Energy-Efficient Power Control and Beamforming for Reconfigurable Intelligent Surface-Aided Uplink IoT Networks

Jiao Wu, Member, IEEE, Seungnyun Kim, Member, IEEE, and Byonghyo Shim, Senior Member, IEEE

Abstract—Recently, reconfigurable intelligent surface (RIS), a planar metasurface consisting of a large number of low-cost reflecting elements, has received much attention due to its ability to improve both the spectrum and energy efficiencies by reconfiguring the wireless propagation environment. In this paper, we propose an RIS phase shift and BS beamforming optimization technique that minimizes the uplink transmit power of the RIS-aided IoT network. Key idea of the proposed scheme, referred to as Riemannian conjugate gradient-based joint optimization (RCG-JO), is to jointly optimize the RIS phase shifts and the BS beamforming vectors using the Riemannian conjugate gradient technique. By exploiting the product Riemannian manifold structure of the sets of unit-modulus phase shifts and unit-norm beamforming vectors, we convert the nonconvex uplink power minimization problem into the unconstrained problem and then find out the optimal solution over the product Riemannian manifold. From the performance analysis and numerical evaluations, we demonstrate that the proposed RCG-JO technique achieves 94% reduction of the uplink transmit power over the conventional scheme without RIS.

Index Terms—Reconfigurable intelligent surface (RIS), internet of things (IoT), energy efficiency, low-rank matrix completion, Riemannian conjugate gradient.

I. INTRODUCTION

As a paradigm to embrace the connection of massive number of devices, such as sensors, wearables, health monitors, and smart appliances, internet of things (IoT) has received much attention recently [2]. Since most of IoT devices are battery-limited, saving the device power is crucial for the dissemination of IoT networks. Recently, reconfigurable intelligent surface (RIS) has been emerging as a potential solution to enhance the sustainability of the IoT network [3]. In a nutshell, RIS is a planar metasurface consisting of a large number of low-cost passive reflecting elements, each of which can independently adjust the phase shift on the incident signal [4], [5]. Due to the capability to modify the wireless channel by controlling the phase shift of each reflecting element, RIS offers various benefits; When the direct link between the IoT device and the base station (BS) is blocked by obstacles, RIS can provide a virtual line-of-sight (LoS) link between them, ensuring the reliable link connection without requiring heavy transmit modules [6]. Due to its simple structure, lightweight, and conformal geometry, RIS can be easily deployed in the desired location. Also, with a small change in wireless standard, RIS can be easily integrated into existing communication systems.

Over the years, various efforts have been made to improve the energy efficiency of IoT networks [7]–[12]. In [7], a joint beamforming technique to minimize the downlink power consumption of wireless networks has been proposed. In [8], a joint power control and beamforming scheme for the RIS-aided device-to-device (D2D) network has been proposed. In [9], the energy-efficient downlink power control and phase shift designs for the Terahertz and MISO systems have been presented. In [10], an RIS phase shift control technique to maximize the downlink energy efficiency of RIS-aided systems has been proposed. In [11], an alternating optimization-based energy efficiency maximization scheme for the uplink RIS-aided systems has been proposed. Also, achievable sum throughput of an RIS-aided wireless powered sensor network (WPSN) has been investigated in [12].

While most of existing works focused on the downlink power control of RIS-aided systems, not much work has been made for the power minimization on the uplink side. Since IoT devices are battery-limited, it is of importance to come up with an energy-efficient uplink power control mechanism based on the RIS technique.

A major problem of the RIS-aided uplink power control is that the constraints on the RIS phase shifts and the BS beamforming vectors are nonconvex. This is because RIS can only change the phase shift of an incident signal so that the passive beamforming coefficients of reflecting elements are subject to the unit-modulus constraints [3]. This, together with the fact that the active beamforming vectors of BS are subject to the unit-norm constraints [13], makes it very difficult to find out the proper RIS phase shifts and BS beamforming vectors.

An aim of this paper is to propose an approach that minimizes the uplink transmit power of an RIS-aided IoT network. Key idea of the proposed scheme, referred to as Riemannian conjugate gradient-based joint optimization (RCG-JO), is to jointly optimize the RIS phase shifts and the BS beamforming vectors using the Riemannian conjugate gradient (RCG) algorithm. Specifically, by exploiting the smooth product Riemannian manifold structure of the sets of unit-modulus...
phase shifts and unit-norm beamforming vectors, we convert the uplink power minimization problem into the unconstrained problem on the Riemannian manifold. Since the optimization over the Riemannian manifold is conceptually analogous to that in the Euclidean space, optimization tools of the Euclidean space, such as the gradient descent method, can be readily used to solve the optimization problem on the Riemannian manifold [14]. In our approach, we employ the Riemannian conjugate gradient (RCG) method to find out the RIS phase shifts and BS beamforming vectors minimizing the uplink transmit power of an RIS-aided IoT network.

In recent years, there have been some efforts exploiting the manifold optimization technique for the RIS phase shifts control [1], [15], [16]. In [1], a Riemannian manifold-based alternating optimization technique for the design of RIS phase shifts and BS beamforming vectors has been proposed. In this scheme, the RIS phase shifts and the BS beamforming vectors are optimized on the Riemannian manifolds in an alternative fashion. In [15], [16], the manifold optimization-based RIS phase shifts control schemes have been proposed to maximize the downlink sum rate.

Our work is distinct from the previous works in the following aspects:

• We propose a Riemannian conjugate gradient-based joint optimization (RCG-JO) algorithm to find out the RIS phase shifts and the BS beamforming vectors minimizing the uplink transmit power of an RIS-aided IoT network. RCG-JO jointly optimizes the RIS phase shifts and the BS beamforming vectors on the product Riemannian manifold of the sets of unit-modulus RIS phase shifts and unit-norm BS beamforming vectors.

• We propose a channel estimation technique for the uplink RIS-aided IoT networks. Specifically, the RIS receives the pilot signals at the active reflecting elements and then feeds them back to the BS. Using the pilot signals, we estimate the partial RIS-aided channel information from which we recover the full RIS-aided channel information via the low-rank matrix completion (LRMC) algorithm.

• We present the convergence analysis of RCG-JO and demonstrate from the numerical experiments that RCG-JO converges to a fixed point within a few number of iterations.

• We provide the empirical simulation results from which we demonstrate that RCG-JO outperforms the conventional power control schemes by a large margin in terms of the uplink transmit power and the computational complexity. For example, when compared to the conventional power control scheme without RIS, RCG-JO saves around 94% of the uplink transmit power. Even when compared to the semidefinite relaxation (SDR)-based power control scheme, RCG-JO saves more than 44% of the uplink transmit power and achieves around 98% reduction in the computational complexity.

The rest of the paper is organized as follows. In Section III, we present the proposed RCG-JO technique. In Section IV, we analyze the convergence and computational complexity of RCG-JO. In Section V, we provide the simulation results and then conclude the paper in Section VI.

Notations: Lower and upper case symbols are used to denote vectors and matrices, respectively. The superscripts (·)*, (·)T, and (·)H denote the conjugate, transpose, and conjugate transpose, respectively. The operation (·) denotes the Hadamard product. ∥x∥ and ∥X∥F denote the Euclidean norm of a vector x and the Frobenius norm of a matrix X, respectively. |x| represents a vector of element-wise absolute values of x. tr(X) is the trace of X. diag(x) and ddiag(X) form diagonal matrices whose diagonal elements are x and diagonal elements of X, respectively. In addition, blkdiag(X1, X2) denotes a block-diagonal matrix with X1 and X2 on the block-diagonal.

II. RIS-AIDED UPLINK IoT SYSTEM MODEL

In this section, we present the system model, the channel estimation, and the data transmission protocols of RIS-aided uplink IoT networks. We then formulate the power minimization problem.

A. RIS-Aided Uplink IoT Network

We consider a single-input multi-output (SIMO) uplink IoT network where K devices with a single antenna transmit signals to the BS equipped with M antennas (see Fig. 1). In this network, an RIS consisting of a planar array of N x N reflecting elements is deployed to assist the uplink transmission. For example, each RIS reflecting element can adjust the phase of the incident signal independently using positive-intrinsic-negative (PIN) diodes [3]. The phase shifts of RIS reflecting elements are configured through the dedicated control link. In the setup, the effective uplink channel between the k-th IoT device and the BS is given by

\[ h_k = d_k + G \text{diag}(\theta) u_k \]  \hspace{1cm} (1)

\[ = d_k + \text{Gdiag}(u_k) \theta \] \hspace{1.5cm} (2)

\[ = d_k + GH_k \theta, \] \hspace{1.5cm} (3)

Fig. 1: Illustration of the RIS-aided uplink IoT network.
Fig. 2: Illustration of uplink RIS-aided channel estimation consisting of two major procedures: 1) estimation of the partial channel information observed at the active reflecting elements and 2) reconstruction of the full RIS-aided channel via the LRMC algorithm.

where $d_k \in \mathbb{C}^{M \times 1}$ is the direct channel from the $k$-th IoT device to the BS, $u_k \in \mathbb{C}^{N \times 1}$ is the channel from the $k$-th IoT device to the RIS, and $G \in \mathbb{C}^{M \times N}$ is the channel from the RIS to BS (see Fig. 1). Also, $H_k = \text{diag}(u_k)$ and $\theta = [\mu_1 \theta_1, \ldots, \mu_N \theta_N]^T$ is the phase shift vector of RIS. In addition, $\mu_n \in [0, 1]$ is the reflection amplitude coefficient and $\phi_n \in [0, 2\pi)$ is the phase shift.

**B. Uplink Channel Estimation**

In this subsection, we propose the uplink channel estimation process in time-division duplexing (TDD)-based RIS-aided IoT networks\(^3\). To enjoy the full potential of RIS, the BS needs to acquire not only the conventional direct channel $d_k$ but also the channels reflected by the RIS (i.e., $G$ and $u_k$). When compared to the direct channel estimation, the estimation of RIS-aided channels is far more difficult since RIS contains a large number of reflecting elements. Recently, the RIS architectures exploiting active reflecting elements that can reflect and also receive the signal have been proposed [3], [6]. Using the pilot signals fed back from the active reflecting elements, the BS can directly estimate the RIS-aided channel components. However, due to the practical limitations such as the implementation cost of RF chains, hardware complexity, and power consumption, only a few active reflecting elements can be employed. That is, the BS can acquire only partial information of the RIS-aided channels from the active reflecting elements.

