with high probability of success.

Index Terms—Blind demixing, grant-free access, signal recovery, Riemannian manifolds, optimization, non-convex.

I. INTRODUCTION

Recent advances in next generation networking technologies are poised to ubiquitously connect the full spectrum of sensors, devices, and computers to facilitate future development of smart cities and smart agriculture, among other applications. These exciting developments, known collectively as Internet of Things (IoT), promise significant benefits in a plethora of fields including health care, farming, environmental science, infrastructure, energy efficiency, transportation, safety and sustainability. In this work, we focus on the technical challenge of grant-free wireless access for a large number of low complexity wireless devices.

Generally, wireless networks are based on either random access (e.g., WiFi networks) or controlled scheduling (e.g., 4G-LTE cellular networks). Contention based random access schemes, such as the CSMA-CA protocol adopted in IEEE 802.11a/g/n/ac, possess the advantage of simplicity but suffer from lower spectrum efficiency due to access collision when the number of active devices is large. Controlled user scheduling based on centralized access grants can achieve high spectrum efficiency but require more elaborate network-user interaction such as random access and contention resolution, and would incur higher energy consumption for many low power devices.

A typical IoT application involves sporadic communications between a significant number transceivers, triggered by external events, in order to save energy. This prompts the need for low-latency communications and the ability to support these

links in the performance-constrained scenario of typical IoT transceivers, in particular, in terms of bandwidth efficiency.

In grant-free access, multiple signals could collide at the receiving node. Although these colliding signals can utilize specialized pilots or training signals as their unique characteristics to be exploited for signal separation, there are at least two problems. First, the training signals would consume precious device energy and network bandwidth to transmit but carry no payload data. Second, to reduce training overhead, shorter pilots should be used. However, there are at most N orthogonal training sequences of length N . Thus, large number of IoT devices pose challenges to both spectrum and energy efficiency.

Blind equalization has been a staple idea in terms of achieving this goal by diminishing the impact of pilots or preambles, aiming to reduce their impact in the overall bandwidth efficiency. Among blind equalization algorithms, the Constant Modulus Algorithm (CMA) presented by Godard [1] in the 1980s is often considered the most widespread technique due to its computational simplicity and practical effectiveness [2], [3]. However, one of its major drawbacks in practical applications is the presence of local extrema - due to the action of additive noise- and its slow convergence [3]–[5]. CMA-based grant free signal recovery typically apply traditional adaptive algorithms such as stochastic descent to find optimum parameters of an underlying linear system for signal recovery. Such solutions would require finely tuning of e.g., normalization and stepsize for satisfactory convergence.

There have been extensive works on CMA and other related formulations aimed at overcoming their drawbacks. Several recent works have proposed different approaches for tackling CMA-based optimization problems. One interesting approach is the transformation of CMA-based equalization to a convex problem, via Semidefinite relaxation [6]–[8], which provides global convergent solutions in a lifted higher dimensional parameter space that are further projected to the original solution space. There are also other relaxation approaches, such as using the trace (nuclear) norm as surrogate for the rank-1 constrain imposed on the CMA problem when defined in terms of matrices [9]. As with any relaxation approach, CMA based on convex relaxation relies on the expectation that the convex problem yields solutions that can be projected to near optimum CMA solutions. Additionally, the problem size grows polynomially with increasing parameter size of the linear system and poses severe practical challenges in many scenarios.

Other line of works include analytical solution to CMA [10], [11]. These solutions and its variants [12], [13] do

This material is based on works supported by the National Science Foundation under Grant No.ECCS-1711823 and No. 2009001. The work of C. Feres is also supported by ANID PFCHA/BCH 72170648 scholarship.

have convergence ambiguity owing to the algebraic solution. However, they are much more complex and cannot work with QAM source signals that do not exhibit the same modulus (magnitude) such as 16-QAM. There are also multistage schemes [14], that depend heavily on the estimation error being close to the MMSE estimate in earlier stages, or the error accumulates through different stages [15].

In the present paper, we present a new Riemannian perspective with which we redefine the orthogonality requirement of different combiners as a Riemannian manifold. Any optimization procedure that solves the blind signal recovery problem is now an unconstrained optimization problem over a Riemannian manifold. The Riemannian geometric formulation has been extensively studied in recent years [16] and has been successfully applied to several domains, such as low-rank matrix decomposition [17], singular value decomposition [18], phase retrieval [19], blind signal demixing [20], dictionary learning [21], among others. As we shall show, it presents a promising direction for improving CMA and related algorithm for blind signal recovery in grant-free network access.

Section II presents the signal model for blind signal recovery, the formulation of the optimization problem, and some comments on the nature of these. Section III introduces the Riemannian geometry two proposed optimization schemes. Section IV details theoretical convergence properties and complexity analysis of this technique. Section V presents numerical simulations on each scenario, and finally Section VI summarizes our conclusions.

Notations: In the following, vectors and matrices will be denoted with small and capital boldface letters, such as $\mathbf{z}$ and $\mathbf{Z}$ respectively. Sets are denoted with calligraphic capital letters. Complex conjugation is denoted with $\bar{z}$. For a complex scalar a , we use $\text{Re}(a)$, $\text{Im}(a)$, $|a|$ and $\angle(a)$ to denote its real part, imaginary part, magnitude and angle, respectively. The transpose, element-wise complex conjugation and conjugate transpose are denoted by $\mathbf{z}^T$, $\bar{\mathbf{z}}$ and $\mathbf{z}^H$, respectively. The Hermitian and skew-Hermitian parts of a matrix $\mathbf{Z}$ are denoted as $\text{herm}(\mathbf{Z}) = 0.5(\mathbf{Z} + \mathbf{Z}^H)$ and $\text{skew}(\mathbf{Z}) = 0.5(\mathbf{Z} - \mathbf{Z}^H)$. The Euclidean norm of vectors and spectral norm of matrices is denoted by $\|\cdot\|$, and the Frobenius norm of matrices is denoted by $\|\cdot\|_F$. Finally, $\text{diag}(\mathbf{z})$ represents a diagonal matrix that uses elements of vector $\mathbf{z}$ on its diagonal, and we define the operator $\text{ddiag}(\mathbf{Z})$ which yields a diagonal matrix which only retains the diagonal elements of $\mathbf{Z}$.

II. SYSTEM MODEL

A. Grant-Free Blind Signal Recovery and Demixing

We consider the signal recovery of multiple users in an access group in a grant-free access system, as depicted in Fig.1. In particular, all potential uplink users in each access group have acquired network timing such that their uplink transmission bursts would span one given set of receiver time slots. Users in each designated access group may randomly transmit within their shared channel in terms of allocated time or frequency resources. Appropriate coding and rate-matching is utilized by all source nodes to have equal number of data symbols K within each access group and burst. Furthermore,

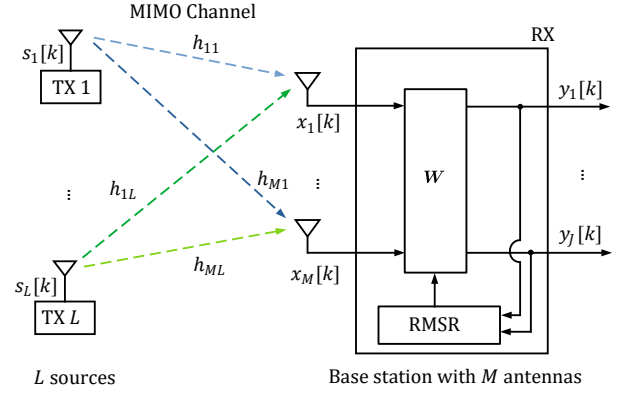


Fig. 1: L sources share a common resource block and transmit independent signals to a host station with M antennas through an unknown physical channel. The host receiver aims to use a linear $\mathbf{W}$ to recover L sources with little mutual interference.

we design systems such that with very high probability or certainty that the number of single-antenna active nodes L shall fall below the number of diversity antennas M at the receiver node. In particular, the receiver node does not necessarily know L . Since the receiver recovers multiple user signals during blind demixing without prior knowledge of their identities, the receiver can utilize user-ID scrambled CRC to check which recovered user signal belongs to which user, similar to the blind detection of PDCCH by users using RNTI-scrambled CRC in LTE or 5G [22], [23].

To summarize, we define the received signal vector $\mathbf{x}_k$, the transmitted signal vector $\mathbf{s}_k$, and the flat fading channel $\mathbf{H}$, respectively, as

$$\mathbf{x}_k = \begin{bmatrix} x_1[k] \\ \vdots \\ x_M[k] \end{bmatrix}, \mathbf{s}_k = \begin{bmatrix} s_1[k] \\ \vdots \\ s_L[k] \end{bmatrix}, \mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1L} \\ \vdots & \cdots & \vdots \\ h_{M1} & \cdots & h_{ML} \end{bmatrix}. \quad (1)$$

Then the received signal vector can be written as

$$\mathbf{x}_k = \mathbf{H}\mathbf{s}_k + \mathbf{n}_k, \quad (2)$$

where the MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{M \times L}$ is assumed to have full column rank L (with $L \leq M$) and $\mathbf{n}_k \in \mathbb{C}^M$ is the vector of additive white Gaussian noises (AWGN) in that resource block, of the same size as $\mathbf{x}_k$ in Eq.(1).

