Joint MSE Constrained Hybrid Beamforming and Reconfigurable Intelligent Surface

Xin He, Jiangzhou Wang, *Fellow, IEEE*, and Yi Gong, *Senior Member, IEEE*

***Abstract*—In this paper, the symbol detection mean squared error (MSE) constrained hybrid analog and digital beamforming is proposed in millimeter wave (mmWave) system, and the reconfigurable intelligent surface (RIS) is proposed to assist the mmWave system. The inner majorization-minimization (iMM) method is proposed to obtain analog transmitter, RIS and analog receivers, and the alternating direction method of multipliers (ADMM) method is proposed to obtain digital transmitter. The proposed iMM and ADMM methods are faster than the semidefinite relaxation (SDR) and interior point methods, respectively. In order to counter against the changing large-scale path loss, the iMM method is helped by channel normalization to reduce the computational complexity, while the ADMM method is helped by the adaptive parameterizations to robust against the changing large-scale path loss. Simulation results show that the computational times of the proposed methods are much faster than other methods, and the proposed methods are robust against the changing large-scale path loss.**

***Index Terms*—Constant modulus, Unit modulus, MU-MIMO, MSE, Hybrid beamforming, Analog beamforming, Reconfigurable Intelligent Surface.**

I. Introduction

In hybrid beamforming design, owing to the expensive radio frequency (RF) chains, the analog beamforming is used to reduce the hardware cost. Since the millimeter wave (mmWave) communication is easy to be blocked by environments, e.g., walls, trees, and buildings, the reconfigurable intelligent surface (RIS) is proposed to assist the mmWave system.

The conventional hybrid beamforming schemes aimed to maximize the sum rate under power and unit modulus constraints [1]–[8]. To maximize the sum rate of the multi-user multiple-input multiple-output (MU-MIMO), an efficient hybrid beamforming scheme was proposed in [1]. Based on matrix factorization, the manifold optimization was proposed to tackle the hybrid beamforming [2]. The semidefinite relaxation (SDR) was proposed to tackle the analog beamforming constraint in [3]. By minimizing the mean squared error (MSE), an efficient majorization-minimization (MM) method was proposed for analog beamforming [4], [5]. The gradient projection method was proposed to solve the unit-modulus least squares problem, and the computational complexity is lower than the SDR method [6]. Another efficient method to tackle the unit-modulus quadratic problem is the alternating direction method of multipliers (ADMM) method, but the nonconvexity makes the convergence property generally not guaranteed [7]. To take the quality of service (QoS) into consideration, the SDR method was proposed to handle the signal-to-interference-noise-ratio (SINR) constraints [9], [10]. However, the SINR constraints were not guaranteed to be satisfied by the SDR method [11]. The inner majorization-minimization (iMM) method was proposed to solve the SINR constrained hybrid beamforming [11]. Although minimizing MSE based hybrid beamforming schemes were extensively studied [12]–[14], the MSE constrained hybrid beamforming has not been studied.

The RIS has been proposed to assist the mmWave system, most of the researches were focus on maximize the sum rate under power and unit modulus constraints [15]–[18]. An approximation method was proposed to maximize the sum rate problem with RIS [15]. The least square approximation was proposed to maximize the sum rate problem with RIS [16]. A singular value decomposition based approximation method was proposed to maximize the capacity with RIS [17]. Based on approximation, a gradient projection method was proposed to solve the sum rate maximization problem with RIS [18]. Similarly, under power and unit modulus constraints, the RIS has been proposed to maximize the signal-to-leakage-ratio [19] or to minimize the SINR outage probability [20]. Without QoS constraints, the complex circle manifold method [21] and the penalty dual decomposition method were proposed in [22], [23]. With only one power transfer constraint and unit-modulus constraint, the MM method and successive convex approximation (SCA) method were proposed in [24]. Only a recent paper [25] took multiple QoS constraints into consideration, and the technique is approximation and manifold optimization.

In this paper, the symbol detection MSE constrained hybrid analog and digital beamforming is proposed in mmWave system, and the RIS is proposed to assist the mmWave system. Instead of the unreliable SDR method, the iMM method is proposed to obtain analog transmitter, RIS and analog receivers. Instead of the expensive interior point method, the ADMM method is proposed to obtain digital transmitter. In order to counter against the changing large-scale path loss, the iMM method is helped by channel normalization to reduce the computational complexity, while the ADMM method is helped by the adaptive parameterizations to robust against the changing large-scale path loss. Simulation results show that the computational times of the proposed methods are much faster than other methods, and the proposed methods are robust

This work was supported in part by the STU Scientific Research Foundation for Talents under Grants NTF21048, and in part by the National Natural Science Foundation of China under Grants 61807018.

Xin He is with the College of Electronics and Information Engineering, Shantou University, China. (e-mail: xinhe@stu.edu.cn). Yi Gong is with Southern University of Science and Technology, Shenzhen, China. (email: gongy@sustech.edu.cn). Jiangzhou Wang is with the School of Engineering, University of Kent, Canterbury, Kent, CT2 7NZ, U.K. (e-mail: J.Z.Wang@kent.ac.uk).

against the changing large-scale path loss. The contributions of the paper are summarized as follows.

- This is the first paper to study the MSE constrained hybrid beamforming.
- The iMM method is unified to obtain the MSE constrained analog transmitter, RIS and analog receivers. The iMM method always provides feasible solution, while the SDR method does not always provide feasible solution.
- The digital transmitter is solved by the ADMM method, which is much faster than the interior point method.
- The proposed iMM and ADMM methods are robust against the changing large-scale path loss.

Notation: In this paper, $\mathbb{E}(\cdot)$, $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote statistical expectation, transposition, conjugate and Hermitian, respectively. The $\|\cdot\|_2$ denotes the norm of a vector, and $\|\cdot\|_F$ stands for the Frobenius norm of a matrix. The notation $\mathrm{Tr}(\mathbf{A})$ is the trace of a square matrix $\mathbf{A}$, and $\mathrm{vec}(\mathbf{G})$ is to vectorize the matrix $\mathbf{G}$ into a column vector. The notation $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ stand for the real part and the imaginary part of a complex number, respectively. The notation $\otimes$ stands for the Kronecker product. For a vector $\boldsymbol{\theta}$, $\boldsymbol{\theta}(n)$ is the n-th element of $\boldsymbol{\theta}$. For a matrix $\mathbf{G}$, $\mathbf{G}(i,m)$ is the element from the i-th row and m-th column of $\mathbf{G}$. The $\mathrm{Diag}(\boldsymbol{\theta})$ is a diagonal matrix with diagonal elements $\boldsymbol{\theta}$. The imaginary unit $j=\sqrt{-1}$. For positive semidefinite matrix $\boldsymbol{\Sigma}$, we have $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{H}{2}}$. The $\mathbf{I}_K$ is a $K\times K$ identity matrix. The elements in $\arg(\mathbf{z})$ are the phases of the input complex vectors.

II. System Model of RIS Aided Hybrid Beamforming

The RIS-aided MU-MIMO transceiver design in mmWave communication system is illustrated in Fig. 1. The base station (BS) is equipped with N transmit antennas, the RIS has N_R reflecting elements, and there are K active users with the k-th user equipped with M_k antennas. The shorter wavelength at mmWave frequencies enables more antennas to be packed in BS and user equipments, the transmit antenna number N is generally large, while the expensive RF chain makes the RF chain number N_{A} at BS small, i.e., $N_{\mathrm{A}}\ll N$. Similarly, the RF chain number at the k-th user M_{A_k} is smaller than the receiver antenna number M_k. Therefore, the transmitter is divided as digital transmitter $\mathbf{G}_{\mathrm{D}}\in\mathbb{C}^{N_{\mathrm{A}}\times L}$ and analog transmitter $\mathbf{G}_{\mathrm{A}}\in\mathbb{C}^{N\times N_{\mathrm{A}}}$, and the k-th receiver is divided into digital receiver $\mathbf{F}_{\mathrm{D}_k}\in\mathbb{C}^{L_k\times M_{\mathrm{A}_k}}$ and analog receiver $\mathbf{F}_{\mathrm{A}_k}\in\mathbb{C}^{M_{\mathrm{A}_k}\times M_k}$.

The channel $\mathbf{H}\in\mathbb{C}^{N_R\times N}$ is the channel from the BS to the RIS, the channel $\mathbf{H}_{r,k}\in\mathbb{C}^{M_k\times N_R}$ is the channel from the RIS to the k-th user, the channel $\mathbf{H}_{d,k}\in\mathbb{C}^{M_k\times N}$ is the direct channel from the BS to the k-th user. The L_k independent data streams are transmitted to the k-th user and the total data stream number is $\sum_{k=1}^{K}L_k=L$. The necessary conditions to guarantee data recovery in MU-MIMO system are

$$L_k\le\min\{M_{\mathrm{A}_k},M_k\},\ L\le\min\{N_{\mathrm{A}},N\}. \tag{1}$$

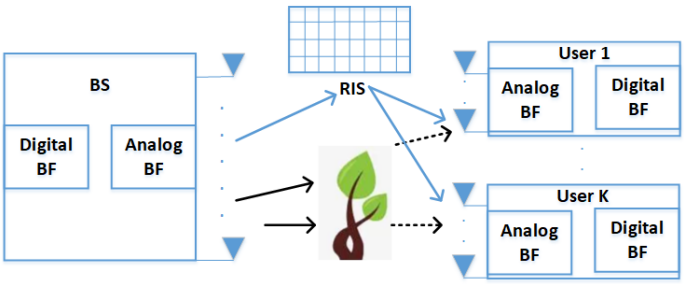

Fig. 1. RIS-aided Hybrid Beamforming in MU-MIMO system.

