Large-Scale Beamforming for Massive MIMO via Randomized Sketching

Hayoung Choi, *Member, IEEE*, Tao Jiang, *Student Member, IEEE*, Yuanming Shi, *Senior Member, IEEE*, Xuan Liu, *Member, IEEE*, Yong Zhou, *Member, IEEE*, and Khaled B. Letaief, *Fellow, IEEE*

***Abstract*—Massive MIMO system yields significant improvements in spectral and energy efficiency for future wireless communication systems. The regularized zero-forcing (RZF) beamforming is able to provide good performance with the capability of achieving numerical stability and robustness to the channel uncertainty. However, in massive MIMO systems, the matrix inversion operation in RZF beamforming becomes computationally expensive. To address this computational issue, we shall propose a novel *randomized sketching* based RZF beamforming approach with low computational complexity. This is achieved by solving a linear system via randomized sketching based on the preconditioned Richard iteration, which guarantees high quality approximations to the optimal solution. We theoretically prove that the sequence of approximations obtained iteratively converges to the exact RZF beamforming matrix linearly fast as the number of iterations increases. Also, it turns out that the system sum-rate for such sequence of approximations converges to the exact one at a linear convergence rate. Our simulation results verify our theoretical findings.**

***Index Terms*—Regularized zero-forcing beamforming, massive MIMO, randomized sketching algorithm, sketching method.**

I. Introduction

With the explosive growth in mobile data traffic and number of mobile devices, as well as the stringent and diverse demands of intelligent mobile services, wireless networks are facing formidable challenges to enable high spectral efficiency and support massive connectivity with low-latency. To satisfy these requirements, network densification becomes the key enabling technology. This is achieved by deploying more base stations (BSs) with storage and computational capabilities, yielding an ultra-dense network (UDN) [2]. In particular, massive multiple-input multiple-output (MIMO) technique provides an alternative to achieve UDN by simply increasing the number of antennas at the existing BS [3], [4]. The key success is based on the fact that deploying large-scale antenna arrays allows for an exceptional array gain and unprecedented spatial resolution such that the wireless communication system is robust to inter-user interference [5]. Furthermore, the large arrays regime provides the opportunities for asymptotic system analysis, e.g., the high-dimensional random matrix theory can provide deterministic approximations for achievable data rates [6], [7].

Transmit beamforming at the BSs is a key method to optimize the network utility function (e.g., sum-rate) in terms of signal-to-interference-plus-noise ratios (SINRs). However, the resulting beamforming optimization problem is generally very difficult to be solved due to the nonconvexity and high-dimensionality. With the known optimal SINRs parameters for maximizing the network utility function, a simple structure for the optimal beamforming can be derived based on the Lagrange duality theory [8]. To find the optimal SINRs parameters, we normally need to solve a sequence of convex subproblems [9]. For instance, in the max-min fairness rate optimization problem, the optimal SINRs parameters can be found via the bi-section method [10], wherein a sequence of convex subproblems are solved. Although the general large-scale convex optimization problem can be solved by the operate splitting method, it still needs to solve a sequence of subspace projection and cone projection problems in the transformed high-dimensional space for the standard cone program [11]. There are three heuristic linear transmit beamforming algorithms in the literature, i.e., matched-filtering, zero-forcing precoding, and regularized zero-forcing (RZF) beamforming. [12]. However, the matched-filtering beamforming algorithm cannot mitigate the interference between the users, and thus achieves a poor performance when the SNR is high. The zero-forcing precoding algorithm suffers from noise inflation, which results in poor performance in the low SNR regime. The RZF precoding algorithm does not have these issues. However, the RZF beamforming needs to compute a matrix inversion with complexity proportional to MK^2, where K is the number of users served by M transmit antennas. This is however computational expensive in the massive MIMO

Manuscript received August 25, 2020; revised January 29, 2021; accepted March 30, 2021. Date of publication April 7, 2021; date of current version June 9, 2021. The work of Hayoung Choi was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) under Grant 2020R1C1C1A01009185. The work of Yong Zhou was supported by the National Natural Science Foundation of China (NSFC) under Grant 62001294. This article was presented in part at the IEEE Global Communications Conference, Waikoloa, HI, USA, Dec. 2019 [1]. The review of this article was coordinated by Dr. Sunwoo Kim. *(Corresponding author: Yong Zhou.)*

Hayoung Choi is with the Department of Mathematics, Kyungpook National University, Daegu 41566, Korea (e-mail: hayoung.choi@knu.ac.kr, hchoi@shanghaitech.edu.cn).

Tao Jiang and Yong Zhou are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: jiangtao1@shanghaitech.edu.cn; zhouyong@shanghaitech.edu.cn).

Yuanming Shi is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China, and also with Yoke Intelligence, Shanghai 201210, China (e-mail: shiym@shanghaitech.edu.cn).

Xuan Liu is with the School of Electrical Engineering and Telecom- munications, The University of New South Wales, Sydney, NSW 2052, Australia (e-mail: xuan.liu@unsw.edu.au).

Khaled B. Letaief is with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Kowloon 999077, Hong Kong (e-mail: eekhaled@ust.hk).

Digital Object Identifier 10.1109/TVT.2021.3071543

0018-9545 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

scenario where $M \gg K \gg 1$. To tackle this issue, [13], [14] proposed to replace the matrix inversion in RZF beamformer by a truncated polynomial expansion, but it is not clear that which degree of the polynomial is needed to guarantee the good performance for the system sum-rate.

In recent years, randomized sketching algorithms [15]–[17] have received a great deal of attention in order to solve large-scale matrix computation problems. The main idea behind randomized sketching algorithms is to compress a given large-scale matrix to a much smaller matrix by multiplying it by a *random matrix* with certain properties. Very expensive computation can then be operated by the smaller matrix efficiently. In particular, several novel randomized algorithms are proposed for the ridge regression problem [18]–[21]. Inspired by these progresses, we propose a randomized sketching based beamforming method to overcome the computational issues for designing beamformers in massive MIMO systems. The main contributions of this paper are summarized as follows.

1) We propose a low computational complexity beamforming scheme based on randomized sketching techniques. The randomized sketching RZF beamforming matrix is achieved by solving a linear system by preconditioned Richard iteration [22] with the randomized sketching techniques [20]. The proposed randomized sketching RZF beamforming method has a computational complexity proportional to LK^2 with $L \ll 2M$ as the sketching matrix size.
2) We prove that the beamforming matrix obtained iteratively converges to the RZF beamforming matrix at a linear convergence rate. Furthermore, we prove that the achievable system sum-rate of the MIMO system with the proposed randomized method converges to the achievable sum-rate given by RZF beamforming linearly as the number of iteration increases.
3) Extensive simulations are conducted to verify our theoretical findings in terms of the convergence rate and the performance guarantee.

The rest of this paper is organized as follows. In Sec. II, the system model and problem statement of estimating the beamforming matrix for a massive MIMO communication system are described. In Sec. III, we propose the randomized sketching method to approximate the beamforming matrix with low-complexity and provide convergence analysis and complexity analysis. In Sec. IV, we prove that the system sum-rate of the randomized sketching based beamformer converges to the sum-rate of the RZF beamforming matrix as the number of iterations increases. We provide the exact rate of convergence as well. In Sec. V, we numerically evaluate the performance of the randomized sketching based beamforming method. Finally, conclusions are drawn in Sec. VI.

Notations: Let $\mathbb{R}$ (resp. $\mathbb{C}$) be the set of real (resp. complex) numbers. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}_{*i}$ (resp. $\boldsymbol{A}_{i*}$) denotes i-th column (resp. row) vector of $\boldsymbol{A}$. $\|\boldsymbol{A}\|_2$ (resp. $\|\boldsymbol{A}\|_F$) denotes the operator (resp. Frobenius) norm. For a vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_2$ denotes the Euclidean norm. The superscript T denotes the transpose operator. $\boldsymbol{A}^H = \bar{\boldsymbol{A}}^T$ is a complex conjugate transpose of $\boldsymbol{A}$. The diagonal matrix whose diagonal entries consist of entries of a vector $\boldsymbol{\lambda}$ is denoted by $\mathrm{diag}\{\boldsymbol{\lambda}\}$. Denote the identity matrix of size K as $\boldsymbol{I}_K$. When the size can be trivially determined by the context, we simply write $\boldsymbol{I}$. Denote the zero matrix with size $K \times M$ as $\boldsymbol{0}_{K\times M}$. Let $\Re(\boldsymbol{A})$ and $\Im(\boldsymbol{A})$ denote the real and imaginary parts of a matrix $\boldsymbol{A}$, respectively. For a matrix $\boldsymbol{Q} \in \mathbb{R}^{2K\times 2M}$ with $M \geq K$ of rank $2K$, its (thin) Singular Value Decomposition (SVD) is the form $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ where $\boldsymbol{U} \in \mathbb{R}^{2K\times 2K}$ is the matrix of the left singular vectors, $\boldsymbol{V} \in \mathbb{R}^{2M\times 2K}$ is the matrix of the right singular vectors, and $\boldsymbol{\Sigma} \in \mathbb{R}^{2K\times 2K}$ is a diagonal matrix whose diagonal entries are the singular value of $\boldsymbol{Q}$. We denote the singular values of a matrix as σ_i. We denote the matrix of the top j left singular vectors as $\boldsymbol{U}_j \in \mathbb{R}^{2K\times j}$ and the matrix of the bottom $2K - j$ left singular vectors as $\boldsymbol{U}_{j,\perp} \in \mathbb{R}^{2K\times(2K-j)}$.

II. System Model and Problem Statement

A. System Model

We consider a single-cell massive MIMO system consisting of one BS equipped with M antennas and K single-antenna users, where $M \geq K$. During the downlink transmission, the received signal at the k-th user is given by

$$y_k = \boldsymbol{h}_k^H \left(\sum_{i=1}^{K} \boldsymbol{w}_i s_i \right) + n_k, \quad k = 1, \ldots, K, \tag{1}$$

where $\boldsymbol{w}_i \in \mathbb{C}^M$ is the transmit beamforming vector from the BS for data symbol s_i to user i, $\boldsymbol{h}_k \in \mathbb{C}^M$ is the channel propagation coefficients from the BS to the k-th user, and $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the additive noise (i.e., n_k is a circularly symmetric complex Gaussian random distribution with mean 0 and variance σ^2). Therefore, with universal frequency reuse, the SINR at the k-th user is given as

$$SINR_k(\boldsymbol{W}) := \frac{|\boldsymbol{h}_k^H \boldsymbol{w}_k|^2}{\sum_{j\neq k} |\boldsymbol{h}_k^H \boldsymbol{w}_j|^2 + \sigma^2}, \tag{2}$$

where $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K] \in \mathbb{C}^{M\times K}$ is the aggregative beamforming matrix with the total transmit power limited by $P > 0$, i.e.,

$$\|\boldsymbol{W}\|_F^2 = \sum_{k=1}^{K} \|\boldsymbol{w}_k\|_2^2 \leq P. \tag{3}$$

The achievable system sum-rate $R(\boldsymbol{W})$ is thus given by

$$R(\boldsymbol{W}) := \sum_{k=1}^{K} \log(1 + SINR_k(\boldsymbol{W})). \tag{4}$$

One of the main goal of transmit beamforming is to maximize the achievable system sum-rate. However, it is generally computationally demanding to find the optimal beamforming matrix $\boldsymbol{W}$ [9].

B. Regularized Zero-Forcing Beamforming

Although there are various precoding techniques such as matched filter, zero forcing, regularized zero-forcing, truncated polynomial expansion, and phased zero forcing [23], this article considers the suboptimal beamforming approach, *regularized*

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

zero-forcing (RZF) beamforming [12], which is known to have the capability achieving robustness and numerical stability to the channel uncertainty [8], [12]. RZF precoder has been considered as the state-of-the-art linear precoder for MIMO wireless communication systems. Since we focus on the computational issues of RZF, we consider the following RZF with equal power allocation for simplification [6]

$$\begin{aligned}\boldsymbol{W}^* &= \beta\left(\boldsymbol{I}_M + \frac{\gamma}{\sigma^2}\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H \\ &= \beta\boldsymbol{H}^H\left(\boldsymbol{I}_K + \frac{\gamma}{\sigma^2}\boldsymbol{H}\boldsymbol{H}^H\right)^{-1}, \end{aligned} \tag{5}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^H \in \mathbb{C}^{K\times M}$ is the channel matrix, $\gamma > 0$ is an optimal regularizer, and $\beta > 0$ is a normalization parameter to satisfy the power constraint (3). In particular, γ can be derived as $\gamma = P/K$ in the symmetric scenario, where the channels are equally strong [12].

