This article has been accepted for publication in IEEE Wireless Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LWC.2023.3291374

Optimal Phase Shift Solution and Diversity Analysis for Discrete RIS-assisted Systems under Rank Deficient Channels

Jia Ye, *Member, IEEE*, Abla Kammoun, *Member, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*

***Abstract*—This paper investigates the design of finite discrete control for each reflecting element of a reconfigurable intelligent surface (RIS) that assists a point-to-point multiple-input single-output communication. Under the assumption that the propagation channel is rank deficient, we propose an optimal algorithm that retrieves the same solution as the exhaustive search method, yet with a polynomial complexity with respect to the number of reflecting elements. Interestingly, we establish that under some specified conditions, the proposed design achieves a signal-to-noise ratio that increases quadratically with the number of reflecting elements, just as the continuously controllable phase shift control. Moreover, we remark that the performance gap between the discrete RIS and the continuous one becomes negligible when the phase shifts of the cascaded channel paths become close to each other. Simulation results confirm the superiority of the proposed design, verify the derived conclusion, and provide us with useful implementation insights.**

***Index Terms*—Finite discrete phase resolution, reconfigurable intelligent surface, rank deficient, signal-to-noise-ratio maximization.**

I. Introduction

The use of reconfigurable intelligent surfaces (RIS) has emerged as a key enabling technology for the upcoming beyond fifth generation (B5G) systems. Comprising an array of passive reflecting elements, RISs have the ability to change the wireless communication environment by adaptively reflecting the incident signals [1]. An important research effort has been carried out to develop efficient algorithms for the control of the phases of each reflecting element in order to satisfy some desired goals such as improving energy efficiency [2], communication coverage [3], and sum rate [4]. Initially, most of the works investigated the use of continuous phase shift designs [2], [4], for which optimal solutions may be derived in some specific cases including single-input single-output systems [5]. However, it is well known that continuous phase shift control cannot be implemented in practice [6], which motivated a subsequent series of research works to investigate the control of the phase shifts under the assumption that they belong to a finite discrete set. Such problems are highly relevant in practice, especially the one-bit control case, since several implementations of reconfigurable intelligent surfaces use only a few or even two states for the phase shifts [7].

The representative literature regarding this research track is represented by the works in [8]–[11]. Several algorithms have been proposed in recent works to maximize the signal-to-noise-ratio (SNR) [6], [8], enhance the sum rate [12], and minimize the transmit power [10], [11] subject to a given SNR constraint. However, some works, such as [8], [10], [12] sub-optimally design the reflecting vector by relaxing the discrete phase shift constraints to continuous ones and then performing the quantization on the obtained continuous solution. The proposed suboptimal approach and additional quantization error result in a performance gap compared to the optimal solution. Although some works, including [6], [11], maintain the discrete phase shift constraint when solving the optimization problem, they still use suboptimal algorithms, as the constraint makes the problem NP-hard and practically solvable only by suboptimal algorithms [8], [11]. Moreover, these algorithms typically require full knowledge of all channel vectors, which poses additional challenges in channel estimation.

Therefore, we are motivated to optimize the RIS-assisted system under the discrete feasible set directly to avoid the quantization error and channel estimation challenges. In contrast to prior studies, we propose a technique that optimally selects the phase shifts with polynomial complexity with respect to the number of reflecting elements. The main premise of the technique is the rank deficiency of the channel, which is the typical property for practical propagation environments [13]. Though traditionally perceived as a major performance limiting factor, the rank deficient channel was also considered in [14], [15] but for reducing the training overhead and complexity of the channel estimation process. Overall, this is the preliminary work that efficiently exploits the channel rank deficiency to enable low-complexity yet optimal algorithms for discrete phase shift design. Moreover, as was particularly shown in [6], dealing with discrete phase shifts still enables a quadratic increase of the SNR with respect to the number of reflecting elements. We also establish that similar to continuous control of the shift phases, our technique allows for the increase of the SNR as the square of the number of reflecting elements. Simulations confirm our findings and demonstrate that our proposed method has better performance than other competing methods under the same phase-shift resolution.

II. Problem Formulation and Algorithm Design

In this paper, we consider a RIS-assisted MISO system where an RIS with N reflecting elements assists communication between a BS equipped with M antennas and a single-antenna user, whose direct link is assumed to be blocked due to obstacles, like tree and buildings. We aim to exploit the channel property to optimally control the phase shifts of RIS under the assumption that they belong to a finite discrete set. To better illustrate our design, we assume for simplicity that there is only one user and is equipped with a single antenna. This work also serves

Jia Ye was with the King Abdullah University of Science and Technology (KAUST) and she is now with the School of Electrical Engineering, Chongqing University, Chongqing, 400044, China (e-mail: yejiaft@163.com).