To estimate the whole channel from the partial information, we exploit the property that the RIS-aided channels are dominated by LoS paths. Since there is only one propagation path, the channel matrix can be readily modeled as a low-rank matrix. When the channel matrix has a low-rank property, BS can reconstruct the full channel matrix from the partial observations using the LRMC algorithm [19]. Once the full RIS-aided channel information is obtained, BS can perform the uplink power allocation and the RIS phase shift control.

1) Sampled Channel Estimation: We assume that the RIS consists of $N$ active reflecting elements where the index set\(^4\) of active reflecting elements is $\Omega$ (see Fig. 2). The uplink channel estimation process consists of $(M + K)$ time slots where the $m$-th BS antenna transmits downlink pilot signal to the RIS at the $m$-th time slot ($m = 1, \ldots, M$) and then the $k$-th IoT device transmits uplink pilot signal to the RIS at the $(M + k)$-th time slot ($k = 1, \ldots, K$).

Let $x_m^{BS}$ be the downlink pilot signal of the $m$-th BS antenna at the $m$-th time slot and $x_k^D$ be the downlink pilot signal of the $k$-th IoT device at the $(M + k)$-th time slot. Then the received signal matrices $Y_m^{BS}, Y_k^D \in \mathbb{C}^{N \times N_y}$ of RIS from the $m$-th BS antenna and the $k$-th IoT device are

$$Y_m^{BS} = (x_m^{BS})^H P\Omega(G_m) + N_m^{BS}, \quad m = 1, \ldots, M,$$

$$Y_k^D = x_k^D P\Omega(U_k) + N_k^D, \quad k = 1, \ldots, K,$$

where $G_m \in \mathbb{C}^{N \times N_y}$ is the channel matrix from the $m$-th BS antenna to the RIS such that $\text{vec}(G_m)$ is the $m$-th row vector of $G$, $U_k \in \mathbb{C}^{N \times N_y}$ is the channel matrix from the $k$-th IoT device to the RIS such that $\text{vec}(U_k) = u_k$, and $N_m^{BS}$ and $N_k^D$ are the additive Gaussian noises. Also, $P\Omega(G_m)$ and $P\Omega(U_k)$ are the sampled matrices\(^5\) of $G_m$ and $U_k$. Note that the sampling operator is used since only active reflecting elements can receive the pilot signals. Then the BS can easily acquire the sampled channel matrices $P\Omega(G_m)$ and $P\Omega(U_k)$ using the conventional least squares (LS) and minimum mean square error (MMSE) estimation techniques.

2) Full Channel Reconstruction: After the sampled channel estimation, the BS reconstructs the full channel matrices $\{G_m\}_{m=1}^M$ and $\{U_k\}_{k=1}^K$ from $\{P\Omega(G_m)\}_{m=1}^M$ and $\{P\Omega(U_k)\}_{k=1}^K$ via the LRMC algorithm [19].

To be specific, $G_m$ can be reconstructed by solving the rank minimization problem:

$$\min_{X \in \mathbb{C}^{N \times N_y}} \text{rank}(X)$$

s.t. $P\Omega(X) = P\Omega(G_m)$.

The solution $X^*$ of (6) is the estimate of $G_m$. Since the rank minimization problem is NP-hard, this problem is computationally intractable. To deal with the problem, we replace the non-convex objective function with its convex surrogate. The

\(^1\)In this paper, we assume the ideal phase shift model where the reflection amplitude and phase shift are independent. In our system, the RIS controller coordinates the switching between two working modes, i.e., receiving mode ($\mu_n = 0$) for environment sensing (e.g., channel estimation) and reflecting mode ($\mu_n = 1$) for scattering the incident signals (e.g., data transmission).

\(^2\)To characterize the performance limits of RIS, we assume the phase shifts vary continuously in $[0, 2\pi)$. Note that the proposed scheme can be readily extended to practical systems with finite level of phase shifts via discrete phase quantization [17].

\(^3\)In the 5G systems and beyond, TDD will be a competitive duplexing technique in time-division duplexing (TDD)-based RIS-aided IoT networks. To enjoy the full potential of RIS, the BS needs to acquire not only the conventional direct channel $d_k$ but also the channels reflected by the RIS (i.e., $G$ and $u_k$). When compared to the direct channel estimation, the estimation of RIS-aided channels is far more difficult since RIS contains a large number of reflecting elements. Recently, the RIS architectures exploiting active reflecting elements that can reflect and also receive the signal have been proposed [3], [6]. Using the pilot signals fed back from the active reflecting elements, the BS can directly estimate the RIS-aided channel components. However, due to the practical limitations such as the implementation cost of RF chains, hardware complexity, and power consumption, only a few active reflecting elements can be employed. That is, the BS can acquire only partial information of the RIS-aided channels from the active reflecting elements.

\(^4\)In this paper, we assume the ideal phase shift model where the reflection amplitude and phase shift are independent. In our system, the RIS controller coordinates the switching between two working modes, i.e., receiving mode ($\mu_n = 0$) for environment sensing (e.g., channel estimation) and reflecting mode ($\mu_n = 1$) for scattering the incident signals (e.g., data transmission).

\(^5\)To characterize the performance limits of RIS, we assume the phase shifts vary continuously in $[0, 2\pi)$. Note that the proposed scheme can be readily extended to practical systems with finite level of phase shifts via discrete phase quantization [17].

\(^6\)In the 5G systems and beyond, TDD will be a competitive duplexing technique in time-division duplexing (TDD)-based RIS-aided IoT networks. To enjoy the full potential of RIS, the BS needs to acquire not only the conventional direct channel $d_k$ but also the channels reflected by the RIS (i.e., $G$ and $u_k$). When compared to the direct channel estimation, the estimation of RIS-aided channels is far more difficult since RIS contains a large number of reflecting elements. Recently, the RIS architectures exploiting active reflecting elements that can reflect and also receive the signal have been proposed [3], [6]. Using the pilot signals fed back from the active reflecting elements, the BS can directly estimate the RIS-aided channel components. However, due to the practical limitations such as the implementation cost of RF chains, hardware complexity, and power consumption, only a few active reflecting elements can be employed. That is, the BS can acquire only partial information of the RIS-aided channels from the active reflecting elements.

\(^4\)In this paper, we assume the ideal phase shift model where the reflection amplitude and phase shift are independent. In our system, the RIS controller coordinates the switching between two working modes, i.e., receiving mode ($\mu_n = 0$) for environment sensing (e.g., channel estimation) and reflecting mode ($\mu_n = 1$) for scattering the incident signals (e.g., data transmission).

\(^5\)To characterize the performance limits of RIS, we assume the phase shifts vary continuously in $[0, 2\pi)$. Note that the proposed scheme can be readily extended to practical systems with finite level of phase shifts via discrete phase quantization [17].

\(^6\)In the 5G systems and beyond, TDD will be a competitive duplexing technique in time-division duplexing (TDD)-based RIS-aided IoT networks. To enjoy the full potential of RIS, the BS needs to acquire not only the conventional direct channel $d_k$ but also the channels reflected by the RIS (i.e., $G$ and $u_k$). When compared to the direct channel estimation, the estimation of RIS-aided channels is far more difficult since RIS contains a large number of reflecting elements. Recently, the RIS architectures exploiting active reflecting elements that can reflect and also receive the signal have been proposed [3], [6]. Using the pilot signals fed back from the active reflecting elements, the BS can directly estimate the RIS-aided channel components. However, due to the practical limitations such as the implementation cost of RF chains, hardware complexity, and power consumption, only a few active reflecting elements can be employed. That is, the BS can acquire only partial information of the RIS-aided channels from the active reflecting elements.
nuclear norm \(\|X\|_s\), the sum of the singular values of \(X\), has been widely used as a convex surrogate of rank(\(X\)):

\[
\begin{align*}
\min_{X \in \mathbb{C}^{N_x \times N_y}} &\quad \|X\|_s \\
\text{s.t.} &\quad P_{\Omega}(X) = P_{\Omega}(G_m). \quad (7b)
\end{align*}
\]

The nuclear norm minimization problem (7) can also be recast as a semidefinite programming (SDP) [19]:

\[
\begin{align*}
\min_{Z} &\quad \text{tr}(Z) \\
\text{s.t.} &\quad \text{tr}(A_{x,y}^HZ) = [G_m]_{x,y}, \quad (x,y) \in \Omega, \quad (8b) \\
&\quad Z \geq 0, \quad (8c)
\end{align*}
\]

where \(Z = \begin{bmatrix} Z_1 & X \\ X^H & Z_2 \end{bmatrix} \in \mathbb{C}^{(N_x+N_y) \times (N_x+N_y)}\) for the Hermitian matrices \(Z_1 \in \mathbb{C}^{N_x \times N_x}\) and \(Z_2 \in \mathbb{C}^{N_y \times N_y}\), \(A_{x,y} = e_xe_y^r + N_s\) is the linear sampling matrix, and \(e_x\) is the \(x\)-th column vector of \(I_{N_x} + N_s\). Since (8) is a convex problem, the solutions \(Z^*\) and \(X^*\) of (8) can be obtained using the convex optimization solvers (e.g., CVX [20]). Finally, \(G_m\) is obtained by \(G_m = X^*\).

Similarly, \(U_k\) can be recovered from \(P_{\Omega}(U_k)\) by solving the low-rank matrix completion problem. Finally, the whole RIS-aided channels \(G\) and \(u_k\) are obtained as \(G = [\text{vec}(G_1) \cdots \text{vec}(G_M)]^T\) and \(u_k = \text{vec}(U_k)\).

C. Uplink Data Transmission

After the channel estimation, BS performs the uplink power allocation and the RIS phase shift control using the acquired channel information. Then BS sends the uplink transmit power indicator \(p_k\) and the RIS phase shift vector \(\theta\) to the \(k\)-th IoT device and RIS, respectively. Finally, each IoT device transmits the data to BS by the uplink channel.