In blind multiple signal demixing and recovery, we are interested in deriving J simultaneous demixers $\mathbf{w}_j \in \mathbb{C}^M$, $j \in \{1, \dots, J\}$ that allow the recovery of J sources with minimal interference, each tuned to a distinct signal. We can also write $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_J]$ as the receiver blind demixing parameter matrix such that

$$\mathbf{y}_k = \begin{bmatrix} \mathbf{w}_1^H \\ \vdots \\ \mathbf{w}_J^H \end{bmatrix} \mathbf{x}_k = \mathbf{W}^H \mathbf{x}_k = \begin{bmatrix} \hat{s}_{\ell_1}[k] \\ \vdots \\ \hat{s}_{\ell_J}[k] \end{bmatrix}, \quad \ell_j \in \{1, \dots, L\}. \quad (3)$$

Note that the receiver has explicit knowledge on neither the unknown channels $\mathbf{H}$ nor the number of active sources L ,

except for the statistical properties and the constellation of each source signal. Additionally, we need to ensure that the demixers do not merely restore the same source signal for only a small subset of sources, possibly with different phases or delays [3]. Therefore, when considering simultaneous multiple signal recovery, additional constraints must be enforced for demixers $\mathbf{w}_j, j \in \{1, \dots, J\}$ to recover different source signals. Without loss of generality, we consider $J \leq L$ and we shall also consider the case when accurate estimation of the number of access users is achieved such that $J = L$.

B. Constant Modulus Algorithm

The problem of blind signal recovery has been extensively studied before. In particular, Godard [1] proposed what was later known [2] as the constant modulus algorithm (CMA) to adaptively find a single optimum demixer $\mathbf{w} \in \mathbb{C}^M$ by minimizing the mean CM cost for equalization:

$$E\{|y_k|^2 - R_2\}^2, \quad R_2 = \frac{E\{|s_\ell[k]|^4\}}{E\{|s_\ell[k]|^2\}^2}. \quad (4)$$

It is known that CMA can be applied to i.i.d. signals using QAM constellations of arbitrary size and magnitude [3]. Moreover, even by setting $R_2 = 1$, CMA still converges such that its output recovers QAM source signal with a simple scalar, without affecting signal integrity. In batch implementation, the single-source CM cost can be rewritten as

$$f(\mathbf{w}) = \frac{1}{2K} \sum_{k=1}^K \left(|\mathbf{x}_k^H \mathbf{w}|^2 - R_2 \right)^2, \quad (5)$$

which is a smooth real-valued nonconvex function of $\mathbf{w}$. Note that f presents phase invariance, i.e., if $\hat{\mathbf{w}}$ is a solution that minimizes $f(\hat{\mathbf{w}})$, then the entire set $\mathcal{W}(\hat{\mathbf{w}}) = \{e^{i\theta} \hat{\mathbf{w}} : \theta \in [0, 2\pi]\}$ contains equivalent solutions that achieve the same minimum $f(\hat{\mathbf{w}})$.

C. Simultaneous Multiple Signal Recovery in Demixing

The CMA has been adapted in the past for simultaneous recovery of multiple independent source signals. In these applications, the first step is to define a cumulative demixing cost consisting of J copies of CM costs:

$$f(\mathbf{W}) = \frac{1}{2K} \sum_{j=1}^J \sum_{k=1}^K \left(|\mathbf{x}_k^H \mathbf{w}_j|^2 - R_2 \right)^2. \quad (6)$$

The full blind demixing cost is a function of the J blind demixers $\mathbf{W}$. The joint blind demixing problem is to optimize multiple solution vectors $\hat{\mathbf{W}} = [\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_J]$ that jointly minimize the cumulative CM cost of (6).

This cumulative CM cost by itself cannot guarantee that the recovered signals are indeed from different sources. In fact, even if every one column vector of $\hat{\mathbf{W}}$ captures the same signal source, the cumulative CM cost of (6) is still minimized and cannot prevent such solutions. For this reason, it is clear that the cumulative CM cost of (6) is non-convex and is in fact multi-modal. Hence, the challenge lies in the practical need that distinct source signals be recovered by the J solution vectors of $\hat{\mathbf{W}}$.

Several approaches that aim to enforce the demixers to recover different signals in what is called MIMO blind recovery. Specifically, many works would add regularization term(s) to the cost function (6) to penalize against the recovery of identical signals by more than one solution vectors in $\hat{\mathbf{W}}$. In [24], [25], for example the authors proposed adding a norm of joint cumulants for such source separation objective. Another MIMO CMA approach [26] uses the real part of equalized signals as regularization.

Despite their demonstrated successes, regularization approaches exhibit some drawbacks. First, the regularizing term typically requires a scalar weight that must be tuned, often by trial and error. There is no performance guarantee under various possible scenarios. Second, different regularization approaches might lead to different solutions and performance, while no solution is consistently better than others. Additionally, regularizing terms often increase the computation complexity as regularized cost functions would either require additional computations or delicate non-convex optimization steps. Finally, regularizing terms proposed in the literature generally are limited to promote pairwise signal orthogonality instead of multi-lateral signal orthogonality, and also require more data samples to successfully suppress mutual interferences.

In our approach based on Riemann Geometry, we enforce signal orthogonality among demixer outputs by directly restricting the solution space. Recall the definition of $\mathbf{W}$ as joint demixer matrix. Due to the phase invariance of the CM cost function, the optimal solution satisfies $\hat{\mathbf{W}}^H \mathbf{H} = \mathbf{P}$ with $\mathbf{P} \in \mathbb{C}^{J \times L}$ a generalized permutation matrix (whose non-zero entries are complex numbers with unit magnitude), i.e., $\mathbf{P} \mathbf{P}^H = \mathbf{I}$. Here $\mathbf{I}$ denotes identity matrix of appropriate size. We therefore write the joint signal recovery constraint as

$$\hat{\mathbf{W}}^H \mathbf{H} \mathbf{H}^H \hat{\mathbf{W}} = \mathbf{I}. \quad (7)$$

However, the blind receiver node has no knowledge of the channel $\mathbf{H}$. Therefore, we can leverage source signal orthogonality and white noise property to estimate $\mathbf{H} \mathbf{H}^H$ from the sample covariance matrix of the data vectors $\mathbf{x}_k$:

$$\mathbf{R}_X = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H \xrightarrow{K \rightarrow \infty} \mathbb{E}\{\mathbf{R}_X\} = \mathbf{H} \mathbf{H}^H + \sigma^2 \mathbf{I}. \quad (8)$$

Note that, in the absence of noise, the rank of matrix $\mathbf{H} \mathbf{H}^H \in \mathbb{C}^{M \times M}$ is L ($L \leq M$), i.e., the rank of $\mathbf{H}$. Thus, we formulate the optimization problem for multiple signal recovery as orthogonal constant modulus algorithm (OCMA):

$$\min \quad f(\mathbf{W}) = \frac{1}{4K} \sum_{k=1}^K \left\| \text{ddiag}(\mathbf{W}^H \mathbf{X}_k \mathbf{W}) - R_2 \mathbf{I} \right\|_F^2 \quad (9a)$$

$$\text{s.t.} \quad \mathbf{W}^H \mathbf{R}_X \mathbf{W} = \mathbf{I}. \quad (9b)$$

We first note that the Euclidean gradient of $f(\mathbf{W})$ is

$$\nabla_{\mathbf{W}} f(\mathbf{W}) = \frac{1}{K} \sum_{k=1}^K \mathbf{X}_k \mathbf{W} \left(\text{ddiag}(\mathbf{W}^H \mathbf{X}_k \mathbf{W}) - R_2 \mathbf{I} \right), \quad (10a)$$

and for a matrix $\mathbf{E}$ of the same size as $\mathbf{W}$, the directional derivative of $f(\mathbf{W})$ in direction $\mathbf{E}$ is

$$\begin{aligned} & \mathbf{D}(\nabla_{\mathbf{W}} f(\mathbf{W}))[\mathbf{E}] \\ &= \frac{1}{K} \left(\sum_{k=1}^K \mathbf{X}_k \mathbf{W} \text{ddiag}(\mathbf{W}^H \mathbf{X}_k \mathbf{E} + \mathbf{E}^H \mathbf{X}_k \mathbf{W}) \right. \\ & \quad \left. + \mathbf{X}_k \mathbf{E} (\text{ddiag}(\mathbf{W}^H \mathbf{X}_k \mathbf{W}) - R_2 \mathbf{I}) \right). \end{aligned} \quad (10b)$$

D. Estimating the Number of Active Sources for Demixing

Because of the number of active sources L may vary in practice, the literature has often assumed that L is known. However, in grant-free access, such assumption would not be practical, since, at best, we would only be able to limit the maximum number of simultaneous users according synchronization and slotted scheduling. Thus we shall first present an approach to estimate the number of active sources.

