Ellipsoidal Manifold Optimization for Distributed Antenna Beamforming

Minhao Zhu and Kaiming Shen
School of Science and Engineering and FNii, The Chinese University of Hong Kong, Shenzhen, China
Email: minhaozhu@link.cuhk.edu.cn, shenkaiming@cuhk.edu.cn

***Abstract*—This paper considers the weighted sum-rate maximization for a cellular network with distributed antenna clusters. This optimization problem is difficult because of the high-dimensional variables as well as the per-cluster power constraints. To overcome these difficulties, we generalize the low-dimensional subspace property, which is known for the sum-power constraint, to the per-cluster power constraint case. By this property, we reformulate the original high-dimensional constrained problem into an unconstrained optimization task over a product of ellipsoidal manifolds, thereby achieving significant dimensionality reduction. We systematically derive the necessary Riemannian geometric structures for this specific manifold, including the tangent space, Riemannian metric, orthogonal projection, retraction, and vector transport. Subsequently, we develop a tailored Riemannian conjugate gradient algorithm to solve the reformulated problem. Numerical simulations demonstrate that the proposed algorithm outperforms the standard benchmarks in the presence of multiple antenna clusters.**

I. Introduction

Distributed antenna systems have emerged as a promising architecture for next-generation wireless networks, where transmit antennas are geographically separated rather than co-located at a single base station (BS) [2], [3]. By shortening the access distance between users and antennas, distributed deployments can improve coverage and enhance received signal strength. In such systems, beamforming design differs fundamentally from the conventional co-located case because the transmit power is typically constrained on each antenna cluster rather than by a single sum-power budget. These per-cluster power constraints, arising from hardware and deployment limitations, introduce additional structure and make weighted sum-rate (WSR) maximization significantly more challenging in dense antenna systems.

While the traditional study [4]–[10] on beamforming focuses on the sum-power constraint, there are increasing research interests in the per-cluster power constraint. Per-cluster power constraint arises naturally in distributed antenna and coordinated multipoint (CoMP) architectures [3], [11], where geographically separated antenna clusters have independent hardware or regulatory power budgets. Existing studies have examined this setting from several perspectives, including stochastic downlink analysis [3], fractional-programming-based integrated sensing and communication (ISAC) beamforming under arbitrary per-cluster constraints [12], minimum-mean-square-error precoding with per-antenna-group constraints [13], and penalty-based convex formulations for per-antenna and per-cluster power limits [14].

For WSR maximization, classical precoders such as maximum ratio transmission (MRT) [15], zero-forcing (ZF) [16], and eigen zero-forcing [17] are low-complexity but generally suboptimal. More effective iterative methods such as weighted minimum mean square error (WMMSE) [4], [5] and fractional programming (FP) [6]–[9] usually have a complexity that scales poorly with N_t. Although reduced-complexity variants have been developed for the sum-power setting [10], [18], their underlying structural simplifications do not directly apply to per-cluster power constraints.

We address the beamforming problem through manifold optimization, a powerful framework for constrained optimization [19], [20]. Recent communication applications include beamforming under a sum-power constraint [21], joint active and passive beamforming for reconfigurable intelligent surface (RIS)-assisted systems [22], and robust beamforming under channel uncertainty [23]. These works mainly rely on spherical manifold formulations, which are convenient when the transmit power is governed by a single global constraint. However, direct optimization on high-dimensional manifolds becomes costly when the number of transmit antennas N_t is large. Moreover, the low-dimensional subspace property in [18], which reduces the search space from N_t to KN_r, where K is the number of users and N_r is the number of receive antennas per user, has only been established for the sum-power case. Motivated by this, we generalize the property to the per-cluster power constraint case and reformulate the beamforming problem as an optimization over an ellipsoidal manifold.

Motivated by the above discussion, the main contributions of this paper are summarized as follows:

- We establish a low-dimensional subspace property for WSR maximization under per-cluster power constraints, showing that stationary beamformers of each antenna cluster lie in the subspace spanned by the corresponding cluster channels.
- Using this property, we reformulate the original problem as optimization over a product of ellipsoidal manifolds and derive the associated Riemannian geometric ingredients, including projection, retraction, and vector transport.
- We develop a reduced Riemannian conjugate gradient algorithm and show numerically that it preserves the so-

Proposition proofs can be found in the long version [1] available at https://arxiv.org/abs/2603.23260. The work was supported in part by the NSFC under Grant 12426306 and in part by the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone.

TABLE I
NOTATIONS AND DEFINITIONS.

Symbol	Definition
$\boldsymbol{A}^{\mathrm{H}}$	conjugate transpose
$\boldsymbol{A}^{\top}$	transpose
$\|\boldsymbol{A}\|_F$	Frobenius norm
$\Re(\boldsymbol{A})$	taking real part of each entry
$\nabla f(x)$	Euclidean gradient of $f(x)$
$\mathcal{M}$	manifold
$T_{\boldsymbol{X}}\mathcal{M}$	tangent space of $\mathcal{M}$ at point $\boldsymbol{X}$
$N_{\boldsymbol{X}}\mathcal{M}$	Normal space of $\mathcal{M}$ at $\boldsymbol{X}$
$\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle_{\boldsymbol{X}}$	Riemannian metric at $\boldsymbol{X}$
$\operatorname{grad} f(\boldsymbol{X})$	Riemannian gradient of $f(\boldsymbol{X})$
$\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_n$	product manifold

lution quality of benchmark methods while significantly improving computational efficiency and scalability.