Let \(x_k = \sqrt{p_k} s_k\) be the data signal of the \(k\)-th IoT device where \(s_k\) and \(p_k(\geq 0)\) are the normalized data symbol and the transmit power of the \(k\)-th IoT device, respectively. Then, the received signal at BS from the \(k\)-th IoT device \(y_k\) is

\[
y_k = w_k^H(h_k x_k + \sum_{j \neq k} h_j x_j + n_k) \\
= \sqrt{p_k} w_k^H(d_k + GH_k \theta) s_k + \sum_{j \neq k} \sqrt{p_j} w_k^H(d_j + GH_j \theta) s_j + w_k^H n_k,
\]

where \(w_k \in \mathbb{C}^{M \times 1}\) is the normalized BS beamforming vector for the \(k\)-th IoT device, i.e., \(\|w_k\|_2 = 1\), and \(n_k \sim \mathcal{CN}(0, \sigma_k^2 I)\) is the additive Gaussian noise. In this setting, the uplink achievable rate \(R_k\) of the \(k\)-th IoT device is given by

\[
R_k = \log_2 \left(1 + \frac{p_k |w_k^H(d_k + GH_k \theta)|^2}{\sum_{j \neq k} p_j |w_k^H(d_j + GH_j \theta)|^2 + \sigma_k^2}\right). \quad (11)
\]

It has been shown that the nuclear norm is the convex envelope of rank function on the set \(\{X : \|X\| \leq 1\}\) [19].

Algorithm 1 Riemannian conjugate gradient-based joint optimization algorithm

Input: Number of iterations \(T\), rate requirement \(R_{\min}\), maximum transmit power \(p_{\max}\)
Output: Uplink transmit power \(P_t\), RIS phase shift vector \(\theta\), BS beamforming matrix \(W\)
Initialize: \(t = 1\), \(p_t = p_{\text{ini}}\), \(\theta = \theta_{\text{ini}}\), \(W_t = W_{\text{ini}}\)
while \(P_{\text{total}}\) does not converge do
Optimize \(\theta_t\) and \(W_t\) simultaneously using Algorithm 2 when \(p_t\) is fixed
Optimize \(p_t\) by solving an LP problem when \(\theta_t\) and \(W_t\) are fixed
\(t = t + 1\)
end

D. Uplink Power Minimization Problem Formulation

The uplink power minimization problem to optimize the RIS phase shift vector \(\theta\), the BS beamforming matrix \(W = [w_1 \cdots w_K]\), and the device power vector \(p = [p_1 \cdots p_K]\) is formulated as

\[
P_1 : \min_{\theta, W, p} \sum_{k=1}^K p_k \quad (12a)
\]

\[
s.t. \quad \sum_{j \neq k} p_j |w_k^H(d_j + GH_j \theta)|^2 + \sigma_k^2 \geq 2 \rho_{\text{min}}^2 - 1, \quad \forall k \in K, \quad (12b) \\
|\theta_n| = 1, \quad \forall n \in N, \quad (12c) \\
|w_k| = 1, \quad \forall k \in K, \\
0 \leq p_k \leq p_{\text{max}}, \quad \forall k \in K, \quad (12d)
\]

where \(K\) and \(N\) are the sets of IoT devices and RIS reflecting elements and \(P_{\text{min}}\) and \(p_{\text{max}}\) are the rate requirement and the maximum transmit power of the \(k\)-th IoT device, respectively. Note that (12c) is the unit-modulus constraint of the RIS phase shift and (12d) is the unit-norm constraint of the BS beamforming vector. Due to the nonconvexity of the unit-modulus and unit-norm constraints, \(P_1\) is a non-convex problem. This, together with the quadratic fractional and coupled structure of the rate function in (12b), makes \(P_1\) very difficult to solve.

III. RIEMANNIAN CONJUGATE GRADIENT-BASED JOINT OPTIMIZATION ALGORITHM

The primal goal of the proposed RCG-JO technique is to find out the RIS phase shifts and the BS beamforming vectors minimizing the uplink transmit power of RIS-aided IoT networks. As mentioned, main obstacles in solving the uplink power minimization problem are the nonconvex unit-modulus constraint of the RIS phase shift and unit-norm constraint of the BS beamforming vector. To handle these issues, we exploit the smooth product Riemannian manifold structure of the sets of unit-modulus phase shifts and unit-norm beamforming vectors. Since the product of two manifolds is also a Riemannian manifold with smooth structure, we can readily convert the uplink power minimization problem \(P_1\) to an unconstrained problem on the product Riemannian manifold. After that,
by using the differential geometry tools, we can design the gradient descent algorithm on the Riemannian manifold and use it to obtain the optimal RIS phase shifts and the BS beamforming vectors minimizing the uplink transmit power of the RIS-aided IoT network.

In a nutshell, the proposed RCG-JO algorithm consists of two major steps: 1) Lagrangian relaxation to move the complicated rate constraint to the objective function and 2) alternating optimization of $p$ and $(\theta, W)$ to minimize the modified objective function on the product Riemannian manifold. To be specific, in the alternating optimization step, we first fix the device power $p$ and then jointly optimize the phase shift vector $\theta$ and the BS beamforming matrix $W$ using the RCG method. Once the optimal $\theta$ and $W$ are obtained, the optimization problem of $p$ is formulated as a linear programming (LP) problem where the optimal solution can be easily obtained using the convex optimization technique. We repeat these procedures until the objective function $P_{\text{total}} = \sum_{k=1}^{K} p_k$ converges (see Algorithm 1).

### A. Notions on Riemannian Manifolds

In this subsection, we briefly introduce properties and operators of the optimization on Riemannian manifold. Roughly speaking, a smooth manifold is a generalization of the Euclidean space on which the notion of differentiability exists [21]. The tangent space $T_X Y$ at a point $X$ of a manifold $Y$ is the set of the tangent vectors of all the curves at $X$, where the curve $\gamma$ of a mapping $\gamma : C \rightarrow Y$. Put it formally, for a given point $X \in Y$, the tangent space of $\gamma$ at $X$ is defined as $T_X Y = \{\gamma'(0) : \gamma \text{ is a curve in } Y, \gamma(0) = X\}$ (see Fig. 3(a)).

A manifold $Y$ together with a smoothly varying inner product $g = \langle \cdot, \cdot \rangle : T_X Y \times T_X Y \rightarrow C$ on the tangent space $T_X Y$ forms a smooth Riemannian manifold, denoted as $(Y, g)$ where $g$ is termed as the Riemannian metric.

When the Riemannian manifold $(Y, g)$ is the cartesian product of two Riemannian manifolds $(Y_1, g_1)$ and $(Y_2, g_2)$, then the Riemannian metric $g$ is defined as $g = g_1 + g_2$ [22]. In the following lemma, we show that the tangent space on the product Riemannian manifold is the direct sum of the tangent spaces on each Riemannian manifold.

**Lemma 1.** For the point $X = X_1 \oplus X_2$ where $X_1 \in Y_1, X_2 \in Y_2$, the tangent space at $X$ of the product Riemannian manifold $Y = Y_1 \times Y_2$ is given by [22]

$$T_X Y = T_{X_1} Y_1 \oplus T_{X_2} Y_2. \quad (13)$$

Given a smooth objective function $f$ on the Riemannian manifold $Y$, minimization of $f$ requires the notion of gradient at every $X \in Y$. To be specific, the Riemannian gradient of $f$ at $X$, denoted by $\nabla_X f(X)$, is defined as a unique vector on $T_X Y$ that yields the steepest-descent of $f$. Put it formally, $\nabla_X f(X)$ is obtained by projecting $\nabla_X f(X)$ onto $T_X Y$ (see Fig. 3(b)).

Using Lemma 1, the Riemannian gradient $\nabla_X f(X)$ is given by

$$\nabla_X f(X) = \nabla_{X_1} f_1(X_1) \oplus \nabla_{X_2} f_2(X_2), \quad (14)$$

where $\nabla_{X_1} f_1(X_1) \in T_{X_1} Y_1$ is the Riemannian gradient of $f_1$ at $X_1 \in Y_1$ for $i = 1, 2$.

#### Definition 1. An orthogonal projection onto the tangent space $T_X Y$ of a matrix $V \in T_X Y$ is a mapping $P_{T_X Y} : C^{n \times b} \rightarrow T_X Y$ such that for a given $U \in C^{n \times b}$, $(U - P_{T_X Y}(U), V) = 0$ for all $V \in T_X Y$. In particular, the orthogonal projection onto the tangent space of product manifold $Y = Y_1 \times Y_2$ is

$$P_{T_X Y}(U) = P_{T_{X_1} Y_1}(U_1) \oplus P_{T_{X_2} Y_2}(U_2), \quad (15)$$

where $P_{T_{X_i} Y_i}(U_i)$ is the projection of $U_i$ onto $T_{X_i} Y_i$ for $i = 1, 2$.

#### Definition 2. The retraction $R_X(V)$ of a matrix $V \in T_X Y$ onto $Y$ is defined as

$$R_X(V) = \arg \min_{Z \in Y} \|X + V - Z\|_F. \quad (16)$$

In particular, the retraction onto the product manifold $Y = Y_1 \times Y_2$ is

$$R_X(V) = R_{X_1}(V_1) \oplus R_{X_2}(V_2), \quad (17)$$

where $R_{X_i}(V_i)$ is the retraction of $V_i \in T_{X_i} Y_i$ onto $Y_i$ for $i = 1, 2$.

#### B. Joint RIS Phase Shift and BS Beamforming Optimization on Product Manifold

In the first step, we fix the device power and then jointly optimize the RIS phase shifts and BS beamforming vectors using the RCG method.

When the device power is given, $P_1$ is reduced to

$$P_2 : \text{Find } \theta, W \quad (18a)$$

s.t. \begin{align}
\frac{p_k}{2P_{\text{min}}} - 1 & \quad A_{k,k}(\theta, w_k) - \sum_{j \neq k} p_j A_{j,k}(\theta, w_k) \geq \sigma_k^2, \\
\forall k & \in K, \\
|\theta_n| & = 1, \quad \forall n \in N, \\
\|w_k\| & = 1, \quad \forall k \in K, \quad (18b, 18c, 18d)
\end{align}
where \( A_{j,k}(\theta, w_k) = |w_k^H (d_j + G \bm{H}, \theta)|^2 \) for \( j, k \in K \). Since the rate expression in (18b) is a joint quadratic function of \( \theta \) and \( W \), (18b) is a nonconvex constraint. To handle this issue, we use the Lagrangian relaxation to move the complicated rate constraints to the objective function. To be specific, the modified objective function is given by

\[
L(\theta, W, \lambda) = \sum_{k=1}^{K} \lambda_k \left( -\frac{p_k}{2\expM^2} A_{k,k}(\theta, w_k) + \sum_{j \neq k} p_j A_{j,k}(\theta, w_k) + \sigma_k^2 \right),
\]

where \( \lambda = [\lambda_1, \ldots, \lambda_K]^T \) is the Lagrangian multiplier obtained by solving the corresponding dual problem. Using \( L(\theta, W, \lambda) \), \( P_2 \) is relaxed into

\[
P_3 : \min_{\theta, W, \lambda} \sum_{k=1}^{K} \lambda_k \left( -\frac{p_k}{2\expM^2} A_{k,k}(\theta, w_k) + \sum_{j \neq k} p_j A_{j,k}(\theta, w_k) + \sigma_k^2 \right)
\]

s.t. \( |\theta_n| = 1, \forall n \in N \), \( |w_k| = 1, \forall k \in K \), \( \lambda_k \geq 0, \forall k \in K \).