Although the beamforming design requires the channel state information (CSI), we remark that the acquisition of the CSI for MIMO system has been widely studied in the literature, e.g., [24], [25], where low complexity algorithms were proposed to solve the MIMO channel estimation problems in different scenarios. In this paper, we only focus on the beamforming design by assuming that the perfect CSI is available, which provides an upper bound on the achievable rate performance.

C. Complexity Issues in Massive MIMO

The main computational complexity for computing (5) lies in computing the matrix inversion directly, which leads $\mathcal{O}(MK^2 + K^3)$ computational complexity. To support ultra-low latency communications in massive MIMO systems, it becomes critical to design large-scale precoding algorithm with low computation complexity [13], [14]. As fast inversions of large-scale matrices in every coherence period needs to be performed, it is desired to find efficient algorithms to reduce the high computational complexity with performance guarantees.

In this paper, we shall develop the randomized sketching based precoding algorithm to compute the large-scale RZF beamforming matrix $\boldsymbol{W}^*$ in (5). This is based on the key observation that the large-scale array regime, i.e., $M \gg K$, offers the opportunity for dimension reduction in (5), thereby reducing the computational complexity while guaranteeing the high performance accuracy. Specifically, we develop the scalable algorithm for computing $\boldsymbol{W}^*$ in (5) based on the principles of *Randomized Numerical Linear Algebra* [17]. In particular, the theoretical guarantees for the achievable system sum-rate (4) using the randomized sketching based beamforming method will be presented in Sec. IV.

III. Randomized Sketching for Large-Scale Beamforming

A. Randomized Sketching Algorithm

Randomized sketching algorithm exploits *randomization* as a computational resource to develop improved algorithms for large-scale matrix computation problems. The key idea of randomized algorithm is to compress a given large-scale matrix to a much smaller matrix by multiplying it by a *random matrix* with certain properties. Very expensive computation can then be performed on the smaller matrix efficiently. For a given matrix $\boldsymbol{A}$ and a random matrix $\boldsymbol{S}$, the technique of replacing $\boldsymbol{A}$ by $\boldsymbol{SA}$ is known as a *sketching technique* and $\boldsymbol{SA}$ is referred to as a *sketch* of $\boldsymbol{A}$. Such $\boldsymbol{S}$ is called a *sketching matrix*.

Sketching technique can be accomplished by *random sampling* or *random projection*. For random sampling method, the sketch consists of a small number of carefully-sampled and rescaled columns/rows of matrix $\boldsymbol{A}$. On the other hand, for random projection method, the sketch consists of a small number of linear combinations of the columns/rows of $\boldsymbol{A}$. We will discuss various construction for the random matrix $\boldsymbol{S}$ in Section III-D.

Sketching technique has been extensively studied for a decade [15]–[17]. Recently, the widespread use of sketching as a tool for matrix computations yields many novel results in many fields, especially in machine learning [18], [19], [26], [27].

B. Randomized Sketching Based RZF Beamforming

The first key observation is that (5) can be expressed as the matrix ridge regression problem as follows [20]:

$$\boldsymbol{W}^* = \underset{\boldsymbol{W}\in\mathbb{C}^{M\times K}}{\operatorname{argmin}} \|\boldsymbol{H}\boldsymbol{W} - \lambda\beta\boldsymbol{I}_K\|_F^2 + \lambda\|\boldsymbol{W}\|_F^2, \tag{6}$$

where $\lambda = \frac{\sigma^2}{\gamma}$. To facilitate algorithm design in real field, we focus on solving the equivalent real counterpart of (6):

$$\boldsymbol{M}^* = \underset{\boldsymbol{M}\in\mathbb{R}^{2M\times K}}{\operatorname{argmin}} \|\boldsymbol{Q}\boldsymbol{M} - \boldsymbol{\Lambda}\|_F^2 + \lambda\|\boldsymbol{M}\|_F^2, \tag{7}$$

where

$$\boldsymbol{M} = \begin{bmatrix}\Re(\boldsymbol{W})\\ \Im(\boldsymbol{W})\end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix}\Re(\boldsymbol{H}) & -\Im(\boldsymbol{H})\\ \Im(\boldsymbol{H}) & \Re(\boldsymbol{H})\end{bmatrix}, \boldsymbol{\Lambda} = \begin{bmatrix}\Re(\lambda\beta\boldsymbol{I}_K)\\ \Im(\lambda\beta\boldsymbol{I}_K)\end{bmatrix}.$$

Note that since $\lambda, \beta > 0$, $\Im(\lambda\beta\boldsymbol{I}_K) = \boldsymbol{0}$. Then the optimal solution of (7) takes the form,

$$\boldsymbol{M}^* = \boldsymbol{Q}^T(\boldsymbol{Q}\boldsymbol{Q}^T + \lambda\boldsymbol{I}_{2K})^{-1}\boldsymbol{\Lambda}. \tag{8}$$

Given the matrix $\boldsymbol{M}^*$, it is trivial to obtain the complex RZF beamforming matrix $\boldsymbol{W}^*$ in (5).

Iterative methods provide the solution to the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ as the limit of a sequence $\boldsymbol{x}^{(j)}$, and usually involve matrix $\boldsymbol{A}$ only through multiplications by given vectors. Generally, any iterative method is based on a suitable splitting of the matrix $\boldsymbol{A}$ with $\boldsymbol{A} = \boldsymbol{E} - \boldsymbol{N}$, where $\boldsymbol{E}$ is nonsingular. Then the sequence $\{\boldsymbol{x}^{(j)}\}$ is generated as follows:

$$\boldsymbol{E}\boldsymbol{x}^{(j+1)} = \boldsymbol{N}\boldsymbol{x}^{(j)} + \boldsymbol{b} \quad \text{for all } j \in \mathbb{N}, \tag{9}$$

where $\boldsymbol{x}^{(0)}$ is a given initial vector. Equivalently, such iteration can be restated as $\boldsymbol{x}^{(j+1)} = \boldsymbol{x}^{(j)} + \boldsymbol{E}^{-1}\boldsymbol{r}^{(j)}$ for all $j \in \mathbb{N}$, where $\boldsymbol{r}^{(j)} := \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(j)}$ is the residual at the step j, where $\boldsymbol{E}$ is called *preconditioner* for $\boldsymbol{A}$. The following iteration is called *preconditioned Richardson iteration* [22]:

$$\boldsymbol{x}^{(j+1)} = \boldsymbol{x}^{(j)} + \alpha_j\boldsymbol{E}^{-1}\boldsymbol{r}^{(j)} \quad \text{for all } j \in \mathbb{N}, \tag{10}$$

where $\alpha_j \neq 0$ is the real acceleration parameter.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Randomized Sketching Based Beamformer.

Input:$\boldsymbol{Q} \in \mathbb{R}^{2K\times 2M}$, $\boldsymbol{\Lambda} \in \mathbb{R}^{2K\times K}$, $\lambda > 0$; number of iterations $t > 0$; sketching matrix $\boldsymbol{S} \in \mathbb{R}^{2M\times L}$;

Initialize: $\boldsymbol{\Lambda}^{(0)} \leftarrow \boldsymbol{\Lambda}$, $\widetilde{\boldsymbol{M}}^{(0)} \leftarrow \boldsymbol{0}_{2M\times K}$, $\boldsymbol{Y} \leftarrow \boldsymbol{0}_{2K\times K}$;

for $j = 1$ **to** t **do**

(i) $\boldsymbol{\Lambda}^{(j)} \leftarrow \boldsymbol{\Lambda}^{(j-1)} - \lambda \boldsymbol{Y}^{(j-1)} - \boldsymbol{Q}\widetilde{\boldsymbol{M}}^{(j-1)}$;

(ii) $\boldsymbol{Y}^{(j)} \leftarrow (\boldsymbol{QSS}^T\boldsymbol{Q}^T + \lambda \boldsymbol{I}_{2K})^{-1}\boldsymbol{\Lambda}^{(j)}$;

(iii) $\widetilde{\boldsymbol{M}}^{(j)} \leftarrow \boldsymbol{Q}^T\boldsymbol{Y}^{(j)}$;

end for

Output: Approximate the solution matrix $\widehat{\boldsymbol{M}}^{(t)} = \sum_{j=1}^{t} \widetilde{\boldsymbol{M}}^{(j)}$.

We present the novel sketching based randomized beamforming in Algorithm 1, which iteratively computes a sequence of matrixes $\widetilde{\boldsymbol{M}}^{(j)} \in \mathbb{R}^{2M\times K}$ for $j = 1, \ldots, t$ and returns the approximation $\widehat{\boldsymbol{M}}^{(t)} = \sum_{j=1}^{t} \widetilde{\boldsymbol{M}}^{(j)}$ to the true solution matrix of (8). In fact, it can be viewed as a preconditioned Richardson iteration. Indeed, for a given $\boldsymbol{Y}^{(j)}$ in Algorithm 1, we denote $\widehat{\boldsymbol{Y}}^{(t)} = \sum_{j=1}^{t} \boldsymbol{Y}^{(j)}$. Note that our solution is $\boldsymbol{M}^{(t)} = \boldsymbol{Q}^T\widehat{\boldsymbol{Y}}^{(t)}$. By (i) and (iii) in Algorithm 1, we have

$$\boldsymbol{\Lambda}^{(j)} = \boldsymbol{\Lambda}^{(j-1)} - (\boldsymbol{QQ}^T + \lambda \boldsymbol{I})\boldsymbol{Y}^{(j-1)}. \tag{11}$$

Applying the recurrence relation (11) successively, it follows that

$$\begin{aligned}
\boldsymbol{\Lambda}^{(j)} &= \boldsymbol{\Lambda}^{(j-2)} - (\boldsymbol{QQ}^T + \lambda \boldsymbol{I})\boldsymbol{Y}^{(j-2)} - (\boldsymbol{QQ}^T + \lambda \boldsymbol{I})\boldsymbol{Y}^{(j-1)} \\
&= \boldsymbol{\Lambda}^{(j-2)} - (\boldsymbol{QQ}^T + \gamma \boldsymbol{I})(\boldsymbol{Y}^{(j-2)} + \boldsymbol{Y}^{(j-1)}) \\
&\vdots \\
&= \boldsymbol{\Lambda}^{(1)} - (\boldsymbol{QQ}^T + \lambda \boldsymbol{I})(\boldsymbol{Y}^{(j-2)} + \cdots + \boldsymbol{Y}^{(1)}) \\
&= \boldsymbol{\Lambda} - (\boldsymbol{QQ}^T + \lambda \boldsymbol{I})\widehat{\boldsymbol{Y}}^{(j-1)}.
\end{aligned}$$

Then it holds that

$$\begin{aligned}
\widehat{\boldsymbol{Y}}^{(t)} &= \widehat{\boldsymbol{Y}}^{(t-1)} + \boldsymbol{Y}^{(t)} \\
&= \widehat{\boldsymbol{Y}}^{(t-1)} + (\boldsymbol{QSS}^T\boldsymbol{Q}^T + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Lambda}^{(t)} \\
&= \widehat{\boldsymbol{Y}}^{(t-1)} + (\boldsymbol{QSS}^T\boldsymbol{Q}^T + \lambda \boldsymbol{I})^{-1}(\boldsymbol{\Lambda} - (\boldsymbol{QQ}^T + \lambda \boldsymbol{I})\widehat{\boldsymbol{Y}}^{(t-1)}).
\end{aligned}$$

Thus, Algorithm 1 can be formulated as a preconditioned Richard iteration to solve the linear system

$$(\boldsymbol{QQ}^T + \lambda \boldsymbol{I}_{2K})\boldsymbol{Y} = \boldsymbol{\Lambda}, \tag{12}$$

with preconditioner $\boldsymbol{E} = (\boldsymbol{QSS}^T\boldsymbol{Q}^T + \lambda \boldsymbol{I}_{2K})$ and $\alpha_j = 1$ for all j in (10).