Table I summarizes the main notation.

II. System Model

Assume that the BS is associated with K downlink user terminals. The BS has N_t transmit antennas while each user terminal has N_r receive antennas. Further, assume that the N_t transmit antennas are divided into C clusters; these clusters are placed at distributed positions. Note that each cluster has its own power constraint P. Each cluster comprises L antennas, so we have $LC = N_t$.

The channel from the cth antenna cluster to user k is denoted by $\boldsymbol{H}_k^{(c)} \in \mathbb{C}^{N_r \times L}$, for $c = 1, \ldots, C$. These channel matrices are stacked horizontally with respect to each user k as

$$\boldsymbol{H}_k = \begin{bmatrix} \boldsymbol{H}_k^{(1)} & \boldsymbol{H}_k^{(2)} & \ldots & \boldsymbol{H}_k^{(C)} \end{bmatrix} \in \mathbb{C}^{N_r \times N_t}. \tag{1}$$

Moreover, the channel matrices are stacked vertically with respect to each antenna cluster c as

$$\boldsymbol{H}^{(c)} = \begin{bmatrix} \boldsymbol{H}_1^{(c)} \\ \vdots \\ \boldsymbol{H}_K^{(c)} \end{bmatrix} \in \mathbb{C}^{KN_r \times L}, \tag{2}$$

which is assumed to be full row rank, i.e., $\operatorname{rank}(\boldsymbol{H}^{(c)}) = KN_r$.

Denote by $d_k \in \min\{N_t, N_r\}$ the number of downlink data streams assigned to user k. Let $\boldsymbol{V}_k^{(c)} \in \mathbb{C}^{L \times d_k}$ be the transmit beamformer of the cth cluster of antennas for user k. The above beamforming matrices are vertically stacked for each user k as

$$\boldsymbol{V}_k = \begin{bmatrix} \boldsymbol{V}_k^{(1)} \\ \vdots \\ \boldsymbol{V}_k^{(C)} \end{bmatrix} \in \mathbb{C}^{N_t \times d_k}, \tag{3}$$

and are horizontally stacked for each antenna cluster c as

$$\boldsymbol{V}^{(c)} = \begin{bmatrix} \boldsymbol{V}_1^{(c)}, \ldots, \boldsymbol{V}_K^{(c)} \end{bmatrix} \in \mathbb{C}^{L \times Kd_k}. \tag{4}$$

Moreover, denote by $\boldsymbol{s}_k \in \mathbb{C}^{d_k}$ the modulated signal vector intended for user k, and $\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \sigma_k^2 \boldsymbol{I})$ the background noise at user k, where σ_k^2 is the noise power. Thus, the received signal of user k can be computed as

$$\boldsymbol{y}_k = \underbrace{\boldsymbol{H}_k \boldsymbol{V}_k \boldsymbol{s}_k}_{\text{desired signal}} + \underbrace{\boldsymbol{H}_k \sum_{j=1, j \neq k}^{K} \boldsymbol{V}_j \boldsymbol{s}_j}_{\text{multi-user interference}} + \boldsymbol{n}_k. \tag{5}$$

An achievable rate R_k for user k is given by

$$R_k = \log \det \left(\boldsymbol{I} + \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{F}_k^{-1} \boldsymbol{H}_k \boldsymbol{V}_k \right), \tag{6}$$

where

$$\boldsymbol{F}_k = \sigma_k^2 \boldsymbol{I} + \sum_{j=1, j \neq k}^{K} \boldsymbol{H}_k \boldsymbol{V}_j \boldsymbol{V}_j^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}}. \tag{7}$$

Each user k is assigned a data rate weight $\omega > 0$ in accordance with its priority. We seek the optimal beamformers $\{\boldsymbol{V}_k\}$ to maximize the weighted sum rates

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{k=1}^{K} \omega_k R_k \tag{8a}$$

$$\text{subject to} \quad \|\boldsymbol{V}^{(c)}\|_F^2 \leq P_c, \text{ for } c = 1, \ldots, C, \tag{8b}$$

where P_c is the power constraint on the cth antenna cluster, and $\boldsymbol{V}$ is the variable tuple $\left[(\boldsymbol{V}^{(1)})^{\top}, \ldots, (\boldsymbol{V}^{(C)})^{\top} \right]^{\top}$.

III. Conversion to Manifold Problems

The goal of this section is to recast problem (8) into a manifold optimization problem in order to get rid of the power constraint. We first show that the power constraint (8b) must be tight for a stationary point solution, so it boils down to limiting the solution to a spherical manifold. Next, we propose further transforming the spherical manifold into an ellipsoidal manifold, thereby reducing the dimension and complexity of the beamforming problem.

A. Spherical Manifold

The key step is to show that the equality must hold in (8b) when we have arrived at a stationary point solution of the problem, as specified in the following proposition.