Given that $\mathbf{H}$ has rank $L \leq M$, the sample covariance matrix $\mathbf{R}_X$ in restriction (9b) is not strictly positive definite in the absence of noise. Thus, the restriction cannot be directly defined as a Riemannian manifold. In noisy scenarios and with several data samples, the sample covariance matrix will likely be positive definite, but would probably be numerically ill-conditioned as its condition number might be very large. However, in both cases we can extract a strictly positive definite matrix from the sample covariance matrix from its rank- L approximation.

We first estimate the number of transmitted signals embedded in the received data via Minka's Laplace method [27], and let the result be L . Let the SVD of the channel matrix $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, with $\mathbf{U} \in \mathcal{U}(M)$, and $\mathbf{V} \in \mathcal{U}(L)$, i.e.,

$$\begin{aligned} \mathbf{H} &= [\mathbf{U}_L \ \mathbf{U}_L^\perp] \begin{bmatrix} \mathbf{\Sigma}_L \\ \mathbf{0}_{(M-L) \times L} \end{bmatrix} \mathbf{V}^H = \mathbf{U}_L \mathbf{\Sigma}_L \mathbf{V}^H, \\ \mathbf{U}_L &\in \text{ST}(M \times L), \mathbf{U}_L^\perp \in \text{ST}(M \times (M-L)), \\ \mathbf{\Sigma}_L &= \text{diag}(\sigma_1, \dots, \sigma_L), \end{aligned} \quad (11)$$

where $\text{ST}(M \times L) = \{\mathbf{A} \in \mathbb{C}^{M \times L} : \mathbf{A}^H \mathbf{A} = \mathbf{I}_L\}$ is the complex Stiefel manifold of orthonormal L -frames in $\mathbb{C}^M$ [18], [28].

First, consider the ergodic noiseless scenario (i.e. $\mathbf{R}_X = \mathbb{E}\{\mathbf{R}_X\}$),

$$\mathbf{H}\mathbf{H}^H = \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^H\mathbf{U}^H = \mathbf{U}_L\mathbf{\Sigma}_L\mathbf{\Sigma}_L^H\mathbf{U}_L^H = \mathbf{U}_L\mathbf{\Lambda}^H\mathbf{U}_L^H. \quad (12)$$

From the above decomposition, we can obtain $\mathbf{U}$ and $\mathbf{\Sigma}$, but not $\mathbf{V}$. Also, note that $\mathbf{\Lambda} = \mathbf{\Sigma}_L\mathbf{\Sigma}_L^H$ is diagonal with positive entries because $\mathbf{H}$ is full-column rank, and both $\mathbf{U}_L, \mathbf{U}_L^\perp$ are full-column rank.

In the noisy case with infinite samples, we have

$$\begin{aligned} \mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I} &= \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^H\mathbf{U}^H + \sigma^2 \mathbf{I} \\ &= \mathbf{U}(\mathbf{\Sigma}\mathbf{\Sigma}^H + \sigma^2 \mathbf{I})^H \mathbf{U}^H = \mathbf{U}\mathbf{\Lambda}_1 \mathbf{U}^H, \end{aligned} \quad (13)$$

and we let $\mathbf{\Lambda}$ equal to the diagonal matrix whose elements are the L largest diagonal components of $\mathbf{\Lambda}_1 - \sigma^2 \mathbf{I}$, with corresponding eigenvector matrix $\mathbf{U}_L$. This approach is very

similar to the so-called probabilistic PCA [29], which obtains the principal components of data and a generative model.

However, even when Minka's Laplace method is known for having satisfactory performance in the limited sample regime, it relies on the assumption of Gaussian signals and might fail to properly estimate the number of sources with discrete modulations. All independent sources contribute with a significant component of the sample covariance matrix, related to its significant eigenvalues, whereas noise will only have minor contributions in other directions as their related eigenvalues are much smaller in high SNR regimes.

For an under-estimated L , the L -rank approximation of the sample covariance matrix would likely fail to capture all relevant directions of the channel, leading to mutual signal interference in signal recovery. Therefore, we compute the normalized L -rank approximation error

$$\frac{\|\mathbf{R}_X - \mathbf{U}_L \mathbf{\Lambda} \mathbf{U}_L^H\|}{\|\mathbf{R}_X\|} \quad (14)$$

for comparison against a preset threshold ϵ_r to decide whether L needs to be increased in a update. We also update the L -rank approximation of the sample covariance matrix. Our test results to be shown later demonstrate the general reliability of this rank estimation method for demixing.

III. ROCMA: A RIEMANNIAN MANIFOLD OPTIMIZATION FRAMEWORK

The Riemannian framework for optimization on manifolds [16] has gained a lot of attention owing to the capability to handle problems with a real-valued objective function defined on a constrained space,

$$\underset{\mathbf{M} \in \mathbb{C}^{m \times n}}{\text{minimize}} \quad f(\mathbf{M}) \quad \text{s.t.} \quad \mathbf{M} \in \mathcal{M}, \quad (15)$$

Note that the (nonlinear) space $\mathcal{M}$ might not be well-defined in terms of addition, continuity, and/or other properties which are typically exploited by regular optimization approaches in Euclidean spaces. The main idea is to redefine the problem as an unconstrained optimization problem over a manifold. Manifolds are topological spaces that, equipped with a metric, locally resemble Euclidean spaces of equal dimension size, but might be remarkably different globally. Some manifold examples include spheres, the set of rotations, the set of positive semidefinite matrices, the set of fixed-rank matrices, and Stiefel manifolds, among many others.

In this section, we first obtain a suitable Riemannian manifold representation of the ROCMA problem (9). Next, we further exploit the obtained Riemannian manifold to define a quotient Riemannian manifold, which allows us to tackle the phase invariance of the demixers directly in the optimization process.

A. Redefining the Geometry of Signal Recovery

Our goal here is to find a suitable geometry that encodes the orthogonality condition of demixers in the search space of Problem (9). Even with a method to estimate the number of sources L , we derive a general version of the geometry where

the receiver attempts to recover $J \leq L$ sources. Considering Eq.(12) in restriction (9b), we have

$$\begin{aligned} I_J &= \hat{\mathbf{W}}^H \mathbf{H} \mathbf{H}^H \hat{\mathbf{W}} = \hat{\mathbf{W}}^H \mathbf{U}_L \Sigma_L \Sigma_L^H \mathbf{U}_L^H \hat{\mathbf{W}} \\ &= (\mathbf{U}_L^H \hat{\mathbf{W}})^H \Lambda (\mathbf{U}_L^H \hat{\mathbf{W}}) = \mathbf{Y}^H \Lambda \mathbf{Y}, \end{aligned} \quad (16)$$

which defines the complex scaled Stiefel manifold of

$$\text{ST}_\Lambda(L \times J) = \{\mathbf{Y} \in \mathbb{C}^{L \times J} : \mathbf{Y}^H \Lambda \mathbf{Y} = I_J\}. \quad (17)$$

This Stiefel manifold defines the set of orthonormal J frames in $\mathbb{C}^L$ through the scaling of Λ , and is a generalization of the complex Stiefel manifold $\text{ST}(L \times J)$ (see [28] for the notion of scaled Stiefel manifold in the real case). Now, recall that $\mathbf{U}_L \in \text{ST}(M \times L)$; and therefore we have

$$\mathbf{Y} = \mathbf{U}_L^H \mathbf{W} \iff \mathbf{W} = \mathbf{U}_L \mathbf{Y}. \quad (18)$$

Hence, by means of this transformation of the optimizing variable $\mathbf{W}$ by $\mathbf{U}_L$, we obtain a Riemannian manifold representation of restriction (9b) as $\overline{\mathcal{M}} = \text{ST}_\Lambda(L \times J)$ that we can use for optimization purposes. From our solution, we obtain the demixer matrix directly with a one-to-one scaling by $\mathbf{U}_L$. The variable transformation (18) implies the need to rewrite the cost function, Euclidean gradient, and directional derivatives of the gradient. Defining $\mathbf{z}_k = \mathbf{U}_L^H \mathbf{x}_k$ and $\mathbf{Z}_k = \mathbf{z}_k \mathbf{z}_k^H = \mathbf{U}_L^H \mathbf{X}_k \mathbf{U}_L$, we have a new cost function

$$g(\mathbf{Y}) = \frac{1}{4K} \sum_{k=1}^K \|\text{ddiag}(\mathbf{Y}^H \mathbf{Z}_k \mathbf{Y}) - R_2 \mathbf{I}\|_F^2, \quad (19)$$

whose Euclidean gradient is

$$\nabla_{\mathbf{Y}} g(\mathbf{Y}) = \frac{1}{K} \sum_{k=1}^K \mathbf{Z}_k \mathbf{Y} (\text{ddiag}(\mathbf{Y}^H \mathbf{Z}_k \mathbf{Y}) - R_2 \mathbf{I}), \quad (20)$$

and the directional derivative of (20) in direction $\mathbf{E}$ is

$$\begin{aligned} D(\nabla_{\mathbf{Y}} g(\mathbf{Y}))[\mathbf{E}] &= \\ &= \frac{1}{K} \sum_{k=1}^K \mathbf{Z}_k \mathbf{Y} \text{ddiag}(\mathbf{Y}^H \mathbf{Z}_k \mathbf{E} + \mathbf{E}^H \mathbf{Z}_k \mathbf{Y}) \\ &\quad + \mathbf{Z}_k \mathbf{E} (\text{ddiag}(\mathbf{Y}^H \mathbf{Z}_k \mathbf{Y}) - R_2 \mathbf{I}). \end{aligned} \quad (21)$$