Abla Kammoun, and Mohamed-Slim Alouini are with KAUST, Thuwal, Makkah Province, Kingdom of Saudi Arabia, 23955-6900 (email: abla.kammoun@kaust.edu.sa; slim.alouini@kaust.edu.sa).

as a fundamental building block for future extensions to more sophisticated scenarios involving multiple receivers and multiple-antenna devices. Therefore, the received signal at the user can be expressed as: $y = \sqrt{P}\mathbf{h}^H\mathbf{\Phi}\mathbf{G}\mathbf{w}x + n$, where P is the transmit power at the BS, $\mathbf{h} \in \mathbb{C}^{N\times 1}$ and $\mathbf{G} \in \mathbb{C}^{N\times M}$ denote the channel between the RIS and the user and that between the BS and the RIS, respectively. $\mathbf{\Phi} = \text{diag}\{\boldsymbol{\phi}\} \in \mathbb{C}^{N\times N}$ denotes the diagonal matrix accounting for the effective phase shifts applied by the RIS reflecting elements with $\boldsymbol{\phi}= [\phi_1, \phi_2, \cdots, \phi_N]^T$, where $\phi_n \in \mathcal{F} \triangleq \left\{\exp\left(\frac{\text{j}2\pi m}{2^b}\right)\right\}_{m=0}^{2^b-1}, n \in \{1, \cdots, N\}$ with b being the phase resolution in number of bits. Besides, $\mathbf{w}$ represents the transmit beamforming vector adopted at the BS, x denotes the downlink transmit symbols for the user satisfying the unit transmit power constraint, and $n \sim \mathcal{CN}(0, N_0)$ is the additive white Gaussian noise (AWGN). The instantaneous SNR at the user can be given by:

$$\gamma = \rho\left|\mathbf{h}^H\mathbf{\Phi}\mathbf{G}\mathbf{w}\right|^2, \tag{1}$$

where $\rho = \frac{P}{N_0}$. It is well known that the maximum instantaneous SNR at the user is achieved by using a maximum ratio transmission (MRT), or equivalently when $\mathbf{w} = \frac{\mathbf{G}^H\mathbf{\Phi}^H\mathbf{h}}{\|\mathbf{G}^H\mathbf{\Phi}^H\mathbf{h}\|}$. Plugging the expression of the optimal $\mathbf{w}$ into (1) yields $\gamma = \rho\boldsymbol{\phi}^T\text{diag}(\mathbf{h}^*)\mathbf{G}\mathbf{G}^H\text{diag}(\mathbf{h})\boldsymbol{\phi}^*$. In this paper, we are interested in maximizing the instantaneous SNR at the user with respect to $\boldsymbol{\phi}$ under the assumption of finite discrete phase shifts for each reflecting element, which amounts to solving the following problem:

$$\max_{\boldsymbol{\phi}\in\mathcal{F}^N} \boldsymbol{\phi}^T\,\text{diag}(\mathbf{h}^*)\,\mathbf{G}\mathbf{G}^H\,\text{diag}(\mathbf{h})\,\boldsymbol{\phi}^*. \tag{2}$$

Mathematically, if $\mathbf{G}$ is full rank, such a problem aiming to maximize a quadratic form over a finite alphabet is known to be NP-hard. In general, the optimal solution can only be computed through exhaustive search, which involves a prohibitively high computational cost, especially for large reflecting surfaces. To the best of our knowledge, as far as the design of RIS is concerned, only sub-optimal strategies have been proposed to solve this problem [12]. In practice, the rank of $\mathbf{G}$ depends only on the number of paths in the propagation environment between the base station and the RIS, and not on the number of the reflecting elements N. Since a limited number of paths is a very common situation in practice, especially in high-frequency scenarios, the channel matrix $\mathbf{G}$ is in practice rank-deficient. More specifically, assuming the BS and the RIS are each equipped with a uniform linear array (ULA) and considering the physical propagation structure of the wireless channel, a possible model for matrix $\mathbf{G}$ is the clustered channel or the extended Saleh-Valenzudela model [16], [17], given by:

$$\mathbf{G} = \sqrt{\frac{NM}{N_G}}\sum_{p=1}^{N_G}\alpha_p\mathbf{a}_N\left(\frac{2d}{\lambda}\sin\left(\varphi_p^{\text{AoA}}\right)\right)\mathbf{a}_M^H\left(\frac{2d}{\lambda}\sin\left(\varphi_p^{\text{AoD}}\right)\right), \tag{3}$$