Proposition 1: For any stationary solution $\boldsymbol{V}^*$ to problem (8), the power constraint must be tight, i.e.,

$$\|(\boldsymbol{V}^{(c)})^*\|_F^2 = P_c, \text{ for } c = 1, \ldots, C. \tag{9}$$

We remark that the special case of the above proposition when $C = 1$ has been shown in the literature [18], [21]. Actually, this special case is quite easy to see: if $\|\boldsymbol{V}^{(1)}\|_F^2 < P_1$, then we can scale up all the $\boldsymbol{V}_k^{(1)}$'s simultaneously to further increase R_k, so the current point cannot be a stationary point. However, the above argument no longer works when $C \geq 2$. Thus, the result of Proposition (1) is a nontrivial generalization.

We then rewrite problem (8) as

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{k=1}^{K} \omega_k R_k \tag{10a}$$

$$\text{subject to} \quad \|\boldsymbol{V}^{(c)}\|_F^2 = P_c, \text{ for } c = 1, \ldots, C. \tag{10b}$$

The new constraint (10b) can be interpreted as a manifold condition. Specifically, each power constraint P_c induces a spherical manifold:

$$\mathcal{S}_c = \left\{ \boldsymbol{V}^{(c)} : \|\boldsymbol{V}_k^{(c)}\|_F^2 = P_c \right\}. \tag{11}$$

Accordingly, (10b) boils down to a product manifold:

$$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_C. \tag{12}$$

As such, the constrained problem (10) can be rewritten as an unconstrained problem with $\boldsymbol{V}$ limited to $\mathcal{S}$:

$$\underset{\boldsymbol{V}\in\mathcal{S}}{\text{minimize}} \quad -\sum_{k=1}^{K} \omega_k R_k. \tag{13}$$

Note that we write this new problem in the minimization form by the convention of manifold optimization [19].

For the single-cluster case with $C = 1$ in [21], $\mathcal{S}$ reduces to a spherical manifold—which has been extensively considered in the literature [18]. For the general product manifold $\mathcal{S}$ with $C \geq 2$, we can extend the existing algorithm for it because it is an embedded submanifold of the ambient Euclidean space $\mathbb{C}^{N_t \times K d_k}$ [19].

B. Ellipsoidal Manifold

From a manifold optimization perspective, the new problem (13) is easier to tackle since the constraint has been removed. We now show that the problem can be further simplified by converting the spherical manifold $\mathcal{S}$ to an ellipsoidal manifold of a lower dimension.

Introduce a new variable $\boldsymbol{X}_k^{(c)} \in \mathbb{C}^{N_r \times d_k}$ that corresponds to each $\boldsymbol{V}_k^{(c)} \in \mathbb{C}^{L \times d_k}$. Similarly, we define

$$\boldsymbol{X}_k = \begin{bmatrix} \boldsymbol{X}_k^{(1)} \\ \vdots \\ \boldsymbol{X}_k^{(C)} \end{bmatrix} \in \mathbb{C}^{CN_r \times d_k} \tag{14}$$

and

$$\boldsymbol{X}^{(c)} = \left[\boldsymbol{X}_1^{(c)}, \ldots, \boldsymbol{X}_K^{(c)} \right] \in \mathbb{C}^{KN_r \times K d_k}, \tag{15}$$

and the tuple $(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(C)})$ is denoted by $\boldsymbol{X}$. By substituting

$$\boldsymbol{V}_k^{(c)} = (\boldsymbol{H}^{(c)})^{\mathrm{H}} \boldsymbol{X}_k^{(c)} \tag{16}$$

into problem (13), we obtain a manifold problem of $\boldsymbol{X}$:

$$\underset{\boldsymbol{X}\in\mathcal{M}}{\text{minimize}} \quad -\sum_{k=1}^{K} \omega_k \hat{R}_k, \tag{17}$$

where the data rate is now computed as

$$\hat{R}_k = \log\det\Bigg(\boldsymbol{I} + \boldsymbol{X}_k^{\mathrm{H}} \boldsymbol{G}_k^{\mathrm{H}} \Bigg(\sigma_k^2 \boldsymbol{I} + \sum_{j=1, j\neq k}^{K} \boldsymbol{G}_k \boldsymbol{X}_j \boldsymbol{X}_j^{\mathrm{H}} \boldsymbol{G}_k^{\mathrm{H}} \Bigg)^{-1} \boldsymbol{G}_k \boldsymbol{X}_k \Bigg) \tag{18}$$

with

$$\boldsymbol{G}_k = \left[\boldsymbol{H}_k^{(1)} (\boldsymbol{H}^{(1)})^{\mathrm{H}}, \ldots, \boldsymbol{H}_k^{(C)} (\boldsymbol{H}^{(C)})^{\mathrm{H}} \right], \tag{19}$$

and the product manifold now becomes

$$\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_C, \tag{20}$$

where each $\mathcal{M}_c$ is an ellipsoidal manifold:

$$\mathcal{M}_c = \left\{ \boldsymbol{X}^{(c)} : \mathrm{tr}\left((\boldsymbol{X}^{(c)})^{\mathrm{H}} \boldsymbol{Q}^{(c)} \boldsymbol{X}^{(c)} \right) = P_c \right\}, \tag{21}$$

where

$$\boldsymbol{Q}^{(c)} = \boldsymbol{H}^{(c)} (\boldsymbol{H}^{(c)})^{\mathrm{H}}. \tag{22}$$

Proposition 2: The set $\mathcal{M}_c$ defined in (21) is a smooth embedded submanifold of the ambient Euclidean space $\mathbb{C}^{KN_r \times K d_k}$.