To optimize $\mathbf{Y}$ over $\overline{\mathcal{M}}$, we need to first define the linear space that approximates the manifold around a point $\mathbf{Y}$, which is called the tangent space at $\mathbf{Y}$ and is denoted as $\text{T}_{\mathbf{Y}} \overline{\mathcal{M}}$. For $\overline{\mathcal{M}} = \text{ST}_\Lambda(L \times J)$, the tangent space is

$$\begin{aligned} \text{T}_{\mathbf{X}} \overline{\mathcal{M}} &= \{\mathbf{G} \in \mathbb{C}^{L \times J} : \mathbf{G} = \mathbf{X} \Omega + \mathbf{X}_\perp \mathbf{A}, \\ &\quad \Omega = -\Omega^H \in \mathbb{C}^{J \times J}, \mathbf{A} \in \mathbb{C}^{(L-J) \times J}\}. \end{aligned} \quad (22)$$

In other words, $\mathbf{X}^H \Lambda \mathbf{G} = \Omega$ is skew-Hermitian.

We can now define length in the tangent space with a Riemannian metric $d_{\mathbf{Y}}$, which is a smooth inner product defined at each element $\mathbf{Y}$ for elements of the tangent space $\text{T}_{\mathbf{Y}} \overline{\mathcal{M}}$. In our case, we use a scaled version of the real-trace metric, given by

$$d_{\mathbf{Y}}(\mathbf{E}, \mathbf{C}) = \text{Re}(\text{Tr}(\mathbf{E}^H \Lambda \mathbf{C})), \quad \mathbf{E}, \mathbf{C} \in \text{T}_{\mathbf{Y}} \overline{\mathcal{M}}. \quad (23)$$

We also define a projection to the tangent space, which allows to restrict optimization only in the directions of interest,

which indeed belong to the tangent space. For $\mathbf{G} \in \overline{\mathcal{M}}$, the projection operator is

$$\text{Proj}_{\mathbf{Y}}^{\text{T}}(\mathbf{G}) = \mathbf{G} - \mathbf{Y} \text{herm}(\mathbf{Y}^H \Lambda \mathbf{G}) \in \text{T}_{\mathbf{Y}} \overline{\mathcal{M}}, \quad (24)$$

which enables us to define the Riemannian gradient and Riemannian Hessian from the Euclidean gradient and its directional derivative, respectively.

For optimization purposes, the motion along the manifold from point $\mathbf{Y}$ in a given direction $\mathbf{E}$ is given by a retraction $\text{R}_{\mathbf{Y}}(\mathbf{E})$, which in our case corresponds to the polar retraction for the complex scaled Stiefel manifold [28]

$$\begin{aligned} \text{R}_{\mathbf{X}}^{\text{St}}(\mathbf{G}) &= (\mathbf{X} + \mathbf{G})((\mathbf{X} + \mathbf{G})^H \Lambda (\mathbf{X} + \mathbf{G}))^{-\frac{1}{2}} \\ &= (\mathbf{X} + \mathbf{G})(\mathbf{I} + \mathbf{G}^H \Lambda \mathbf{G})^{-\frac{1}{2}}. \end{aligned} \quad (25)$$

B. Riemannian Quotient Geometry

In the context of Riemannian manifold optimization, quotient Riemannian manifolds are used to define a manifold that presents invariance of the cost function or the representation of the manifold itself [30]. It can be defined by equipping the original or ambient manifold with an equivalence relation between its points to describe the aforementioned invariance.

Let $\sim$ be such an equivalence relation, i.e., $\mathbf{Y} \sim \mathbf{Y}_0$ denotes that $\mathbf{Y}$ and $\mathbf{Y}_0$ are equivalent in terms of the invariance of interest. Thus, we can identify equivalent points to $\mathbf{Y}$ as one single set known as equivalence class, denoted as

$$[\mathbf{Y}] = \{\mathbf{Y}_0 \in \mathcal{M} : \mathbf{Y}_0 \sim \mathbf{Y}\}. \quad (26)$$

The Riemannian quotient manifold is the set of equivalence classes:

$$\mathcal{M} = \overline{\mathcal{M}} / \sim = \{[\mathbf{Y}] : \mathbf{Y} \in \overline{\mathcal{M}}\}. \quad (27)$$

A quotient manifold is an abstract space whose elements are subsets of the ambient manifold. However, the use of quotient manifolds in Riemannian optimization has additional advantages, such as the ability of obtaining a strictly positive definite Hessian by neglecting directions related to the cost function invariance, and potential reduction of problem dimensionality by applying a simple representation of the elements in the space. Even in a case when there is no such representation and the ambient manifold is used for computational purposes, the quotient geometry is theoretically important to establish convergence properties of second-order methods that rely on the positive definiteness of the Hessian on the manifold.

Now, recall that the cost function (9a) presents unimodular phase invariance in each demixer such that for each demixer $\mathbf{w}_j$, a rotated demixer $e^{i\theta} \mathbf{w}_j$, $\theta \in [0, 2\pi]$ yields the same cost value. When considering multiple demixers in $\mathbf{W}$, we want to describe unimodular phase invariance on each of the J demixers simultaneously. Let $\mathcal{U}(1)^{\times J}$ the group of diagonal unitary matrices of size J , i.e.

$$\mathcal{U}(1)^{\times J} = \{\mathbf{D} \in \mathcal{U}(J) : \mathbf{D} = \text{diag}(e^{i\theta_1} \dots e^{i\theta_J})\}.$$

Thus, the group action of $\mathcal{U}(1)^{\times J}$ defines an equivalence relation between demixer matrices. The corresponding equivalence class is then

$$[\mathbf{W}] = \{\mathbf{W} \mathbf{D} : \mathbf{D} \in \mathcal{U}(1)^{\times J}\},$$

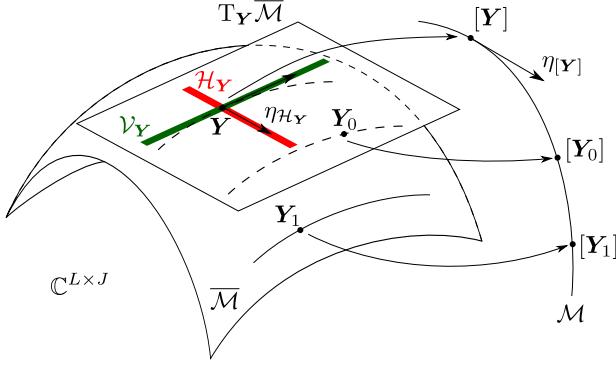


Fig. 2: Representation of the ambient manifold $\bar{\mathcal{M}}$ and quotient manifold $\mathcal{M}$. The tangent space $T_Y \bar{\mathcal{M}}$ is divided into a vertical space $\mathcal{V}_Y$ (in green) and a horizontal space $\mathcal{H}_Y$ (in red), which contains the relevant search directions $\eta_{\mathcal{H}_Y}$. These directions correspond to tangent directions $\eta_{[Y]}$ at the point $[Y]$ in the quotient manifold.

and by means of (18), we have that $\mathbf{W}\mathbf{D} = \mathbf{U}_L \mathbf{Y}\mathbf{D}$. In other words, we can rewrite the equivalence class in terms of $\mathbf{Y}$ as

$$[\mathbf{Y}] = \{\mathbf{Y}\mathbf{D} : \mathbf{D} \in \mathcal{U}(1)^{\times J}\}, \quad (28)$$

and we obtain a Riemannian quotient manifold that considers the cost function invariance as

$$\mathcal{M} = \bar{\mathcal{M}}/\mathcal{U}(1)^{\times J}. \quad (29)$$

The quotient manifold $\mathcal{M}$ is an abstract space, and requires matrix representations in the computational space $\bar{\mathcal{M}}$. Fortunately, an element $\mathbf{Y}_q$ on the quotient manifold can be represented by an element $\mathbf{Y}$ in the computational space. Thus, every geometry-related operation over the quotient manifold can be defined in terms of elements and operations in the computational space.