where N_G is the number of paths between the BS and the RIS known also as the channel's degrees of freedom, α_p denotes the complex gain of the p-th path between the BS and the RIS, φ_p^{AoA} and φ_p^{AoD} represent the p-th angle of arrival (AoA) from the BS and angle of departure (AoD) to the RIS, whereas d denotes the antenna spacing, λ is the carrier wavelength, and $\mathbf{a}_X(\varphi)$, $X \in \{N, M\}$ is the array steering vector, given by: $\mathbf{a}_X(\varphi) = \frac{1}{\sqrt{X}}\left[1, e^{\jmath\pi\varphi}, \cdots, e^{\jmath\pi\varphi(X-1)}\right]^H$. A poorly scattered channel is known to be associated with a low capacity, which precludes achieving high data rates. Interestingly, we prove in this paper that the structure of these channels, while being a limiting performance factor, can be exploited to enable the implementation of optimal designs with affordable complexity. Considering the problem in (2), we note that when N_G is small, the matrix $\mathbf{H} := \text{diag}(\mathbf{h}^*)\,\mathbf{G}$ becomes rank-deficient. Thus, problem (2) writing as:

$$\max_{\boldsymbol{\phi}\in\mathcal{F}^N}\boldsymbol{\phi}^T\mathbf{H}\mathbf{H}^H\boldsymbol{\phi}^* = \max_{\boldsymbol{\phi}\in\mathcal{F}^N}\left\|\boldsymbol{\phi}^T\mathbf{H}\right\| = \max_{\boldsymbol{\phi}\in\mathcal{F}^N}\left\|\boldsymbol{\phi}^H\mathbf{H}^*\right\|. \tag{4}$$

The second equation works because the received SNR is a real number, and the norm-2 operation is a function from a real or complex vector space to the non-negative real numbers. In other words, taking the conjugate on the received SNR does not change the objective function at all. The optimization problem thus turns out to be that of optimizing a quadratic form involving a rank-deficient matrix. It should be noted that solving the optimization problem only requires the availability of channel state information (CSI) of the cascaded channel, which is assumed to be perfectly known by adopting the existing channel estimation techniques [13], [15].

III. Algorithm Developments and Complexity

Such a problem has been studied in the reference [18], in which the authors leverage rank-deficiency of the matrix to obtain a fully parallelizable and rank-scalable algorithm. The main idea behind that algorithm consists in introducing auxiliary hyperspherical coordinates and converting the original problem to a double maximization of a linear form over a multidimensional continuous set, that is,

$$\max_{\boldsymbol{\phi}\in\mathcal{F}^N}\ \max_{\boldsymbol{\xi}\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right]^{2N_G-2}\times\left(-\frac{\pi}{2^b},\frac{\pi}{2^b}\right]}\mathcal{R}\{\boldsymbol{\phi}^H\mathbf{H}^*\mathbf{c}(\boldsymbol{\xi})\}, \tag{5}$$

where $\boldsymbol{\xi} \in \left(-\frac{\pi}{2},\frac{\pi}{2}\right]^{2N_G-2}\times\left(-\frac{\pi}{2^b},\frac{\pi}{2^b}\right]$ is the introduced $2N_G-1$ auxiliary hyperspherical coordinates, which gives the $2N_G \times 1$ hyperspherical real vector with unit radial coordinate as $\tilde{\mathbf{c}}(\boldsymbol{\xi}) = \Big[\sin\xi_1, \sin\xi_2\cos\xi_1, \cdots, \sin\xi_{2N_G-1}\prod_{i=1}^{2N_G-2}\cos\xi_i,$ $\cos\xi_{2N_G-1}\prod_{i=1}^{2N_G-2}\cos\xi_i\Big]^T$, and $N_G \times 1$ hyperspherical complex vector $\mathbf{c}(\boldsymbol{\xi}) = \tilde{\mathbf{c}}_{2:2:2NG}(\boldsymbol{\xi}) + \jmath\tilde{\mathbf{c}}_{1:2:2N_G-1}(\boldsymbol{\xi})$. The equality of (5) and (4) can be proved based on the fact that $\mathcal{R}\{\boldsymbol{\phi}^H\mathbf{H}^*\mathbf{c}(\boldsymbol{\xi})e^{\jmath\hat{\theta}}\} \le \left|\boldsymbol{\phi}^H\mathbf{H}^*\mathbf{c}(\boldsymbol{\xi})\right| \le \left\|\boldsymbol{\phi}^H\mathbf{H}^*\right\|$. By interchanging the maximization in (5), the original optimization problem thus can be equivalently transferred to

$$\max_{\boldsymbol{\xi}\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right]^{2N_G-2}\times\left(-\frac{\pi}{2^b},\frac{\pi}{2^b}\right]}\sum_{n=1}^{N}\max_{\phi_n\in\mathcal{F}^N}\mathcal{R}\{\phi_n^*\mathbf{H}_{n,1:M}^*\mathbf{c}(\boldsymbol{\xi})\}, \tag{6}$$

where $\mathbf{H}_{n,1:M}$ denotes the n-th row of $\mathbf{H}$. Therefore, the unit-vector maximization problem is decomposed into symbol-by-symbol maximizations for a given auxiliary hyperspherical coordinates $\boldsymbol{\xi}$. We can find the optimal reflecting vector if we scan the entire multidimensional continuous set. In this case, the reflecting parameter ϕ_n is only determined by the corresponding row of the cascaded channel $\mathbf{H}^*$ when the decision boundary is not crossed. Due to the structure of $\mathcal{F}$, there are