The relaxed problem \( P_3 \) looks simpler than \( P_2 \), but it is still nonconvex and difficult to solve since the objective function of \( P_3 \) is a joint quadratic function of \( \theta \) and \( W \). Additionally, we need to deal with the unit-modulus constraints (20b) and the unit-norm constraints (20c). To handle the problem, we jointly optimize \( \theta \) and \( W \) on the product manifold of the unit-modulus phase shifts and the unit-norm beamforming vectors using the RCG method. Once we obtain \( \theta \) and \( W \), we update the Lagrangian multiplier \( \lambda \). We repeat these procedures until \( \theta \) and \( W \) converge.

1) Joint RIS Phase Shift and BS Beamforming Optimization: For a given \( \lambda \), \( P_3 \) is reduced to the unconstrained problem on the product manifold:

\[
P(\theta, W) : \min_{(\theta, W) \in M_\theta \times M_W} L(\theta, W),
\]

where \( M_\theta \) and \( M_W \) are the complex circle manifold and complex oblique manifold given by

\[
M_\theta = \{ \theta \in \mathbb{C}^{N \times 1} : |\theta_n| = 1, \forall n \in N \},
\]

\[
M_W = \{ W \in \mathbb{C}^{M \times K} : \text{diag}(W^H W) = I_K \},
\]

with the inner products defined as \( g_\theta(z_1, z_2) = \langle z_1, z_2 \rangle = \text{Re}(z_1^H z_2) \) and \( g_W(Z_1, Z_2) = \langle Z_1, Z_2 \rangle = \text{Re}(\text{tr}(Z_1^H Z_2)) \), respectively.

By combining \( \theta \) and \( W \) into \( \Sigma = \text{blkdiag}(\theta, W) \), \( P(\theta, W) \) is re-expressed as

\[
P_\Sigma : \min_{\Sigma \in M} L(\Sigma),
\]

where \( M = M_\theta \times M_W \) is the product manifold with the inner product defined as \( g_\Sigma = g_\theta + g_W \). In the following lemma, we provide the tangent space of the product manifold \( M \).

**Lemma 2.** The tangent space \( T_\Sigma M \) of the product manifold \( M \) at the point \( \Sigma \) is given by

\[
T_\Sigma M = \Sigma_{0} M_\theta \oplus T_{\Sigma_{0}} M_W,
\]

where \( \Sigma_{0} \) is the tangent space of \( M_\theta \) at \( \theta \) and \( T_{\Sigma_{0}} M_W \) is the tangent space of \( M_W \) at \( W \).

In order to minimize the objective function \( L(\Sigma) \) in \( P_\Sigma \), we need a Riemannian gradient which is obtained by projecting the Euclidean gradient of \( L(\Sigma) \) on \( \Sigma \) onto \( T_\Sigma M \).

**Lemma 3.** The orthogonal projection \( P_{T_\Sigma M}(U) \) of \( U = \text{blkdiag}(u, U) \) onto \( T_\Sigma M \) is

\[
P_{T_\Sigma M}(U) = P_{\Sigma_{0} M_\theta}(u) + P_{T_{\Sigma_{0}} M_W}(U),
\]

where \( P_{\Sigma_{0} M_\theta}(u) = u - \text{Re}(\theta^* \otimes u) \otimes \theta \) is the orthogonal projection of \( d \) onto \( \Sigma_{0} M_\theta \) and \( P_{T_{\Sigma_{0}} M_W}(U) = U - W \text{diag}(\text{Re}(W^H U)) \) is the orthogonal projection of \( U \) on \( T_{\Sigma_{0}} M_W \).

To make sure that the point \( \Sigma \) is updated in the direction of the tangent space \( T_\Sigma M \) while staying on \( M \), a retraction operation, a mapping from \( T_\Sigma M \) to \( M \), is needed.

**Lemma 4.** The retraction \( R_{\Sigma}(V) \) of \( V = \text{blkdiag}(v, V) \) in \( T_\Sigma M \) is

\[
R_{\Sigma}(V) = R_\theta(v) + R_W(V),
\]

where \( R_\theta(v) = (\theta + v) \otimes \frac{1}{\theta^* + v} \) is the retraction of \( v \in \Sigma_{0} M_\theta \) and \( R_W(V) = \text{diag}(W^H V) \) is the retraction of \( V \in T_{\Sigma_{0}} M_W \).

To find out \( \Sigma \) minimizing \( L(\Sigma) \) on the product manifold \( M \), we exploit the RCG method, an extension of the conjugate gradient (CG) method to the Riemannian manifold. In this approach, the update equations of the conjugate direction \( D \) and the point \( \Sigma \) are given by

\[
D_i = -\text{grad}_M L(\Sigma_i) + \beta_i P_{T_{\Sigma_i} M}(D_{i-1}),
\]

\[
\Sigma_{i+1} = R_{\Sigma_i}(\alpha_i D_i),
\]

where \( \text{grad}_M L(\Sigma_i) \) is the Riemannian gradient of \( L(\Sigma_i) \) on \( \Sigma_i \), \( \beta_i \) is the Fletcher-Reeves conjugate gradient parameter, and \( \alpha_i \) is the step size [23].

We note that the RCG method is distinct from the conventional CG method in three respects: 1) the projection of the previous conjugate direction \( D_{i-1} \) onto the tangent space \( T_{\Sigma_i} M \) is needed before performing a linear combination of \( \text{grad}_M L(\Sigma) \) and \( D_{i-1} \) since they lie on two different spaces \( T_{\Sigma_i} M \) and \( T_{\Sigma_{i-1}} M \), 2) the Riemannian gradient \( \text{grad}_M L(\Sigma) \) is used instead of the Euclidean gradient \( \nabla_\Sigma L(\Sigma) \) since we need to find out the search direction on the tangent space of \( M \), and 3) the retraction is required to ensure that the updated point \( \Sigma_{i+1} \) lies on \( M \). Specifically, the Riemannian gradient \( \text{grad}_M L(\Sigma) \) is obtained by projecting

\[
\nabla_\Sigma L(\Sigma) = \text{proj}_{T_{\Sigma} M}(\nabla_{\Sigma_{0}} L(\Sigma_{0})),
\]

where \( \nabla_{\Sigma_{0}} L(\Sigma_{0}) \) is the Euclidean gradient of \( L(\Sigma_{0}) \) on \( \Sigma_{0} \) and \( \beta_i \) is conjugate update parameter.

---

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
the Euclidean gradient onto the tangent space. In the following lemma, we provide the closed-form expression of $\nabla_M L(\Sigma)$ of $L(\Sigma)$ on $\mathcal{M}$.

**Lemma 5.** The Riemannian gradient $\nabla_M L(\Sigma)$ of $L(\Sigma)$ on $\mathcal{M}$ is

$$
\nabla_M L(\Sigma) = \nabla_M \theta L(\theta) \oplus \nabla_M w L(W). \tag{30}
$$

**Specifically, the Riemannian gradient $\nabla_M \theta L(\theta)$ of $L(\theta)$ on $\mathcal{M}_\theta$ is

$$
\nabla_M \theta L(\theta) = \mathcal{P}_{\mathcal{M}_\theta}(\nabla_\theta L(\theta)) = \nabla_\theta L(\theta) - \Re \{\Theta^* \odot \nabla_\theta L(\theta)\} \odot \Theta, \tag{31}
$$

where $\nabla_\theta L(\theta)$ is the Euclidean gradient of $L(\theta)$ given by (see (19))

$$
\nabla_\theta L(\theta) = \sum_{k=1}^{K} \lambda_k \left( -\frac{p_k}{2\mathbb{R}_+^m} \frac{\partial A_{k,k}(\theta, w_k)}{\partial \theta} + \sum_{j \neq k}^{K} p_j \frac{\partial A_{j,k}(\theta, w_k)}{\partial \theta} \right), \tag{32}
$$

and

$$
\frac{\partial A_{j,k}(\theta, w_k)}{\partial \theta} = H_j^H \mathbf{G}_w w_k^H (d_j + \mathbf{G}_j \Theta). \tag{33}
$$

Also, the Riemannian gradient $\nabla_M w L(W)$ of $L(W)$ on $\mathcal{M}_w$ is

$$
\nabla_M w L(W) = \mathcal{P}_{\mathcal{M}_w}(\nabla_w L(W)) = \nabla_w L(W) - W \text{diag} \{\Re \{W^H \nabla_w L(W)\}\}, \tag{34}
$$

where $\nabla_w L(W) = \left[ \frac{\partial L(W)}{\partial w_1}, \cdots, \frac{\partial L(W)}{\partial w_K} \right]$ is the Euclidean gradient of $L(W)$ given by (see (19))

$$
\frac{\partial L(W)}{\partial w_k} = \lambda_k \left( -\frac{p_k}{2\mathbb{R}_+^m} \frac{A_{k,k}(\theta, w_k)}{w_k} + \sum_{j \neq k}^{K} p_j \frac{\partial A_{j,k}(\theta, w_k)}{\partial w_k} \right), \tag{35}
$$

and

$$
\frac{\partial A_{j,k}(\theta, w_k)}{\partial w_k} = (d_j + \mathbf{G}_j \Theta)(d_j + \mathbf{G}_j \Theta)^H w_k. \tag{36}
$$

Once $\Sigma$ is determined, we can obtain the phase shift vector $\theta$ and beamforming matrix $W$ by decomposing $\Sigma$. The optimization steps of $\theta$ and $W$ are summarized in Algorithm 2.