We now look for a representation of the tangent space of the quotient manifold $T_Y \mathcal{M}$ using the tangent vectors of the ambient manifold $\bar{\mathcal{M}}$. We accomplish this by characterizing the tangent space $T_Y \bar{\mathcal{M}}$ as the direct sum of two orthogonal spaces: the *vertical* space $\mathcal{V}_Y$, which contains the directions tangent to the equivalence classes, and the *horizontal* space $\mathcal{H}_Y$, which contains the tangent directions orthogonal to the vertical space. That is, the horizontal space contains the directions of interest in terms of optimization, and the tangent vectors of the quotient manifold can be represented as vectors of the horizontal space of the ambient manifold. This forms a Riemannian submersion from the quotient manifold to the computational space, thereby defining a correspondence between elements of the quotient space and elements of the computational space [16]. Figure 2 shows a depiction of the quotient manifold geometry and its relation to the ambient manifold.

Let $\mathbf{D} : \mathbb{R} \rightarrow \mathcal{U}(1)^{\times J}$ be a path in the equivalence class such that $\mathbf{D}(0) = \mathbf{I}$. The vertical space is given by vectors of the form $\mathbf{X}\mathbf{D}'(0)$ where vectors $\mathbf{D}'(t)$ are tangent to $\mathcal{U}(1)^{\times J}$, whose tangent set corresponds to the Lie algebra

of unitary diagonal matrices $\tau(J)$, consisting of diagonal imaginary matrices of size $J \times J$. Therefore,

$$\begin{aligned} \mathcal{V}_Y &= \{\mathbf{Y}\mathbf{T} : \mathbf{T} \in \tau(J)\} \\ &= \{\mathbf{Y}\mathbf{T} : \mathbf{T} \in \mathbb{C}^{J \times J} \text{ imaginary diagonal}\}, \end{aligned} \quad (30)$$

and the horizontal space is then given by

$$\begin{aligned} \mathcal{H}_Y &= (\mathcal{V}_Y)^\perp \\ &= \{\mathbf{G} \in T_Y \bar{\mathcal{M}} : \langle \mathbf{G}, \mathbf{F} \rangle = 0 \quad \forall \mathbf{F} \in \mathcal{V}_Y\} \\ &= \{\mathbf{G} \in T_Y \bar{\mathcal{M}} : \langle \mathbf{G}, \mathbf{Y}\mathbf{T} \rangle = 0 \quad \forall \mathbf{T} \in \tau(J)\} \\ &= \{\mathbf{G} \in T_Y \bar{\mathcal{M}} : \text{Re}(\text{Tr}(\mathbf{G}^H \mathbf{\Lambda} \mathbf{Y} \mathbf{T})) = 0 \quad \forall \mathbf{T} \in \tau(J)\} \end{aligned} \quad (31)$$

and thus $\mathbf{Y}^H \mathbf{\Lambda} \mathbf{G}$ is skew-Hermitian with zero diagonal, to be orthogonal to any $\mathbf{T} \in \tau(J)$. This is equivalent to state that the projection to horizontal space is given by

$$\begin{aligned} \text{Proj}_Y^H(\mathbf{E}) &= \text{Proj}_Y^T(\mathbf{E}) - \mathbf{Y} \text{ddiag}(\mathbf{Y}^H \mathbf{\Lambda} \text{Proj}_Y^T(\mathbf{E})) \\ &= \mathbf{E} - \mathbf{Y} \text{herm}(\mathbf{Y}^H \mathbf{\Lambda} \mathbf{E}) \\ &\quad - \mathbf{Y} \text{ddiag}(\text{skew}(\mathbf{Y}^H \mathbf{\Lambda} \mathbf{E})). \end{aligned} \quad (32)$$

Finally, the retraction for the quotient manifold corresponds to the retraction on the ambient manifold, restricted to the horizontal space. We can see that the polar retraction depends only on the equivalence class:

$$\begin{aligned} \text{R}_{[Y]}^{\text{St}}(\mathbf{G}) &= \text{R}_{\mathbf{Y}\mathbf{D}}^{\text{St}}(\mathbf{G}) \\ &= (\mathbf{Y}\mathbf{D} + \mathbf{G}\mathbf{D})(\mathbf{I}_J + (\mathbf{G}\mathbf{D})^H \mathbf{\Lambda} \mathbf{G}\mathbf{D})^{-\frac{1}{2}} \\ &= \text{R}_Y^{\text{St}}(\mathbf{G})\mathbf{D} = [\text{R}_Y^{\text{St}}(\mathbf{G})]. \end{aligned} \quad (33)$$

Consequently, we can effectively optimize over the quotient manifold $\mathcal{M}$ using representatives from the ambient manifold $\bar{\mathcal{M}}$. Table I summarizes the geometric definitions of the quotient manifold $\mathcal{M}$ used in ROCMA. Readers interested in additional details of the quotient manifold discussions may refer to [30, Section 9.9].

C. Riemannian Optimization for Blind Signal Recovery

We use a Riemannian Trust-Region (RTR) algorithm, which is a second-order optimization approach with superlinear convergence rate [31]. First, the algorithm searches a direction $\mathbf{E}$ on the horizontal space $\mathcal{H}_Y$. At each iteration we solve the trust-region subproblem with $\mathbf{Y} \in \mathcal{M}$

$$\begin{aligned} \mathcal{Q} : \quad & \underset{\mathbf{E} \in \mathcal{H}_Y}{\text{minimize}} && q(\mathbf{E}) \\ & \text{s.t.} && d_Y(\mathbf{E}, \mathbf{E}) \leq \delta^2 \end{aligned} \quad (34)$$

where $\mathbf{E}$ is in the horizontal space of iterate $\mathbf{Y}$ and δ denotes the trust region radius. The cost function is given by

$$q(\mathbf{E}) = d_Y(\mathbf{E}, \text{grad}_Y g(\mathbf{Y})) + \frac{1}{2} d_Y(\mathbf{E}, \text{Hess}_Y g(\mathbf{Y})[\mathbf{E}]) \quad (35)$$

in which $\text{grad}_Y g(\mathbf{Y})$ denotes the Riemannian gradient and $\text{Hess}_Y g(\mathbf{Y})[\mathbf{E}]$ denotes the Riemannian Hessian, each obtained by projecting their Euclidean counterparts to the horizontal space of the ambient manifold.

We can now define the ROCMA algorithm as summarized in Algorithm 1. Succinctly, we first initialize by estimating

TABLE I: Riemannian geometry definitions required for manifold optimization of ROCMA.

Name	Definition
Computational space $\overline{\mathcal{M}}$	$\text{ST}_{\mathbf{A}}(L \times L)$
Quotient space $\mathcal{M} = \overline{\mathcal{M}}/\sim$	$\text{ST}_{\mathbf{A}}(L \times L)/\text{U}(1)^{\times L}$
Riemannian metric $d_{\mathbf{Y}}$	$d_{\mathbf{Y}}(\mathbf{E}, \mathbf{C}) = \text{Re}(\text{Tr}(\mathbf{E}^{\text{H}} \mathbf{\Lambda} \mathbf{C}))$
Horizontal space $\mathcal{H}_{\mathbf{Y}}$	$\mathcal{H}_{\mathbf{Y}} = \{\mathbf{Y} \mathbf{T} : \mathbf{T} \in \mathfrak{t}(L)\}$
Horizontal space projection $\text{Proj}_{\mathbf{Y}}^{\text{H}}$	$\text{Proj}_{\mathbf{Y}}^{\text{H}}(\mathbf{E}) = \mathbf{E} - \mathbf{Y} \text{herm}(\mathbf{Y}^{\text{H}} \mathbf{\Lambda} \mathbf{E}) - \mathbf{Y} \text{ddiag}(\text{skew}(\mathbf{Y}^{\text{H}} \mathbf{\Lambda} \mathbf{E}))$
Riemannian gradient $\text{grad}_{\mathbf{Y}} g$	$\text{grad}_{\mathbf{Y}} g(\mathbf{E}) = \text{Proj}_{\mathbf{Y}}^{\text{H}}(\nabla_{\mathbf{Y}} g(\mathbf{E}))$
Riemannian Hessian $\text{Hess}_{\mathbf{Y}} g$	$\text{Hess}_{\mathbf{Y}} g(\mathbf{Y})[\mathbf{E}] = \text{Proj}_{\mathbf{Y}}^{\text{H}}(\text{D}_{\mathbf{Y}} g(\mathbf{Y})[\mathbf{E}])$
Retraction $\text{R}_{\mathbf{Y}}$	$\text{R}_{\mathbf{Y}}(\mathbf{E}) = (\mathbf{X} + \mathbf{E})(\mathbf{I}_J + \mathbf{E}^{\text{H}} \mathbf{\Lambda} \mathbf{E})^{-0.5}$

the number of sources L , perform an L -rank eigendecomposition that removes noise contribution in eigenvalues, and by corroborating that the L -rank approximation is close to the sample covariance matrix to adjust L if needed. After scaling the data vectors, we define cost function, quotient Riemannian manifold, and geometry operations. Thereafter, we determine Riemannian Trust Regions: in each iteration we solve the trust-regions subproblem $\mathcal{Q}$ in the horizontal space of the current iterate, obtaining a descent direction $\mathbf{E}$ in the horizontal space $\mathcal{H}_{\mathbf{Y}}$, whose magnitude is given by the size of the accepted trust region [16]. The subsequent solution iterate is computed using the retraction of $\mathbf{E}$, which brings the result back to the manifold. Once the algorithm converges, we compute the demixer matrix by scaling the obtained solution with $\mathbf{U}_L$.