2^{b-1} decision boundaries, that is, to make the imaginary part of $e^{-\jmath\pi\frac{2k+1}{2^b}}\mathbf{H}^*_{n,1:M}\mathbf{c}\left(\boldsymbol{\xi}\right), k=0,1,\cdots,\frac{2k+1}{2^b}$ to be zeros, which means $\tilde{\mathbf{H}}_{l,1:2N_G}\tilde{\mathbf{c}}\left(\boldsymbol{\xi}\right)=0, l=1,2,\cdots,2^{b-1}N$, where $\tilde{\mathbf{H}}_{l,1:2N_G}$ is the l-th row of $\tilde{\mathbf{H}}$ with size $2^{b-1}N\times 2N_G$, which is defined as $\tilde{\mathbf{H}}_{:,1:2:2N_G-1}=\Re(\hat{\mathbf{H}})$ and $\tilde{\mathbf{H}}_{:,2:2:2N_G}=\Im(\hat{\mathbf{H}})$ with $\hat{\mathbf{H}}=\mathbf{H}^*\otimes\left[\begin{array}{c} e^{j\frac{\pi}{2^b}}, e^{j\frac{3\pi}{2^b}},\cdots,e^{j\frac{(2^b-1)\pi}{2^b}}\end{array}\right]^T$ and $\otimes$ denotes the Kronecker product. The $2N_G-1$ hypercube $\left(-\frac{\pi}{2},\frac{\pi}{2}\right]^{2N_G-2}\times\left(-\frac{\pi}{2^b},\frac{\pi}{2^b}\right]$ can be divided into several noninterleaving cells by the selected $2N_G-1$ hypersurfaces from set $\{\tilde{\mathbf{H}}_{1,1:2N_G},\cdots,\tilde{\mathbf{H}}_{2^{b-1}N,1:2N_G}\}$. Following the **proposition 1** in [18], the combinations of the intersections of the selected $2N_G-1$ hypersurfaces where at most two surfaces originate from the same row have a unique intersection point that leads the cell division. Each cell gives a distinct candidate reflecting vector, where one of them is the optimal reflecting vector. As claimed in [18], the number of interested cells is $\mathcal{O}\left(\left(2^{b-1}N\right)^{2N_G-1}\right)$. The size of the feasible set is reduced from exponential to polynomial. As we need to visit each cell to calculate the candidate reflecting vectors, it brings $\mathcal{O}\left(2^{b-1}N\right)$ computational complexity. Here, we omitted many derivations and proof processes due to the page limit, whose details can be found in [18]. Therefore, the SNR maximization problem under rank deficient channels can be optimally solved in a parallelizable and rank-scalable manner of complexity $\mathcal{O}\left(\left(2^{b-1}N\right)^{2N_G}\right)$, which represents a considerable complexity saving compared to the naive exhaustive search approach. It is worth noting that in the special case when $b=1$, and $N_G=1$, the complexity can be further reduced to $\mathcal{O}(N\log(N))$ instead of $\mathcal{O}(N^2)$ [19].

IV. Discussion

An important question that arises is whether the optimal solution of (2) under discrete phase shifts allows for a quadratic increase of the SNR with respect to the number of reflecting elements, just as in the case of optimized continuous phase shifts. In what follows, we establish that the proposed design allows for the sought-for quadratic increase of the SNR. Additionally, we carry out a performance comparison between optimal designs of discrete and continuous phase shifts when the rank of matrix $\mathbf{G}$ is one.

Quadratic increase of the SNR with the number of RIS elements. To ease the notations, for $p=1,\cdots,N_G$, and $T\in\{N,M\}$ we define $\mathbf{a}_{X,p}=\mathbf{a}_X\left(\frac{2d}{\lambda}\sin\left(\varphi_p^{\rm AoA}\right)\right)$. Furthermore, we consider the following assumptions:

- We adopt the channel model in (3) and assume an asymptotically favorable propagation environment as described in [20]. This implies that for M sufficiently large, we may argue that for $p\neq q$, $\left|\mathbf{a}^H_{M,p}\mathbf{a}_{M,q}\right|\leq\epsilon$, where ϵ is a small scalar, assumed to be possibly taken smaller than $\frac{1}{2}\min_{1\leq p\leq N_G}\frac{|\alpha_p|^2}{N_G-1}$.
- We consider the case in which the maximum of the absolute value of α_p is bounded below and above by constants $C_{\min}$ and $C_{\max}$, respectively, as expressed by the inequality $C_{\min}\leq\max_{1\leq p\leq N_G}|\alpha_p|\leq C_{\max}$. This assumption is reasonable because α_p, the complex gain of each channel path, is typically modeled as a complex Gaussian random variable with zero mean and a certain variance [16], and in some cases can be set to a fixed value for a given propagation environment [21]. Consequently, we can assert that $\max_{1\leq p\leq N_G}|\alpha_p|$ is of $\mathcal{O}(1)$, as it is solely determined by the propagation environment and not affected by other system parameters.
- Moreover, denoting by h_n the n-th element of vector $\mathbf{h}$, we shall assume that $\sum_{n=1}^N|h_n|$ scales with N, that is: $\sum_{n=1}^N|h_n|\geq CN$, for some constant C.

Under these notations, the SNR γ can be written as:

$$\gamma=\frac{\rho NM}{N_G}\boldsymbol{\phi}^T\operatorname{diag}\left(\mathbf{h}^*\right)\sum_{p=1}^{N_G}\alpha_p^2\mathbf{a}_{N,p}\mathbf{a}_{N,p}^H\operatorname{diag}\left(\mathbf{h}\right)\boldsymbol{\phi}^*$$
$$+\frac{\rho NM}{N_G}\boldsymbol{\phi}^T\operatorname{diag}\left(\mathbf{h}^*\right)\sum_{p=1}^{N_G}\sum_{q\neq p}^{N_G}\mathbf{a}_{N,p}\mathbf{a}_{M,p}^H\mathbf{a}_{M,q}\mathbf{a}_{N,q}^H\operatorname{diag}\left(\mathbf{h}\right)\boldsymbol{\phi}^*. \tag{7}$$

For the sake of simplicity, we establish a quadratic increase of the SNR with the number of reflecting elements for one-bit RIS phase resolution. For higher resolutions, the quadratic increase follows directly since in this case the SNR is larger than that for one-bit resolution. The feasible set $\mathcal{F}$ thus reduced to $\{\pm1\}$, which indicates that the conjugate operation of $\boldsymbol{\phi}$ can be ignored, that is, $\boldsymbol{\phi}^*=\boldsymbol{\phi}$.

Therefore, based on the first assumption and using the relation $\mathbf{x}^T\mathbf{c}\mathbf{d}^T\mathbf{x}+\mathbf{x}^T\mathbf{d}\mathbf{c}^T\mathbf{x}\geq-\mathbf{x}^T\mathbf{c}\mathbf{c}^T\mathbf{x}-\mathbf{x}^T\mathbf{d}\mathbf{d}^T\mathbf{x}$ which holds for any $N\times1$ vectors $\mathbf{c}$, $\mathbf{d}$ and $\mathbf{x}$, we may lower-bound the SNR achieved by optimal $\boldsymbol{\phi}\in\{\pm1\}$ as:

$$\begin{aligned}\max_{\boldsymbol{\phi}\in\{\pm1\}^N}\gamma\geq&\max_{\boldsymbol{\phi}\in\{\pm1\}}\frac{\rho NM}{N_G}\boldsymbol{\phi}^T\operatorname{diag}\left(\mathbf{h}^*\right)\\&\times\sum_{p=1}^{N_G}(\alpha_p^2-(N_G-1)\,\epsilon)\mathbf{a}_{N,p}\mathbf{a}_{N,p}^H\operatorname{diag}\left(\mathbf{h}\right)\boldsymbol{\phi}\\\geq&\frac{\rho NMC_{\min}^2}{2N_G}\max_{\boldsymbol{\phi}\in\{\pm1\}}\left|\mathbf{a}_{N,p_0}^H\operatorname{diag}\left(\mathbf{h}\right)\boldsymbol{\phi}\right|^2,\end{aligned} \tag{8}$$

where p_0 is such that $\alpha_{p_0}=\max_{1\leq p\leq N_G}|\alpha_p|$. Let $\mathbf{g}=\operatorname{diag}\left(\mathbf{h}^*\right)\mathbf{a}_{N,p_0}$, and define $\mathcal{L}=\max_{\boldsymbol{\phi}\in\{\pm1\}^N}\left|\boldsymbol{\phi}^T\mathbf{g}\right|^2$. This problem involves the maximization of a quadratic form of a rank-two matrix over binary constraints. Following a similar approach to that in [18], we introduce an auxiliary variable $\theta\in[-\pi,\pi)$ and convert the optimization problem to:

$$\begin{aligned}\mathcal{L}&=\max_{\boldsymbol{\phi}\in\{\pm1\}^N}\max_{\theta\in[-\pi,\pi)}\left|\operatorname{Re}\left\{\boldsymbol{\phi}^T\mathbf{g}e^{-\jmath\theta}\right\}\right|^2\\&=\max_{\theta\in[-\pi,\pi)}\left|\sum_{n=1}^N|g_n|\max_{\boldsymbol{\phi}\in\{\pm1\}^N}\left\{\phi_n\cos\left(\theta_n-\theta\right)\right\}\right|^2\\&=\max_{\theta\in[-\pi,\pi)}\left|\sum_{n=1}^N|g_n|\left|\cos\left(\theta_n-\theta\right)\right|\right|^2,\end{aligned} \tag{9}$$