The following proposition is the building block of this section:

Proposition 3: Problem (17) is equivalent to problem (10) in the sense that $\boldsymbol{X}$ is a stationary point solution of problem (17) if and only if the corresponding $\boldsymbol{V}$ is a stationary point solution of problem (10).

We remark that the above theorem is foreshadowed by the result of [18], which aims to reduce the variable dimension for the WMMSE algorithm when $C = 1$. Most importantly, because $N_r \ll L$, the dimension and complexity of the manifold optimization in (17) are much lower than those of the spherical manifold case in (13).

IV. Proposed Manifold Optimization Algorithm

We propose to use the Riemannian conjugate gradient method [19] to address the manifold optimization problem in (17). In this section, we first specify the geometric elements and gradient computation for the ellipsoidal product manifold $\mathcal{M}$ in (20), and then summarize the proposed algorithm.

A. Geometry of the Ellipsoidal Manifold

For each component manifold $\mathcal{M}_c$, the tangent space at $\boldsymbol{X}^{(c)} \in \mathcal{M}_c$ is

$$T_{\boldsymbol{X}^{(c)}}\mathcal{M}_c = \left\{ \boldsymbol{\xi}^{(c)} : \Re\left\{ \mathrm{tr}\left(\boldsymbol{Q}^{(c)} \left(\boldsymbol{\xi}^{(c)} (\boldsymbol{X}^{(c)})^{\mathrm{H}} \right) \right) \right\} = 0 \right\}, \tag{23}$$

and the normal space is

$$N_{\boldsymbol{X}^{(c)}}\mathcal{M}_c = \left\{ \boldsymbol{\eta}^{(c)} : \boldsymbol{\eta}^{(c)} = \lambda \boldsymbol{Q}^{(c)} \boldsymbol{X}^{(c)}, \lambda \in \mathbb{R} \right\}. \tag{24}$$

For any two tangent vectors $\boldsymbol{\xi}^{(c)}, \boldsymbol{\eta}^{(c)} \in T_{\boldsymbol{X}^{(c)}}\mathcal{M}_c$, the Riemannian metric is

$$\left\langle \boldsymbol{\xi}^{(c)}, \boldsymbol{\eta}^{(c)} \right\rangle_{\boldsymbol{X}^{(c)}} = \Re\left\{ \mathrm{tr}\left((\boldsymbol{\xi}^{(c)})^{\mathrm{H}} \boldsymbol{\eta}^{(c)} \right) \right\}. \tag{25}$$

The orthogonal projection of an arbitrary vector $\boldsymbol{Z}^{(c)} \in \mathbb{C}^{N_r \times K d_k}$ onto the tangent space $T_{\boldsymbol{X}^{(c)}}\mathcal{M}_c$ is given by

$$\mathcal{P}_{\boldsymbol{X}^{(c)}}(\boldsymbol{Z}^{(c)}) = \boldsymbol{Z}^{(c)} - \lambda \boldsymbol{Q}^{(c)} \boldsymbol{X}^{(c)}, \tag{26}$$

where λ is chosen such that $\mathcal{P}_{\boldsymbol{X}^{(c)}}(\boldsymbol{Z}^{(c)}) \in T_{\boldsymbol{X}^{(c)}}\mathcal{M}_c$, i.e.,

$$\lambda = \frac{\mathrm{tr}\left(\boldsymbol{Q}^{(c)} \left(\boldsymbol{Z}^{(c)} (\boldsymbol{X}^{(c)})^{\mathrm{H}} + \boldsymbol{X}^{(c)} (\boldsymbol{Z}^{(c)})^{\mathrm{H}} \right) \right)}{2\,\mathrm{tr}\left(\boldsymbol{Q}^{(c)} \boldsymbol{Q}^{(c)} \boldsymbol{X}^{(c)} (\boldsymbol{X}^{(c)})^{\mathrm{H}} \right)}. \tag{27}$$

$$\boldsymbol{X}_\ell^{(c)} = \mathcal{R}(\boldsymbol{\eta}_\ell^{(c)}, \alpha_\ell) = \sqrt{P_c \Big/ \operatorname{tr}\left(\boldsymbol{Q}^{(c)} \left(\boldsymbol{X}_{\ell-1}^{(c)} + \alpha_\ell \boldsymbol{\eta}_\ell^{(c)}\right)\left(\boldsymbol{X}_{\ell-1}^{(c)} + \alpha_\ell \boldsymbol{\eta}_\ell^{(c)}\right)^{\mathrm{H}}\right)}\left(\boldsymbol{X}_{\ell-1}^{(c)} + \alpha_\ell \boldsymbol{\eta}_\ell^{(c)}\right) \tag{29}$$

$$\lambda' = \frac{\operatorname{tr}\left(\boldsymbol{Q}^{(c)}\left(\boldsymbol{\xi}^{(c)}(\boldsymbol{X}^{(c)} + \boldsymbol{\eta}^{(c)})^{\mathrm{H}} + (\boldsymbol{X}^{(c)} + \boldsymbol{\eta}^{(c)})(\boldsymbol{\xi}^{(c)})^{\mathrm{H}}\right)\right)}{2\operatorname{tr}\left(\boldsymbol{Q}^{(c)}\boldsymbol{Q}^{(c)}(\boldsymbol{X}^{(c)} + \boldsymbol{\eta}^{(c)})(\boldsymbol{X}^{(c)} + \boldsymbol{\eta}^{(c)})^{\mathrm{H}}\right)} \tag{31}$$