Algorithm 1 iteratively computes a sequence of matrices $\widetilde{\boldsymbol{M}}^{(j)}$ for $j = 1, \ldots, t$ and returns the approximation $\widehat{\boldsymbol{M}}^{(t)} = \sum_{j=1}^{t} \widetilde{\boldsymbol{M}}^{(j)}$ to the true solution $\boldsymbol{M}^*$ in (8). Equivalently, it computes the approximation $\widehat{\boldsymbol{W}}^{(t)} = \sum_{j=1}^{t} \widetilde{\boldsymbol{W}}^{(j)}$ to the true solution $\boldsymbol{W}^*$ in (5). We call such approximation $\widehat{\boldsymbol{W}}^{(t)}$ *a randomized sketching based beamformer*.

Algorithm 1 uses the sketching matrix for the preconditioner in order to improve the rate of convergence and reduce the computational complexity. Specifically, using the sketching matrix $\boldsymbol{S} \in \mathbb{R}^{2M\times L}$ with $L \ll 2M$, the preconditioner $\boldsymbol{E} = (\boldsymbol{QSS}^T\boldsymbol{Q}^T + \lambda \boldsymbol{I}_{2K})$ can be computed by matrix $\boldsymbol{QS}$ with much smaller size.

C. Convergence Analysis

The convergence analysis depends on the selected sketching matrix, which satisfies the constraint (13). Theorem 1 presents a quality-of-approximation result under the assumption that the sketching matrix satisfies the constraint (13).

Theorem 1: Assume that for some constant $0 < \varepsilon < 1$, the sketching matrix $\boldsymbol{S} \in \mathbb{R}^{2M\times L}$ satisfies the following constraint

$$\|\boldsymbol{V}^T\boldsymbol{SS}^T\boldsymbol{V} - \boldsymbol{I}_{2K}\|_2 \le \frac{\varepsilon}{2}, \tag{13}$$

where $\boldsymbol{V} \in \mathbb{R}^{2M\times 2K}$ is the matrix of right singular vectors of $\boldsymbol{Q}$. Then, after t number of iterations, the approximation $\widehat{\boldsymbol{W}}^{(t)}$ returned by Algorithm 1 satisfies

$$\|\widehat{\boldsymbol{W}}^{(t)} - \boldsymbol{W}^*\|_F \le \varepsilon^t \|\boldsymbol{W}^*\|_F,$$

where $\boldsymbol{W}^*$ is the true value of the RZF beamforming matrix in (5) in the complex version.

Proof: Note that by [28], (8) can be also expressed as

$$\boldsymbol{M}^* = (\boldsymbol{Q}^T\boldsymbol{Q} + \lambda \boldsymbol{I}_{2M})^{-1}\boldsymbol{Q}^T\boldsymbol{\Lambda}. \tag{14}$$

Then each column of $\boldsymbol{M}^*$ can be considered as the solution of the following optimization problem

$$\arg\min_{\boldsymbol{M}_{*i}\in\mathbb{R}^{2M}} \|\boldsymbol{QM}_{*i} - \boldsymbol{\Lambda}_{*i}\|_2^2 + \lambda\|\boldsymbol{M}_{*i}\|_2^2, \tag{15}$$

for each $i = 1, \ldots, K$. Recall that $\boldsymbol{M}_{i*}$ and $\boldsymbol{\Lambda}_{i*}$ is the i-th column of $\boldsymbol{M}$ and $\boldsymbol{\Lambda}$, respectively. By Theorem 1 in [20], it follows that

$$\|\widehat{\boldsymbol{M}}_{*i}^{(t)} - (\boldsymbol{M}^*)_{*i}\|_2 \le \varepsilon^t \|(\boldsymbol{M}^*)_{*i}\|_2,$$

for all $i = 1, \ldots, K$. Then we have

$$\begin{aligned}
\|\widehat{\boldsymbol{M}}^{(t)} - \boldsymbol{M}^*\|_F^2 &= \sum_{i=1}^{K} \|\widehat{\boldsymbol{M}}_{*i}^{(t)} - (\boldsymbol{M}^*)_{*i}\|_2^2 \\
&\le \varepsilon^{2t} \sum_{i=1}^{K} \|(\boldsymbol{M}^*)_{*i}\|_2^2 \\
&\le \varepsilon^{2t} \|\boldsymbol{M}^*\|_F^2.
\end{aligned}$$

Clearly, $\|\widehat{\boldsymbol{W}}^{(t)} - \boldsymbol{W}^*\|_F = \|\widehat{\boldsymbol{M}}^{(t)} - \boldsymbol{M}^*\|_F$ and $\|\boldsymbol{W}^*\|_F = \|\boldsymbol{M}^*\|_F$. ■

To check whether a sketching matrix $\boldsymbol{S}$ satisfies (13), a number of columns L that is proportional to $2K\log(2K)$ is required (see Theorem 3). Thus, the running time of any algorithm that computes the sketch $\boldsymbol{QS}$ is also proportional to $2K\log(2K)$. To reduce the running time, it would be much better to use a

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

parameter which is significantly smaller than $2K$. For simplicity of exposition, we will assume that the rank of $\boldsymbol{Q}$ is $2K$.

In the context of ridge regression, a much more important quantity than the rank of $\boldsymbol{Q}$ is the *(effective) degrees of freedom* of $\boldsymbol{Q}$ as follows [29]:

$$d_\lambda = \sum_{i=1}^{2K} \frac{\sigma_i^2}{\sigma_i^2 + \lambda}, \tag{16}$$

where σ_i are the singular values of $\boldsymbol{Q}$ and $\lambda = \frac{\sigma^2}{\gamma}$. Equivalently, the (effective) degrees of freedom is defined as the trace of the matrix $\boldsymbol{Q}(\boldsymbol{Q}^T\boldsymbol{Q} + \lambda \boldsymbol{I}_{2K})^{-1}\boldsymbol{Q}^T$ [16]. Since $\lambda > 0$, it is trivial that $d_\lambda \leq 2K$. That is, the degrees of freedom d_λ is upper bounded by the rank of $\boldsymbol{Q}$.

Define a diagonal matrix $\boldsymbol{\Sigma}_\lambda \in \mathbb{R}^{2K\times 2K}$ whose i-th diagonal entry is given by

$$(\boldsymbol{\Sigma}_\lambda)_{ii} = \sqrt{\frac{\sigma_i^2}{\sigma_i^2 + \lambda}}, \quad i = 1, \ldots, 2K, \tag{17}$$

where σ_i is the i-th singular value of $\boldsymbol{Q}$ and $\lambda = \frac{\sigma^2}{\gamma}$.

Now we provide a weaker constraint with the effective degrees of freedom.

Theorem 2: Assume that for some constant $0 < \varepsilon < 1$, the sketching matrix $\boldsymbol{S} \in \mathbb{R}^{2M\times L}$ satisfies the following constraint

$$\|\boldsymbol{\Sigma}_\lambda \boldsymbol{V}^T \boldsymbol{S}\boldsymbol{S}^T \boldsymbol{V}\boldsymbol{\Sigma}_\lambda - \boldsymbol{\Sigma}_\lambda^2\|_2 \leq \frac{\varepsilon}{4\sqrt{2}}, \tag{18}$$

where $\boldsymbol{V} \in \mathbb{R}^{2M\times 2K}$ is the matrix of right singular vectors of $\boldsymbol{Q}$. Then, after t number of iterations, the approximation $\widehat{\boldsymbol{W}}^{(t)}$ returned by Algorithm 1 satisfies

$$\|\widehat{\boldsymbol{W}}^{(t)} - \boldsymbol{W}^*\|_F \leq \frac{\varepsilon^t}{\sqrt{2}} \left(\|\boldsymbol{W}^*\|_F^2 + \frac{1}{2\lambda}\|\boldsymbol{U}_{\xi,\perp}^T \boldsymbol{\Lambda}\|_F^2 \right)^{\frac{1}{2}}, \tag{19}$$

where ξ is an integer number such that $\sigma_{\xi+1}^2 \leq \lambda \leq \sigma_\xi^2$, $\boldsymbol{U}_{j,\perp} \in \mathbb{R}^{2K\times(2K-j)}$ is the matrix of the bottom $2K - j$ left singular vectors of the matrix $\boldsymbol{Q}$, and $\boldsymbol{W}^*$ is the true value of the RZF beamforming matrix in (5) in the complex version.

Proof: Since each column of $\boldsymbol{M}^*$ can be considered as the solution of (15), by Theorem 2 in [20], it follows that

$$\|\widehat{\boldsymbol{M}}_{*i}^{(t)} - \boldsymbol{M}_{*i}^*\|_2 \leq \frac{\varepsilon^t}{2} \left(\|\boldsymbol{M}_{*i}^*\|_2 + \frac{1}{\sqrt{2\lambda}}\|\boldsymbol{U}_{\xi,\perp}^T \boldsymbol{\Lambda}_{*i}\|_2 \right),$$

for all $i = 1, \ldots, K$. Then we have

$$\begin{aligned}
\|\widehat{\boldsymbol{M}}^{(t)} - \boldsymbol{M}^*\|_F^2 &= \sum_{i=1}^{K} \|\widehat{\boldsymbol{M}}_{*i}^{(t)} - \boldsymbol{M}_{*i}^*\|_2^2 \\
&\leq \sum_{i=1}^{K} \frac{\varepsilon^{2t}}{4} \left(\|\boldsymbol{M}_{*i}^*\|_2 + \frac{1}{\sqrt{2\lambda}}\|\boldsymbol{U}_{\xi,\perp}^T \boldsymbol{\Lambda}_{*i}\|_2 \right)^2 \\
&\leq \sum_{i=1}^{K} \frac{\varepsilon^{2t}}{2} \left(\|\boldsymbol{M}_{*i}^*\|_2^2 + \frac{1}{2\lambda}\|\boldsymbol{U}_{\xi,\perp}^T \boldsymbol{\Lambda}_{*i}\|_2^2 \right) \\
&\leq \frac{\varepsilon^{2t}}{2} \left(\|\boldsymbol{M}^*\|_F^2 + \frac{1}{2\lambda}\|\boldsymbol{U}_{\xi,\perp}^T \boldsymbol{\Lambda}\|_F^2 \right).
\end{aligned}$$

■

This improved dependency on d_λ instead of the rank of matrix $\boldsymbol{Q}$ results in a mild loss in accuracy. λ can be thought of as regularizing the bottom $2K - \xi$ singular values of the matrix $\boldsymbol{Q}$, since it dominates them. Theorem 2 presents a quality-of-approximation result, which uses a relative-additive error approximation. The term $\|\boldsymbol{U}_{\xi,\perp}^T \boldsymbol{\Lambda}\|_F$ is a norm of the part of matrix $\boldsymbol{\Lambda}$ that lies on the regularized component of $\boldsymbol{Q}$. The quality of the approximation worsens as this quantity increases. The error decreases exponentially fast with the number of iterations.

The bounds of (13) and (18) guarantee high-quality approximations to the optimal solution. Constraint (13) can be satisfied by constructing the sampling-and-rescaling matrix $\boldsymbol{S}$ whose size depends on the rank of matrix $\boldsymbol{Q}$, and Theorem 1 guarantees relative error approximations. The second constraint (18) can be satisfied by sampling with respect to the ridge leverage scores, which construct the sampling-and-rescaling matrix $\boldsymbol{S}$ whose size depends on the degrees of freedom d_λ, and Theorem 2 guarantees relative error approximations.