Letting $\mathbf{x}\in\mathbb{C}^{L\times 1}$ be the transmitted symbol with zero-mean and $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\}=\mathbf{I}_L$, the received signal at the k-th mobile user is

$$\mathbf{y}_k=\mathbf{H}_k\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\mathbf{x}+\mathbf{n}_k, \tag{2}$$

$$\mathbf{H}_k=\mathbf{H}_{r,k}\mathrm{Diag}(\boldsymbol{\theta})\mathbf{H}+\mathbf{H}_{d,k}, \tag{3}$$

where $\mathbf{H}_k$ is the effective channel from BS to the k-th user, and the noise $\mathbf{n}_k$ is circularly-symmetric complex Gaussian, its distribution information is described as $\mathbf{n}_k\sim\mathcal{CN}(\mathbf{0},\boldsymbol{\Sigma}_k)$ with covariance matrix $\boldsymbol{\Sigma}_k\succ 0$. Owing to the complexity limitation of hardware implementation in RIS design, the constraint on RIS element $\boldsymbol{\theta}$ is

$$|\boldsymbol{\theta}(n)|=1,\forall n\in[1,N_R], \tag{4}$$

where only phase rotation is allowed in $\boldsymbol{\theta}$.

At the k-th mobile user's receiver, the analog receiver $\mathbf{F}_{\mathrm{A}_k}$ and digital receiver $\mathbf{F}_{\mathrm{D}_k}$ are used to filter the received signal $\mathbf{y}_k$. Then the recovered $L_k\times 1$ data stream is

$$\hat{\mathbf{x}}_k=\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\mathbf{x}+\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{n}_k. \tag{5}$$

Since the transmitted data $\mathbf{x}$ is independent of the noise $\mathbf{n}_k$, the symbol detection MSE of the k-th user is

$$\mathrm{MSE}_k=\mathbb{E}_{\mathbf{x},\mathbf{n}_k}\left(\mathrm{Tr}\{(\mathbf{D}_k\mathbf{x}-\hat{\mathbf{x}}_k)(\mathbf{D}_k\mathbf{x}-\hat{\mathbf{x}}_k)^H\}\right), \tag{6}$$

$$=\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}-\mathbf{D}_k\|_F^2+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2, \tag{7}$$

where the matrix $\mathbf{D}_k=[\mathbf{0}_{L_k\times\sum_{i=1}^{k-1}L_i}\ \mathbf{I}_{L_k}\ \mathbf{0}_{L_k\times\sum_{i=k+1}^{K}L_i}]\in\mathbb{R}^{L_k\times L}$ is to select the k-th user's data streams.

The MSE constrained hybrid beamforming is to minimize the transmit power at the BS, subject to K users' MSE requirements and the unit modulus constraints on analog transceivers and RIS, i.e.,

$$\begin{aligned}\min_{\mathbf{G}_{\mathrm{D}},\mathbf{G}_{\mathrm{A}},\boldsymbol{\theta},\{\mathbf{F}_{\mathrm{A}_k},\mathbf{F}_{\mathrm{D}_k}\}_{k=1}^K}\ &\|\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\|_F^2\\ \text{s.t. }\ &\mathrm{MSE}_k\le\varepsilon_k,\ \forall k\in[1,K],\\ &|\mathbf{G}_{\mathrm{A}}(i,m)|=1,\ \forall i\in[1,N],m\in[1,N_{\mathrm{A}}],\\ &|\mathbf{F}_{\mathrm{A}_k}(i,m)|=1,\ \forall i\in[1,M_{\mathrm{A}_k}],m\in[1,M_k],k\in[1,K],\\ &|\boldsymbol{\theta}(n)|=1,\forall n\in[1,N_R],\end{aligned} \tag{8}$$

where $\varepsilon_k>0$ is the k-th user's MSE target.

Since all constraints of the problem (8) are nonconvex constraints, the block coordinate descent (BCD) methodology is used to find $\mathbf{G}_{\mathrm{D}},\mathbf{G}_{\mathrm{A}},\boldsymbol{\theta},\{\mathbf{F}_{\mathrm{A}_k},\mathbf{F}_{\mathrm{D}_k}\}_{k=1}^K$ sequentially.

III. Analog Beamforming

A. Analog Transmitter Design

With digital transmitter and receivers being fixed, the optimal analog transmitter is obtained by,

$$\begin{aligned}\min_{\mathbf{G}_\mathrm{A}}\ & \|\mathbf{G}_\mathrm{A}\mathbf{G}_\mathrm{D}\|_F^2\\ \text{s.t. }& \|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_\mathrm{A}\mathbf{G}_\mathrm{D}-\mathbf{D}_k\|_F^2+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{\Sigma}_k^{\frac{1}{2}}\|_F^2\le\varepsilon_k,\forall k,\\ & |\mathbf{G}_\mathrm{A}(i,m)|=1,\ \forall i,m,\end{aligned}\tag{9}$$

which is a quadratically constrained quadratic program (QCQP) with extra constant modulus constraints. Therefore, the iMM method in [11] can be applied to solve problem (9).

To get the vectorized quadratic form, the objective function is reformulated as,

$$\begin{aligned}\|\mathbf{G}_\mathrm{A}\mathbf{G}_\mathrm{D}\|_F^2 &= \|\mathrm{vec}(\mathbf{G}_\mathrm{A}\mathbf{G}_\mathrm{D})\|_2^2 &(10)\\ &= \|(\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g}\|_2^2 &(11)\end{aligned}$$

where $\mathbf{g}:=\mathrm{vec}(\mathbf{G}_\mathrm{A})$. Similarly, the first term of the MSE function in (7) is reformulated as,

$$\begin{aligned}&\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_\mathrm{A}\mathbf{G}_\mathrm{D}-\mathbf{D}_k\|_F^2 &(12)\\ &=\|\mathrm{vec}(\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_\mathrm{A}\mathbf{G}_\mathrm{D}-\mathbf{D}_k)\|_2^2 &(13)\\ &=\|\underbrace{(\mathbf{G}_\mathrm{D}^T\otimes(\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k))}_{\mathbf{C}_k}\mathrm{vec}(\mathbf{G}_\mathrm{A})-\mathrm{vec}(\mathbf{D}_k)\|_2^2 &(14)\\ &=\|\mathbf{C}_k\mathbf{g}-\mathbf{d}_k\|_2^2 &(15)\end{aligned}$$

where $\mathbf{d}_k:=\mathrm{vec}(\mathbf{D}_k)$. Therefore, the problem (9) is the same as

$$\begin{aligned}\min_{\mathbf{g}}\ & \|(\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g}\|_2^2\\ \text{s.t. }& \|\mathbf{C}_k\mathbf{g}-\mathbf{d}_k\|_2^2+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{\Sigma}_k^{\frac{1}{2}}\|_F^2\le\varepsilon_k,\forall k\\ & |\mathbf{g}(n)|=1,\ \forall n\in[1,N_\mathrm{A}N],\end{aligned}\tag{16}$$

The majorizers of the cost function and constraint functions are illustrated as follows. Firstly,

$$\begin{aligned}\|(\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g}\|_2^2 &= \mathbf{g}^H(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g} &(17)\\ \|\mathbf{C}_k\mathbf{g}-\mathbf{d}_k\|_2^2 &= \mathbf{g}^H\mathbf{C}_k^H\mathbf{C}_k\mathbf{g}+\|\mathbf{d}_k\|_2^2-2\mathrm{Re}(\mathbf{g}^H\mathbf{C}_k^H\mathbf{d}_k). &(18)\end{aligned}$$

Secondly, let $\mathbf{g}_i$ be the vectorized feasible solution $\mathbf{G}_\mathrm{A}$ of (8), one majorizer is constructed as [26, e.q. (26)]

$$\begin{aligned}\mathbf{g}^H(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g} &\le 2\mathrm{Re}(\mathbf{g}^H(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N-t_0\mathbf{I})\mathbf{g}_i)\\ &\quad+t_0\mathbf{g}^H\mathbf{I}\mathbf{g}+\mathbf{g}_i^H(t_0\mathbf{I}-\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g}_i &(19)\\ \mathbf{g}^H\mathbf{C}_k^H\mathbf{C}_k\mathbf{g} &\le 2\mathrm{Re}(\mathbf{g}^H(\mathbf{C}_k^H\mathbf{C}_k-t_k\mathbf{I})\mathbf{g}_i)\\ &\quad+t_k\mathbf{g}^H\mathbf{I}\mathbf{g}+\mathbf{g}_i^H(t_k\mathbf{I}-\mathbf{C}_k^H\mathbf{C}_k)\mathbf{g}_i &(20)\end{aligned}$$

where $t_0=N\mathrm{Tr}(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T), t_k=\mathrm{Tr}(\mathbf{C}_k^H\mathbf{C}_k)$. Therefore, a majorizer of the MSE function under $\{|\mathbf{g}(n)|=1\}_{n=1}^{N_R}$ is constructed as

$$\mathrm{MSE}_k\le\overbrace{2\mathrm{Re}(\mathbf{g}^H[(\mathbf{C}_k^H\mathbf{C}_k-t_k\mathbf{I})\mathbf{g}_i-\mathbf{C}_k^H\mathbf{d}_k])+c_k}^{\overline{\mathrm{MSE}}_k}\tag{21}$$

where $c_k=\|\mathbf{d}_k\|_2^2+2NN_\mathrm{A}t_k-\mathbf{g}_i^H\mathbf{C}_k^H\mathbf{C}_k\mathbf{g}_i+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{\Sigma}_k^{\frac{1}{2}}\|_F^2$. After those constructions, the problem (16) is solved by a series of subproblems,

Algorithm 1 K-bisection method to find $\boldsymbol{\nu}$.