$$\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{Z}) = \begin{bmatrix} \mathcal{P}_{\boldsymbol{X}^{(1)}}(\boldsymbol{Z}^{(1)}) \\ \vdots \\ \mathcal{P}_{\boldsymbol{X}^{(C)}}(\boldsymbol{Z}^{(C)}) \end{bmatrix}, \quad \mathcal{R}(\boldsymbol{\eta}, \alpha_\ell) \;= \begin{bmatrix} \mathcal{R}(\boldsymbol{\eta}^{(1)}, \alpha_\ell) \\ \vdots \\ \mathcal{R}(\boldsymbol{\eta}^{(C)}, \alpha_\ell) \end{bmatrix}, \quad \mathcal{T}(\boldsymbol{\eta}, \alpha_\ell) = \begin{bmatrix} \mathcal{T}(\boldsymbol{\eta}^{(1)}, \alpha_\ell) \\ \vdots \\ \mathcal{T}(\boldsymbol{\eta}^{(C)}, \alpha_\ell) \end{bmatrix} \tag{32}$$

$$\beta_\ell = \frac{\langle \operatorname{grad} f(\boldsymbol{X}_{\ell-1}),\ \operatorname{grad} f(\boldsymbol{X}_{\ell-1}) - \mathcal{T}(\operatorname{grad} f(\boldsymbol{X}_{\ell-2}), \alpha_{\ell-2})\rangle_{\boldsymbol{X}_{\ell-1}}}{\langle \mathcal{T}(\boldsymbol{\eta}_{\ell-1}, \alpha_{\ell-2}),\ \operatorname{grad} f(\boldsymbol{X}_{\ell-1}) - \mathcal{T}(\operatorname{grad} f(\boldsymbol{X}_{\ell-2}), \alpha_{\ell-2})\rangle_{\boldsymbol{X}_{\ell-1}}} \tag{38}$$

Given a search direction $\boldsymbol{\eta}_\ell^{(c)} \in T_{\boldsymbol{X}^{(c)}}\mathcal{M}_c$ and stepsize $\alpha_\ell > 0$ in ℓth iteration, we update $\boldsymbol{X}_\ell^{(c)}$ as

$$\hat{\boldsymbol{X}}_\ell^{(c)} = \boldsymbol{X}_{\ell-1}^{(c)} + \alpha_\ell \boldsymbol{\eta}_\ell^{(c)}. \tag{28}$$

We then retract $\hat{\boldsymbol{X}}_\ell^{(c)}$ back onto $\mathcal{M}_c$ using the retraction defined in (29) at the top of the page.

To transport $\boldsymbol{\eta}_{\ell-1}^{(c)}$ from $T_{\boldsymbol{X}_{\ell-2}^{(c)}}\mathcal{M}_c$ to $T_{\boldsymbol{X}_{\ell-1}^{(c)}}\mathcal{M}_c$, we compute the transport as

$$\begin{aligned} \mathcal{T}(\boldsymbol{\eta}_{\ell-1}^{(c)}, \alpha_{\ell-1}) &= \mathcal{P}_{\mathcal{R}(\boldsymbol{\eta}_{\ell-1}^{(c)}, \alpha_{\ell-1})}\left(\boldsymbol{\eta}_{\ell-1}^{(c)}\right) \\ &= \boldsymbol{\eta}_{\ell-1}^{(c)} - \lambda' \boldsymbol{Q}^{(c)}\left(\boldsymbol{X}_{\ell-1}^{(c)} + \boldsymbol{\eta}_{\ell-1}^{(c)}\right), \end{aligned} \tag{30}$$

where λ' is given in (31) at the top of the page. For the product manifold $\mathcal{M}$, the projection, retraction, and transport are computed component-wise in (32). The Riemannian metric on the product manifold $\mathcal{M}$ is defined as

$$\langle \boldsymbol{\xi}, \boldsymbol{\eta}\rangle_{\boldsymbol{X}} = \sum_{c=1}^{C} \left\langle \boldsymbol{\xi}^{(c)}, \boldsymbol{\eta}^{(c)} \right\rangle_{\boldsymbol{X}^{(c)}} = \sum_{c=1}^{C} \Re\left\{\operatorname{tr}\left((\boldsymbol{\xi}^{(c)})^{\mathrm{H}}\boldsymbol{\eta}^{(c)}\right)\right\}. \tag{33}$$

B. Riemannian Gradient

Define the objective function in (17) as $f(\boldsymbol{X}) := -\sum_{k=1}^{K}\omega_k \hat{R}_k$. Recall the tuple $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K]$, we first compute the Euclidean gradient of $f(\boldsymbol{X})$ with respect to each variable $\boldsymbol{X}_k$. The Euclidean gradient is $\nabla f(\boldsymbol{X}) = [\nabla f(\boldsymbol{X}_1), \ldots, \nabla f(\boldsymbol{X}_K)]$, where $\nabla f(\boldsymbol{X}_k)$ is given by

$$\nabla f(\boldsymbol{X}_k) = -\omega_k \nabla_{\boldsymbol{X}_k}\hat{R}_k - \sum_{i\neq k}^{K} \omega_i \nabla_{\boldsymbol{X}_k}\hat{R}_i. \tag{34}$$