D. Sketching Matrices

Matrix sketching attempts to reduce the size of large matrices while minimizing the loss of spectral information that is useful in tasks like linear regression. Matrix sketching algorithms use a typically randomized procedure to compress $\boldsymbol{Q} \in \mathbb{R}^{2K\times 2M}$ into an approximation (or "sketch") $\boldsymbol{C} \in \mathbb{R}^{2K\times L}$ with many fewer columns $(L \ll 2M)$. Matrix sketching can be accomplished by *random sampling* or *random projection*. Random projection algorithms construct $\boldsymbol{C}$ by forming L random linear combinations of the columns in $\boldsymbol{Q}$. On the other hand, random sampling algorithms construct $\boldsymbol{C}$ by selecting and possibly rescaling a L columns in $\boldsymbol{Q}$. In the latter case, we call a sketching matrix $\boldsymbol{S}$ as the *sampling-and-rescaling matrix*.

Sampling itself is simple and extremely efficient. A simple way to perform this random sampling would be to select those columns uniformly at random in i.i.d. trials, which mean $p_1 = p_2 = \cdots = p_{2M} = \frac{1}{2M}$. A more sophisticated and much more powerful way to do this would be to construct an important sampling routines which select columns using carefully chosen, non-uniform probabilities $\{p_i\}_{i=1}^n$. It is known that variations on the standard "statistical leverage scores" give probabilities that are provably sufficient for approximations such as low-rank approximation. Many of these probabilities are modifications on the standard statistical leverage scores.

Definition 1: The *(statistical) leverage score* of the i^{th} column $\boldsymbol{Q}_{*i}$ of $\boldsymbol{Q}$ is defined as:

$$\tau_i = \boldsymbol{Q}_{*i}^T (\boldsymbol{Q}\boldsymbol{Q}^T)^\dagger \boldsymbol{Q}_{*i}, \tag{20}$$

for $i = 1, 2, \ldots, 2M$.

Here, $\dagger$ denotes the Moore-Penrose pseudoinverse of a matrix. When $\boldsymbol{Q}\boldsymbol{Q}^T$ is full rank, $(\boldsymbol{Q}\boldsymbol{Q}^T)^\dagger = (\boldsymbol{Q}\boldsymbol{Q}^T)^{-1}$. τ_i measures how important $\boldsymbol{Q}_{*i}$ is in composing the range of $\boldsymbol{Q}$. It is maximized at 1 when $\boldsymbol{Q}_{*i}$ is linearly independent from $\boldsymbol{Q}$'s other columns and decreases when many other columns approximately align with $\boldsymbol{Q}_{*i}$ or when $\|\boldsymbol{Q}_{*i}\|_2$ is small.

Leverage score sampling sets p_i proportional to the (exact or approximate) leverage scores τ_i of $\boldsymbol{Q}$. The leverage scores are

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2: Construct Sampling-And-Rescaling Matrix.

Input Sampling probabilities p_i, $i = 1, \ldots, 2M$; integer $L \ll 2M$;
$\boldsymbol{S} \leftarrow \boldsymbol{O}_{2M\times L}$;
for $j = 1$ **to** L **do**
Pick $i_j \in \{1, \ldots, \mathbf{2}M\}$ **with** $\mathbb{P}(i_j = i) = p_i$;
$\boldsymbol{S}_{i_j,j} \leftarrow (Lp_{i_j})^{-\frac{1}{2}}$;
end for
Output: Sampling-and-rescaling matrix $\boldsymbol{S}$;

used in fast sketching algorithms for linear regression and matrix preconditioning [30]–[32].

Notably, leverage scores are defined in terms of $\boldsymbol{Q}_{*i}$, which is not always unique and regardless can be sensitive to matrix perturbations. As a result, the scores can change drastically when $\boldsymbol{Q}$ is modified slightly or when only partial information about the matrix is known. This largely limits the possibility of quickly approximating the scores with sampling algorithms, and motivates our adoption of a new leverage score. Rather than using leverage scores based on $\boldsymbol{Q}_{*i}$, we employ regularized scores called *ridge leverage scores*, which have been used for approximate kernel ridge regression [33] and in works on iteratively computing standard leverage scores [34], [34]. For a given regularization parameter λ, we define the λ-ridge leverage score as:

$$\tau_i^{\lambda} = \boldsymbol{Q}_{*i}^T(\boldsymbol{Q}\boldsymbol{Q}^T + \lambda \boldsymbol{I}_{2K})^{-1}\boldsymbol{Q}_{*i}. \tag{21}$$

Let $\boldsymbol{Q}_\ell$ be the best low-rank approximation for $\boldsymbol{Q}$ with respect to the Frobenius norm. In other words,

$$\boldsymbol{Q}_\ell = \underset{\boldsymbol{X}:\mathrm{rank}(\boldsymbol{X})\le\ell}{\arg\min} \|\boldsymbol{Q} - \boldsymbol{X}\|_F.$$

Note that $\boldsymbol{Q}_\ell$ can be expressed as $\boldsymbol{U}_\ell\boldsymbol{U}_\ell^T\boldsymbol{Q}$. That is, the best rank ℓ approximation can be found by projecting $\boldsymbol{Q}$ onto the span of its top ℓ singular vectors. We will always set $\lambda = \|\boldsymbol{Q} - \boldsymbol{Q}_\ell\|_F^2/\ell$ as follows.

Definition 2: The *ridge leverage score* of the i^{th} column $\boldsymbol{Q}_{*i}$ of $\boldsymbol{Q}$ with respect to the ridge parameter $\lambda > 0$ is defined as:

$$\bar{\tau}_i = \boldsymbol{Q}_{*i}^T\left(\boldsymbol{Q}\boldsymbol{Q}^T + \frac{\|\boldsymbol{Q} - \boldsymbol{Q}_\ell\|_F^2}{\ell}\boldsymbol{I}_{2K}\right)^{-1}\boldsymbol{Q}_{*i}, \tag{22}$$

for $i = 1, 2, \ldots, 2M$.

Note that the ridge leverage score can also be expressed as

$$\bar{\tau}_i = \|(\boldsymbol{V}\boldsymbol{\Sigma}_\lambda)_{i*}\|_2^2 \quad \text{for all } i = 1, 2, \ldots, 2M,$$

where $\boldsymbol{V} \in \mathbb{R}^{2M\times 2K}$ is the matrix of right singular vectors of $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}_\lambda$ is defined as (17). The constraint (18) can also be satisfied by sampling with respect to the ridge leverage scores [33]. The difference is that, instead of having the column size L of the matrix $\boldsymbol{S}$ depend on $2K$, it now depends on d_λ, which could be considerably smaller. Indeed, it follows that by sampling-and-rescaling $\mathcal{O}(d_\lambda \ln d_\lambda)$ from the design matrix $\boldsymbol{Q}$ (using either exact or approximate ridge leverage scores).

In this article we only consider the *sampling-and-rescaling matrix* for a sketching matrix $\boldsymbol{S}$. Algorithm 2 provides the construction of it. The following theorems show how many sampled columns guarantee that the sketching matrix holds the constraint (13). This theorem is adopted from Theorem 3 in [20], so the proof is omitted.

Theorem 3: Let $\boldsymbol{V} \in \mathbb{R}^{2M\times 2K}$ be the matrix of right singular vectors of $\boldsymbol{Q}$. Let $\boldsymbol{S}$ be constructed by Algorithm 2 with the sampling probabilities $p_i = \|\boldsymbol{V}_{i*}\|_2^2/2K$ for $i = 1, \ldots, 2M$. Let δ be a failure probability and let $0 < \varepsilon \le 1$ be an accuracy parameter. If the number of sampled columns L satisfies

$$L \ge \frac{16K}{3\varepsilon^2}\log\left(\frac{4(1+2K)}{\delta}\right), \tag{23}$$

then, with probability at least $1 - \delta$,

$$\|\boldsymbol{V}^T\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{V} - \boldsymbol{I}_{2K}\|_2 \le \varepsilon. \tag{24}$$

The sampling probabilities $p_i = \|\boldsymbol{V}_{i*}\|_2^2/2K$ are the column *leverage scores* [20] of the channel matrix $\boldsymbol{Q}$. Setting $L = \mathcal{O}(\varepsilon^{-2}K\ln K)$ suffices to satisfy the condition (13). [35] demonstrated a construction for such $\boldsymbol{S}$ with $L = \mathcal{O}(\varepsilon^{-2}K)$ columns such that, for $\boldsymbol{Q} \in \mathbb{R}^{2M\times 2K}$, the product $\boldsymbol{Q}\boldsymbol{S}$ can be computed in time $\mathcal{O}(nnz(\boldsymbol{Q})) + \mathcal{O}(K^3/\varepsilon^{\lambda})$ for some constant λ. Here $nnz(\boldsymbol{Q})$ is the number of nonzero entries of $\boldsymbol{Q}$.

Additionally, there are a variety of sketching matrix constructions for $\boldsymbol{S}$ that can satisfy (24). The running time of sketch $\boldsymbol{Q}\boldsymbol{S}$ depends on the dimension of $\boldsymbol{S}$, and the construction of $\boldsymbol{S}$ sampling with respect to *leverage scores* is proportional to $2K$ (we assume that $\mathrm{rank}(\boldsymbol{Q}) = 2K$), which means the running time of $\boldsymbol{Q}\boldsymbol{S}$ is also proportional to $2K$. Therefore, we let $\boldsymbol{S}$ dimensionality depend on the degrees of freedom d_λ of the ridge regression problem, as opposed to the rank of matrix $\boldsymbol{Q}$. In this way, the running time would result in significant savings.

To achieve the reduction in running time, the column size L of matrix $\boldsymbol{S}$ is thus better designed proportional to degrees of freedom d_λ which depends on the distribution of the singular value of $\boldsymbol{Q}$ and λ instead of proportional to $2K$ for $d_\lambda \le 2K$, which could be significantly smaller than $2K$.

Theorem 4: Let $\boldsymbol{V} \in \mathbb{R}^{2M\times 2K}$ be the matrix of right singular vectors of $\boldsymbol{Q}$. Let $\boldsymbol{S}$ be constructed by Algorithm 2 with the sampling probabilities $p_i = \|(\boldsymbol{V}\boldsymbol{\Sigma}_\lambda)_{i*}\|_2^2/d_\lambda$ for $i = 1, \ldots, 2M$. Let δ be a failure probability and let $0 < \varepsilon \le 1$ be an accuracy parameter. If the number of sampled columns L satisfies

$$L \ge \frac{8d_\lambda}{3\varepsilon^2}\log\left(\frac{4(1+d_\lambda)}{\delta}\right), \tag{25}$$

then, with probability at least $1 - \delta$,

$$\|\boldsymbol{\Sigma}_\lambda\boldsymbol{V}^T\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{V}\boldsymbol{\Sigma}_\lambda - \boldsymbol{\Sigma}_\lambda^2\|_2 \le \varepsilon. \tag{26}$$

The sampling probabilities $p_i = \|(\boldsymbol{V}\boldsymbol{\Sigma}_\lambda)_{i*}\|_2^2/d_\lambda$ are the column *ridge leverage scores* [33], [36] of the channel matrix $\boldsymbol{Q}$. Similarly to the constraint of (24), setting $L = \mathcal{O}(d_\lambda \ln d_\lambda)$ suffices to satisfy the condition (26). Note that while the leverage scores which construct the sampling-and-rescaling matrix $\boldsymbol{S}$ with the column size L depend on the rank of $\boldsymbol{Q}$, the ridge leverage scores constructing $\boldsymbol{S}$ depend on d_λ, which could be considerably smaller than the rank of $\boldsymbol{Q}$. Hence, it could surely achieve time saving. If the rank of the channel matrix $\boldsymbol{Q}$ is smaller that $2K$, e.g., the channels of some users are highly correlated, the ridge leverage scores scheme has a smaller sketch

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

size according to Theorem 3 and Theorem 4, and thus enjoys a lower computational complexity. However, the running time savings would lead to a drop in accuracy as shown in Theorem 2.