Algorithm 1 Riemannian Orthogonal CMA (ROCMA)

Given: $\mathbf{x}_k \in \mathbb{C}^M$, $k \in \{1, \dots, K\}$, trust region radius δ , low-rank approximation tolerance ϵ_r

A) Source estimation:

- 1: Estimate number of independent sources L with Minka's Laplace method
- 2: Obtain L largest eigenvalues and corresponding eigenvectors of sample covariance matrix $\mathbf{R}_{\mathbf{X}}$ to construct L -rank approximation $\mathbf{R}_{\mathbf{X}} = \sum_k \mathbf{x}_k \mathbf{x}_k^{\text{H}} \approx \mathbf{U}_L \mathbf{\Lambda} \mathbf{U}_L^{\text{H}}$
- 3: **while** $\|\mathbf{R}_{\mathbf{X}} - \mathbf{U}_L \mathbf{\Lambda} \mathbf{U}_L^{\text{H}}\| > \epsilon_r \|\mathbf{R}_{\mathbf{X}}\|$ **do**
- 4: $L = L + 1$
- 5: Update $\mathbf{\Lambda}$ and $\mathbf{U}_L$ with next eigenvalue/eigenvector
- 6: **end while**

B) Initialization:

- 7: Define variables $\mathbf{z}_k = \mathbf{U}_L^{\text{H}} \mathbf{x}_k$ and objective function g
- 8: Define Riemannian manifold $\mathcal{M} = \text{ST}_{\mathbf{A}}(L \times J)/\text{U}(1)^{\times J}$ with metric $d_{\mathbf{Y}}$, projection $\text{Proj}_{\mathbf{Y}}^{\text{H}}$, retraction $\text{R}_{\mathbf{Y}}$

C) Riemannian Trust Regions:

- 9: **while** not converged **do**
 - 10: Obtain descent direction $\mathbf{E}_t$ by solving $\mathcal{Q}$ in $\mathcal{H}_{\mathbf{Y}_t}$
 - 11: $\mathbf{Y}_{t+1} = \text{R}_{\mathbf{Y}_t}(\mathbf{E}_t)$
 - 12: **end while**
 - 13: $\mathbf{W}_{\text{final}} = \mathbf{U}_L \mathbf{Y}_{\text{final}}$
-

A known algorithm to solve the trust-region subproblem $\mathcal{Q}$ based in a truncated Conjugate Gradient approach is available as Algorithm 11 in [16, Section 7.3]. The manifold optimization toolbox Manopt [32] implements a variation of this algorithm. We use this open-source toolbox Manopt in our implementation of Algorithm 1 by leveraging its flexibility for

selectable choices of stopping criteria, tolerances, and other parameters.

IV. PERFORMANCE AND THEORETICAL ANALYSIS

A. Convergence Conditions and Properties of CMA

The global convergence properties of CMA for PAM and QAM modulations in noiseless scenarios are well known [3, Chapters 4, 7]. The case of SIMO-CMA blind equalizers, also known as fractionally-spaced CMA or CMA-FSE (when applied to blind equalization scenarios), correspond to the case of recovering the transmitted signal via multiple antennas (for blind beamforming) or an oversampled equalizer (for blind equalization). The CMA-FSE equalizer has guaranteed global convergence as long as the subchannels have no common zeros, when using an equalizer with memory length larger than the order of the channel [33].

MIMO-CMA equalizers are an extension of CMA-FSE, where multiple sources are transmitting independent sources, and we adaptively find an equalizer that recovers one signal with minimal multi-user interference and minimum ISI [25], [34]. Global convergence of MIMO-CMA equalizers have similar requirements as the case of CMA-FSE equalizers, which in turn is equivalent to have the channel convolution matrix $\mathbf{H}$ with full column rank. Thus, channel matrix $\mathbf{H}$ of full-column rank provides guaranteed global convergence in noiseless scenarios.

Multiple source recovery is a special case of the multiple source recovery scheme presented [25] with zero-ISI subchannels, and therefore global convergence is also guaranteed under similar conditions. Hence, in the following we will always assume that the channel matrix $\mathbf{H}$ has full-column rank. Moreover, under noisy transmissions, it is well known that one effect of low additive channel noise is the addition of local minima to the cost function in the vicinity of the global solution [35], [36].

B. Known Results and Uses of Riemannian Optimization

Riemannian manifold optimization with different solvers, such as Riemannian Trust-Regions, has well-known convergence guarantees [31] over several classical manifolds, such as the Stiefel manifold and the generalized Stiefel manifold, [28], the Grassmannian manifold [16], and many others. These properties also apply to the quotient Riemannian manifolds, as the computational space is still the ambient manifold [16].

Since the geometry of the CMA cost function is well-behaved, in terms of its strong convexity near desired solutions and bounded curvature [37], the proposed Riemannian optimization will also have the well known convergence guarantees both globally and locally. In particular, Riemannian Trust regions will converge superlinearly [31].

Additionally, some existing works have also analyzed the particular case of cost functions that are mathematically similar to Eq.(19). These works present scenarios closely related to the constant modulus portion of the OCMA problem, but did not exploit the (scaled) orthogonality of several solutions in the problem geometry. In [38] the authors optimize over the Stiefel manifold to maximize the diagonal terms of a matrix quadratic form for joint diagonalization, which is similar to the CMA cost function by setting $R_2 = 0$. Another work [39] tackles the phase retrieval problem by defining a manifold geometry with the so-called fixed-norms manifold. Our proposed ROCMA generalizes the existing works and leverage the desired convergence properties of Riemannian manifold optimization.

C. Computational Complexity

We can estimate the computational complexity of the ROCMA algorithm by analyzing each step of Algorithm 1. In particular, the Riemannian Trust-Region step considers the iterations needed for convergence and also the iterations of each call to the trust-region subproblem algorithm. In the following, we refer to the former as outer iterations, and the latter as inner iterations.

- 1) The source estimation step is dominated by the computational cost of Minka's Laplace method, with a cost of $\mathcal{O}(ML)$, and a number of eigendecompositions, which can be obtained iteratively with a cost of $\mathcal{O}(M^3)$.
- 2) The computational cost of initialization is dominated by obtaining z_k via scaling, with a computational cost of $\mathcal{O}(MLK)$. Other computations have negligible cost, used mainly for defining geometry operations.
- 3) The cost of an inner iteration is dominated by the computation of Riemannian Hessian, which has a cost of $\mathcal{O}(L^2K)$. Other operations are linear over the samples with a cost of $\mathcal{O}(LK)$, or linear over averages at a cost of $\mathcal{O}(L)$.
- 4) The cost of an outer iteration is dominated by the cost of the total inner iterations required in the particular outer iteration I_t . Thus, this dominant component requires $\mathcal{O}(L^2KI_t)$. The Riemannian retraction R_Y has a cost of $\mathcal{O}(L^3)$, whereas other operations are linear over scalars and have a cost of $\mathcal{O}(1)$. The final scaling by U_L has a cost of $\mathcal{O}(ML^2)$.

Considering the major dominant computation of the aforementioned steps, Table II summarizes the complexity of ROCMA in each step.

V. SIMULATION RESULTS

A. Definitions

We test our proposed ROCMA in a multi-user signal recovery setup. We consider L sources, each transmitting K

TABLE II: Computational complexity of ROCMA.

ROCMA Steps	Total cost
Estimate L	$\mathcal{O}(ML)$
Iterative eigendecomposition R_Y	$\mathcal{O}(M^3)$
Define z_k	$\mathcal{O}(MLK)$
RTR outer iteration	$\mathcal{O}(L^3 + L^2KI_t)$
Final scaling	$\mathcal{O}(ML^2)$

independent symbols from a regular QAM constellation with unit average energy. The central receiver node has M receive antennas whereas each transmit node as a single antenna. The channels $\mathbf{H}$ are modeled as stationary Rayleigh with i.i.d. entries $H_{ml} \sim \mathcal{N}(0, \frac{1}{2}) + i\mathcal{N}(0, \frac{1}{2})$, under i.i.d. additive white Gaussian noise $\mathbf{n}_k$. All noise are independent of the channel and data signals. Our tests are performed over different values of average SNR at the receiver.

To measure the performance of the algorithm, we define the total normalized interference (NTI) for each of the recovered sources, defined by

$$\text{NTI}_j = \frac{\sum_i |C_{ij}|^2 - \max_i |C_{ij}|^2}{\max_i |C_{ij}|^2} \quad (36)$$

where $\mathbf{C} = \mathbf{H}^H \mathbf{W}$ represents the final demixed channel matrix, with its columns corresponding to each channel-demixer pair. Unless otherwise stated, we average 1000 runs per SNR value.