Note that $\nabla f(\boldsymbol{X}_k)$ is vertically stacked as

$$\nabla f(\boldsymbol{X}_k) = \left[\nabla f(\boldsymbol{X}_k^{(1)})^\top, \nabla f(\boldsymbol{X}_k^{(2)})^\top, \ldots, \nabla f(\boldsymbol{X}_k^{(C)})^\top\right]^\top. \tag{35}$$

The Riemannian gradient is then obtained by projection:

$$\operatorname{grad} f(\boldsymbol{X}) = \mathcal{P}_{\boldsymbol{X}}(\nabla f(\boldsymbol{X})). \tag{36}$$

Algorithm 1 Proposed Distributed Antenna Beamforming

1: Initialize $\boldsymbol{X}_0 \in \mathcal{M}$, set $\ell = 1$, and choose parameters $\sigma, p \in (0,1)$, $\alpha^0 > 0$.
2: Set the initial search direction $\boldsymbol{\eta}_0 = -\operatorname{grad} f(\boldsymbol{X}_0)$.
3: **repeat**
4: Compute the Riemannian gradient $\operatorname{grad} f(\boldsymbol{X}_{\ell-1})$ using (36).
5: **if** $\ell = 1$ **then**
6: Set the search direction $\boldsymbol{\eta}_1 = -\operatorname{grad} f(\boldsymbol{X}_0)$.
7: **else**
8: Compute the vector transport $\mathcal{T}(\boldsymbol{\eta}_{\ell-1}, \alpha_{\ell-1})$.
9: Compute β_ℓ using (39).
10: Update the search direction $\boldsymbol{\eta}_\ell$ using (37).
11: **end if**
12: Find stepsize α_ℓ via backtracking line search with the Armijo condition (40).
13: Retract onto the manifold using (32) and obtain $\boldsymbol{X}_\ell$.
14: Increment $\ell \leftarrow \ell + 1$.
15: **until** the value of $f(\boldsymbol{X}_\ell)$ converges

C. Conjugate Search Direction and Line Search

We construct the search direction $\boldsymbol{\eta}_\ell$ as

$$\boldsymbol{\eta}_\ell = -\operatorname{grad} f(\boldsymbol{X}_{\ell-1}) + \beta_\ell \mathcal{T}(\boldsymbol{\eta}_{\ell-1}, \alpha_{\ell-1}), \tag{37}$$

where β_ℓ is the Hestenes-Stiefel parameter [24] given in (38) at the top of page. To ensure descent, we set

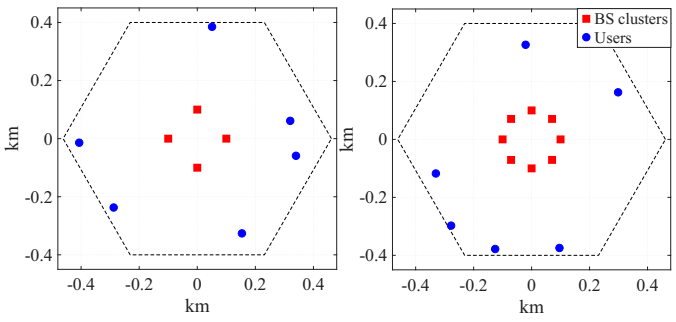
$$\beta_\ell = \begin{cases} \max\{\beta_\ell^{HS}, 0\}, & \text{if } \langle \operatorname{grad} f(\boldsymbol{X}_{\ell-1}), \boldsymbol{\eta}_\ell\rangle_{\boldsymbol{X}_{\ell-1}} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

The stepsize α_ℓ is determined by Armijo backtracking: find the smallest non-negative integer m such that

$$\begin{aligned} f(\mathcal{R}(\boldsymbol{\eta}_\ell, \sigma^m\alpha^0)) \leq f(\boldsymbol{X}_{\ell-1}) \\ + \sigma^m p\ \alpha^0 \langle \operatorname{grad} f(\boldsymbol{X}_{\ell-1}), \boldsymbol{\eta}_\ell\rangle_{\boldsymbol{X}_{\ell-1}}, \end{aligned} \tag{40}$$

and set $\alpha_\ell = \sigma^m\alpha^0$.

The proposed method is summarized in Algorithm 1. The



(a) $C = 4$ clusters (b) $C = 8$ clusters

Fig. 1. Network topology with hexagonal cell for $C = 4$ clusters and $C = 8$ clusters.

convergence analysis can be found in [1].

V. Numerical Results

A. Simulation Setup

In this section, we evaluate the proposed algorithm in a single-cell downlink with a hexagonal cell of radius 400 m, where users are randomly placed as illustrated in Fig. 1. The C BS clusters are uniformly placed on a circle around the cell center. Each user is equipped with $N_r = 4$ receive antennas and receives $d_k = 4$ streams, while each cluster has $L = 128$ transmit antennas, so $N_t = CL$; the parameters are chosen such that $N_t \geq CKN_r$. The path-loss is modeled as $128.1 + 37.6 \log_{10} d + \psi$ dB, where d is the BS-to-user distance in km and ψ is a zero-mean Gaussian random variable with standard deviation 8 dB; Rayleigh fading is assumed for small-scale fading. The noise power is $\sigma_k^2 = -80$ dBm, all user weights are set to $\omega_k = 1$, and Armijo backtracking uses $\sigma = 0.6$, $p = 0.1$, and $t^0 = 10^{10}$. All results are averaged over 1000 independent realizations.