E. Time Complexity

We now discuss the time complexity of Algorithm 1. Note that each column $\boldsymbol{M}$ in (8) can be computed by each column of $\boldsymbol{\Lambda}$, separately. We consider $\boldsymbol{\Lambda}$ as a column vector. Let $\boldsymbol{\Theta} = \boldsymbol{QSS}^T\boldsymbol{Q}^T + \lambda \boldsymbol{I}_{2K}$. Note that to find $\boldsymbol{\Theta}^{-1}$, it suffices to compute the singular value decomposition of $\boldsymbol{QS}$. Since the singular values of $\boldsymbol{\Theta}$ can be computed through $\boldsymbol{\Sigma}_{\boldsymbol{QS}} + \lambda \boldsymbol{I}_{2K}$, where $\boldsymbol{\Sigma}_{\boldsymbol{A}}$ denotes the singular value of $\boldsymbol{A}$. And the left and right singular vectors of $\boldsymbol{\Theta}$ are the same as the left singular vectors of $\boldsymbol{QS}$. We store it implicitly by storing its left (and right) singular vector $\boldsymbol{U}_{\boldsymbol{\Theta}}$ and its singular values $\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}$, before we just compute all the necessary matrix-vector products using this implicit representation of $\boldsymbol{\Theta}^{-1}$. The above analysis shows that we do not need to compute $\boldsymbol{\Theta}^{-1}$ directly. Thus computing $\boldsymbol{\Theta}^{-1}$ takes $\mathcal{O}(LK^2)$ time.

Updating each $\boldsymbol{\Lambda}^{(j)}$, $\boldsymbol{Y}^{(j)}$, and $\widetilde{\boldsymbol{M}}^{(j)}$ is dominated by the aforementioned running times, as all updates amount to just matrix-vector products. Thus, summing over all t iterations, the running time of Algorithm 1 is given by

$$\mathcal{O}(t \cdot nnz(\boldsymbol{Q})) + \mathcal{O}(LK^2). \tag{27}$$

Thus the time complexity is reduced evidently. Note that the complexity of computing the matrix inversion (8) is $\mathcal{O}(MK^2)$.

IV. The System Sum-Rate Analysis With Approximate RZF Beamformers

In this section, we show that the system sum-rate of the randomized sketching based beamformer converges to the sum-rate of the RZF beamforming matrix as the number of iterations increases. Moreover, if an approximation sequence converges to the true beamforming matrix with the rate of convergence $\mathcal{O}(\beta_t)$, then the system sum-rate of the approximation sequence converges with the same rate of convergence $\mathcal{O}(\beta_t)$. Before stating our main results, we introduce the extra notation, ϕ_{kj}, to cast SINR at the k-th user in (2) with a simpler form. From now on, we assume that the channel matrix $\boldsymbol{H}$ is fixed and the beamforming matrix $\boldsymbol{W}$ is considered as complex variables. Then we can easily deal with the system sum-rate for any approximate beamforming matrix.

For each k, j, let a function $\phi_{kj} : \mathbb{C}^{M\times K} \to [0, +\infty)$ be defined by $\phi_{kj}(\boldsymbol{W}) = |\boldsymbol{h}_k^H \boldsymbol{w}_j|^2$ for all $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K] \in \mathbb{C}^{M\times K}$. Note that

$$\begin{aligned}\phi_{kj}(\boldsymbol{W}) = &\left(\Re(\boldsymbol{h}_k)^T\Re(\boldsymbol{w}_j) - \Im(\boldsymbol{h}_k)^T\Im(\boldsymbol{w}_j)\right)^2 \\ &+ \left(\Im(\boldsymbol{h}_k)^T\Re(\boldsymbol{w}_j) + \Re(\boldsymbol{h}_k)^T\Im(\boldsymbol{w}_j)\right)^2 \geq 0.\end{aligned}$$

The system sum-rate of a given variables $\boldsymbol{W}$ can thus be rewritten as

$$R(\boldsymbol{W}) = \sum_{k=1}^{K} \log\left(1 + \frac{\phi_{kk}(\boldsymbol{W})}{\sum_{j\neq k}\phi_{kj}(\boldsymbol{W}) + \sigma^2}\right). \tag{28}$$

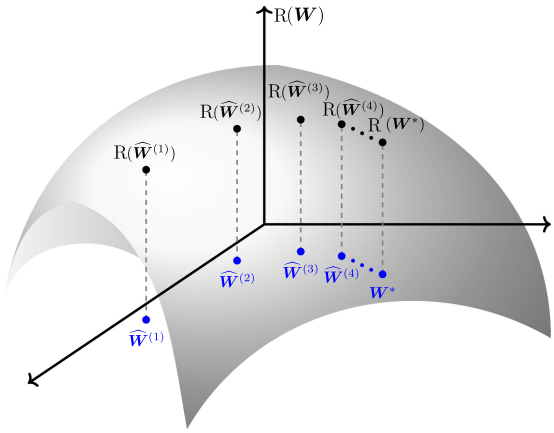


Fig. 1. The system sum-rate $R(\boldsymbol{W})$.

That is, R can be viewed as a function from $\mathbb{C}^{M\times K}$ to $[0, +\infty)$, as shown in Fig. 1.

Let V be a nonempty open subset of $\mathbb{R}^n$, $f : V \to \mathbb{R}^m$, and $p \in \mathbb{N}$. Recall that a function f is said to be $\mathcal{C}^p$ on V if each partial derivative of f of order $k \leq p$ exists and is continuous on V. f is said to be $\mathcal{C}^\infty$ on V if f is $\mathcal{C}^p$ on V for all $p \in \mathbb{N}$. In other words, a $\mathcal{C}^\infty$-mapping is a function that is differentiable for all degrees of differentiation.

Lemma 1: The system sum-rate R is a $\mathcal{C}^\infty$-mapping on $\mathbb{R}^{2M\times K}$.

Proof: Note that the complex variables $\boldsymbol{W} \in \mathbb{C}^{M\times K}$ can be considered as real variables $\boldsymbol{M} \in \mathbb{R}^{2M\times K}$. We use $\boldsymbol{M}$ and $\boldsymbol{W}$ interchangeably. Let $\boldsymbol{M} \in \mathbb{R}^{2M\times K}$ be real variables. Then it is easy to check that that ϕ_{kj} is a multivariate polynomial in $\mathbb{R}[\boldsymbol{M}]$, i.e., the ring of polynomials with real coefficients over variables $\boldsymbol{M}$. Thus ϕ_{kj} is C^∞-mapping on $\mathbb{R}^{2M\times 2K}$. Since the logarithm function are C^∞-mapping, the function R is a C^∞-mapping on $\mathbb{R}^{2M\times 2K}$, provided $\sum_{j\neq k}\phi_{kj}(\boldsymbol{W}) + \sigma^2 \neq 0$. Let $\boldsymbol{H} \in \mathbb{C}^{K\times M}$ be a given channel matrix. Considering the beamforming matrix $\boldsymbol{M} \cong \boldsymbol{W}$ as real variables in $\mathbb{R}^{2M\times K}(\cong \mathbb{C}^{M\times K})$, the system sum-rate in (4) can be considered as a function $R : \mathbb{R}^{2M\times K} \longrightarrow [0, +\infty)$ defined by

$$\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K] \cong \begin{bmatrix}\Re(\boldsymbol{w}_1) & \cdots & \Re(\boldsymbol{w}_K) \\ \Im(\boldsymbol{w}_1) & \cdots & \Im(\boldsymbol{w}_K)\end{bmatrix} \longmapsto R(\boldsymbol{W}).$$

In other words, ϕ_{kj} can be considered as a function from $\mathbb{R}^{2M\times 2K} \longrightarrow [0, +\infty)$. Moreover, it is easy to check that ϕ_{kj} is a multivariate polynomial in $\mathbb{R}[\boldsymbol{M}]$, which is C^∞-mapping on $\mathbb{R}^{2M\times 2K}$. Since it can be rewritten as

$$R(\boldsymbol{W}) = \sum_{k=1}^{K} \log\left(1 + \frac{\phi_{kk}(\boldsymbol{W})}{\sum_{j\neq k}\left(\phi_{kj}(\boldsymbol{W}) + \sigma^2\right)}\right),$$

and the logarithm function is C^∞-mapping, the function R is a C^∞-mapping on $\mathbb{R}^{2M\times 2K}$, provided $\sum_{j\neq k}(\phi_{kj}(\boldsymbol{W}) + \sigma^2) \neq 0$. ■

Denote the true solution of the regularized RZF problem (8) in the complex version as $\boldsymbol{W}^*$. By Lemma 1, R is continuous. By Theorem 1, each entry of the approximation converges to the

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

entry of the true solution, respectively, i.e.,

$$\Re(\widehat{\boldsymbol{W}}_{ij}^{(t)}) \longrightarrow \Re(\boldsymbol{W}_{ij}^{*}) \quad \text{as} \quad t \longrightarrow \infty,$$

$$\Im(\widehat{\boldsymbol{W}}_{ij}^{(t)}) \longrightarrow \Im(\boldsymbol{W}_{ij}^{*}) \quad \text{as} \quad t \longrightarrow \infty.$$

Note that the image of a convergent sequence under a continuous function converges to the image of limit. Thus, the following holds.

Proposition 5: Assume that for some constant $0 < \varepsilon < 1$, the sketching matrix $\boldsymbol{S} \in \mathbb{R}^{2M \times L}$ satisfies the constraint (13). Let t be the number of iterations. Then, the system sum-rate of the approximation $R(\widehat{\boldsymbol{W}}^{(t)})$ converges to the system sum-rate of the true solution $R(\boldsymbol{W}^*)$ as the number of iterations increases. That is,

$$R\left(\widehat{\boldsymbol{W}}^{(t)}\right) \longrightarrow R\left(\boldsymbol{W}^*\right) \quad \text{as} \quad t \longrightarrow \infty. \tag{29}$$

The next theorem is our key result. It shows that the error of the system sum-rate is bounded by the error of an approximation of beamforming matrix. Using this result, the rate of convergence for the system sum-rate of an approximation can be obtained.

Theorem 6: Let $\boldsymbol{H}$ be a given channel matrix, and let $\widehat{\boldsymbol{W}}$ (resp. $\boldsymbol{W}^*$) be the approximation (resp. true) RZF beamforming matrix. Then it holds that

$$\begin{aligned}&\left|R\left(\widehat{\boldsymbol{W}}\right) - R\left(\boldsymbol{W}^*\right)\right| \\ &\leq C\|\boldsymbol{H}\|_F^2 \left(\|\widehat{\boldsymbol{W}} - \boldsymbol{W}^*\|_F^2 + 2\|\widehat{\boldsymbol{W}} - \boldsymbol{W}^*\|_F \|\boldsymbol{W}^*\|_F\right),\end{aligned}$$

where C is constant independent to $\widehat{\boldsymbol{W}}$.

Proof: See Appendix A. ∎

Suppose a sequence $\{\beta_t\}_{t=1}^{\infty}$ converges to zero, and $\{\alpha_t\}_{t=1}^{\infty}$ converges to a number α. Recall that $\{\alpha_t\}_{t=1}^{\infty}$ converges to α with *rate of convergence* $\mathcal{O}(\beta_t)$ if a positive constant K exists with $|\alpha_t - \alpha| \leq K|\beta_t|$ for sufficiently large t.

The following shows that if an approximation sequence $\{\widehat{\boldsymbol{W}}^{(t)}\}_{t=1}^{\infty}$ converges to the true beamforming matrix $\boldsymbol{W}^*$ with the rate of convergence $\mathcal{O}(\beta_t)$, then $\{R(\widehat{\boldsymbol{W}}^{(t)})\}_{t=1}^{\infty}$ converges to $R(\boldsymbol{W}^*)$ with the same rate of convergence $\mathcal{O}(\beta_t)$.

Theorem 7: Let $\eta \geq 0$. If an approximation sequence $\{\widehat{\boldsymbol{W}}^{(t)}\}_{t=1}^{\infty}$ converges to $\boldsymbol{W}^*$ such that $\|\widehat{\boldsymbol{W}}^{(t)} - \boldsymbol{W}^*\|_F \leq |\beta_t|(\|\boldsymbol{W}^*\|_F + \eta)$, then

$$\left|R\left(\widehat{\boldsymbol{W}}^{(t)}\right) - R\left(\boldsymbol{W}^*\right)\right| \leq 3C|\beta_t|\|\boldsymbol{H}\|_F^2(\|\boldsymbol{W}^*\|_F + \eta)^2,$$

provided sufficiently large t.