initialization: set $i=0$ and use the latest finite $\boldsymbol{\nu}$ as $\boldsymbol{\nu}[0]$, otherwise $\boldsymbol{\nu}[0]=[0,0,\cdots,0]$. For analog transmitter problem, $\overline{f}_k(\boldsymbol{\nu})$ is $\overline{\mathrm{MSE}}_k(\boldsymbol{\nu})$. For RIS problem, $\overline{f}_k(\boldsymbol{\nu})$ is $\breve{\mathrm{MSE}}_k(\boldsymbol{\nu})$. For notational simplicity, $\overline{f}_k(a)$ with scalar input a means $\overline{f}_k(\boldsymbol{\nu})|_{\nu_k=a}$.
repeat
  for each $k\in[1,K]$ **do**
    if $\overline{f}_k(0)\le\varepsilon_k$, **then**
      $\nu_k=0$
    else if $|\lim_{\nu_k\to\infty}\overline{f}_k(\boldsymbol{\nu})-\varepsilon_k|\le\epsilon_0$ **then**
      $\nu_k=\infty$. Stop Algorithm 1.
    else
      Let $\nu_k^l=0,\nu_k^u=1$
      if $\overline{f}_k(1)\le\varepsilon_k$ **then**
        $\nu_k^u=1$
      else
        repeat
          $\nu_k^u=2\nu_k^u$
        until $\overline{f}_k(\nu_k^u)\le\varepsilon_k$
        $\nu_k^l=\nu_k^u/2$
      end if
      repeat
        Let $\nu_k=\frac{\nu_k^l+\nu_k^u}{2}$. If $\overline{f}_k(\nu_k)>\varepsilon_k$, then $\nu_k^l=\nu_k$. Otherwise, $\nu_k^u=\nu_k$.
      until $|\overline{f}_k(\nu_k)-\varepsilon_k+\epsilon/2|\le\epsilon/2$
    end if
  end for
  Let $\boldsymbol{\nu}[i+1]=[\nu_1,\nu_2,\cdots,\nu_K]$, and set $i=i+1$.
until $\|\boldsymbol{\nu}[i-1]-\boldsymbol{\nu}[i]\|_2\le\epsilon_1$

$$\begin{aligned}\min_{\mathbf{g}}\ & 2\mathrm{Re}(\mathbf{g}^H(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N-t_0\mathbf{I})\mathbf{g}_i)+c_0\\ \text{s.t. }& \overline{\mathrm{MSE}}_k\le\varepsilon_k,\forall k\in[1,K]\\ & |\mathbf{g}(n)|=1,\ \forall n\in[1,N_\mathrm{A}N],\end{aligned}\tag{22}$$

where $c_0=2NN_\mathrm{A}t_0-\mathbf{g}_i^H(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g}_i$. The strong duality holds between the nonconvex problem (22) and its dual problem, with dual variable $\boldsymbol{\nu}=[\nu_1,\cdots,\nu_K]^T$, under the following mild conditions [11]:

- The complex numbers in $(t_k\mathbf{I}-\mathbf{C}_k^H\mathbf{C}_k)\mathbf{g}_i+\mathbf{C}_k^H\mathbf{d}_k$ are not zeros.
- The element-wise phases of $(t_k\mathbf{I}-\mathbf{C}_k^H\mathbf{C}_k)\mathbf{g}_i+\mathbf{C}_k^H\mathbf{d}_k$ are not equal to the corresponding phases of $\sum_{l\neq k}^K[(\mathbf{C}_l^H\mathbf{C}_l-t_l\mathbf{I})\mathbf{g}_i-\mathbf{C}_l^H\mathbf{d}_l]\nu_l+(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N-t_0\mathbf{I})\mathbf{g}_i$.
- The hyperplane $\overline{\mathrm{MSE}}_k(\mathbf{g})=0$ is not parallel to the hyperplane $\sum_{l\neq k}^K\overline{\mathrm{MSE}}_l(\mathbf{g})\nu_l+2\mathrm{Re}(\mathbf{g}^H(\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N-t_0\mathbf{I})\mathbf{g}_i)=0$.

Therefore, the optimal solution of problem (22) is obtained from the Karush–Kuhn–Tucker (KKT) conditions,

$$\begin{aligned}\mathbf{g}(\boldsymbol{\nu})=\exp\Big(j\arg\big(&\sum_{k=1}^K[(t_k\mathbf{I}-\mathbf{C}_k^H\mathbf{C}_k)\mathbf{g}_i+\mathbf{C}_k^H\mathbf{d}_k]\nu_k\\ &+(t_0\mathbf{I}-\mathbf{G}_\mathrm{D}^*\mathbf{G}_\mathrm{D}^T\otimes\mathbf{I}_N)\mathbf{g}_i\big)\Big), &(23)\\ 0\le\nu_k\le\infty,\ \overline{\mathrm{MSE}}_k(\boldsymbol{\nu})\le\varepsilon_k,\ \forall k\in[1,K], &(24)\end{aligned}$$

Algorithm 2 iMM method to find analog transmitter $\mathbf{G}_{\mathrm{A}}$ in problem (9), RIS $\boldsymbol{\theta}$ in the problem (27), analog receivers $\{\mathbf{F}_{\mathrm{A_k}}\}_{k=1}^K$ in problem (39).

1: **initialization:** set $i=0$, $\mathbf{g}_i, \boldsymbol{\theta}_i, \{\mathbf{f}_{k,i}\}_{k=1}^K$ are the latest feasible solutions
2: **repeat**
3: If the problem is (9), use Algorithm 1 to find $\mathbf{g}(\boldsymbol{\nu})$ in (23) to (25).
4: If the problem is (27), use Algorithm 1 to find $\boldsymbol{\theta}(\boldsymbol{\nu})$ in (35) to (37).
5: If the problem is (39), calculate (48).
6: Denote the objective value of the corresponding problem as $f[i+1]$. Set $i=i+1$, let $\mathbf{g}_i=\mathbf{g}(\boldsymbol{\nu})$, $\boldsymbol{\theta}_i=\boldsymbol{\theta}(\boldsymbol{\nu})$, $\{\mathbf{f}_{k,i}=\mathbf{f}_k\}_{k=1}^K$.
7: **until** $i\geq 2$ and $(f[i-1]-f[i])/f[i]\leq\epsilon_2$

$$\nu_k(\overline{\mathrm{MSE}}_k(\boldsymbol{\nu})-\varepsilon_k)=0,\ \forall k\in[1,K], \tag{25}$$

where

$$\overline{\mathrm{MSE}}_k(\boldsymbol{\nu})=2\mathrm{Re}(\mathbf{g}(\boldsymbol{\nu})^H[(\mathbf{C}_k^H\mathbf{C}_k-t_k\mathbf{I})\mathbf{g}_i-\mathbf{C}_k^H\mathbf{d}_k]+c_k. \tag{26}$$

The dual variable $\boldsymbol{\nu}$ in (23) to (25) is solved in Algorithm 1. With optimal solution of the subproblem (22), the iMM method to solve problem (16) is described in Algorithm 2, which is monotonically converged to the KKT point of problem (9) [11].

B. RIS Design

With fixed transmitter and receivers, the optimization problem to get the optimal RIS is obtained by minimizing the total MSE of all K users under individual MSE constraints and RIS constraint,

$$\begin{aligned}\min_{\boldsymbol{\theta}}\ &\sum_{k=1}^K\mathrm{MSE}_k\\ \text{s.t. }&\mathrm{MSE}_k\leq\varepsilon_k,\ \forall k\in[1,K],\\ &|\boldsymbol{\theta}(n)|=1,\ \forall n\in[1,N_R].\end{aligned} \tag{27}$$

For better presentation, the first term of the MSE function in (7) is reformulated as,

$$\begin{aligned}&\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_{r,k}\mathrm{Diag}(\boldsymbol{\theta})\overbrace{\mathbf{H}\mathbf{G}}^{\mathbf{H}_G}+\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_{d,k}\mathbf{G}-\mathbf{D}_k\|_F^2\\ =&\|\underbrace{\begin{bmatrix}\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_{r,k}\mathrm{Diag}(\mathbf{h}_{G,1})\\ \cdots\\ \mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_{r,k}\mathrm{Diag}(\mathbf{h}_{G,L})\end{bmatrix}}_{\mathbf{A}_k}\boldsymbol{\theta}+\underbrace{\mathrm{vec}(\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_{d,k}\mathbf{G}-\mathbf{D}_k)}_{\mathbf{b}_k}\|_2^2\end{aligned} \tag{28}$$

$$=\|\mathbf{A}_k\boldsymbol{\theta}+\mathbf{b}_k\|_2^2, \tag{29}$$

where $\mathbf{h}_{G,l}$ is the l-th column vector of $\mathbf{H}_G:=\mathbf{H}\mathbf{G}$. Therefore, the problem (27) is the same as

$$\begin{aligned}\min_{\boldsymbol{\theta}}\ &\sum_{k=1}^K\|\mathbf{A}_k\boldsymbol{\theta}+\mathbf{b}_k\|_2^2+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2\\ \text{s.t. }&\|\mathbf{A}_k\boldsymbol{\theta}+\mathbf{b}_k\|_2^2+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2\leq\varepsilon_k,\forall k\in[1,K]\\ &|\boldsymbol{\theta}(n)|=1,\ \forall n\in[1,N_R].\end{aligned} \tag{30}$$

The quadratic function in (30) is reformulated as

$$\|\mathbf{A}_k\boldsymbol{\theta}+\mathbf{b}_k\|_2^2=\boldsymbol{\theta}^H\mathbf{A}_k^H\mathbf{A}_k\boldsymbol{\theta}+\|\mathbf{b}_k\|_2^2+2\mathrm{Re}(\boldsymbol{\theta}^H\mathbf{A}_k^H\mathbf{b}_k). \tag{31}$$

At the latest feasible solution $\boldsymbol{\theta}=\boldsymbol{\theta}_i$, one majorizer of the quadratic term in (31) is constructed as [26, e.q. (26)]

$$\begin{aligned}\boldsymbol{\theta}^H\mathbf{A}_k^H\mathbf{A}_k\boldsymbol{\theta}\leq&\bar{t}_k\boldsymbol{\theta}^H\mathbf{I}\boldsymbol{\theta}+2\mathrm{Re}(\boldsymbol{\theta}^H(\mathbf{A}_k^H\mathbf{A}_k-\bar{t}_k\mathbf{I})\boldsymbol{\theta}_i)\\ &+\boldsymbol{\theta}_i^H(\bar{t}_k\mathbf{I}-\mathbf{A}_k^H\mathbf{A}_k)\boldsymbol{\theta}_i\end{aligned} \tag{32}$$

where $\bar{t}_k=\mathrm{Tr}(\mathbf{A}_k^H\mathbf{A}_k)$. Therefore, a majorizer of the MSE function under $\{|\boldsymbol{\theta}(n)|=1\}_{n=1}^{N_R}$ is constructed as

$$\mathrm{MSE}_k\leq\overbrace{2\mathrm{Re}(\boldsymbol{\theta}^H[(\mathbf{A}_k^H\mathbf{A}_k-\bar{t}_k\mathbf{I})\boldsymbol{\theta}_i+\mathbf{A}_k^H\mathbf{b}_k])+\bar{c}_k}^{\breve{\mathrm{MSE}}_k} \tag{33}$$

where $\bar{c}_k=\|\mathbf{b}_k\|_2^2+2N_Rt_k-\boldsymbol{\theta}_i^H\mathbf{A}_k^H\mathbf{A}_k\boldsymbol{\theta}_i+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2$. After those constructions, the problem (30) can be solved by a series of subproblems,

$$\begin{aligned}\min_{\boldsymbol{\theta}}\ &\sum_{k=1}^K\breve{\mathrm{MSE}}_k\\ \text{s.t. }&\breve{\mathrm{MSE}}_k\leq\varepsilon_k,\forall k\in[1,K]\\ &|\boldsymbol{\theta}(n)|=1,\ \forall n\in[1,N_R].\end{aligned} \tag{34}$$