We benchmark our proposed algorithm against the following algorithms.

- **WMMSE** [5]: a standard BCD benchmark generalized to per-cluster power constraints by solving the precoder update under convex constraints.
- **RWMMSE** [18]: a reduced version of WMMSE algorithm adapted to the per-cluster power constraints through the low-dimensional reformulation.
- **Riemannian conjugate gradient** [21]: the conventional Riemannian conjugate gradient method applied to the original product manifold formulation (10).
- **EZF** [17]: a low-complexity non-iterative linear precoding baseline by performing singular value decomposition (SVD) on individual user channels.

B. Performance Under Clustered Antenna Configurations

We evaluate clustered settings with $C \in \{4, 8\}$ and $K = 6$, where $N_t = C \times L$. The cluster power budgets are $\boldsymbol{P} = [100, 500, 100, 500]$ for $C = 4$ and $\boldsymbol{P} = [100, 500, 100, 500, 100, 500, 100, 500]$ for $C = 8$. Fig. 2(a) and Fig. 2(b) depict the WSR evolution versus elapsed time for the cases of $C = 4$ and $C = 8$, respectively. In both settings, the proposed algorithm converges to essentially the same local optimum as WMMSE and the conventional Riemannian conjugate gradient method, which verifies that the low-dimensional reformulation does not compromise solution quality. By contrast, RWMMSE converges to a lower WSR. More importantly, the proposed method reaches the converged WSR with less runtime than the conventional Riemannian conjugate gradient method. Although it may require a comparable or slightly larger number of iterations, each iteration is substantially cheaper because the optimization is carried out in the reduced-dimensional variable space. This runtime advantage becomes more pronounced when the number of clusters increases from $C = 4$ in Fig. 2(a) to $C = 8$ in Fig. 2(b), since the dimension of the original beamforming variable grows with $N_t = CL$, whereas the reduced variable remains much smaller.

To compare solution quality under a fixed computational budget, Fig. 2(c) plots the cumulative distribution function (CDF) of the achieved WSR when the running time is limited to 2.5 seconds for $C = 8$. WMMSE is omitted because its high per-iteration complexity and CVX calls prevent meaningful progress within this budget. The proposed algorithm achieves the highest WSR over most of the CDF range, confirming that its lower per-iteration cost translates into better real-time performance.

Finally, Fig. 3(a) and Fig. 3(b) compare the convergence time of the proposed algorithm and Riemannian conjugate gradient, which is defined as the elapsed time until $\left|\mathrm{WSR}^{(\ell)} - \mathrm{WSR}^{(\ell-1)}\right| < 10^{-3}$. In Fig. 3(a), the proposed algorithm is faster for all tested values of K, with larger gains when K is small and the reduced variable dimension is much smaller than that of Riemannian conjugate gradient. In Fig. 3(b), the proposed algorithm is also faster for all tested values of L; moreover, the runtime of Riemannian conjugate gradient increases much more sharply with L, whereas that of the proposed algorithm grows only mildly. These results further demonstrate the scalability benefit of the reduced formulation.

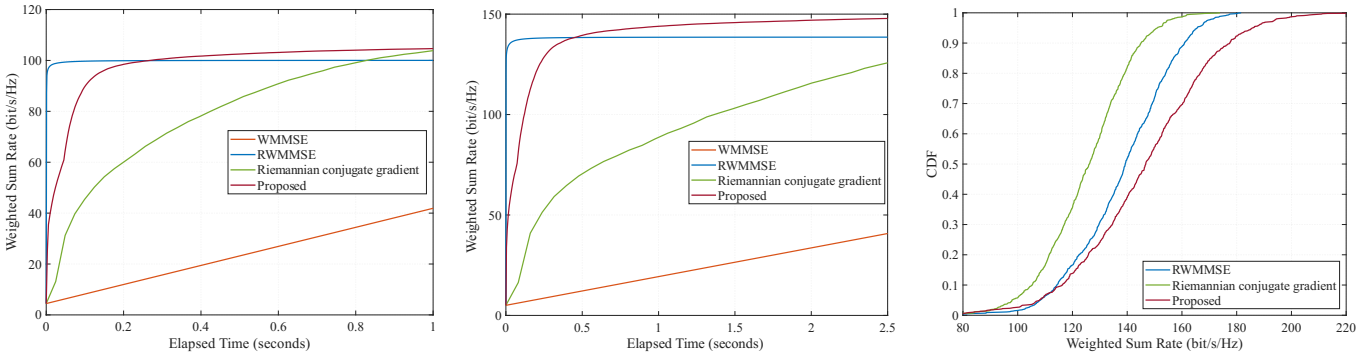
VI. Conclusion

This paper studied WSR maximization for distributed antenna beamforming under per-cluster power constraints. By extending the low-dimensional subspace property to this setting, we reformulated the original problem as optimization over a product of ellipsoidal manifolds and developed a distributed antenna beamforming algorithm. Numerical results showed that the proposed method achieves the same local optima as benchmark algorithms such as WMMSE and conventional Riemannian conjugate gradient, while offering substantially lower runtime and better scalability as the number of antenna clusters or antennas per cluster increases.