Proof: Since β_t converges to 0, there exists $T \in \mathbb{N}$ such that $t \geq T$ implies $|\beta_t| < 1$. By Theorem 6 it follows that

$$\begin{aligned}\left|R\left(\widehat{\boldsymbol{W}}\right) - R\left(\boldsymbol{W}^*\right)\right| &\leq C\|\boldsymbol{H}\|_F^2 \Big(|\beta_t|^2 \left(\|\boldsymbol{W}^*\|_F + \eta\right)^2 \\ &\quad + 2|\beta_t| \left(\|\boldsymbol{W}^*\|_F + \eta\right) \|\boldsymbol{W}^*\|_F\Big).\end{aligned}$$

Since $\eta \geq 0$, we have

$$\begin{aligned}&|\beta_t|^2 \left(\|\boldsymbol{W}^*\|_F + \eta\right)^2 + 2|\beta_t| \left(\|\boldsymbol{W}^*\|_F + \eta\right) \|\boldsymbol{W}^*\|_F \\ &\leq |\beta_t|^2 \left(\|\boldsymbol{W}^*\|_F + \eta\right)^2 + 2|\beta_t| \left(\|\boldsymbol{W}^*\|_F + \eta\right)^2 \\ &\leq 3|\beta_t| \left(\|\boldsymbol{W}^*\|_F + \eta\right)^2,\end{aligned}$$

provided $t \geq T$. ∎

Using Theorem 7, one can find the rate of convergence for the system sum-rate of the approximation sequence generated by Algorithm 1.

Corollary 8:

(i) Assume that for $0 < \varepsilon < 1$, the sketching matrix $\boldsymbol{S}$ satisfies the constraint (13). Then, after t number of iterations, the approximation $\widehat{\boldsymbol{W}}^{(t)}$ returned by Algorithm 1 holds

$$\left|R\left(\widehat{\boldsymbol{W}}^{(t)}\right) - R\left(\boldsymbol{W}^*\right)\right| \leq 3C\varepsilon^t\|\boldsymbol{H}\|_F^2\|\boldsymbol{W}^*\|_F^2.$$

(ii) Assume that for $0 < \varepsilon < 1$, the sketching matrix $\boldsymbol{S}$ satisfies the constraint (18). Then, after t number of iterations, the approximation $\widehat{\boldsymbol{W}}^{(t)}$ returned by Algorithm 1 holds

$$\begin{aligned}\left|R\left(\widehat{\boldsymbol{W}}^{(t)}\right) - R\left(\boldsymbol{W}^*\right)\right| &\leq 3C\varepsilon^t\|\boldsymbol{H}\|_F^2\Big(\|\boldsymbol{W}^*\|_F \\ &\quad + \frac{1}{\sqrt{2\lambda}}\|\boldsymbol{U}_{2K,\perp}^T\boldsymbol{\Lambda}\|_2\Big)^2.\end{aligned}$$

Here, $\boldsymbol{W}^*$ is the true value in (5).

Proof: (i) It holds from Theorem 7 with $\eta = 0$ and Theorem 1. (ii) It holds from Theorem 7 with $\eta = \frac{1}{\sqrt{2\lambda}}\|\boldsymbol{U}_{2K,\perp}^T\boldsymbol{\Lambda}\|_2$ and Theorem 1. ∎

V. Simulations

In this section, we simulate the performance of the proposed randomized sketching based beamformer in Algorithm 1. We consider the following channel model between the BS and the k-th user:

$$\boldsymbol{h}_k = 10^{-\tilde{L}(d_k)/20}\sqrt{\varphi_k s_k}\boldsymbol{f}_k, \tag{30}$$

where $\tilde{L}(d_k)$ is the path-loss at distance d_k, s_k is the shadowing coefficients, φ_k is the antenna gain, and $\boldsymbol{f}_k$ is the small fading coefficient. We use the standard cellular network parameters as shown in [37, Table I]. We consider a single cell massive MIMO system with $M = 1500$ antennas at the BS and $K = 50$ single-antenna users uniformly and independently distributed in the square region $[-5000, 5000] \times [5000, 5000]$ meters.

First, we compare three different sampling-and-rescaling methods whose random matrices are generated by Algorithm 2 with the following sampling probabilities $\{p_i\}_{i=1}^{2M}$:

- (**Uniformly at random**) Calculate $p_i = \frac{1}{2M}$ for $i = 1, \ldots, 2M$.
- (**Leverage scores**) Calculate $p_i = \frac{\|\boldsymbol{V}_{i*}\|_2^2}{2K}$ for $i = 1, \ldots, 2M$.
- (**Ridge leverage scores**) Calculate $p_i = \frac{\|(\boldsymbol{V}\boldsymbol{\Sigma}_\lambda)_{i*}\|_2^2}{d_\lambda}$ for $i = 1, \ldots, 2M$.

Here, $\boldsymbol{V}$ denotes the right singular value of matrix $\boldsymbol{Q}$, and $\boldsymbol{\Sigma}_\lambda$ denotes the diagonal matrix given by (17).

Fig. 2 shows the sum-rate error for three different sampling-and-rescaling methods. We fix the iteration numbers $t = 15$.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

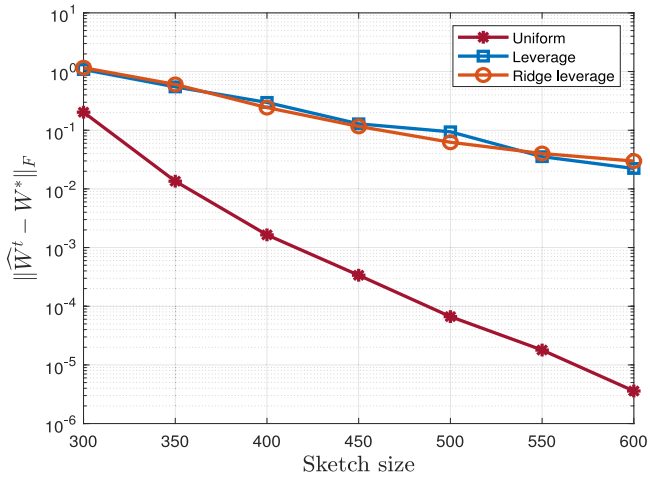


Fig. 2. Solution error vs. sketch size.

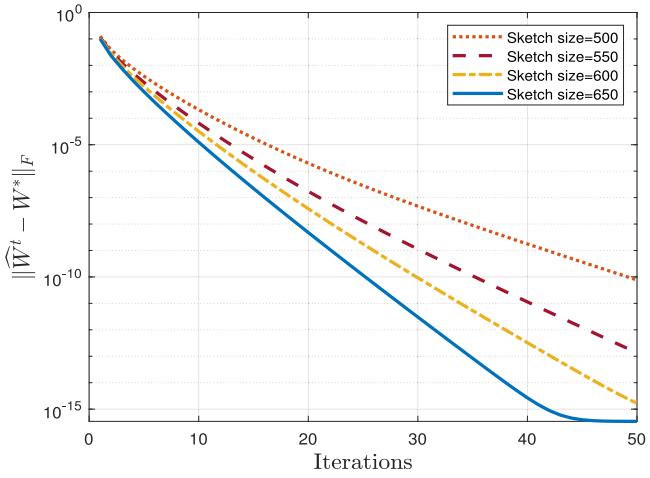


Fig. 4. Solution error vs. iteration.

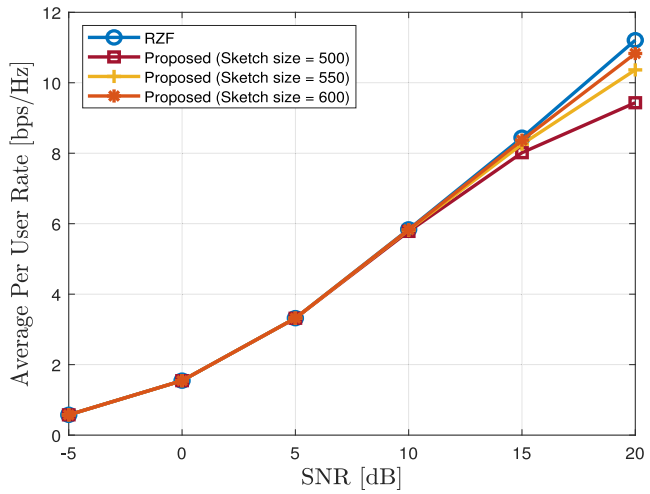


Fig. 3. Average per user rate vs. SNR.

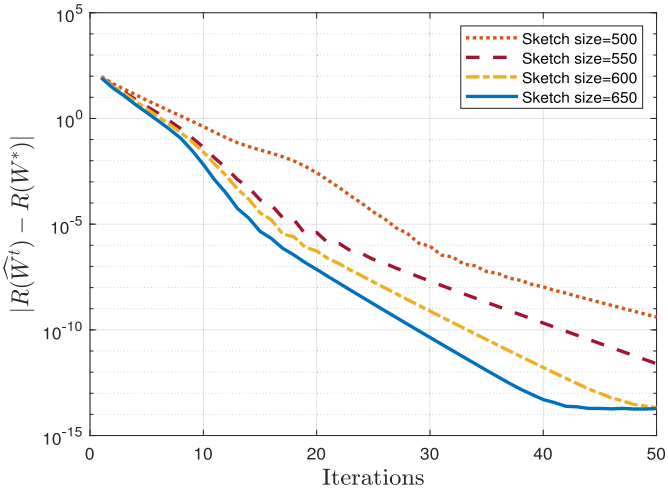


Fig. 5. Sum-rate error vs. iteration.

Each graphs present the average of 200 replicated runs. It clearly illustrates that the uniformly at random method achieves better accuracy than the other two sampling-and-rescaling methods. We thus generate sampling matrix uniformly at random in the following simulations.

We compare the proposed randomized sketching based beamformers under various sketch sizes and different SNR which is defined as the transmit power at the BS over the received noise power at all the users. We generate sketching matrices with different sizes, and terminate Algorithm 1 after 10 iterations. As shown in Fig. 3, the randomized sketching based beamformer performs closely to RZF beamforming in terms of the average per user rate as the sketch size increases.

Fig. 4 illustrates that the iterative solution converges to the RZF beamforming matrix at a linear convergence rate as shown in Theorem 1 by plotting the trend of $\|\widehat{\boldsymbol{W}}^{(t)} - \boldsymbol{W}^*\|_F$ up to 50 iterations with SNR being 5. Fig. 5 illustrates that the achievable sum-rate of the randomized beamforming converges to the achievable sum-rate given by RZF beamforming linearly as shown in Corollary 8. It demonstrates the trend of $|R(\widehat{\boldsymbol{W}}^{(t)}) - R(\boldsymbol{W}^*)|$ within 50 iterations. It is clear that in Fig. 4 and 5 the error decreases fast with the number of iterations, and larger sketch size leads to faster convergence rate.

Finally, we compare the proposed randomized sketching method with the polynomial expansion based method [13] given the channel matrix $\boldsymbol{H}$. Note that the polynomial expansion based method [13] consists of two steps, i.e., seeking the polynomial coefficients based on the channel matrix $\boldsymbol{H}$ and then computing the beamforming matrix using the polynomial coefficients. We follow the simulation settings in [13] and consider the case $M = 1000$, $K = 50$. The sketch size is set to be 500, and we terminate our proposed sketch method after 10 iterations. The results averaged for 100 times are presented in Fig. 6. For the polynomial expansion based method, we present the results with polynomial expansion degree equals $2, 3$ and 4, respectively. As can be seen from Fig. 6, our proposed method and the truncated polynomial expansion method with degree 3 achieve similar performance as the RZF method. To compute the beamforming matrix, the average running time of our proposed randomized sketching method is $0.0214s$, while it costs $2.0950s$

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

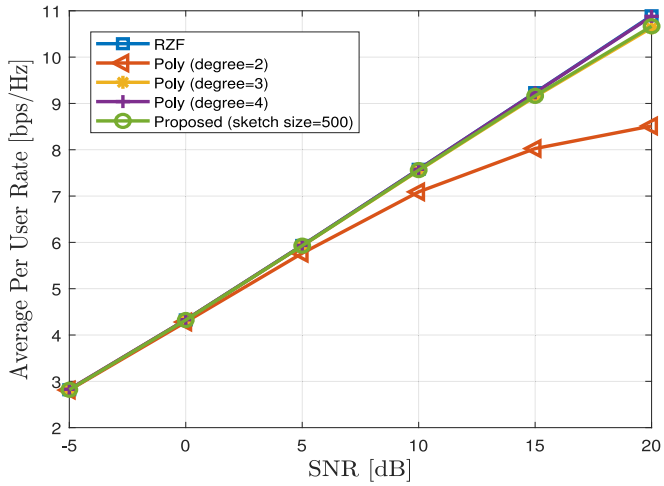


Fig. 6. Compassion with truncated polynomial expansion method [13].