2) **Lagrangian Multiplier Update:** After updating the RIS phase shift vector $\theta$ and the BS beamforming matrix $W$, we update the Lagrangian multiplier $\lambda$. To be specific, $\lambda$ is updated in the direction of maximizing the dual function $D(\lambda) = \min_{\theta, W} L(\theta, W, \lambda)$ as

$$
\lambda = \arg \max_{\lambda \geq 0} D(\lambda). \tag{38}
$$

Since $D(\lambda)$ is the optimal value of the optimization problem $\mathcal{P}_3$, we cannot take derivative with respect to $\lambda$, meaning that we cannot directly use the conventional gradient ascent method in finding out the optimal value of $\lambda$.

**Algorithm 2.** Optimization of $(\theta, W)$ on product manifold $\mathcal{M}$

**Input:** Tolerance $\epsilon$, number of iterations $T$, $c_1 = 0.0001$, $c_2 = 0.1$

**Output:** $(\theta, W)$

**Initialize:** $i = 1$, $\Sigma_1 = \text{blkdiag}(\theta_1, W_1) \in \mathcal{M}$, $D_1 = -\nabla_M \theta L(\Sigma_1)$, $\beta_1 = 0$

$\text{while } i \leq T \text{ do}$

\[ \begin{align*}
\nabla_M \theta L(\Sigma_i) &= \mathcal{P}_{\mathcal{M}_\theta}(\nabla_\theta L(\Sigma_i)) \\
\beta_i &= \frac{||\nabla_M \theta L(\Sigma_{i-1})||^2}{(||\nabla_M \theta L(\Sigma_i)||^2 + ||D_{i-1}||^2)} \\
D_i &= -\nabla_M \theta L(\Sigma_i) + \beta_i \mathcal{P}_{\mathcal{M}_\theta}(D_{i-1}) \\
\text{Find a step size } \alpha_i > 0 \text{ such that} \\
L(R_{\Sigma_i}(\alpha_i D_i)) &\leq L(\Sigma_i) + c_1 \alpha_i (||\nabla_M \theta L(\Sigma_i)||, D_i) \\
(||\nabla_M \theta L(\Sigma_i)||, D_i) &\leq c_2 (||\nabla_M \theta L(\Sigma_i)||, D_i) \\
\Sigma_{i+1} &= R_{\Sigma_i}(\alpha_i D_i) \\
\text{if } ||\Sigma_{i+1} - \Sigma_i||^2 < \epsilon \text{ then} \\
\theta_{i+1} &= \Sigma_{i+1}(1 : N, 1) \\
W_{i+1} &= \Sigma_{i+1}(N + 1 : N + M, 2 : K + 1) \\
\text{Exit from the while loop}
\end{align*} \]

$i = i + 1$

**end**

To address this problem, we use the subgradient method, a generalized concept of gradient method for the non-smooth convex functions [24]. In particular, the subgradient $g = [g_1, \cdots, g_K]^T$ of the dual function $D(\lambda)$ is given by

$$
g_k = -\frac{p_k}{2\mathbb{R}_+^m} A_{k,k}(\theta, w_k) + \sum_{j \neq k}^{K} p_j A_{j,k}(\theta, w_k) + \sigma_k^2. \tag{39}
$$

Using $g_k$, we obtain the update equation of $\lambda$:

$$
\lambda_{i+1} = \max(\lambda_i + \eta_i g_i, 0), \tag{40}
$$

where $\eta_i$ is the step size [24].

**C. Uplink Transmit Power Minimization**

Once the RIS phase shift vector $\theta$ and the BS beamforming matrix $W$ are obtained, we next find out the device power vector $p$ minimizing the total uplink transmit power. When $\theta$ and $W$ are given, the original uplink transmit power minimization problem $\mathcal{P}_1$ is reduced to

$$
\mathcal{P}_4 : \min_p \sum_{k=1}^{K} p_k \tag{41a}
$$

s.t. $\frac{p_k}{2\mathbb{R}_+^m} A_{k,k}(\theta, w_k) - \sum_{j \neq k}^{K} p_j A_{j,k}(\theta, w_k) \geq \sigma_k^2$, \quad $\forall k \in K$, \tag{41b}

$0 \leq p_k \leq p_{\text{max}}, \quad \forall k \in K$. \tag{41c}

Since $\mathcal{P}_4$ is an LP optimization problem, the optimal solution can be easily obtained using the convex optimization tools (e.g., CVX [20]).

**IV. CONVERGENCE AND COMPUTATIONAL COMPLEXITY ANALYSIS OF RCG-JO**

In this section, we analyze the convergence and the computational complexity of RCG-JO.
TABLE I: Comparison of computational complexity \( (K = 1, M = 4) \)

| Method          | Complexity for various numbers of reflecting elements \( N \) |
|-----------------|-----------------------------------------------------------|
|                 | \( N = 4 \)     | \( N = 16 \)   | \( N = 32 \)   | \( N = 64 \)   |
| RCG-JO          | \( O(K^2N^2M + K^2 + K^2N^3 + K^2M^2) \)                |
|                 | \( 1.45 \times 10^4 \)        | \( 5.14 \times 10^4 \)        | \( 3.69 \times 10^4 \)        | \( 2.79 \times 10^4 \)        |
| SDR-based scheme| \( O(K^2N^2M + N^3 + K^2M^2) \)                        |
|                 | \( 8.26 \times 10^4 \)        | \( 1.68 \times 10^4 \)        | \( 1.07 \times 10^4 \)        | \( 6.87 \times 10^4 \)        |

**A. Convergence Analysis of RCG-JO Algorithm**

Recall that the proposed RCG-JO technique consists of two major iterations: 1) inner iteration for jointly optimizing \( \theta \) and \( W \) on the product Riemannian manifold (Algorithm 2) and 2) outer iteration for alternately updating \( (\theta, W) \) and \( p \) (Algorithm 1). In case of Algorithm 1, due to the alternating optimization operations, it is very difficult to prove its convergence analytically so that we demonstrate its convergence from the numerical results. In case of Algorithm 2, we show that it converges to a local minimizer in this subsection.

We first explain the strong Wolfe conditions used to determine the step size \( \alpha_i \).

**Definition 3.** A step size \( \alpha_i \) is said to satisfy the strong Wolfe conditions, restricted to the conjugate direction \( D_i \), if the following two inequalities hold [21]:

\[
L(R_{\Sigma_i}(\alpha_i D_i)) \leq L(\Sigma_i) + c_1 \alpha_i \langle \nabla L(\Sigma_i), D_i \rangle, \quad (42)
\]

\[
|\langle \nabla L(\Sigma_i), P_{T_{\Sigma_i}}(\alpha_i D_i) \rangle| \leq c_2 \langle \nabla L(\Sigma_i), D_i \rangle. \quad (43)
\]

The first condition, known as Armijo’s rule, ensures that the step size decreases the cost function \( L(\Sigma) \) sufficiently. The second condition, known as curvature condition, ensures that the Riemannian gradient converges to zero. Note that since \( \nabla L(\Sigma_i) \in T_{\Sigma_i}(\alpha_i D_i)M \), the second Wolfe condition (43) can be converted to

\[
|\langle \nabla L(\Sigma_i), P_{T_{\Sigma_i}}(\alpha_i D_i) \rangle| \leq c_2 \langle \nabla L(\Sigma_i), D_i \rangle. \quad (44)
\]

It has been shown that the step size satisfying the strong Wolfe conditions always exists if \( \nabla L(\Sigma_i) \) is Lipschitz continuous along \( D \) [25, Proposition 1]. In the following proposition, we show the Lipschitz continuity of the Riemannian gradient.

**Proposition 1.** The objective function \( L(\Sigma) \) is bounded below, meaning that there exists a constant \( L^* \in \mathbb{R} \) such that for all \( \Sigma \in \mathcal{M} \), \( L^* \leq L(\Sigma) \). Also, the Riemannian gradient \( \nabla L(\Sigma_i) \) is Lipschitz continuous along \( D \). That is, for every \( \Sigma \in \mathcal{M} \), there exists a \( K > 0 \) such that for all \( D \in T_{\Sigma}M \),

\[
\|\nabla L(\Sigma_i) - \nabla L(\Sigma(0))\| \leq K\|D\|. \quad (45)
\]

**Proof.** See Appendix A. \( \square \)

Since the Riemannian gradient does not converge if its derivative is unbounded, the Lipschitz continuity of the Riemannian gradient is crucial for the convergence of RCG method. Using the strong Wolfe conditions and the Lipschitz continuity of Riemannian gradient, we can show that the angle between the Riemannian gradient and the conjugate direction is bounded.

**Theorem 1.** (Zoutendijk condition) Let \( \Sigma_i \in \mathcal{M} \) be the point and \( D_i \in T_{\Sigma_i}M \) be the conjugate direction of \( i \)-th RCG iteration. Then

\[
\sum_{i=1}^{\infty} \frac{\langle \nabla L(\Sigma_i), D_i \rangle^2}{\|D_i\|^2} < \infty. \quad (46)
\]

**Proof.** See Appendix B. \( \square \)

Next, in the following proposition, we prove that the inner product of Riemannian gradient \( \nabla L(\Sigma) \) and the conjugate direction \( D \) is bounded.

**Proposition 2.** Let \( \Sigma_i \in \mathcal{M} \) be the point and \( D_i \in T_{\Sigma_i}M \) be the conjugate direction of \( i \)-th RCG iteration with the conjugate gradient parameter \( \beta_i = \frac{\|\nabla L(\Sigma_i)\|}{\|\nabla L(\Sigma_i)\|^2} \). If the step size \( \alpha_i \) satisfies the strong Wolfe conditions (42) and (43) with \( c_2 < 1/2 \), then for every \( i \in \mathbb{N} \),

\[
\frac{1}{1-c_2} - \frac{\langle \nabla L(\Sigma_i), D_i \rangle}{\|\nabla L(\Sigma_i)\|^2} \leq 2c_2 - 1. \quad (47)
\]

**Proof.** See Appendix C. \( \square \)

Finally, by combining Theorem 1 and Proposition 2, we show the convergence of RCG method.