Our setting does not include initialization, and so ROCMA is initialized randomly. We also note that the Manopt RTR algorithm does not require a predetermined number of iterations or a fixed stepsize since it uses a stopping criterion and backtracking for the trust-regions subproblem.

B. Numerical Test Results

We test the source recovery capabilities of ROCMA, under different system sizes, different number of samples, and different QAM constellations.

Figure 3 presents the probability of successful recovery of a multiple sources with respect to the number of samples, for different system sizes and an SNR of 20dB. We test both QPSK and 16-QAM modulations. After ROCMA, we define success as the event that all demixers attain an NTI smaller than -20dB. RSMR achieves successful recovery with high probability with a reasonable number of samples for both modulation schemes, even for a rather large system size.

The number of samples that guarantee recovery with high probability grows with both modulation order and system size. Clearly, the system size demonstrates a stronger impact on the required number of data samples.

Using the minimum number of samples that achieves 100% of success probability in each setting, we compute the average (normalize) total interference of all demixers per outer iteration of ROCMA. Because of fading channels, random noises, and random signals, we are also interested in examining the distribution of the computational complexity over Monte Carlo simulations. Thus, we decide to use the number of iterations where the algorithm achieves successful recovery of all sources (as defined above) to characterize the speed of convergence. In particular, we define the *fastest recovery*

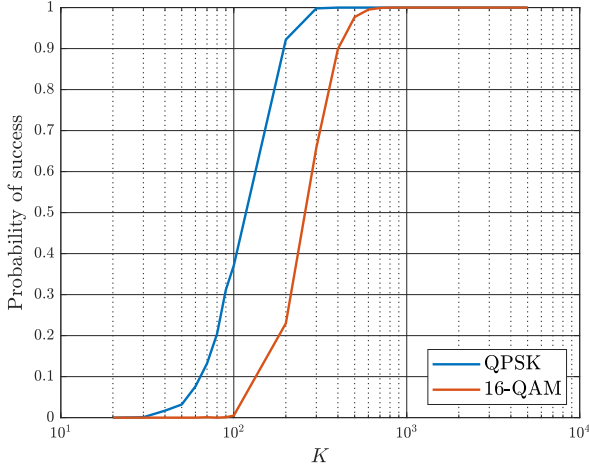
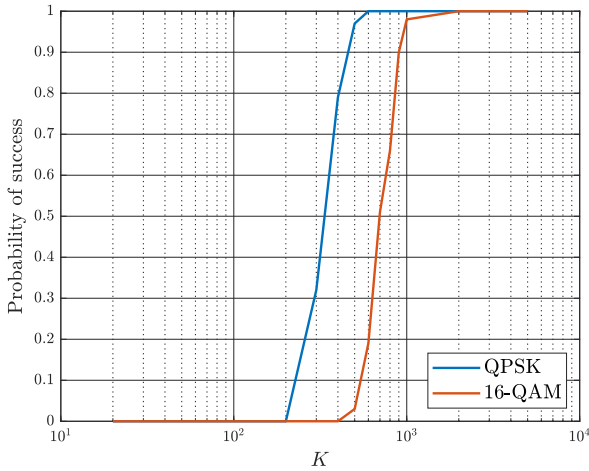
(a) $M \times L = 8 \times 4$.(b) $M \times L = 16 \times 9$.

Fig. 3: Probability of successful recovery for all demixers using ROCMA vs. number of samples K . Each scenario is simulated 1000 times for each system size and modulation scheme.

set as the tests that attain successful recovery with the least number of iterations (representing 3-5% of the simulations), the *slowest recovery set* as the tests that take the most number of iterations (representing 2-13% of the simulations). We also consider the *faster convergence subset* as the subset by discarding the slowest recovery subset (the faster 87-97% of all recovery results).

Figures 4 and 5 present these numerical evaluation for two systems of sizes $M \times L = 8 \times 4$ and 16×9 , respectively. In all cases, the average over all simulations closely resembles the average over the faster subset, which indicates that the bias of the *slowest recovery subset* is rather insignificant in terms of average convergence. When comparing the fastest and slowest recovery sets, our results appear to demonstrate significant variability in computational speed. However, the number of outer iterations does not vary drastically.

In comparison the traditional gradient descent implementa-

tions of CMA, which usually require thousands of iterations but at least hundreds of samples (in either batch or stochastic implementation) to achieve similar performance, the ROCMA is quite efficient and steady in computation complexity and reliably recovers the QAM data sources.

C. Constellation of QAM Signal Recovery

To further visualize the performance of our proposed ROCMA, Figure 6 illustrates the recovered constellations of all active sources at one time instant upon convergence. The system size is $M \times L = 16 \times 9$ and the 9 signals are all 16-QAM modulation. We selected an experiment with median performance among all tests using the number of samples K that guarantees 100% of successful recovery of all signals (up to a phase rotation) based on results in Fig. 3. As seen from Figure 6, all 9 QAM sources have been successfully recovered within 12 total iterations. Their 16-QAM constellations illustrate open-eye diagrams to guarantee error free recovery of the 9 different sources.

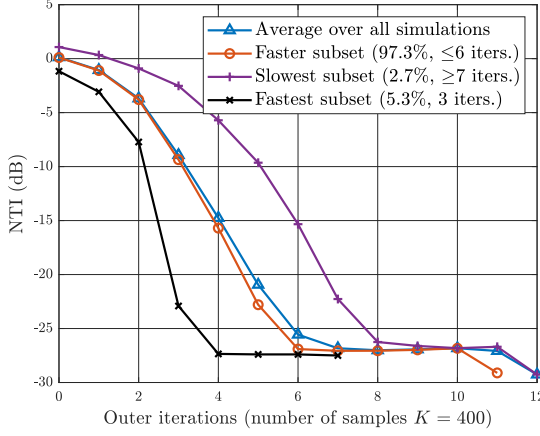
VI. CONCLUSION

In this paper, we present a new approach for blind multiple signal recovery that trades off computational complexity with speed and high probability of successful recovery. This is achieved by means of minimizing a Constant-Modulus Algorithm cost function with Riemannian optimization, such that the orthogonality of different demixers is embedded in the geometry of a Riemannian manifold. We derive this geometry and obtain the geometrical definitions that allow to minimize over the manifold as the search space of the optimization problem. The results of our approach show high probability of successful recovery of all sources with a reasonable number of samples, for rather large system sizes and different modulation schemes.

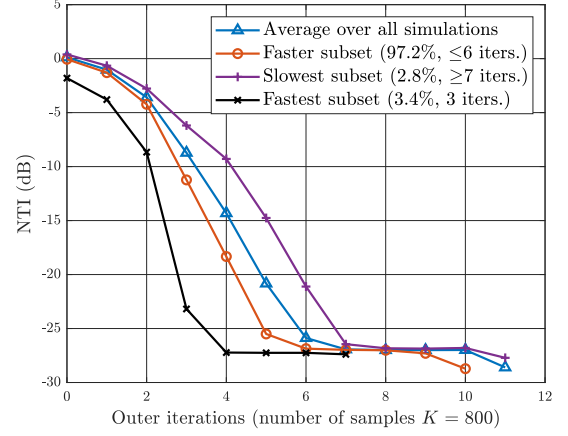
Future research paths include the adaptation for multiple source recovery and equalization over ISI fading channels, a stochastic or mini-batch reformulation of the algorithm, and the definition of new geometrical perspectives that could exploit information from forward error correction procedures.

REFERENCES

- [1] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, November 1980.
- [2] J. Treichler and M. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 420–431, 1985.
- [3] Z. Ding and Y. G. Li, *Blind Equalization and Identification*. Marcel Dekker, Inc., december 2000.
- [4] H. H. Zeng, L. Tong, and C. R. Johnson, "An analysis of constant modulus receivers," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 2990–2999, Nov 1999.
- [5] O. Dabeer and E. Masry, "Convergence analysis of the constant modulus algorithm," *IEEE Transactions on Information Theory*, vol. 49, no. 6, pp. 1447–1464, June 2003.
- [6] B. Mariere, Z.-Q. T. Luo, and T. N. Davidson, "Blind constant modulus equalization via convex optimization," *IEEE Trans. Signal Processing*, vol. 51, pp. 805–818, 2003.
- [7] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

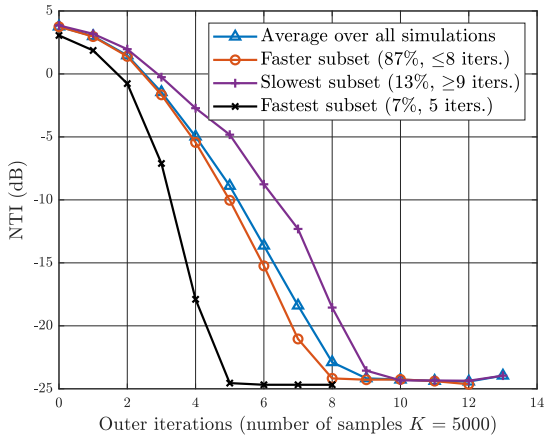


(a) QPSK.