TABLE I
COMPARISON FOR TIME COMPLEXITY

Approaches	Time (seconds)
Proposed (sketch size = 500)	0.0214
Poly (degree = 2)	0.6728
Poly (degree = 3)	2.0950
Poly (degree = 4)	5.2871

via the truncated polynomial expansion method with degree 3 (see TABLE I). This is because the time complexity is high when we compute the polynomial coefficients. Therefore, given the channel matrix $\boldsymbol{H}$, the proposed sketching method is more efficient compared to the polynomial expansion based method.

VI. CONCLUSION

We proposed a randomized sketching based RZF beamforming approach to tackle the computational challenges of precoding in massive MIMO systems. This was achieved by solving the linear system for the matrix inversion via randomized sketching based on the preconditioned Richard iteration. The computational complexity of our proposed method scales with LK^2, where $L \ll 2M$ is the sketching matrix size. Furthermore, we proved that the proposed algorithm iteratively converges to the RZF beamforming matrix at a linear convergence rate. Also, the achievable sum-rate with the randomized sketching based RZF beamformer linearly converges to the achievable sum-rate with the RZF beamformer as the number of iteration increases. Simulation results were demonstrated to verify our theoretical findings.

This paper shows the benefit of exploiting randomized numerical linear algebra for the very large-scale RZF beamforming problem. The key insight is that the channel matrix contains redundancy for designing the beamforming matrix if the number of antennas at the BS goes to very large. The idea can potentially be leveraged in various future 6 G wireless networks design, where the problem dimension goes increasingly large [38]. For example, in wireless federated systems [39], the high dimensional local model parameters are required to be aggregated at an edge server, which leads to a communication bottleneck. By using the randomized numerical linear algebra, we may reduce the communication overhead during the model aggregation in federated learning [40]. In addition, the reconfigurable intelligent surface (RIS) assisted massive MIMO system suffers from high dimensional problem optimization due to the large number of reflecting elements at the RIS [41]. Therefore, it is also interesting to exploit the potential use of randomized numerical linear algebra in the RIS systems.

For future research, it is important to extend the randomized sketching beamforming method for other utility maximization problems. One of our key observations is that the solution structure of the matrix ridge regression problem is the same as the RZF beamforming solution, which connects the randomized numerical algorithms to the RZF beamforming. For other utility function maximization problems, e.g., minimum rate maximization, it remains open to apply the randomized numerical algorithms to reduce the computational complexity. It is also interesting to further reduce the computational complexity by exploiting the statistical information of the channels. For example, we can design the sketching matrix based on the statistical information of the channel matrix [42], so as to avoid constructing the sketching matrix for every channel realization.

APPENDIX A
PROOF OF THEOREM 6

Lemma 2: For all $a, b > 0$, it holds that $|\log(1+a) - \log(1+b)| \leq |a-b|$.

Proof: Let $f(x) = \log(1+x)$ and $a, b > 0$ with $a \neq b$. By the Mean Value Theorem, there exists $\xi \in (a, b)$ such that

$$\left|\frac{f(a)-f(b)}{a-b}\right| = |f'(\xi)| < 1.$$

■

Lemma 3: Let $a, b > 0$ be given. Then it holds that

$$\left|\frac{y}{x+b} - \frac{a}{b}\right| \leq \frac{1}{|b|}|y-a| + \frac{a}{b^2}|x|, \tag{31}$$

for all $x, y \geq 0$.

Proof: By the triangle inequality, it holds that

$$\left|\frac{y}{x+b} - \frac{a}{b}\right| = \frac{|b(y-a) - ax|}{|x+b||b|} \leq \frac{|b(y-a)| + |ax|}{|b|^2},$$

for all $x, y \geq 0$. ■

Lemma 4: For each k, j it holds that

$$\left|\phi_{kj}(\widehat{\boldsymbol{W}}) - \phi_{kj}(\widetilde{\boldsymbol{W}})\right|$$
$$\leq \left\|\widehat{\boldsymbol{w}}_j - \widetilde{\boldsymbol{w}}_j\right\|_2 \left\|\boldsymbol{h}_k\right\|_2^2 \left(\left\|\widehat{\boldsymbol{w}}_j - \widetilde{\boldsymbol{w}}_j\right\|_2 + 2\left\|\widetilde{\boldsymbol{w}}_j\right\|_2\right).$$

Proof: Using the fact $|\boldsymbol{h}_k^H \boldsymbol{w}_j|^2 = \boldsymbol{w}_j^H \boldsymbol{h}_k \boldsymbol{h}_k^H \boldsymbol{w}_j$, by the triangle inequality, we have that

$$\left|\phi_{kj}(\widehat{\boldsymbol{W}}) - \phi_{kj}(\widetilde{\boldsymbol{W}})\right|$$
$$= \left|\widehat{\boldsymbol{w}}_j^H \boldsymbol{h}_k \boldsymbol{h}_k^H \widehat{\boldsymbol{w}}_j - \widetilde{\boldsymbol{w}}_j^H \boldsymbol{h}_k \boldsymbol{h}_k^H \widetilde{\boldsymbol{w}}_j\right|$$
$$= \left|\widehat{\boldsymbol{w}}_j^H \boldsymbol{h}_k \boldsymbol{h}_k^H (\widehat{\boldsymbol{w}}_j - \widetilde{\boldsymbol{w}}_j) + (\widehat{\boldsymbol{w}}_j^H - \widetilde{\boldsymbol{w}}_j^H) \boldsymbol{h}_k \boldsymbol{h}_k^H \widetilde{\boldsymbol{w}}_j\right|$$

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

$$\leq \left|\widehat{\boldsymbol{w}}_j^H \boldsymbol{h}_k \boldsymbol{h}_k^H (\widehat{\boldsymbol{w}}_j - \widetilde{\boldsymbol{w}}_j)\right| + \left|(\widehat{\boldsymbol{w}}_j^H - \widetilde{\boldsymbol{w}}_j^H)\boldsymbol{h}_k \boldsymbol{h}_k^H \widetilde{\boldsymbol{w}}_j\right|$$
$$\leq \left\|\widehat{\boldsymbol{w}}_j - \widetilde{\boldsymbol{w}}_j\right\|_2 \left\|\boldsymbol{h}_k \boldsymbol{h}_k^H\right\|_2 \left(\left\|\widehat{\boldsymbol{w}}_j\right\|_2 + \left\|\widetilde{\boldsymbol{w}}_j\right\|_2\right).$$

By the triangle inequality and the definition of operator norm, it holds that

$$\left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j\right\|_2 \left\|\boldsymbol{h}_k \boldsymbol{h}_k^H\right\|_2 \left(\left\|\widehat{\boldsymbol{w}}_j\right\|_2 + \left\|\boldsymbol{w}_j\right\|_2\right)$$
$$\leq \left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j\right\|_2 \left\|\boldsymbol{h}_k \boldsymbol{h}_k^H\right\|_2 \left(\left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j\right\|_2 + 2\left\|\boldsymbol{w}_j\right\|_2\right)$$
$$\leq \left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j\right\|_2 \left\|\boldsymbol{h}_k\right\|_2^2 \left(\left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j\right\|_2 + 2\left\|\boldsymbol{w}_j\right\|_2\right).$$

■

Now we are ready to prove Theorem 6.

Since $SINR_k > 0$ for all k, the triangle inequality and Lemma 2 imply that

$$\left|R\left(\widehat{\boldsymbol{W}}\right) - R\left(\boldsymbol{W}^*\right)\right| \leq \sum_{k=1}^{K} \left|SINR_k(\widehat{\boldsymbol{W}}) - SINR_k(\boldsymbol{W}^*)\right|.$$

Then by Lemma 3 it follows that

$$\sum_{k=1}^{K} \left|SINR_k(\widehat{\boldsymbol{W}}) - SINR_k(\boldsymbol{W}^*)\right|$$
$$= \sum_{k=1}^{K} \left|\frac{\phi_{kk}(\widehat{\boldsymbol{W}})}{\sum_{j\neq k}\phi_{kj}(\widehat{\boldsymbol{W}}) + \sigma^2} - \frac{\phi_{kk}(\boldsymbol{W}^*)}{\sum_{j\neq k}\phi_{kj}(\boldsymbol{W}^*) + \sigma^2}\right|$$
$$\leq \sum_{k=1}^{K} \left[\frac{1}{\sum_{j\neq k}\phi_{kj}(\boldsymbol{W}^*) + \sigma^2}\left|\phi_{kk}(\widehat{\boldsymbol{W}}) - \phi_{kk}(\boldsymbol{W}^*)\right|\right.$$
$$\left. + \frac{\phi_{kk}(\boldsymbol{W}^*)}{(\sum_{j\neq k}\phi_{kj}(\boldsymbol{W}^*) + \sigma^2)^2}\left|\sum_{j\neq k}\phi_{kj}(\widehat{\boldsymbol{W}}) - \sum_{j\neq k}\phi_{kj}(\boldsymbol{W}^*)\right|\right]$$
$$\leq \frac{C}{2}\sum_{k=1}^{K}\left[\left|\phi_{kk}(\widehat{\boldsymbol{W}}) - \phi_{kk}(\boldsymbol{W}^*)\right| + \sum_{j\neq k}\left|\phi_{kj}(\widehat{\boldsymbol{W}}) - \phi_{kj}(\boldsymbol{W}^*)\right|\right]$$
$$\leq C\|\boldsymbol{H}\|_F^2 \sum_{k=1}^{K}\left(\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2^2 + 2\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2\left\|\boldsymbol{w}_k^*\right\|_2\right)$$
$$\leq C\|\boldsymbol{H}\|_F^2\left(\left\|\widehat{\boldsymbol{W}} - \boldsymbol{W}^*\right\|_F^2 + 2\left\|\widehat{\boldsymbol{W}} - \boldsymbol{W}^*\right\|_F\left\|\boldsymbol{W}^*\right\|_F\right),$$

where $C = 2\max_k\left\{\frac{1}{\sum_{j\neq k}\phi_{kj}(\boldsymbol{W}^*)+\sigma^2}, \frac{\phi_{kk}(\boldsymbol{W}^*)}{(\sum_{j\neq k}\phi_{kj}(\boldsymbol{W}^*)+\sigma^2)^2}\right\}$.