Under mild conditions [11], the optimal solution of problem (34) comes from the KKT conditions,

$$\boldsymbol{\theta}(\boldsymbol{\nu})=\exp(j\arg(\sum_{k=1}^K[(\bar{t}_k\mathbf{I}-\mathbf{A}_k^H\mathbf{A}_k)\boldsymbol{\theta}_i-\mathbf{A}_k^H\mathbf{b}_k](1+\nu_k))), \tag{35}$$

$$0\leq\nu_k\leq\infty,\ \breve{\mathrm{MSE}}_k(\boldsymbol{\nu})\leq\varepsilon_k,\ \forall k\in[1,K], \tag{36}$$

$$\nu_k(\breve{\mathrm{MSE}}_k(\boldsymbol{\nu})-\varepsilon_k)=0,\ \forall k\in[1,K], \tag{37}$$

where

$$\breve{\mathrm{MSE}}_k(\boldsymbol{\nu})=2\mathrm{Re}(\boldsymbol{\theta}(\boldsymbol{\nu})^H[(\mathbf{A}_k^H\mathbf{A}_k-\bar{t}_k\mathbf{I})\boldsymbol{\theta}_i+\mathbf{A}_k^H\mathbf{b}_k]+\bar{c}_k. \tag{38}$$

The dual variable $\boldsymbol{\nu}$ in (35) to (37) is solved in Algorithm 1. With optimal solution of the subproblem (34), the iMM method to solve problem (27) is described in Algorithm 2, which is monotonically converged to the KKT point of problem (27) [11].

C. Analog Receiver Design

With transmitter and digital receivers being fixed, the k-th user's analog receiver is obtained by minimizing the k-th user's MSE,

$$\begin{aligned}\min_{\mathbf{F}_{\mathrm{A}_k}}\ &\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}-\mathbf{D}_k\|_F^2+\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2\\ \text{s.t. }&|\mathbf{F}_{\mathrm{A}_k}(i,m)|=1,\ \forall i,m,\end{aligned} \tag{39}$$

The first term of the objective function in (39) is reformulated as,

$$\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}-\mathbf{D}_k\|_F^2 \tag{40}$$

$$=\|\underbrace{((\mathbf{H}_k\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}})^T\otimes\mathbf{F}_{\mathrm{D}_k})}_{\tilde{\mathbf{C}}_k}\mathrm{vec}(\mathbf{F}_{\mathrm{A}_k})-\mathrm{vec}(\mathbf{D}_k)\|_2^2 \tag{41}$$

$$=\|\tilde{\mathbf{C}}_k\mathbf{f}_k-\mathbf{d}_k\|_2^2, \tag{42}$$

where $\mathbf{f}_k := \mathrm{vec}(\mathbf{F}_{\mathrm{A}_k})$. Similarly, the second term of the objective function in (39) is reformulated as,

$$\|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2 = \|(\boldsymbol{\Sigma}_k^{\frac{T}{2}} \otimes \mathbf{F}_{\mathrm{D}_k})\mathbf{f}_k\|_2^2. \tag{43}$$

Therefore, the problem (39) is the same as

$$\begin{aligned} \min_{\mathbf{f}_k} \ & \|\tilde{\mathbf{C}}_k\mathbf{f}_k - \mathbf{d}_k\|_2^2 + \|(\boldsymbol{\Sigma}_k^{\frac{T}{2}} \otimes \mathbf{F}_{\mathrm{D}_k})\mathbf{f}_k\|_2^2 \\ \text{s.t. } & |\mathbf{f}_k(n)| = 1, \ \forall n \in [1, M_{\mathrm{A}_k}M_k]. \end{aligned} \tag{44}$$

The majorizer of the cost function is illustrated as follows. Firstly,

$$\begin{aligned} \|\tilde{\mathbf{C}}_k\mathbf{f}_k - \mathbf{d}_k\|_2^2 + \|(\boldsymbol{\Sigma}_k^{\frac{T}{2}} \otimes \mathbf{F}_{\mathrm{D}_k})\mathbf{f}_k\|_2^2 = \|\mathbf{d}_k\|_2^2 - 2\mathrm{Re}(\mathbf{f}_k^H\tilde{\mathbf{C}}_k^H\mathbf{d}_k) \\ + \mathbf{f}_k^H(\tilde{\mathbf{C}}_k^H\tilde{\mathbf{C}}_k + \boldsymbol{\Sigma}_k^T \otimes (\mathbf{F}_{\mathrm{D}_k}^H\mathbf{F}_{\mathrm{D}_k}))\mathbf{f}_k. \end{aligned} \tag{45}$$

With the latest feasible point $\mathbf{f}_{k,i}$, a majorizer of the objective function is constructed as [26, e.q. (26)]

$$\begin{aligned} &\|\tilde{\mathbf{C}}_k\mathbf{f}_k - \mathbf{d}_k\|_2^2 + \|(\boldsymbol{\Sigma}_k^{\frac{T}{2}} \otimes \mathbf{F}_{\mathrm{D}_k})\mathbf{f}_k\|_2^2 \le \|\mathbf{d}_k\|_2^2 - 2\mathrm{Re}(\mathbf{f}_k^H\tilde{\mathbf{C}}_k^H\mathbf{d}_k) \\ &+ 2\mathrm{Re}(\mathbf{f}_k^H(\tilde{\mathbf{C}}_k^H\tilde{\mathbf{C}}_k + \boldsymbol{\Sigma}_k^T \otimes (\mathbf{F}_{\mathrm{D}_k}^H\mathbf{F}_{\mathrm{D}_k}) - \tilde{t}_k\mathbf{I})\mathbf{f}_{k,i}) \\ &+ \tilde{t}_k\mathbf{f}_k^H\mathbf{I}\mathbf{f}_k + \mathbf{f}_{k,i}^H(\tilde{t}_k\mathbf{I} - \tilde{\mathbf{C}}_k^H\tilde{\mathbf{C}}_k - \boldsymbol{\Sigma}_k^T \otimes (\mathbf{F}_{\mathrm{D}_k}^H\mathbf{F}_{\mathrm{D}_k}))\mathbf{f}_{k,i} \end{aligned} \tag{46}$$

where $\tilde{t}_k = \mathrm{Tr}(\tilde{\mathbf{C}}_k^H\tilde{\mathbf{C}}_k + \boldsymbol{\Sigma}_k^T \otimes (\mathbf{F}_{\mathrm{D}_k}^H\mathbf{F}_{\mathrm{D}_k}))$. Under the constraint $\{|\mathbf{f}_k(n)| = 1\}_{n=1}^{N_R}$, the quadratic term $\tilde{t}_k\mathbf{f}_k^H\mathbf{I}\mathbf{f}_k$ turns into a constant term. After removing the terms unrelated to the optimization variable $\mathbf{f}_k$ in the majorizer (46), the problem (44) is tackled by a series of subproblems,

$$\begin{aligned} \min_{\mathbf{f}_k} \ & \mathrm{Re}(\mathbf{f}_k^H[(\tilde{\mathbf{C}}_k^H\tilde{\mathbf{C}}_k + \boldsymbol{\Sigma}_k^T \otimes (\mathbf{F}_{\mathrm{D}_k}^H\mathbf{F}_{\mathrm{D}_k}) - \tilde{t}_k\mathbf{I})\mathbf{f}_{k,i} - \tilde{\mathbf{C}}_k^H\mathbf{d}_k]) \\ \text{s.t. } & |\mathbf{f}_k(n)| = 1, \ \forall n \in [1, M_{\mathrm{A}_k}M_k], \end{aligned} \tag{47}$$

The optimal solution of the problem (47) is

$$\begin{aligned} \mathbf{f}_k = \exp\Big(j\arg\big([\tilde{t}_k\mathbf{I} - \tilde{\mathbf{C}}_k^H\tilde{\mathbf{C}}_k - \boldsymbol{\Sigma}_k^T \otimes (\mathbf{F}_{\mathrm{D}_k}^H\mathbf{F}_{\mathrm{D}_k})]\mathbf{f}_{k,i} \\ + \tilde{\mathbf{C}}_k^H\mathbf{d}_k\big)\Big), \end{aligned} \tag{48}$$

The iMM method to solve problem (39) is described in Algorithm 2, which is monotonically converged to the KKT point of problem (39) [11].

IV. Digital Beamforming

A. Digital Transmitter Design

With fixed analog transmitter and receivers, the optimal digital transmitter $\mathbf{G}_{\mathrm{D}}$ is obtained as,

$$\begin{aligned} &\min_{\mathbf{G}_{\mathrm{D}}} \ \|\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\|_F^2 \\ &\text{s.t.} \|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k\|_F^2 + \|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2 \le \varepsilon_k, \forall k, \end{aligned} \tag{49}$$

which is a convex QCQP problem, and can be solved by the standard interior point method with computational complexity order $\mathcal{O}(N_{\mathrm{A}}^{3.5})$ [27], [28]. To reduce the computational complexity, the ADMM method is proposed to solve it. Since the cost function $\|\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\|_F^2$ makes the normalized beamforming vectors and power vectors cannot be decoupled in the fixed point method [29], which cannot be used to tackle problem (49).