where for $n=1,\cdots,N$, g_n and θ_n stands for the amplitude and argument of the n-th element of $\mathbf{g}$, respectively. Using the fact that $|x|\geq x^2$ for all $x\in[-1,1]$, $\mathcal{L}$ can be lower-bounded as:

$$\begin{aligned}\mathcal{L}&\geq\max_{\theta\in[-\pi,\pi)}\left|\sum_{n=1}^N|g_n||\cos(\theta_n-\theta)|^2\right|^2\\&\geq\max_{\theta\in[-\pi,\pi)}\left|\sum_{n=1}^N|g_n|\frac{1+\cos\left(2(\theta_n-\theta)\right)}{2}\right|^2\end{aligned}$$

$$= \max_{\theta\in[-\pi,\pi)} \left|\sum_{n=1}^{N}\frac{1}{2}|g_n| + \sum_{n=1}^{N}\cos(2\theta_n)\frac{|g_n|}{2}\cos(2\theta) + \sum_{n=1}^{N}\sin(2\theta_n)\frac{|g_n|}{2}\sin(2\theta)\right|^2$$

$$\overset{(a)}{\geq} \left|\sum_{n=1}^{N}\frac{1}{2}|g_n|\right|^2 \overset{(b)}{=} \frac{1}{N}\left|\sum_{n=1}^{N}\frac{1}{2}|h_n|\right|^2. \quad (10)$$

The inequality (a) follows from the fact that we can always find θ such that $\sum_{n=1}^{N}\frac{|g_n|}{2}\cos(2(\theta_n-\theta))$ is positive. Indeed, for that it suffices to choose θ such that $\cos(2\theta)$ and $\sin(2\theta)$ have the same sign as $\sum_{n=1}^{N}\cos(2\theta_n)\frac{|g_n|}{2}$ and $\sum_{n=1}^{N}\sin(2\theta_n)\frac{|g_n|}{2}$, respectively. Furthermore, the equality in (b) is because the absolute value of the elements of $\mathbf{a}_{N,p_0}$ are equal to $\frac{1}{\sqrt{N}}$. Plugging (10) into (8) and using the third assumption, we thus conclude that the SNR increases quadratically with N.

Case of rank-one channel matrices. Consider the channel model in (3) and assume that $N_G = 1$. In this case, $\mathbf{G}$ is a rank-one channel. Assuming the elements of $\boldsymbol{\phi}$ can take any complex value with amplitude 1, the SNR achieved by optimal continuously controllable $\boldsymbol{\phi}$ is given by:

$$\gamma_{\text{op}}^{\text{cont}} := \max_{|\phi_n|=1,n=1,\cdots,N} \rho N M\alpha_1^2\left|\boldsymbol{\phi}^T\operatorname{diag}(\mathbf{h}^*)\,\mathbf{a}_{N,1}\right|^2 = \rho M\alpha_1^2\left|\sum_{n=1}^{N}|h_n|\right|^2, \quad (11)$$

where the last equality follows by noticing that the optimal phase shifts for the continuous case are given by $\phi_n = \exp\left\{-j\left(\arg(h_n)+\frac{2d}{\lambda}\sin\left(\phi_1^{\text{AoA}}\right)(n-1)\right)\right\}$. On the other hand, by reference to (9), the optimal solution for the 1-bit phase shift case writes as:

$$\gamma_{\text{op}}^{1-\text{bit}} := \rho M\alpha_1^2\left|\sum_{n=1}^{N}|h_n|\,|\cos(\theta_n-\theta)|\right|^2. \quad (12)$$

Remark 1: Comparing (11) with (12), we can easily understand that the one-bit control induces in general a loss since $|\cos(\theta_n-\theta)|\leq 1$, unless all θ_n are equal to the same value, in which case both designs lead to the same performance. We may thus expect the difference in performance between the continuous and 1-bit design to become negligible when all AoAs are close to each other.

Remark 2: From the above discussion, it follows that the rank-deficient quadratic form based algorithm retrieves the optimal solution with a much lower complexity than the exhaustive search and at the same time allows for a quadratic increase of the SNR in the same way as the optimal continuous phase shift designs.