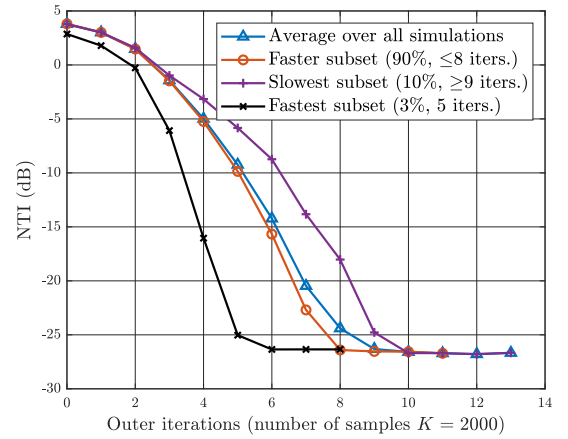


(b) 16-QAM.

Fig. 4: Average total interference for all demixers using ROCMA vs. number of outer iterations. We average 1000 scenarios with $L = 4$, $M = 8$, and different modulation schemes, using in each case the number of samples that guarantees 100% of successful recovery. The different subsets are marked with the percentage they represent over the total set of simulations, and the number of outer iterations for successful recovery to which that set corresponds.



(a) QPSK.



(b) 16-QAM.

Fig. 5: Average total interference for all demixers using ROCMA vs. number of outer iterations. We average 1000 scenarios with $L = 9$, $M = 16$, and different modulation schemes, using in each case the number of samples that guarantees 100% of successful recovery. The different subsets are marked with the percentage they represent over the total set of simulations, and the number of outer iterations for successful recovery to which that set corresponds.

- [8] K. Wang, "Semidefinite relaxation based blind equalization using constant modulus criterion," *CoRR*, vol. abs/1808.07232, 2018. [Online]. Available: <http://arxiv.org/abs/1808.07232>
- [9] A. Adler and M. Wax, "Constant modulus algorithms via low-rank approximation," *Signal Processing*, vol. 160, pp. 263 – 270, 2019.
- [10] B. Hassibi, A. Paulraj, and T. Kailath, "On a closed form solution to the constant modulus factorization problem," in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, vol. 2, Oct 1994, pp. 775–779.
- [11] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.
- [12] A.-J. van der Veen, "An adaptive version of the algebraic constant modulus algorithm [blind source separation applications]," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4, March 2005, pp. iv/873–iv/876 Vol. 4.
- [13] V. Zarzoso and P. Comon, "Optimal step-size constant modulus algorithm," *IEEE Transactions on Communications*, vol. 56, no. 1, pp. 10–13, January 2008.
- [14] J. J. Shynk and R. P. Gooch, "Performance analysis of the multistage cma adaptive beamformer," in *Proceedings of MILCOM '94*, Oct 1994, pp. 316–320 vol.2.
- [15] D. Liu and L. Tong, "An analysis of constant modulus algorithm for array signal processing," *Signal Processing*, vol. 73, no. 1, pp. 81 – 104, 1999.
- [16] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton University Press, 2008.
- [17] K. Yang, Y. Shi, and Z. Ding, "Low-rank matrix completion for mobile edge caching in fog-ran via riemannian optimization," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.
- [18] H. Sato, "Riemannian conjugate gradient method for complex singular value decomposition problem," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 5849–5854.
- [19] W. Huang, K. A. Gallivan, and X. Zhang, "Solving phaselift by low-

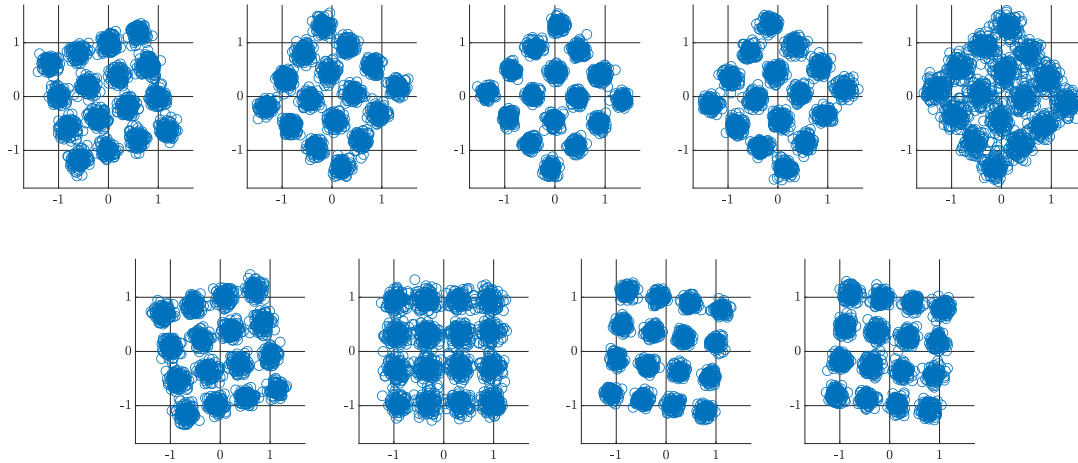


Fig. 6: Recovered constellations for 16-QAM, $M \times L = 16 \times 9$, the number of samples that guarantees 100% of successful recovery ($K = 2000$), and different modulation schemes.

- rank riemannian optimization methods1,” *Procedia Computer Science*, vol. 80, pp. 1125 – 1134, 2016, international Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA.
- [20] J. Dong, K. Yang, and Y. Shi, “Blind demixing for low-latency communication,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 897–911, Feb 2019.
- [21] A. Cherian and S. Sra, “Riemannian dictionary learning and sparse coding for positive definite matrices,” 2015.
- [22] *Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding*, 3GPP Tech. Specification TS 36.212 (v13.2.0), 2016.
- [23] A. Jalali and Z. Ding, “Joint detection and decoding of polar coded 5g control channels,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2066–2078, 2020.
- [24] Y. Chen, C. Nikias, and J. G. Proakis, “Crimmo: criterion with memory nonlinearity for blind equalization,” in *[1991] Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems Computers*, Nov 1991, pp. 694–698.
- [25] Ye Li and K. J. R. Liu, “Adaptive blind source separation and equalization for multiple-input/multiple-output systems,” *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2864–2876, Nov 1998.
- [26] A. Ikhlef and D. Le Guennec, “A simplified constant modulus algorithm for blind recovery of mimo qam and psk signals: A criterion with convergence analysis,” *EURASIP Journal on Wireless Communications and Networking*, vol. 56, no. 6, December 2007.
- [27] T. P. Minka, “Automatic choice of dimensionality for pca,” ser. NIPS’00. Cambridge, MA, USA: MIT Press, 2000, p. 577–583.
- [28] A. Edelman, T. A. Arias, and S. T. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, p. 303–353, Apr. 1999.
- [29] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*, 1st ed. Springer Publishing Company, Incorporated, 2016.
- [30] N. Boumal, “An introduction to optimization on smooth manifolds,” Available online, Nov 2020. [Online]. Available: <http://www.nicolasboumal.net/book>
- [31] P. A. Absil, C. G. Baker, and K. A. Gallivan, “Trust-region methods on riemannian manifolds,” *Found. Comput. Math.*, vol. 7, no. 3, pp. 303–330, Jul. 2007.
- [32] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, “Manopt, a matlab toolbox for optimization on manifolds,” *Journal of Machine Learning Research*, vol. 15, pp. 1455–1459, 2014.
- [33] Ye Li and Zhi Ding, “Global convergence of fractionally spaced godard equalizers,” in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, vol. 1, Oct 1994, pp. 617–621 vol.1.
- [34] S. Mayrargue, “A blind spatio-temporal equalizer for a radio-mobile channel using the constant modulus algorithm (cma),” in *Proceedings of ICASSP ’94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. iii, April 1994, pp. III/317–III/320 vol.3.
- [35] H. H. Zeng, L. Tong, and C. R. Johnson, “Relationships between the constant modulus and wiener receivers,” *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1523–1538, 1998.
- [36] H. H. Zeng, L. Tong, and C. Johnson, “An analysis of constant modulus receivers,” *IEEE transactions on Signal Processing*, vol. 47, no. 11, pp. 2990–2999, 1999.
- [37] C. Feres and Z. Ding, “Wirtinger Flow Meets Constant Modulus Algorithm: Revisiting Signal Recovery for Grant-Free Access,” *IEEE Transactions on Signal Processing*, p. Accepted, 2020.
- [38] F. J. Theis, T. P. Cason, and P. A. Absil, “Soft dimension reduction for ica by joint diagonalization on the stiefel manifold,” in *Independent Component Analysis and Signal Separation*, T. Adali, C. Jutten, J. M. T. Romano, and A. K. Barros, Eds. Springer Berlin Heidelberg, 2009, pp. 354–361.
- [39] T. Bendory, Y. C. Eldar, and N. Boumal, “Non-convex phase retrieval from stft measurements,” *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 467–484, 2018.