For better presentation, the first term of the MSE function in (7) is reformulated as,

$$\|\overbrace{\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\mathbf{H}_k\mathbf{G}_{\mathrm{A}}}^{\mathbf{B}_k}\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k\|_F^2 = \|\mathbf{B}_k\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k\|_F^2 \tag{50}$$

Therefore, the problem (49) is the same as

$$\begin{aligned} \min_{\mathbf{G}_{\mathrm{D}}} \ & \|\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\|_F^2 \\ \text{s.t. } & \|\mathbf{B}_k\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k\|_F^2 \le \varepsilon_k - \|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2, \forall k \end{aligned} \tag{51}$$

With slack variables $\{\mathbf{E}_k \in \mathbb{C}^{L_k \times L}\}_{k=1}^K$, the problem (51) is equivalent to

$$\begin{aligned} \min_{\mathbf{G}_{\mathrm{D}}, \{\mathbf{E}_k\}_{k=1}^K} \ & \|\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\|_F^2 \\ \text{s.t. } & \|\mathbf{E}_k\|_2^2 \le \varepsilon_k - \|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2, \forall k \in [1, K] \\ & \mathbf{E}_k = \mathbf{B}_k\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k, \forall k \in [1, K]. \end{aligned} \tag{52}$$

Furthermore, we can construct an indicator function $\mathbb{I}_k(\mathbf{E}_k)$ to represent the k-th MSE constraint as

$$\mathbb{I}_k(\mathbf{E}_k) = \begin{cases} 0, & \text{If } \|\mathbf{E}_k\|_F^2 \le \varepsilon_k - \|\mathbf{F}_{\mathrm{D}_k}\mathbf{F}_{\mathrm{A}_k}\boldsymbol{\Sigma}_k^{\frac{1}{2}}\|_F^2, \\ \infty, & \text{Otherwise.} \end{cases} \tag{53}$$

Therefore, the problem (52) is equivalent to

$$\begin{aligned} \min_{\mathbf{G}_{\mathrm{D}}, \{\mathbf{E}_k\}_{k=1}^K} \ & \|\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\|_F^2 + \sum_{k=1}^K \mathbb{I}_k(\mathbf{E}_k) \\ \text{s.t. } & \mathbf{E}_k = \mathbf{B}_k\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k, \forall k \in [1, K], \end{aligned} \tag{54}$$

which is a two-block convex ADMM formulation [30].

With Lagrange multiplier $\{\boldsymbol{\Lambda}_k \in \mathbb{C}^{L_k \times L}\}_{k=1}^K$ and penalty factor μ, the augmented Lagrange of the problem (54) is

$$\begin{aligned} &\mathcal{L}(\mathbf{G}_{\mathrm{D}}, \{\mathbf{E}_k, \boldsymbol{\Lambda}_k\}_{k=1}^K) \\ &= \|\mathbf{G}_{\mathrm{A}}\mathbf{G}_{\mathrm{D}}\|_F^2 + 2\sum_{k=1}^K \mathrm{Re}(\mathrm{Tr}(\boldsymbol{\Lambda}_k^H(\mathbf{B}_k\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k - \mathbf{E}_k))) \\ &\quad + \sum_{k=1}^K \mathbb{I}_k(\mathbf{E}_k) + \mu\sum_{k=1}^K \|\mathbf{B}_k\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k - \mathbf{E}_k\|_F^2. \end{aligned} \tag{55}$$

To update the primal and dual variables sequentially, the ADMM iteration steps are

$$\mathbf{G}_{\mathrm{D}} \leftarrow \arg\min_{\mathbf{G}_{\mathrm{D}}} \mathcal{L}(\mathbf{G}_{\mathrm{D}}, \{\mathbf{E}_k, \boldsymbol{\Lambda}_k\}_{k=1}^K), \tag{56}$$

$$\mathbf{E}_k \leftarrow \arg\min_{\mathbf{E}_k} \mathcal{L}(\mathbf{G}_{\mathrm{D}}, \{\mathbf{E}_k, \boldsymbol{\Lambda}_k\}_{k=1}^K), \tag{57}$$

$$\boldsymbol{\Lambda}_k \leftarrow \boldsymbol{\Lambda}_k + \mu(\mathbf{B}_k\mathbf{G}_{\mathrm{D}} - \mathbf{D}_k - \mathbf{E}_k), \tag{58}$$

Algorithm 3 ADMM method to find digital transmitter $\mathbf{G}$ in problem (49).

1: **initialization:** set $\{\mathbf{E}_k, \mathbf{\Lambda}_k\}_{k=1}^K$ as the previous values, otherwise $\{\mathbf{E}_k = 0, \mathbf{\Lambda}_k = 0\}_{k=1}^K$. Precompute $\mathbf{B} = (\mathbf{G}_{\mathrm{A}}^H \mathbf{G}_{\mathrm{A}} + \mu \sum_{k=1}^K \mathbf{B}_k^H \mathbf{B}_k)^{-1}$.
2: **repeat**
3: Update $\mathbf{G}_{\mathrm{D}}$ in (59) with precomputed $\mathbf{B}$
4: Update $\mathbf{E}_k$ in (60) for all $k \in [1, K]$
5: Update $\mathbf{\Lambda}_k$ in (58) for all $k \in [1, K]$
6: **until** The termination criteria below (60) are satisfied.

where the left arrow notation $b \leftarrow a$ means that the numeric value of a is assigned to the symbol b. The closed-form solutions of (56) and (57) are listed as follows,

$$\mathbf{G}_{\mathrm{D}} = \underbrace{(\mathbf{G}_{\mathrm{A}}^H \mathbf{G}_{\mathrm{A}} + \mu \sum_{k=1}^K \mathbf{B}_k^H \mathbf{B}_k)^{-1}}_{\mathbf{B}} \sum_{k=1}^K (\mathbf{B}_k^H (\mu(\mathbf{D}_k + \mathbf{E}_k) - \mathbf{\Lambda}_k)), \tag{59}$$

$$\mathbf{E}_k = \begin{cases} \hat{\mathbf{E}}_k, & \text{If } \|\hat{\mathbf{E}}_k\|_F^2 \le \varepsilon_k - \|\mathbf{F}_{\mathrm{D}_k} \mathbf{F}_{\mathrm{A}_k} \mathbf{\Sigma}_k^{\frac{1}{2}}\|_F^2 \\ \hat{\mathbf{E}}_k / \|\hat{\mathbf{E}}_k\|_F \cdot \sqrt{\varepsilon_k - \|\mathbf{F}_{\mathrm{D}_k} \mathbf{F}_{\mathrm{A}_k} \mathbf{\Sigma}_k^{\frac{1}{2}}\|_F^2}, & \text{Otherwise} \end{cases}, \tag{60}$$

where $\hat{\mathbf{E}}_k = \mathbf{B}_k \mathbf{G}_{\mathrm{D}} - \mathbf{D}_k + \mathbf{\Lambda}_k / \mu$. The termination criteria of the ADMM method are $\{\|\mathbf{B}_k \mathbf{G}_{\mathrm{D}} - \mathbf{D}_k - \mathbf{E}_k\|_F \le \epsilon^p\}_{k=1}^K$, and the Frobenius norm of the difference between two $\mathbf{G}_{\mathrm{D}}$ in successive iterations is smaller than ϵ^d, where ϵ^p and ϵ^d are the tolerances.

The proposed ADMM method is described in Algorithm 3. If the transmitter problem (49) is strictly feasible, the two functions in the objective function of (54) are closed, proper, and convex. Furthermore, the strong duality ensures that the unaugmented Lagrangian of (54) has a saddle point. Therefore, the limit solution of the proposed ADMM method converges to the optimal solution under the mild strictly feasible condition [30, p.16].

Since there is only one matrix inversion to calculate $\mathbf{B}$ in (59), its major computational complexity comes from the matrix multiplications. Therefore, the computational complexity order of the proposed ADMM approach is $\mathcal{O}(N_{\mathrm{A}}^2 L)$, which is lower than that of the interior point method with $\mathcal{O}(N_{\mathrm{A}}^{3.5})$.

B. Digital Receiver Design

With transmitter and analog receivers being fixed, the k-th user's digital receiver is obtained by minimizing the k-th user's MSE,

$$\min_{\mathbf{F}_{\mathrm{D}_k}} \|\mathbf{F}_{\mathrm{D}_k} \mathbf{F}_{\mathrm{A}_k} \mathbf{H}_k \mathbf{G}_{\mathrm{A}} \mathbf{G}_{\mathrm{D}} - \mathbf{D}_k\|_F^2 + \|\mathbf{F}_{\mathrm{D}_k} \mathbf{F}_{\mathrm{A}_k} \mathbf{\Sigma}_k^{\frac{1}{2}}\|_F^2. \tag{61}$$

Since the MSE function is a convex quadratic function of $\mathbf{F}_{\mathrm{D}_k}$, its optimal solution occurs at the stationary point,

$$\mathbf{F}_{\mathrm{D}_k} = \mathbf{D}_k \mathbf{H}_{\mathrm{E}_k}^H (\mathbf{H}_{\mathrm{E}_k} \mathbf{H}_{\mathrm{E}_k}^H + \mathbf{F}_{\mathrm{A}_k} \mathbf{\Sigma}_k \mathbf{F}_{\mathrm{A}_k}^H)^{-1}, \tag{62}$$

where $\mathbf{H}_{\mathrm{E}_k} = \mathbf{F}_{\mathrm{A}_k} \mathbf{H}_k \mathbf{G}_{\mathrm{A}} \mathbf{G}_{\mathrm{D}}$ is an effective channel. The computational complexity order of the receiver design is $\mathcal{O}(M_{\mathrm{A}_k}^3)$.

Algorithm 4 Joint RIS, analog and digital transceivers.