**Theorem 2.** (Convergence of RCG method) Let \( \Sigma_i \in \mathcal{M} \) be the point and \( D_i \in T_{\Sigma_i}M \) be the conjugate direction of \( i \)-th RCG iteration with the conjugate gradient parameter \( \beta_i = \frac{\|\nabla L(\Sigma_i)\|}{\|\nabla L(\Sigma_i)\|^2} \). If the step size \( \alpha_i \) satisfies the strong Wolfe conditions (42) and (43) with \( c_2 < 1/2 \), then

\[
\lim_{i \to \infty} \|\nabla L(\Sigma_i)\| = 0. \quad (48)
\]

**Proof.** See Appendix D. \( \square \)

**Remark 1.** Theorem 2 shows that the RCG iteration of proposed scheme converges to a local minimizer \( \Sigma^* \) of objective function \( L(\Sigma) \) on the product manifold \( \mathcal{M} \).

**B. Computational Complexity Analysis of RCG-JO Algorithm**

In our analysis, we measure the complexity in terms of the number of floating point operations (flops). We first provide the complexity analysis of the joint optimization of \( \theta \) and \( W \) in Algorithm 2.

**Lemma 6.** The total computational complexity \( C(\theta, W) \) of Algorithm 2 is given by

\[
C(\theta, W) = O(K^2N^2M + K^2N^3 + K^2M^2). \quad (49)
\]

**Proof.** See Appendix E. \( \square \)
After updating the RIS phase shift vector \( \theta \) and the BS beamforming matrix \( \mathbf{W} \), we update the Lagrangian multiplier \( \lambda \) using the subgradient method. The complexity of updating \( \lambda \) is \( \mathcal{O}(K^2N^2M) \).

Once \( \theta, \mathbf{W}, \) and \( \lambda \) are updated, the optimization of the uplink power vector \( \mathbf{p} \) is achieved by solving an LP problem in (41). Note that the numbers of flops to compute \( A_{jk}(\theta, \mathbf{W}) \) and to solve the LP problem are \( N^2M \) and \( K^3 \), respectively. Thus, the overall complexity to optimize \( \mathbf{p} \) is \( \mathcal{O}(K^2N^2M + K^3) \).

In conclusion, the computational complexity \( \mathcal{C}_{\text{RCG-JO}} \) of the proposed RCG-JO scheme is

\[
\mathcal{C}_{\text{RCG-JO}} = \mathcal{C}(\theta, \mathbf{W}) + \mathcal{C}_\lambda + \mathcal{C}_p = \mathcal{O}(K^2N^2M + K^3 + K^2N^3 + K^2M^2). \tag{50}
\]

For comparison, we also discuss the complexity of the conventional SDR-based scheme, which consists of three major steps: 1) optimization of \( \theta \) using SDR for fixed \( \mathbf{w} \) and \( \mathbf{p} \), 2) optimization of \( \mathbf{w} \) using SDR for fixed \( \theta \) and \( \mathbf{p} \), and 3) optimization of \( \mathbf{p} \) by solving an LP problem with the obtained \( \theta \) and \( \mathbf{W} \). It has been shown that the worst-case complexity of the SDR method for optimizing \( \theta \in \mathbb{C}^{N \times 1} \) is \( \mathcal{O}(N^6) \) [7]. Similarly, the worst-case complexity of optimizing \( \mathbf{W} \in \mathbb{C}^{M \times K} \) using SDR is \( \mathcal{O}(K^6M^6) \). In addition, the optimization of \( \mathbf{p} \) is achieved by solving the LP problem, so that the complexity is \( \mathcal{O}(K^2N^2M + K^3) \). In summary, the overall complexity \( \mathcal{C}_{\text{SDR}} \) of the SDR-based scheme is

\[
\mathcal{C}_{\text{SDR}} = \mathcal{O}(K^2N^2M + N^6 + K^6M^6). \tag{51}
\]

In Table I, we compare the computational complexities of the proposed RCG-JO and the conventional SDR-based schemes. When the number of reflecting elements \( N \) is 4, the complexity of RCG-JO is significantly lower than that of the SDR-based scheme (around 98%). This is because the essential operations of RCG-JO are just matrix additions and multiplications whereas the SDR-based scheme needs a convex optimization technique to solve the SDP problem.

V. SIMULATION RESULTS

A. Simulation Setup

In this section, we present the numerical results to evaluate the performance of proposed RCG-JO algorithm. Our simulation setup is based on the uplink IoT network where \( K = 2 \) single-antenna devices transmit signals to the BS equipped with \( M = 4 \) receiving antennas (see Fig. 4). This uplink transmission is assisted by the RIS equipped with \( N = 64 \) reflecting elements, which is randomly located at the circumference of the circle centered at the BS with the radius \( R = 65 \text{ m} \). Also, \( K \) devices are uniformly distributed on a line that is perpendicular to the line connecting the BS and the RIS. The maximal vertical distance between the IoT devices and the line is \( d_v = 3 \text{ m} \). We set the horizontal distance between the BS and IoT devices to \( d_h = 57 \text{ m} \). Throughout the simulations, we set the rate requirement \( h_{k_{\text{min}}} \) and noise power \( \sigma_n^2 \) to 0.3 bps/Hz and \(-60 \text{ dBm} \), respectively.

We use the Rician fading channel models for \( d_k, u_k \) and \( G \) with the Rician factors \( \kappa_d = 0, \kappa_k = 10, \) and \( \kappa_G = \infty \), respectively. First, the channel matrix from the RIS to BS \( G \) is given by

\[
G = \sqrt{\beta_G} \left( \sqrt{\frac{\kappa_G}{\kappa_G + 1}} G^\text{LoS} + \sqrt{\frac{1}{\kappa_G + 1}} G^\text{NLoS} \right), \tag{52}
\]

where \( G^\text{LoS} = a_{\text{BS}}(\psi_G) e_r^H(\varphi_G, \varphi_G) = \text{LoS component} \) with \( \psi_G, \varphi_G, \) and \( \varphi_G \) being the angle of arrival (AoA) at the BS, the azimuth and elevation of the angles of departure (AoDs) at the RIS, respectively, \( G^\text{NLoS} \) is the NLoS component generated from complex Gaussian distribution, \( \beta_G \) is the path loss between RIS and BS, and \( \kappa_G \geq 0 \) is the Rician factor. Here, \( a_{\text{BS}}(\psi_G) = [1, \ldots, e^{-j\pi(M-1)\sin \varphi_G}]^T \in \mathbb{C}^{M \times 1} \) is the BS array steering vector and \( a_{\text{RIS}}(\psi_G, \varphi_G) = a_{\text{RIS},x}(\psi_G, \varphi_G) \otimes a_{\text{RIS},y}(\psi_G, \varphi_G) \in \mathbb{C}^{N \times 1} \) is the RIS array steering vector where

\[
a_{\text{RIS},x}(\psi_G, \varphi_G) = [1, \ldots, e^{-j\pi(N_x-1)\cos \varphi_G \sin \varphi_G}]^T \quad \text{and} \quad a_{\text{RIS},y}(\psi_G, \varphi_G) = [1, \ldots, e^{-j\pi(N_y-1)\sin \varphi_G \sin \varphi_G}]^T.
\]

Second, the channel vector from the \( k \)-th IoT device to the RIS \( u_k \) is expressed as

\[
u_k = \sqrt{\beta_k} \left( \sqrt{\frac{\kappa_k}{\kappa_k + 1}} u_k^\text{LoS} + \sqrt{\frac{1}{\kappa_k + 1}} u_k^\text{NLoS} \right), \tag{53}
\]

where \( u_k^\text{LoS} = a_{\text{RIS}}(\psi_k, \varphi_k) \) is the LoS component with \( \psi_k \) and \( \varphi_k \) being the azimuth and elevation of the AoAs at the RIS, respectively, \( u_k^\text{NLoS} \sim \mathcal{CN}(0, I_N) \) is the NLoS component, \( \beta_k \) is the path loss, and \( \kappa_k \geq 0 \) is the Rician factor. The channel vector from the \( k \)-th IoT device to BS \( d_k \) is modeled similarly with \( u_k \). In general, RIS is deployed at the position where the LoS links with the BS and IoT devices are guaranteed, so \( G \) and \( u_k \) are dominated by the LoS paths. The path loss is \( \beta = C_0 (d/D_0)^{-\alpha} \), where \( d \) is the distance, \( \alpha \) is the path loss exponent, and \( C_0 = -30 \text{ dB} \) is the path loss at the reference distance \( D_0 = 1 \text{ m} \) [7]. For the channels \( d_k, u_k \) and \( G \), we set \( \alpha \) to be 3.8, 2.8, and 2, respectively. In each point of the plots, the simulation results are averaged over 1,000 independent channel realizations. The simulation parameters are summarized in Table II.

For comparison, we test the following benchmark techniques: 1) SDR-based scheme where \( \theta \) and \( \mathbf{W} \) are optimized alternately using SDR [7], 2) difference-of-convex (DC)-

\[\text{Note that in the SDR-based scheme, after solving the relaxed SDP problem, the Gaussian randomization technique is employed to find out the feasible rank-one solution.}\]
based scheme where $\theta$ and $W$ are optimized alternately using the DC programming [26], 3) deep reinforcement learning (DRL)-based scheme where $\theta$ and $W$ are jointly optimized by leveraging the DRL technique [27], 4) random phase shifts where $\theta$ is randomly generated and the maximum ratio transmission (MRT) is used for $W$, and 5) conventional system without RIS where $W$ is generated randomly.

### B. Simulation Results

In Fig. 5, we plot the uplink transmit power as a function of the number of RIS reflecting elements $N$. From the simulation results, we observe that RCG-JO outperforms the conventional schemes using SDR and random phase shifts. For example, when the number of RIS reflecting elements is $N = 64$, the proposed scheme achieves 44% and 74% reduction in power over the SDR-based scheme and the conventional scheme using random phase shifts, respectively. Also, we see that the performance gap between RCG-JO and the SDR-based scheme increases gradually with $N$. Furthermore, we see that the power saving gain of the conventional scheme using random phase shifts does not change with $N$. This is because without the optimization of RIS phase shifts and BS beamforming vectors, the RIS reflected signal power is comparable to the signal power transmitted from the direct link, so that the gain obtained from the joint active and passive beamforming is marginal. Also, we compare the uplink transmit power of RCG-JO with the DRL-based scheme. We observe that RCG-JO achieves 57% of power reduction over the DRL-based scheme. Note, for the DRL-based approach, it is not easy to find out the optimal decision (i.e., RIS phase shifts and BS beamforming vectors) minimizing the uplink transmit power and at the same time satisfying the rate requirements of the IoT devices. This is because the goal of DRL is to learn the decision policy maximizing the cumulative reward so that the minimization of uplink transmit power and the rate requirements might not be satisfied simultaneously.