Note that the first inequality holds from Lemma 3. The last second inequality holds from the following by Lemma 4.

$$\sum_{k=1}^{K}\sum_{j\neq k}\left|\phi_{kj}(\widehat{\boldsymbol{W}}) - \phi_{kj}(\boldsymbol{W}^*)\right|$$
$$\leq \sum_{j=1}^{K}\sum_{k=1}^{K}\left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j^*\right\|_2\left\|\boldsymbol{h}_k\right\|_2^2\left(\left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j^*\right\|_2 + 2\left\|\boldsymbol{w}_j^*\right\|_2\right)$$
$$\leq \|\boldsymbol{H}\|_F^2 \sum_{j=1}^{K}\left(\left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j^*\right\|_2^2 + 2\left\|\widehat{\boldsymbol{w}}_j - \boldsymbol{w}_j^*\right\|_2\left\|\boldsymbol{w}_j^*\right\|_2\right),$$

$$\sum_{k=1}^{K}\left|\phi_{kk}(\widehat{\boldsymbol{W}}) - \phi_{kk}(\boldsymbol{W}^*)\right|$$
$$\leq \sum_{k=1}^{K}\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2\left\|\boldsymbol{h}_k\right\|_2^2\left(\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2 + 2\left\|\boldsymbol{w}_k^*\right\|_2\right)$$
$$\leq \sum_{k=1}^{K}\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2\left\|\boldsymbol{H}\right\|_F^2\left(\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2 + 2\left\|\boldsymbol{w}_k^*\right\|_2\right)$$
$$\leq \|\boldsymbol{H}\|_F^2 \sum_{k=1}^{K}\left(\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2^2 + 2\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2\left\|\boldsymbol{w}_k^*\right\|_2\right).$$

The last inequality holds from the Cauchy-Schwartz inequality as follows:

$$\sum_{k=1}^{K}\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2\left\|\boldsymbol{w}_k^*\right\|_2$$
$$\leq \left(\sum_{k=1}^{K}\left\|\widehat{\boldsymbol{w}}_k - \boldsymbol{w}_k^*\right\|_2^2\right)^{\frac{1}{2}}\left(\sum_{k=1}^{K}\left\|\boldsymbol{w}_k^*\right\|_2^2\right)^{\frac{1}{2}}$$
$$\leq \left\|\widehat{\boldsymbol{W}} - \boldsymbol{W}^*\right\|_F\left\|\boldsymbol{W}^*\right\|_F.$$

REFERENCES

[1] H. Choi, T. Jiang, W. Li, and Y. Shi, "Randomized sketching based beamforming for massive MIMO," in *Proc. IEEE Glob. Commun. Conf.*, Waikoloa, Hawaii, USA, Dec. 2019, pp. 1–6.

[2] Y. Shi, J. Zhang, W. Chen, and K. B. Letaief, "Generalized sparse and low-rank optimization for ultra-dense networks," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 42–48, Jun. 2018.

[3] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality-what is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.

[5] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[6] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[7] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.

[8] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[9] E. Björnson *et al.*, "Optimal resource allocation in coordinated multi-cell systems," *Found. Trends Commun. Inf. Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

[10] R. Zakhour and S. V. Hanly, "Base station cooperation on the downlink: Large system analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012.

[11] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, Sep. 2015.

[12] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

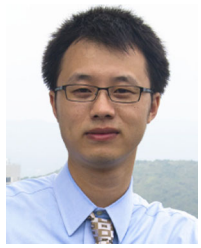
[13] A. Mueller, A. Kammoun, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: Reducing complexity in massive MIMO," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, pp. 1–22, Feb. 2016.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

[14] A. Benzin, G. Caire, Y. Shadmi, and A. M. Tulino, "Low-complexity truncated polynomial expansion DL precoders and UL receivers for massive MIMO in correlated channels," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1069–1084, Feb. 2019.
[15] N. Halko, P. Martinsson, and J. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, pp. 217–288, May 2011.
[16] D. P. Woodruff, "Sketching as a tool for numerical linear algebra," *Found. Trends Theor. Comput. Sci.*, vol. 10, no. 1-2, pp. 1–157, 2014.
[17] P. Drineas and M. W. Mahoney, "Randnla: Randomized numerical linear algebra," *Commun. ACM*, vol. 59, pp. 80–90, May 2016.
[18] S. Wang, A. Gittens, and M. W. Mahoney, "Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging," *J. Mach. Learn. Res.*, vol. 18, pp. 1–50, May 2018.
[19] Y. Yang, M. Pilanci, and M. J. Wainwright, "Randomized sketches for kernels: Fast and optimal nonparametric regression," *Ann. Statist.*, vol. 45, pp. 991–1023, Jun. 2017.
[20] A. Chowdhury, J. Yang, and P. Drineas, "An iterative, sketching-based framework for ridge regression," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2018, pp. 988–997.
[21] J. Lin and V. Cevher, "Optimal rates of sketched-regularized algorithms for least-squares regression over hilbert spaces," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2018, pp. 3102–3111.
[22] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*. Berlin, Germany: Springer Science & Business Media, vol. 23, 2008.
[23] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3920–3931, Dec. 2018.
[24] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmWave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.
[25] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, M. Debbah *et al.* "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, to be published, doi: 10.1109/TCOMM.2021.3063236.
[26] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, "Coherent matrix completion," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2014, pp. 674–682.
[27] W. Bi and J. T. Kwok, "Efficient multi-label classification with many labels," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2013, pp. 405–413.
[28] C. Saunders, A. Gammerman, and V. Vovk, "Ridge regression learning algorithm in dual variables," in *Proc. Int. Conf. Mach. Learn.*, Jul. 1998, pp. 515–521.
[29] T. K. Dijkstra, "Ridge regression and its degrees of freedom," *Qual. Quantity*, vol. 48, pp. 3185–3193, Nov. 2014.
[30] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, "Sampling algorithms for l2 regression and applications," in *Proc. ACM-SIAM Symp. Discrete Algorithm*, Jan. 2006, pp. 1127–1136.
[31] M. B. Cohen, Y. T. Lee, C. Musco, C. Musco, R. Peng, and A. Sidford, "Uniform sampling for matrix approximation," in *Proc. Conf. Innov. Theor. Comput. Sci.*, Jan. 2015, pp. 181–190.
[32] T. Sarlos, "Improved approximation algorithms for large matrices via random projections," in *Proc. IEEE Symp. Found. Comput. Sci.*, Oct. 2006, pp. 143–152.
[33] A. E. Alaoui and M. W. Mahoney, "Fast randomized Kernel ridge regression with statistical guarantees," in *Proc. Neural Inf. Process. Syst.*, Dec. 2015, pp. 775–783.
[34] M. Kapralov, Y. T. Lee, C. Musco, C. Musco, and A. Sidford, "Single pass spectral sparsification in dynamic streams," *SIAM J. Comput.*, vol. 46, no. 1, pp. 456–477, 2017.
[35] M. B. Cohen, J. Nelson, and D. P. Woodruff, "Optimal approximate matrix product in terms of stable rank," in *Proc. Int. Colloq. Automata, Lang., Program.*, vol. 55, pp. 1–14, Jul. 2016.
[36] M. B. Cohen, C. Musco, and C. Musco, "Input sparsity time low-rank approximation via ridge leverage score sampling," in *Proc. 28th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2017, pp. 1758–1777.
[37] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.
[38] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.
[39] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.
[40] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-efficient edge AI: Algorithms and systems," *IEEE Commun. Surv. Tut.*, vol. 22, no. 4, pp. 2167–2191, Jul. 2020.
[41] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
[42] Y. Shi, J. Zhang, W. Chen, and K. B. Letaief, "Enhanced group sparse beamforming for green cloud-RAN: A random matrix approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2511–2524, Apr. 2018.

Hayoung Choi (Member, IEEE) received the B.S. and M.S. degrees from the Department of Mathematics, Kyungpook National University, Deagu, South Korea, in 2006 and 2012, respectively, and the Ph.D. degree from the Department of Mathematics, University of Wyoming, Laramie, WY, USA, in 2015. He was a Postdoctoral Researcher with the School of Information Science and Technology, ShanghaiTech University in China, Shanghai, China. He is currently an Assistant Professor with the Department of Mathematics, Kyungpook National University, Daegu, South Korea. His research interests include matrix analysis, numerical linear algebra, large-scale data analysis, and applications of these topics to problems in machine learning, data mining, and communications.

Tao Jiang (Student Member, IEEE) received the B.S. degree in communication engineering from Xidian University, Xi'an, China, in 2017 and the M.S. degree in computer science from ShanghaiTech University, Shanghai, China, in 2020. His main research interests include optimization theory, machine learning, and wireless communications.



Yuanming Shi (Senior Member, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2011 and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2015. Since September 2015, he has been with the School of Information Science and Technology in ShanghaiTech University, Shanghai, China, where he is currently a tenured Associate Professor. From October 2016 to February 2017, he visited University of California, Berkeley, Berkeley, CA, USA. His research interests include optimization, statistics, machine learning, signal processing, and their applications to 6 G, IoT, and AI. He was the recipient of the 2016 IEEE Marconi Prize Paper Award in Wireless Communications, and the 2016 Young Author Best Paper Award by the IEEE Signal Processing Society. He is the Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

Xuan Liu (Member, IEEE) received the B.S. degree in telecommunication from Jilin University, Changchun, China, in 2010, the M.S. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2012, and the Ph.D. degree in electronic and information engineering from The University of Sydney, Sydney, NSW, Australia, in 2017. She is currently a Senior Research Associate with the School of Electrical Engineering and Telecommunications, University of New South Wales, Kensington, NSW, Australia. Her research interests include network information theory, large-scale optimization, big data analysis, and Internet of Things.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.

Yong Zhou (Member, IEEE) received the B.Sc. and M.Eng. degrees from Shandong University, Jinan, China, in 2008 and 2011, respectively, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2015. From November 2015 to January 2018, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada. He is currently an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. His research interests include Internet of Things, machine learning, and reconfigurable intelligent surface.

Khaled B. Letaief (Fellow, IEEE) received the B.S. degree with distinction in electrical engineering, and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in December 1984, August 1986, and May 1990, respectively. From 1990 to 1993, he was a Faculty Member with the University of Melbourne, Melbourne, VIC, Australia. Since 1993, he has been with the Hong Kong University of Science & Technology (HKUST), Hong Kong. He held many administrative positions, including the Dean of Engineering, Head of the Electronic and Computer Engineering Department, Director of the Wireless IC Design Center, founding Director of Huawei Innovation Laboratory, and Director of the Hong Kong Telecom Institute of Information Technology with HKUST. He has authored or coauthored more than 630 papers with more than 38600 citations and an h-index of 87 along with 15 patents, including 11 US inventions in his research fields. He is currently an internationally recognized Leader of wireless communications and networks, and his research interests include artificial intelligence, big data analytics systems, mobile cloud and edge computing, tactile Internet, 5G systems and beyond.

He is a Member of the United States National Academy of Engineering, Fellow of Hong Kong Institution of Engineers, Hong Kong, and Member of the Hong Kong Academy of Engineering Sciences, Hong Kong. He is also recognized by Thomson Reuters as an ISI Highly Cited Researcher and was listed among the 2020 top 30 of AI 2000 Internet of Things Most Influential Scholars.

He was the recipient of many distinguished awards and honors including the 2019 Distinguished Research Excellence Award by HKUST School of Engineering (Highest research award and only one recipient or three years is honored for his or her contributions), the 2019 IEEE Communications Society and Information Theory Society Joint Paper Award, the 2018 IEEE Signal Processing Society Young Author Best Paper Award, the 2017 IEEE Cognitive Networks Technical Committee Publication Award, the 2016 IEEE Signal Processing Society Young Author Best Paper Award, the 2016 IEEE Marconi Prize Paper Award in Wireless Communications, the 2011 IEEE Wireless Communications Technical Committee Recognition Award, the 2011 IEEE Communications Society Harold Sobol Award, the 2010 Purdue University Outstanding Electrical and Computer Engineer Award, the 2009 IEEE Marconi Prize Award in Wireless Communications, the 2007 IEEE Communications Society Joseph LoCicero Publications Exemplary Award, and more than 16 IEEE Best Paper Awards.

He was a Consultant for different organizations including Huawei, ASTRI, ZTE, Nortel, PricewaterhouseCoopers, and Motorola. He is the founding Editor-in-Chief of the prestigious IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and was the Editor-in-Chief on the Editorial Board of other premier journals including the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS - WIRELESS SERIES. He was also involved in organizing many flagship international conferences. He is well recognized for his dedicated service to professional societies and IEEE, where he was in many leadership positions. These include Treasurer of the IEEE Communications Society, IEEE Communications Society Vice-President for Conferences, the Chair of IEEE Committee on Wireless Communications, elected Member of IEEE Product Services and Publications Board, and IEEE Communications Society Vice-President for Technical Activities. From 2018 to 2019, he was also the President of the IEEE Communications Society, the world's leading organization for communications professionals with headquarter in New York City and a Member in 162 countries.

Authorized licensed use limited to: ShanghaiTech University. Downloaded on August 03,2021 at 04:01:33 UTC from IEEE Xplore. Restrictions apply.