1: **initialization:** use initialization scheme (63), and set $i_4 = 0$.
2: **repeat**
3: Use Algorithm 3 to find digital transmitter $\mathbf{G}_{\mathrm{D}}$, denote the transmit power as $P[i_4 + 1]$.
4: Use Algorithm 2 to find following beamformers in sequence, analog transmitter $\mathbf{G}_{\mathrm{A}}$, RIS $\boldsymbol{\theta}$, analog receivers $\{\mathbf{F}_{\mathrm{A_k}}\}_{k=1}^K$.
5: Use (62) as the digital receiver, and set $i_4 = i_4 + 1$.
6: **until** $i_4 \ge 2$ and $(P[i_4 - 1] - P[i_4])/P[i_4] \le \epsilon_4$

TABLE 1
COMPUTATIONAL COMPLEXITY ORDERS OF DIFFERENT ALGORITHMS

	Interior point method	ADMM	iMM
Analog Tx	$\mathcal{O}((N_{\mathrm{A}} N)^{6.5})$(SDR)	$\times$	$\mathcal{O}((N_{\mathrm{A}} N)^2)$
Analog Rx	$\mathcal{O}((M_{\mathrm{A}_k} M_k)^{6.5})$(SDR)	$\times$	$\mathcal{O}((M_{\mathrm{A}_k} M_k)^2)$
RIS	$\mathcal{O}(N_R^{6.5})$(SDR)	$\times$	$\mathcal{O}(N_R^2)$
Digital Tx	$\mathcal{O}(N_{\mathrm{A}}^{3.5})$(convex QCQP)	$\mathcal{O}(N_{\mathrm{A}}^2 L)$	$\times$

V. JOINT ALGORITHM & COMPLEXITY ANALYSIS

The BCD methodology requires an initial $\mathbf{G}_{\mathrm{A}}, \boldsymbol{\theta}, \{\mathbf{F}_{\mathrm{A}_k}, \mathbf{F}_{\mathrm{D}_k}\}_{k=1}^K$ to make the problem (8) feasible. Therefore, we propose a simple initialization scheme: the phases of elements in $\mathbf{G}_{\mathrm{A}}, \boldsymbol{\theta}, \{\mathbf{F}_{\mathrm{A}_k}\}_{k=1}^K$ are generated uniformly and independently in $[0, 2\pi]$, and the digital receivers are

$$\{\mathbf{F}_{\mathrm{D}_k} = \sqrt{\frac{\varepsilon_k}{2\mathrm{Tr}(\mathbf{F}_{\mathrm{A}_k} \mathbf{\Sigma}_k \mathbf{F}_{\mathrm{A}_k}^H)}} \mathbf{I}_{M_{\mathrm{A}_k}}(1 : L_k, :)\}_{k=1}^K, \tag{63}$$

where $\mathbf{I}_{M_{\mathrm{A}_k}}(1 : L_k, :)$ is a matrix filled with the first L_k rows of identity matrix $\mathbf{I}_{M_{\mathrm{A}_k}}$. If the elements in $\{\mathbf{H}_{r,k}, \mathbf{H}_{d,k}\}_{k=1}^K, \mathbf{H}$ are independent continuous random variables, then with probability one, the problem (8) is feasible and all MSE constraints are strictly feasible under the initialization scheme (63).

The joint algorithm to solve problem (8) is described in Algorithm 4. The MSE minimizations in (27), (39) and (61) make the resulting MSEs smaller than the MSE targets $\{\varepsilon_k\}_{k=1}^K$. Then, any $\mathrm{MSE}_k < \varepsilon_k$ ensures that the k-th user's digital beamformer vectors can be scaled down without violating any constraints. Therefore, the next optimal digital transmitter must reduce the transmission power, i.e., $P[i_4 + 1] \le P[i_4]$. Since the transmit power is monotonically nonincreasing and the transmit power is bounded below from zero, Algorithm 4 converges.

The computational complexity orders of the proposed method are compared with the interior point method in Table 1. It can be seen from Table 1 that the combination of the proposed ADMM and iMM methods has lower computational complexity order than other methods.

VI. ADAPTIVE ALGORITHMS FOR CHANGING PATH LOSS

Since the mobile users are usually moving around, it is important to design the algorithm, which considers the large-

scale path loss and is adaptive to the changing path loss. The large-scale path loss is modelled as,

$$PL = PL_0(\frac{D}{D_0})^{-\alpha}, \tag{64}$$

where $PL_0 = -30\text{dB}$, $D_0 = 1$ m, α is the path-loss exponent, and D is the distance between transmitter and receiver.

A. Adaptive Parametrization for ADMM

When the user is moving away from the BS, the large-scale path loss also changes quickly. Therefore, the proposed algorithm must be adaptive to the channel path loss automatically, instead of tuning parameters constantly. For example, a good penalty factor makes the ADMM method converge quickly. Otherwise, if $\mu = 10^3, \alpha = 4$ and the BS-user distance is 100 meters, the magnitude of $\mu\sum_{k=1}^{K}\mathbf{B}_k^H\mathbf{B}_k$ is around 10^{-6}. Since magnitude of the diagonal elements of $\mathbf{G}_{\text{A}}^H\mathbf{G}_{\text{A}}$ is N, the contribution of $\mu\sum_{k=1}^{K}\mathbf{B}_k^H\mathbf{B}_k$ in the matrix $\mathbf{B} = (\mathbf{G}_{\text{A}}^H\mathbf{G}_{\text{A}} + \mu\sum_{k=1}^{K}\mathbf{B}_k^H\mathbf{B}_k)^{-1}$ is diminishing. Therefore, we propose the following adaptive parametrization for the ADMM method,

$$\mu = 10N/\|\sum_{k=1}^{K}\mathbf{B}_k^H\mathbf{B}_k\|_F. \tag{65}$$

By using above penalty factor, the magnitude of $\mu\sum_{k=1}^{K}\mathbf{B}_k^H\mathbf{B}_k$ is about ten times larger than that of $\mathbf{G}_{\text{A}}^H\mathbf{G}_{\text{A}}$. This process coincides the common sense that $\mu\sum_{k=1}^{K}\mathbf{B}_k^H\mathbf{B}_k$ is the major factor in the ADMM iteration, while $\mathbf{G}_{\text{A}}^H\mathbf{G}_{\text{A}}$ is a regularization parameter.

B. Channel Normalization for iMM

If all users are 100 meters away from the BS and $\alpha = 4$, the channel magnitude is about 10^{-7}. To compensate for this large path loss, the optimized power $\|\mathbf{G}_{\text{A}}\mathbf{G}_{\text{D}}\|_F^2$ should be very large. Since the optimal dual variable $\boldsymbol{\nu}$ is the sensitivity of the optimal transmit power with respect to constraint perturbation, the dual variable $\boldsymbol{\nu}$ must be large under large path loss.

Furthermore, from (7), we know that if $(\{\mathbf{H}_k\}_{k=1}^K, \mathbf{G}_{\text{A}}\mathbf{G}_{\text{D}})$ is an optimal pair, then $(\{\mathbf{H}_k\}_{k=1}^K/s, \mathbf{G}_{\text{A}}\mathbf{G}_{\text{D}} \cdot s)$ is also an optimal pair, where s is a positive scaling factor. Therefore, we can normalize the channel such that the obtained power $\|\mathbf{G}_{\text{A}}\mathbf{G}_{\text{D}}\|_F^2$ is relatively small, then the search space for dual variable $\boldsymbol{\nu}$ is smaller and the computational complexity is reduced.

The scaling factor is chosen as $s = \min\{\|\mathbf{H}_{d,k}\|_F\}_{k=1}^K$. The channels $\{\mathbf{H}_k\}_{k=1}^K$ are replaced by $\{\mathbf{H}_k\}_{k=1}^K/s$, and the optimized transmit power will be divided by s^2.

VII. Simulation Results

The small scale channel is modelled to be Rician fading,

$$\hat{\mathbf{H}} = \sqrt{\frac{\beta}{\beta+1}}\hat{\mathbf{H}}_{\text{LoS}} + \sqrt{\frac{1}{\beta+1}}\hat{\mathbf{H}}_{\text{NLoS}}, \tag{66}$$

where β is the Rician factor. The line-of-sight channel $\hat{\mathbf{H}}_{\text{LoS}}$ comes from the standard mmWave channel [1], while the non-line-of-sight channel $\hat{\mathbf{H}}_{\text{NLoS}}$ is a Rayleigh fading channel [31], [32], where all elements in $\hat{\mathbf{H}}_{\text{NLoS}}$ are independently distributed as $\mathcal{CN}(0,1)$. The Rician factor in channel $\mathbf{H}_d$ is $\beta = \infty$, and that in other channels are $\beta = 1$. The position of BS, users and RIS are illustrated in Fig. 2. The BS is always placed in the origin with coordinate (0,0). The x-axis coordinate of users is 100m, and the y-axis coordinate of users is randomly generated in [-0.5m, 0.5m]. The y-axis coordinate of RIS is 2m, and the x-axis coordinate of RIS is 100m unless stated otherwise.

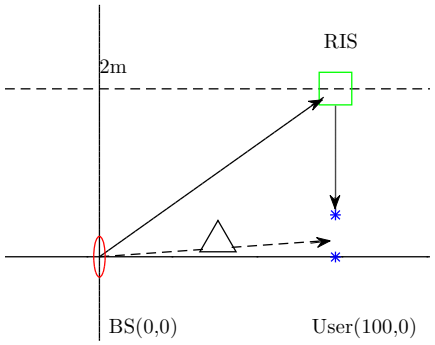

Fig. 2. The position of BS, users and RIS.

Unless stated otherwise, the path-loss exponents of the BS-User, BS-RIS, RIS-User channels are $\alpha = 4$, $\alpha = 2$, $\alpha = 2$, respectively. The transmit antenna number is $N = 10$, the receiver antenna numbers are $\{M_k = 10\}_{k=1}^K$, the RIS element number is $N_R = 40$, the transmit and receive RF chain numbers are $N_{\text{A}} = 4, \{M_{\text{A}_k} = 4\}_{k=1}^K$, respectively. The user number is $K = 2$ and the data stream numbers are $\{L_k = 1\}_{k=1}^K$. The MSE targets are fixed as $\{\varepsilon_k = 0.01\}_{k=1}^K$, and the noise covariance matrices are $\{\boldsymbol{\Sigma}_k = 0.01\mathbf{I}_{M_k}\}_{k=1}^K$.