We next evaluate the uplink transmit power of the proposed RCG-JO algorithm and benchmark schemes as a function of the rate requirement of IoT devices $R_k^{\text{min}}$. As shown in Fig. 6, we observe that the uplink transmit power increases when the rate requirement of IoT devices becomes more strict. Also, when compared to the case without RIS, the rate requirement of IoT devices can be satisfied with lower uplink transmit power in the proposed scheme. For instance, when the rate requirement of devices is $R_k^{\text{min}} = 0.4$ bps/Hz, RCG-JO achieves around 92% of uplink power reduction over the conventional scheme without RIS. This is because through the joint active and passive beamforming at the BS and RIS, we can improve the signal power and reduce the interference so that the rate requirements of IoT devices can be satisfied with lower transmit power. Furthermore, we see that RCG-JO significantly reduces the uplink transmit power over the benchmark schemes using SDR, DRL, and random phase shift. Even when compared to the DC-based scheme, RCG-JO achieves 24% of the uplink power saving gain. Note that in the DC-based scheme, due to the non-smoothness of matrix spectral norm, the subgradient method should be employed to minimize the objective function. Since the subgradient does not guarantee the steepest-descent of objective function, the performance degradation over the conventional gradient descent method is unavoidable. Whereas, in the proposed RCG-JO, we can directly minimize the objective function on the smooth Riemannian manifold.
using the Riemannian gradient descent method.

In Fig. 7, we plot the uplink transmit power as a function of the horizontal distance between the BS and IoT devices $d_h$. We observe that RCG-JO outperforms the benchmark schemes as $d_h$ increases. For example, when the horizontal distance between the BS and devices is $d_h = 55$ m, RCG-JO saves 29% and 40% of the uplink transmit power over the SDR-based scheme and the conventional scheme using random phase shift, respectively. In the system where the RIS is not employed, we see that the uplink transmit power is considerable due to the signal attenuation, in particular when the devices are located far away from the BS. Whereas, in the RIS-aided IoT networks, we see that the uplink transmit power increases initially and then decreases as the horizontal distance increases. Basically, when the IoT device is far from the BS, large uplink transmit power is needed to satisfy the rate requirement of a device. In this case, the IoT device is getting close to the RIS so that the RIS-aided channel gain increases gradually. As a result, the rate requirement can be satisfied with relatively lower uplink transmit power.

In Fig. 8, we investigate the uplink transmit power of the proposed scheme and benchmark schemes as a function of the noise power $\sigma_k^2$. From the simulation results, we observe that the uplink transmit power increases with the noise power $\sigma_k^2$ and RCG-JO outperforms the benchmark schemes. For example, when the noise power $\sigma_k^2 = -55$ dBm, the proposed scheme achieves 51% and 70% of uplink power reductions over the conventional schemes using SDR and the random phase shift, respectively. We also see that by using the RIS, the uplink transmit power of RCG-JO is significantly reduced (more than 80%) over the conventional scheme without RIS.

In Fig. 9, we evaluate the uplink transmit power of RCG-JO and benchmark schemes as a function of the number of IoT devices $K$. We observe that RCG-JO outperforms the benchmark schemes in all tested scenarios. For example, when $K = 11$, RCG-JO saves 35% and 78% of the uplink transmit power over the SDR-based scheme and the conventional scheme without RIS, respectively. We also see that the power saving gain of RCG-JO over the SDR-based scheme increases with $K$. For instance, when we change $K$ from 5 to 11, the power saving gain of RCG-JO over the SDR-based scheme increases from 18% to 35%. While RCG-JO solves the unconstrained optimization problem on the product manifold, the SDR-based scheme needs to find out the feasible rank-one solution after solving the SDP problem so that the

Fig. 7: Uplink transmit power vs. BS-devices horizontal distance $d_h$.

Fig. 8: Uplink transmit power vs. noise power $\sigma_k^2$.

Fig. 9: Uplink transmit power vs. number of IoT devices $K$.

Fig. 10: Uplink transmit power vs. percentage of active reflecting elements.
performance degradation is inevitable. This, together with the result of computational complexity analysis in Section IV.B, demonstrates the effectiveness of RCG-JO.

To evaluate the effectiveness of proposed channel estimation technique, we investigate the uplink transmit power of RCG-JO using the perfect channel information, the estimated channel information, and the sampled channel information when \( N = 100 \). In Fig. 10, we observe that when the percentage of active reflecting elements is larger than 20%, RCG-JO using the estimated channel information performs close to RCG-JO using the genie channel information. This is because the RIS-aided channel matrix can be readily modeled as a low-rank matrix and thus the LRMC algorithm can effectively reconstruct the full channel matrix (see Section II.B).

In Fig. 11, we investigate the cumulative distributions of the number of iterations required for the convergence of outer iteration (Algorithm 1) and inner iteration (Algorithm 2). We test the number of iterations required for the convergence when using 10,000 independent channel realizations. In all tested cases, we observe that Algorithm 1 converges within 15 iterations and Algorithm 2 converges within 30 iterations.

VI. CONCLUSION

In this paper, we proposed an RIS phase shift and BS beamforming optimization technique to minimize the uplink transmit power of an RIS-aided IoT network. Key idea of the proposed RCG-JO algorithm is to jointly optimize the RIS phase shifts and BS beamforming vectors using the Riemannian conjugate gradient method. By exploiting the product Riemannian manifold structure of the sets of unit-modulus RIS phase shift and unit-norm BS beamforming vector, we converted the uplink power minimization problem to an unconstrained problem on the Riemannian manifold. Then, we employed the Riemannian conjugate gradient method to find out the optimal RIS phase shifts and the BS beamforming vectors simultaneously. We demonstrated from the performance analysis and numerical evaluations that the proposed RCG-JO algorithm is effective in saving the uplink transmit power of RIS-aided IoT networks. In our work, we assumed the ideal phase shift model where the reflection amplitude and phase shift are independent, but an extension to the realistic scenarios with the phase-dependent reflection amplitude would also be an interesting research direction worth pursuing.

APPENDIX A

PROOF OF PROPOSITION 1

One can easily see that due to the unit-modulus constraint of \( \theta \) (12c) and the unit-norm constraint of \( \textbf{W} \) (12d), the objective function \( L(\Sigma) \) is bounded.

Recall that the Riemannian gradient can be decomposed as \( \text{grad}_{\mathcal{M}_d}L(\textbf{R}_d) = \text{grad}_{\mathcal{M}_d}L(\textbf{R}_d(d)) \oplus \text{grad}_{\mathcal{M}_w}L(\textbf{R}_w(D)) \) where \( \textbf{D} = \textbf{d} \oplus \textbf{D} \). Thus, we firstly find out the constants \( K_\theta \) and \( K_w \) satisfying the Lipschitz conditions for \( \text{grad}_{\mathcal{M}_d}L(\textbf{R}_d(d)) \) and \( \text{grad}_{\mathcal{M}_w}L(\textbf{R}_w(D)) \), respectively. We then obtain the constant \( K = \max(K_\theta, K_w) \) satisfying the Lipschitz condition for \( \text{grad}_{\mathcal{M}}L(\textbf{R}_\Sigma(D)) \).

When \( \textbf{W} \) is given, \( L(\theta) \) is a quadratic function with respect to \( \theta \):

\[
L(\theta) = \sum_{k=1}^{K} \lambda_k \left( -\frac{p_k}{2n} - 1 \right) A_k(k, \theta, \textbf{w}_k) + \sum_{j \neq k} p_j A_{j,k}(\theta, \textbf{w}_k) + \sigma_k^2 \]

\[
= \theta^H \textbf{B} \theta + \textbf{b}^H \theta + \theta^H \textbf{b} + b, \tag{54}
\]

where \( \textbf{B} = \sum_{k=1}^{K} \lambda_k \left( -\frac{p_k}{2n} - 1 \right) \textbf{H}_k^H \textbf{G}_k \textbf{w}_k \textbf{G}_k^H \textbf{w}_k + \sum_{j \neq k} p_j \textbf{H}_j^H \textbf{G}_k \textbf{w}_k \textbf{G}_k^H \textbf{w}_j \textbf{H}_j \textbf{G}_j \textbf{w}_j \textbf{H}_j^H \right) \) and \( \textbf{b} = \sum_{k=1}^{K} \lambda_k \left( -\frac{p_k}{2n} - 1 \right) \sum_{j \neq k} p_j \textbf{H}_j^H \textbf{G}_k \textbf{w}_k \textbf{G}_k^H \textbf{w}_j \textbf{H}_j \textbf{d}_j \right) \). Then, \( \text{grad}_{\mathcal{M}_d}L(\textbf{R}_d(d)) \) is expressed as

\[
\text{grad}_{\mathcal{M}_d}L(\textbf{R}_d(d)) = \text{P}_{\mathcal{T}_d} \text{P}_{\mathcal{M}_d} \left( \nabla_\theta L(\textbf{R}_d(d)) \right)
= \text{P}_{\mathcal{T}_d} \text{P}_{\mathcal{M}_d} (\text{BR}_\theta(d) + \textbf{b})
= \text{BR}_\theta(d) + \textbf{b}
- \text{Re}(\theta \odot (\text{BR}_\theta(d) + \textbf{b})) \odot \theta. \tag{55}
\]

Using the triangle inequality, we have

\[
\|\text{grad}_{\mathcal{M}_d}L(\textbf{R}_d(d)) - \text{grad}_{\mathcal{M}_d}L(\textbf{R}_d(0))\| \\
\leq \|\textbf{B}(\textbf{R}_\theta(d) - \textbf{R}_\theta(0))\| + \|\text{Re}(\theta \odot (\text{BR}_\theta(d) - \text{BR}_\theta(0)))\| \\
\leq 2\|\textbf{B}(\textbf{R}_\theta(d) - \textbf{R}_\theta(0))\| \\
\leq 2\|\textbf{B}\| \|\textbf{R}_\theta(d) - \textbf{R}_\theta(0)\|. \tag{56}
\]

Since \( \textbf{B} \) is a quadratic function of \( \textbf{W} \) and the elements of \( \textbf{W} \) are bounded by the unit-norm constraints, \( \|\textbf{B}\| \) is bounded on \( \textbf{W} \in \mathcal{M}_w \).