V. Numerical Results

In this section, we present numerical results to illustrate and discuss the performance of the discrete RIS-assisted system under the proposed algorithm. We adopt the channel model described in Section II to generate the channel with antenna spacing $d = \lambda/2$. In general, the AoAs and AoDs follow the uniform distribution within $[0, 2\pi]$, while the path gain follows the complex Gaussian distribution, $\mathcal{CN}(0,1)$ [17]. The proposed rank deficient based algorithm is named 'RD' while the exhaustive search method is denoted as 'ES' in the plotted figures. As a benchmark, we consider the algorithm given in [6] which aims to maximize the received SNR of the RIS-assisted system under finite bit resolution for the phase-shifts and denote it by 'Ref' in all data plots.

In the first experiment, we investigate the achievable SNR [1] of the finite discrete control RIS-assisted system as a function of the transmit power for the proposed scheme and that of [6]. As seen from Fig. 1, the performance of the proposed algorithm coincides with the exhaustive search method. Using the same bit resolution $b = 1$, the algorithm proposed in [6] achieves worse performance than the proposed algorithm. Such a loss becomes more important as the transmit power increases. Complexity-wise, the proposed algorithm presents a complexity of order $O(\left(2^{b-1}N\right)^{2N_G})$ which, as mentioned in Section II, can be reduced to $O(N\log(N))$ when $N_G = 1$ and $b = 1$, while the algorithm proposed in [6] presents a complexity $O(N)$. However, the proposed algorithm is parallelizable and it only requires the information of the cascaded channel $\mathbf{H}$, while the existing algorithm cannot be utilized without knowing the full CSI of $\mathbf{h}$ and $\mathbf{G}$. In practice, one can apply our algorithm for the maximization of the quadratic form associated with the one-rank deficient matrix produced by the strongest path, so that the complexity becomes $O(N\log(N))$, and as such comparable to [6]. However, this will be at the cost of losing optimality. Consideration of the impact of such a procedure is beyond the scope of this paper but will be pursued in the future. Furthermore, the SNR gap between the proposed algorithm and the reference one gets smaller when the resolution increases, which indicates the rank deficient based algorithm will be more suitable for RIS-assisted systems under low bit resolution. In addition, as expected, the performance improves with the increase of the resolution, where the performance of the discrete RIS with $b = 3$ almost coincides with that of the optimal continuous RIS.

In a second experiment, we verify that the SNR of the proposed 1-bit phase shift control achieves an SNR that increases quadratically with the number of elements. As shown in Fig. 2, the SNRs achieved by the proposed algorithm and the one given in [6] grow linearly with N^2. Comparison between both schemes shows that the proposed scheme has a gain that becomes more important as the number of reflecting elements increases. Moreover, to verify the insights we concluded in Remark 1, we generate the AOAs of the channel $\mathbf{G}$ and $\mathbf{h}$ randomly within the range $[\pi-\delta, \pi+\delta]$ and plot the SNRs achieved by continuous RIS and 1-bit control RIS in Fig. 3 for different values of δ. It can be observed that the SNR gap between the two kinds of RIS designs decreases as the AoAs get closer to each other, and totally disappear once δ equals 0, that is, all AoAs are equal to the same value. The observed phenomenon is in agreement with Remark 1, which indicates that the 1-bit control RIS works well as the continuous one when all the angles of propagation paths become closer to each other.

VI. Conclusions

This paper has addressed the problem of selecting the optimal phase shifts for a point-to-point MISO RIS-assisted system under the constraint of finite phase shift resolution. The main idea behind the proposed algorithm lies in exploiting the rank

[1]For the sake of clarity, the SNRs in all the following figures are expressed in their raw form (not in decibels), representing the numerical ratio of the signal power to the noise power, without any specific units attached to them.

This article has been accepted for publication in IEEE Wireless Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LWC.2023.3291374

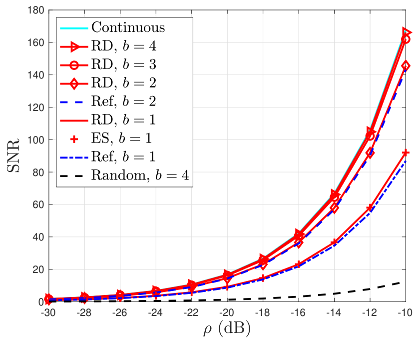

Fig. 1. SNR versus ρ with $M = 8$, $N = 16$, and $N_G = 3$.

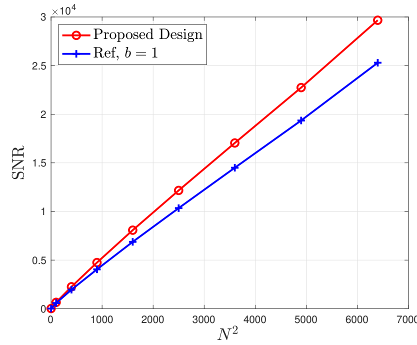

Fig. 2. SNR versus N^2 with $M = 16$, $\rho = 0$ dB, and $N_G = 2$.