The simulation parameters of different algorithms are described as follows. In Algorithm 1, the bisection accuracy level is $\epsilon = 10^{-8}, \epsilon_0 = 10^{-4}$, and the termination threshold is $\epsilon_1 = 10^{-4}$. In Algorithm 2, the iMM termination threshold is $\epsilon_2 = 2 \times 10^{-4}$. In Algorithm 3, the ADMM tolerances are $\epsilon^p = \epsilon^d = 10^{-8}$, and the penalty factor is set in (65). In Algorithm 4, the power termination threshold is $\epsilon_4 = 10^{-2}$.

All the simulation results are averaged over 300 random channel realizations. All optimization problems are solved on a laptop PC with 3.6 GHz CPU and 128GB RAM, and the simulation software is Matlab. The results of the SDR method [33], [34] with interior point algorithm are obtained with the solver CVX [35], and the Gaussian randomization number is 10^5. The parameters in the penalty convex-concave procedure (CCP) method [36] are $\tau_0 = 1, \tau_{max} = 10^5, \mu = 2$.

A. Comparison of Different Methods

The transmit powers, $10\log_{10}(\|\mathbf{G}_{\text{A}}\mathbf{G}_{\text{D}}\|_F^2/\boldsymbol{\Sigma}_1(1,1))$ (dB), of the proposed method and other methods are compared in Fig. 3. It can be seen that the transmit power of the proposed iMM method is monotonically decreasing, which validates the convergence analysis for Algorithm 4. Furthermore, the

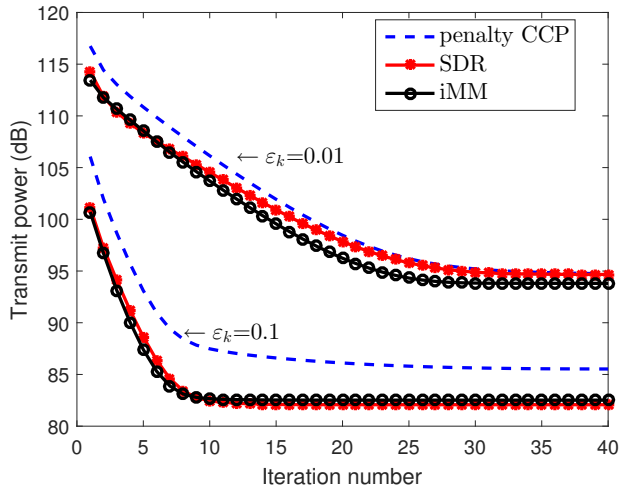

Fig. 3. Transmit powers under different methods.

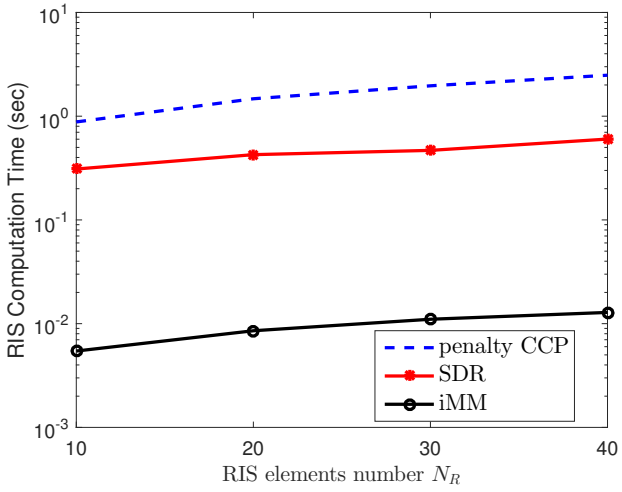

Fig. 5. RIS computational times of different methods.

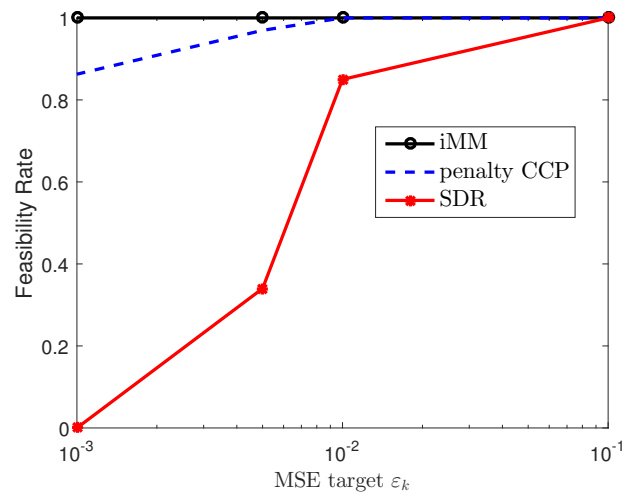

Fig. 4. Feasibility rate of different methods.

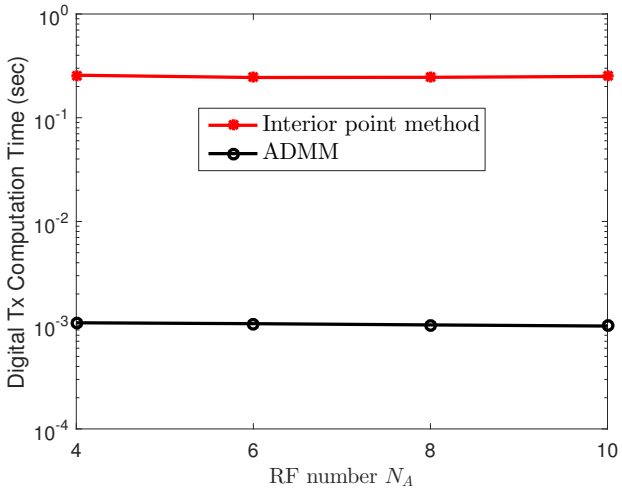

Fig. 6. Digital transmitter computational times of different methods.

transmit powers of the proposed method with different MSE targets are comparable to that of SDR and penalty CCP methods, which reveal the good performance of the proposed method.

The feasibility rates of different methods are compared in Fig. 4. It can be seen from Fig. 4 that the feasibility rate of the iMM method is always one, while that of the SDR method is monotonically decreasing when the MSE target is decreasing. A few instances of the penalty CCP method also fail to be feasible, this is due to the numeric instability of the penalty methodology. Therefore, the iMM method can always provide feasible solution when the elements of the wireless channels are independent, while the randomization based SDR method fails to provide feasible solution when the MSE target is small.

To compare the RIS computational times, the proposed method is compared with the other methods in Fig. 5. It can be seen from Fig. 5 that the computational times of the iMM method are almost 10 times faster than that of the SDR method, which validates the complexity analysis in Table 1. It is also seen from Fig. 5 that the computational time of the penalty CCP method is the largest among all methods, which is owing to the large number of convex QCQP subproblems in the penalty method. Therefore, Figs. 3 and 5 reveal that the proposed iMM method achieves similar transmit power but with a much lower computational complexity than other methods.

The digital transmitter computational times of different methods are depicted in Fig. 6. It is seen that the computational times of the proposed ADMM method are much smaller than the interior point method, which validates the complexity analysis in Table 1. Furthermore, the proposed ADMM method is not sensitive to the change of RF number.

B. Impact of Path Loss and RIS Location

The transmit powers under different RIS locations are depicted in Fig. 7. When the path-loss exponent from BS to

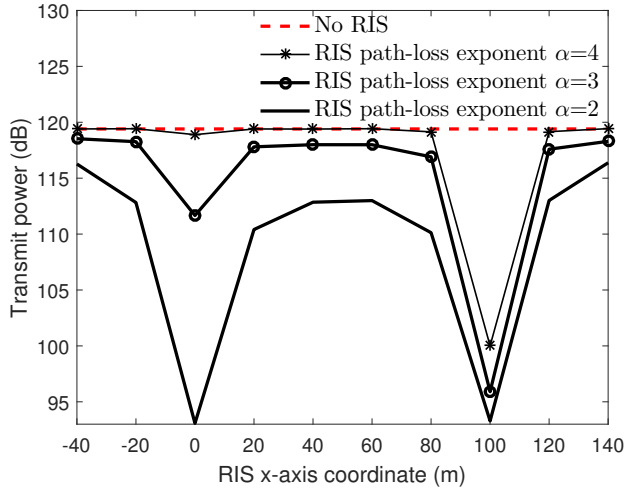

Fig. 7. Transmit powers under different RIS locations and RIS-user path-loss exponents.

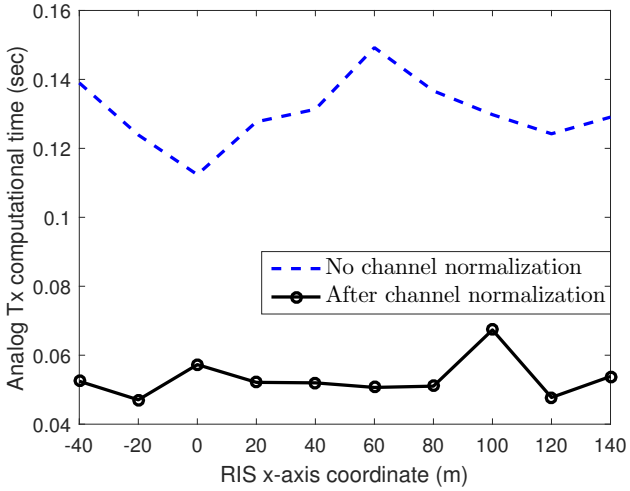

Fig. 9. Analog transmitter computational times under different channels.

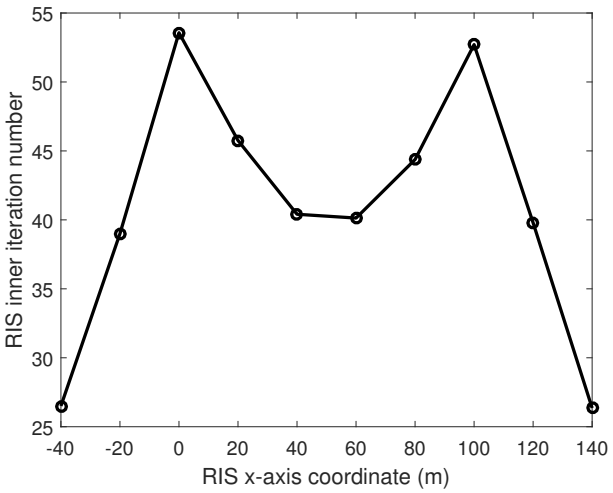

Fig. 8. RIS inner iteration number under different RIS locations.