References

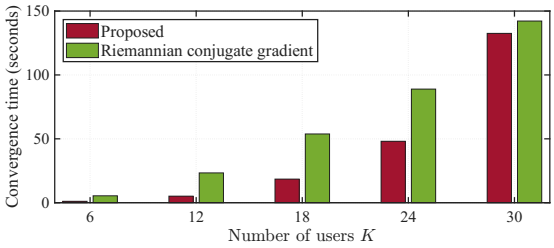
[1] M. Zhu and K. Shen, "Ellipsoidal manifold optimization for distributed antenna beamforming," Mar. 2026, [Online]. Available: https://arxiv.org/abs/2603.23260.

[2] K. Kerpez, "A radio access system with distributed antennas," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 265–275, 1996.

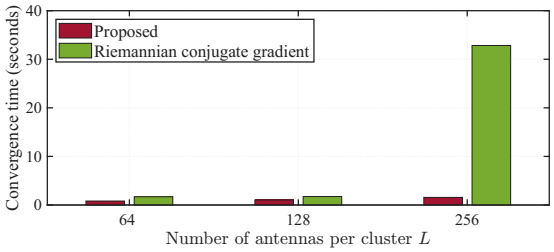


(a) Convergence in elapsed time when $C = 4$. (b) Convergence in elapsed time when $C = 8$. (c) CDF of WSR when $C = 8$.

Fig. 2. Convergence performance with $L = 128, K = 6, N_r = 4, d_k = 4$. Fig. 2(a) and Fig. 2(b) show the convergence of WSR in elapsed time under cluster configurations with $C = 4$ and $C = 8$, respectively. Fig. 2(c) shows the CDF of the WSR achieved by different algorithms with a fixed running time.



(a) Convergence time with varying K.



(b) Convergence time with varying L.

Fig. 3. Convergence time under per-cluster power constraints with $C = 8, N_r = 2, d_k = 2$ and varying K and L.

[3] Y. Lin and W. Yu, "Downlink spectral efficiency of distributed antenna systems under a stochastic model," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6891–6902, 2014.

[4] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.

[5] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[6] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, 2018.

[7] ——, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, 2018.

[8] K. Shen, W. Yu, L. Zhao, and D. P. Palomar, "Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization–maximization approach," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 2164–2177, 2019.

[9] K. Shen, Z. Zhao, Y. Chen, Z. Zhang, and H. Victor Cheng, "Accelerating quadratic transform and WMMSE," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 11, pp. 3110–3124, 2024.

[10] Y. Chen, Y. Feng, X. Li, L. Zhao, and K. Shen, "Fast fractional programming for multi-cell integrated sensing and communications," *IEEE Trans. Wireless Commun.*, vol. 24, no. 8, pp. 6797–6812, 2025.

[11] J. Zhao, T. Q. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, p. 2762–2775, 06 2013.

[12] D. Kim and S.-H. Park, "Accelerated fractional programming for multi-user ISAC beamforming optimization with per-group power constraints," *IEEE Wireless Commun. Lett.*, vol. 14, no. 4, pp. 1174–1178, 2025.

[13] M. Mazrouei-Sebdani and W. A. Krzymień, "On MMSE vector-perturbation precoding for MIMO broadcast channels with per-antenna-group power constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3745–3751, 2013.

[14] K. Wang and X. Zhang, "Penalty function-based precoding for downlink multiuser MIMO systems," *AEU-Int. J. Electron. Commun.*, vol. 67, no. 2, pp. 167–173, 2013.

[15] T. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, 1999.

[16] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, 2011, pp. 1–5.

[17] L. Sun and M. R. McKay, "Eigen-based transceivers for the MIMO broadcast channel with semi-orthogonal user selection," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5246–5261, 2010.

[18] X. Zhao, S. Lu, Q. Shi, and Z.-Q. Luo, "Rethinking WMMSE: Can its complexity scale linearly with the number of BS antennas?" *IEEE Trans. Signal Process.*, vol. 71, pp. 433–446, 2023.

[19] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[20] N. Boumal, *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.

[21] R. Sun, C. Wang, A.-A. Lu, X. Gao, and X.-G. Xia, "Precoder design for massive MIMO downlink with matrix manifold optimization," *IEEE Trans. Signal Process.*, vol. 72, pp. 1065–1080, 2024.

[22] E. Shtaiwi, H. Zhang, A. Abdelhadi, A. L. Swindlehurst, Z. Han, and H. V. Poor, "Sum-rate maximization for RIS-assisted integrated sensing and communication systems with manifold optimization," *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4909–4923, 2023.

[23] C. Wang, A.-A. Lu, X. Gao, and Z. Ding, "Robust precoding for 3d massive MIMO configuration with matrix manifold optimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3423–3437, 2022.

[24] Z. Salleh and A. Alhawarat, "An efficient modification of the hestenes-stiefel nonlinear conjugate gradient method with restart property," *J. Inequal. Appl.*, vol. 2016, no. 1, p. 110, Apr 2016.