Now, what we need to show is that the retraction operator \( \textbf{R}_\theta \) is Lipschitz continuous. To do so, we prove that each element of \( \textbf{R}_\theta \) is Lipschitz continuous. Let \( \theta_n \) and \( d_n \) be the
We use the mathematical induction to prove Proposition 2. When \( i = 1 \), \( \mathbf{D}_1 = -\text{grad}_M L(\Sigma_1) \) so that we can easily see that (47) holds. Now suppose that (47) holds for \( i \geq 1 \). By using the update equation of \( \mathbf{D}_i \) in (28) and the Fletcher-Reeves conjugate gradient parameter \( \beta_i = \frac{\|\text{grad}_M L(\Sigma_i)\|^2}{\|\text{grad}_M L(\Sigma_{i+1})\|^2} \), we have

\[
\frac{\langle \text{grad}_M L(\Sigma_{i+1}), \mathbf{D}_{i+1} \rangle}{\|\text{grad}_M L(\Sigma_{i+1})\|^2} = \frac{\langle \text{grad}_M L(\Sigma_i) - \beta_i \mathbf{D}_i, \mathbf{D}_i \rangle}{\|\text{grad}_M L(\Sigma_i)\|^2} = 1 + \frac{\langle \text{grad}_M L(\Sigma_{i+1}), \mathbf{D}_i \rangle}{\|\text{grad}_M L(\Sigma_i)\|^2} \geq 1 - \frac{\beta_i}{2} \|\text{grad}_M L(\Sigma_i)\|^2.
\]

Then, using the second Wolfe condition (44) and (63), we obtain

\[
-1 + c_2 \left( \frac{\langle \text{grad}_M L(\Sigma_i), \mathbf{D}_i \rangle}{\|\text{grad}_M L(\Sigma_i)\|^2} \right) \left( \text{grad}_M L(\Sigma_{i+1}) \right) \leq \left( \frac{\langle \text{grad}_M L(\Sigma_i), \mathbf{D}_i \rangle}{\|\text{grad}_M L(\Sigma_i)\|^2} \right) \left( \text{grad}_M L(\Sigma_{i+1}) \right) \leq 1 - c_2 \|\text{grad}_M L(\Sigma_i)\|^2.
\]

By employing the induction hypothesis (47) for \( i \), we see that (47) holds for \( i + 1 \), which establishes the proposition 2.
By sequentially applying (66) until \( i = 1 \), we obtain
\[
\|D_i\|^2 \leq \frac{1 + c_2}{1 - c_2} \|\grad_A M(L(\Sigma))\|^4 \sum_{j=1}^{i} \frac{1}{\|\grad_A M(L(\Sigma_j))\|^2}.
\] (67)

If we assume \( \lim_{i \to \infty} \|\grad_A M(L(\Sigma_i))\| \neq 0 \), meaning that \( \epsilon(>0) \) such that \( \|\grad_A M(L(\Sigma_i))\| \geq \epsilon \) for all \( i \in \mathcal{N} \) exists, then we have
\[
\|D_i\|^2 \leq \frac{1 + c_2}{1 - c_2} \|\grad_A M(L(\Sigma_i))\|^4 \sum_{j=1}^{i} \frac{1}{\epsilon^2}.
\] (68)

This implies that
\[
\sum_{i=1}^{\infty} \frac{\|\grad_A M(L(\Sigma_i))\|^4}{\|D_i\|^2} \geq \frac{1 - c_2}{1 + c_2} \sum_{i=1}^{\infty} \frac{1}{i} = \infty.
\] (69)

However, Theorem 1 and Proposition 2 imply that \( \sum_{i=1}^{\infty} \frac{\|\grad_A M(L(\Sigma_i))\|^4}{\|D_i\|^2} < \infty \), which is a contradiction to (69). Therefore, we obtain the desired result
\[
\lim_{i \to \infty} \|\grad_A M(L(\Sigma_i))\| = 0.
\]

### APPENDIX E

**PROOF OF LEMMA 6**

In the first step of Algorithm 2, we compute \( \grad_A M(L(\Sigma)) \), which is given by
\[
\grad_A M(L(\Sigma)) = (\nabla_A L(\theta_i) - \Re\{\theta_i \odot \nabla_A L(\theta_i)\} \odot \theta_i) \oplus (\nabla_EL(W_i) - \Re\{W_i \odot \nabla_EL(W_i)\}).
\] (70)

Note that the numbers of flops required for computing the Euclidean gradients \( \nabla_A L(\theta_i) \) and \( \nabla_EL(W_i) \) are \( K^2 N^2 M + K^2 N^3 \) and \( K^2 N^2 M + K^2 M^2 \), respectively. Thus, the complexity \( C_1 \) of computing \( \grad_A M(L(\Sigma)) \) is
\[
C_1 = O(K^2 N^2 M + K^2 N^3 + K^2 M^2).
\] (71)

Using the Riemannian gradient \( \grad_A M(L(\Sigma)) \), we then compute the RCG coefficient \( \beta_i = \frac{\|\grad_A M(L(\Sigma))\|^2}{\|\grad_A M(L(\Sigma_{i-1}))\|^2} \). The complexity \( C_2 \) of computing \( \beta_i \) is
\[
C_2 = O(K^2 N + K^2 M).
\] (72)

Then the complexity \( C_3 \) for updating the Riemannian conjugate direction \( D_i = -\grad_A M(L(\Sigma)) + \beta_i \beta_i \Re(P_{\Sigma_i}(D_{i-1})) \) is
\[
C_3 = O(K^2 M + K N).
\] (73)

Next, we find out the step size \( \alpha_i \) via the line search which consists of the computation of the following elements: 1) \( RS_{\Sigma_i}(\alpha_i D_i) \), 2) \( \grad_A M(L(\Sigma)) \), 3) \( L(RS_{\Sigma_i}(\alpha_i D_i)) \), and 4) \( L(\Sigma_j) \). To compute the retraction \( RS_{\Sigma_i}(\alpha_i D_i) \), the required number of flops is \( K^2 N^3 M + K^2 N^3 \). To compute \( \grad_A M(L(\Sigma)) \), the required number of flops is \( K^2 N^2 M + K^2 N^3 + K^2 M^2 \). Also, the numbers of flops required for computing \( L(RS_{\Sigma_i}(\alpha_i D_i)) \) and \( L(\Sigma_j) \) are both \( K^2 N^2 M \). Thus, the complexity \( C_4 \) of the line search for updating \( \alpha_i \) is
\[
C_4 = O(K^2 N^2 M + K^2 N^3 + K^2 M^2).
\] (74)

Finally, the complexity \( C_5 \) of updating \( \Sigma \) is
\[
C_5 = O(K^2 M + N).
\] (75)

In summary, the complexity \( C(\theta, \mathbf{w}) \) of Algorithm 2 is
\[
C(\theta, \mathbf{w}) = C_1 + C_2 + C_3 + C_4 + C_5 = O(K^2 N^2 M + K^2 N^3 + K^2 M^2).
\] (76)
[22] C. Udriste, Convex functions and optimization methods on Riemannian manifolds. Springer Science & Business Media, vol. 297, 2013.

[23] Y.-H. Dai and Y. Yuan, “Nonlinear conjugate gradient methods”, Shanghai Science and Technology Publisher, Shanghai, 2000.

[24] S. Boyd, L. Xiao, and A. Mutapcic, “Subgradient methods”, lecture notes of EE392o, Stanford University, Autumn Quarter, vol. 2004, pp. 2004–2005, 2003.

[25] W. Ring and B. Wirth, “Optimization methods on Riemannian manifolds and their application to shape space”, SIAM J. Opt., vol. 22, no. 2, pp. 596–627, 2012.

[26] M. Fu, Y. Zhou, and Y. Shi, “Intelligent reflecting surface for downlink non-orthogonal multiple access networks”, in Proc. IEEE Global Commun. Conf. (Globecom) Workshops, pp. 1-6, 2019.

[27] C. Huang, R. Mo, and C. Yuen, “Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning”, IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1839–1850, 2020.

Jiao Wu received the B.S. degree in communication engineering from North China Electric Power University, China, in 2015, and the M.S. degree in electronics and communication engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2018. She is currently working toward the Ph.D. degree in Electrical and Computer Engineering with Seoul National University, Korea. Her research interests include signal processing and convex optimization techniques for the 5G and 6G wireless communications.

Seungnyun Kim received the B.S. in the Department of Electrical and Computer Engineering from Seoul National University, Seoul, Korea, in 2016. He is currently working toward the Ph.D. degree in Electrical and Computer Engineering with Seoul National University, Korea. His research interests include signal processing and deep learning techniques for the 5G and 6G wireless communications.

Byonghyo Shim (SM’09) received the B.S. and M.S. degree in Control and Instrumentation Engineering (currently Electrical Eng.) from Seoul National University (SNU), Seoul, Korea, in 1995 and 1997, respectively, and the M.S. degree in Mathematics and the Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2004 and 2005, respectively. From 1997 and 2000, he was with the Department of Electronics Engineering at the Korean Air Force Academy as an Officer (First Lieutenant) and an Academic Full-time instructor. He also had a short time research position in the LG Electronics, Texas Instruments, and Samsung Electronics in 1997 and 2004, 2019, respectively. From 2005 to 2007, he was with the Qualcomm Inc., San Diego, CA as a Staff Engineer working on CDMA systems. From 2007 to 2014, he was with the School of Information and Communication, Korea University, Seoul, Korea, as an associate professor. Since September 2014, he has been with the Dept. of Electrical and Computer Engineering, Seoul National University, where he is currently a Professor. His research interests include signal processing for wireless communications, statistical signal processing, machine learning, compressed sensing, and information theory. Dr. Shim was the recipient of the M. E. Van Valkenburg Research Award from the ECE Department of the University of Illinois (2005), the Hadong Young Engineer Award from IEIE (2010), the Irwin Jacobs Award from Qualcomm and KICS (2016), the Shinyang Research Award from the Engineering College of SNU (2017), the Okawa Foundation Research Award (2020), and the IEEE COMSOC AP Outstanding Paper Award (2021). He was a technical committee member of Signal Processing for Communications and Networking (SPCOM), and currently serving as an associate editor of IEEE Transactions on Signal Processing (TSP), IEEE Transactions on Communications (TCOM), IEEE Transactions on Vehicular Technology (TVT), IEEE Wireless Communications Letters (WCL), Journal of Communications and Networks (JCN), and a guest editor of IEEE Journal of Selected Areas in Communications (location awareness for radios and networks).