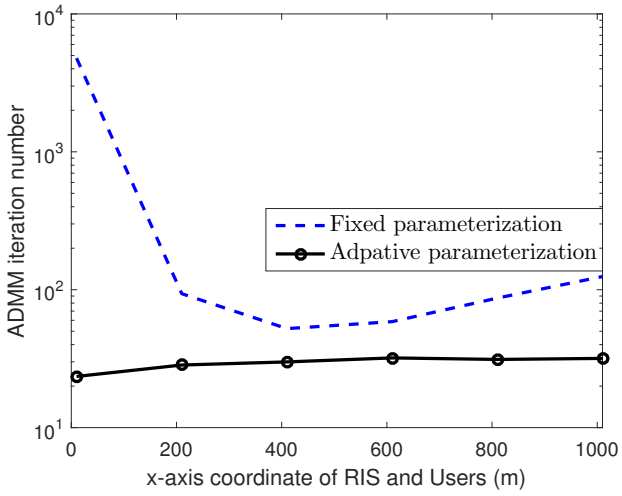

Fig. 10. ADMM iteration number under different user locations.

RIS is the same as that from RIS to user, i.e., $\alpha = 2$, it is seen that the transmit power is the smallest when RIS is close to BS or users, which shows that the RIS should be placed near to the users or BS. When the path-loss exponent from BS to RIS is smaller than that from RIS to user, i.e., $\alpha > 2$, it is seen that the transmit power is the smallest when RIS is close to users, which shows that the RIS should be placed near to the users. It is also seen from the figure that the RIS aided system provides lower transmit power than the system without RIS, which reveals the fact that RIS can assist the mmWave system to conquer the blockage issue.

The RIS inner iteration number under different RIS locations is described in Fig. 8. It is seen that the iteration number of RIS design is large when RIS is close to BS or user. This is due to the fact that the space to reduce the transmit power is large when RIS is close to BS or user, which is shown in Fig. 7.

C. Impact of Adaptive Algorithms

The computational times of the analog transmitter under different channels are described in Fig. 9. It is seen that the computational time of the analog transmitter with channel normalization is about a half of that without channel normalization. This reveals that channel normalization can efficiently speedup the dual computation in analog transmitter, which validates the analysis in Section VI-B.

The iteration number of the digital transmitter under different parameterizations are described in Fig. 10. It is seen that the iteration number of the ADMM method under adaptive parameterizations is stable when the locations of RIS and users are changed from 10 meters to 1010 meters, while the iteration number of the ADMM method with fixed parameterization ($\mu = 10^{10}$) changes a lot when the users are moving. Therefore, the proposed ADMM method is adaptive to the changing of the users' location.

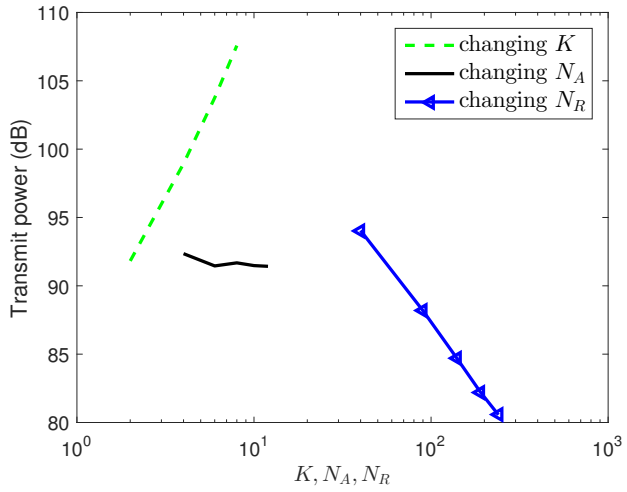

Fig. 11. Transmitter powers under different parameters.

D. Impact of Different Parameters

The transmitter powers under different parameters are described in Fig. 11. It is seen from the figure that the transmit power is monotonically increasing when the user number K is increased from 2 to 10, the transmit power is decreased when the RF chain number N_A is increased, and the transmit power is monotonically decreasing when the RIS elements number N_R is increased. This shows that RIS without expensive RF chains can be used to save transmit power in mmWave systems.

VIII. Conclusions

In this paper, the RIS aided hybrid beamforming with MSE constraints was proposed and solved. The iMM method has been proposed to obtain analog transceivers and RIS, and the ADMM method has been proposed to obtain digital transmitter. The proposed iMM and ADMM methods are simpler than the SDR and interior point methods, respectively. In order to counter against the changing large-scale path loss, the iMM method is helped by channel normalization to reduce the computational complexity, while the ADMM method is helped by the adaptive parameterizations to robust against the changing large-scale path loss. Simulation results showed that the computational time of proposed methods are much faster than other methods, and the proposed methods are robust against the changing large-scale path loss.

References

[1] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.

[2] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, 2016.

[3] L. Jiang and H. Jafarkhani, "Multi-user analog beamforming in millimeter wave MIMO systems based on path angle information," *IEEE Trans. on Wireless Communications*, vol. 18, no. 1, pp. 608–619, 2019.

[4] A. Arora, C. G. Tsinos, B. Shankar Mysore R, S. Chatzinotas, and B. Ottersten, "Majorization-minimization algorithms for analog beamforming with large-scale antenna arrays," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019, pp. 1–5.

[5] A. Arora, C. G. Tsinos, B. S. M. R. Rao, S. Chatzinotas, and B. Ottersten, "Hybrid transceivers design for large-scale antenna arrays using majorization-minimization algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 701–714, 2020.

[6] J. Tranter, N. D. Sidiropoulos, X. Fu, and A. Swami, "Fast unit-modulus least squares with applications in beamforming," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2875–2887, 2017.

[7] C. G. Tsinos and B. Ottersten, "An efficient algorithm for unit-modulus quadratic programs with application in beamforming for wireless sensor networks," *IEEE Signal Processing Letters*, vol. 25, no. 2, pp. 169–173, 2018.

[8] X. Xue, Y. Wang, L. Dai, and C. Masouros, "Relay hybrid precoding design in millimeter-wave massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 8, pp. 2011–2026, 2018.

[9] Z. Lin, M. Lin, B. Champagne, W.-P. Zhu, and N. Al-Dhahir, "Secrecy-energy efficient hybrid beamforming for satellite-terrestrial integrated networks," *IEEE Transactions on Communications*, pp. 1–1, 2021.

[10] L. F. Abanto-Leon, M. Hollick, and G. H. Sim, "Hybrid precoding for multi-group multicasting in mmWave systems," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–7.

[11] X. He and J. Wang, "QCQP with extra constant modulus constraints: Theory and application to SINR constrained mmWave hybrid beamforming," *IEEE Transactions on Signal Processing (Early Access)*, pp. 1–15, 2022.

[12] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. Ben Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3693–3708, 2019.

[13] J. Cong, T. Lin, and Y. Zhu, "Hybrid MMSE beamforming for multiuser millimeter-wave communication systems," *IEEE Communications Letters*, vol. 22, no. 11, pp. 2390–2393, 2018.

[14] S. Gong, C. Xing, P. Yue, L. Zhao, and T. Q. S. Quek, "Hybrid analog and digital beamforming for RIS-assisted mmwave communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[15] S. H. Hong, J. Park, S.-J. Kim, and J. Choi, "Hybrid beamforming for intelligent reflecting surface aided millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7343–7357, 2022.

[16] J.-C. Chen, "Joint transceiver and intelligent reflecting surface design for mmwave massive MIMO systems," *IEEE Systems Journal*, pp. 1–12, 2022.

[17] E. E. Bahingayi and K. Lee, "Low-complexity beamforming algorithms for IRS-aided single-user massive MIMO mmwave systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[18] M. Shi, X. Li, T. Fan, J. Liu, and S. Lv, "A low complexity algorithm for achievable rate maximization in mmwave systems aided by IRS," *IEEE Wireless Communications Letters*, vol. 11, no. 10, pp. 2215–2219, 2022.

[19] J. Zhang, X. Hu, and C. Zhong, "Phase calibration for intelligent reflecting surfaces assisted millimeter wave communications," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1026–1040, 2022.

[20] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. D. Renzo, "Stochastic learning-based robust beamforming design for ris-aided millimeter-wave systems in the presence of random blockages," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 1057–1061, 2021.

[21] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, 2020.

[22] M. M. Zhao, Q. Wu, M. J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Two-timescale beamforming optimization,," *arXiv preprint arXiv:1912.01818*, 2019.

[23] M. M. Zhao, Q. Wu, M. J. Zhao, and R. Zhang, "Exploiting amplitude control in intelligent reflecting surface aided wireless communication with imperfect CSI,," *arXiv preprint arXiv:2005.07002*, 2020.

[24] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. on Selected Areas in Communications*, vol. 38, no. 8, pp. 1719–1734, 2020.

[25] H. Niu, Z. Chu, F. Zhou, C. Pan, D. W. K. Ng, and H. X. Nguyen, "Double intelligent reflecting surface-assisted multi-user MIMO mmwave

systems with hybrid precoding," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1575–1587, 2022.
[26] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.
[27] A. Nemirovski, *Interior Point Polynomial Time Methods in Convex Programming*. Lecture Notes, 1996.
[28] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1426–1438, 2006.
[29] S. Shuying, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: MMSE balancing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3702–3712, 2008.
[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
[31] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems-part I: chunk allocation," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2734–2744, 2009.
[32] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems-part II: Joint chunk, power and bit allocation," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 499–509, 2012.
[33] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.
[34] X. He, L. Huang, and J. Wang, "Novel relax-and-retract algorithm for intelligent reflecting surface design," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1995–2000, 2021.
[35] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.
[36] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optim. Eng.*, vol. 17, pp. 263–287, 2016.