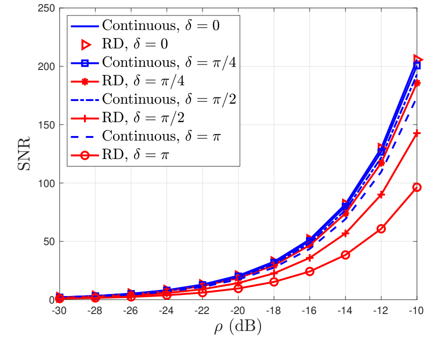

Fig. 3. SNR versus ρ with $M = 8$, $N = 16$, $\rho = 0$ dB.

deficiency of the channel matrix between the transmitter and the RIS in order to obtain the optimal phase shifts with a polynomial complexity in the number of elements of the RIS instead of the exponential complexity required by the exhaustive search method. We establish that the proposed design allows for a quadratic increase of the SNR with the number of RIS elements, like the infinite bit resolution case. Comparison with existing methods under the same bit resolution shows that the proposed technique is parallelizable and has a higher gain, but presents higher complexity. Just as the RISs have been proposed to modify the propagation channel, the proposed design can be viewed as an algorithmic solution that exploits poorly scattered channels in order to obtain optimal designs with reasonable complexity. However, the proposed algorithm might not be suitable for large reflecting surfaces with high phase resolutions as the complexity remains prohibitively high. It thus motivates us to solve the same problem in [22] through the Kronecker decomposition technique considering the rank-deficient mmWave channel with an affordable computational complexity but at the cost of performance loss. The skipped channel estimation problem in this work is also addressed in [22], where the cascaded channel response can be estimated in an efficient and accurate manner by invoking the sparse channel structure and the atomic norm minimization framework.

References

[1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2019.

[2] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[3] S. Zeng, H. Zhang, B. Di, Z. Han, and L. Song, "Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization," *IEEE Communications Letters*, vol. 25, no. 1, pp. 269–273, 2020.

[4] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3064–3076, 2020.

[5] J. Ye, A. Kammoun, and M.-S. Alouini, "Spatially-distributed RISs vs relay-assisted systems: A fair comparison," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 799–817, 2021.

[6] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "On the optimality of reconfigurable intelligent surfaces (RISs): Passive beamforming, modulation, and resource allocation," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4347–4363, 2021.

[7] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *IEEE Conference on Computer Communications, INFOCOM*, Honolulu, HI, USA, 2018, pp. 270–278.

[8] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2604–2620, 2020.

[9] Y. Zhang, C. Zhong, Z. Zhang, and W. Lu, "Sum rate optimization for two way communications with intelligent reflecting surface," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1090–1094, 2020.

[10] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 7830–7833.

[11] ——, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, 2019.

[12] J. Zuo, Y. Liu, E. Basar, and O. A. Dobre, "Intelligent reflecting surface enhanced millimeter-wave NOMA systems," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2632–2636, 2020.

[13] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user mmWave MIMO systems," *IEEE Transactions on Wireless Communications*, 2023.

[14] Y. Jin, J. Zhang, X. Zhang, H. Xiao, B. Ai, and D. W. K. Ng, "Channel estimation for semi-passive reconfigurable intelligent surfaces with enhanced deep residual networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 11 083–11 088, 2021.

[15] J. Mirza and B. Ali, "Channel estimation method and phase shift design for reconfigurable intelligent surface assisted MIMO networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 441–451, 2021.

[16] L. Pang, Y. Li, Y. Zhang, M. Shang, Y. Chen, and A. Wang, "MGGAN-based hybrid beamforming design for massive MIMO systems against rank-deficient channels," *IEEE Communications Letters*, vol. 26, no. 11, pp. 2804–2808, 2022.

[17] X. Qi, G. Xie, Y. Liu, and M. Zhu, "A family of practical hybrid precoding algorithms for rank-deficient millimeter-wave MIMO systems," *IEEE Access*, vol. 7, pp. 185 740–185 749, 2019.

[18] A. T. Kyrillidis and G. N. Karystinos, "Rank-deficient quadratic-form maximization over M-phase alphabet: Polynomial-complexity solvability and algorithmic developments," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 3856–3859.

[19] G. N. Karystinos and A. P. Liavas, "Efficient computation of the binary vector that maximizes a rank-deficient quadratic form," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3581–3593, 2010.

[20] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, 2014, pp. 76–80.

[21] E. Basar and I. Yildirim, "SimRIS channel simulator for reconfigurable intelligent surface-empowered communication systems," in *IEEE Latin-American Conference on Communications (LATINCOM)*, 2020, pp. 1–6.

[22] J. Ye, A. Kammoun, and M.-S. Alouini, "Optimal phase shift solution and diversity analysis for discrete RIS-assisted systems under rank deficient channels," *Submitted to IEEE Transactions on Wireless Communications*, 2022.