

# Hybrid Precoding for Millimeter Wave MIMO Systems With Finite-Alphabet Inputs

Juening Jin\*, Yahong Rosa Zheng†, Wen Chen\* and Chengshan Xiao§

\*Department of Electronic Engineering, Shanghai Jiao Tong University

Shanghai, 200240, China, Email: jueningjin@gmail.com, wenchen@sjtu.edu.cn

†Department of Electrical and Computer Engineering, Missouri University of Science and Technology  
Rolla, MO 65409, USA, Email: zhengyr@mst.edu

§Department of Electrical and Computer Engineering, Lehigh University  
Bethlehem, PA 18015, USA, Email: chx417@lehigh.edu

**Abstract**—This paper investigates the hybrid precoding design for millimeter wave (mmWave) multiple-input multiple-output (MIMO) systems with finite alphabet inputs. The precoding problem is a joint optimization of analog and digital precoders, and it imposes nonconvex constant modulus constraints on the analog precoder. We treat this problem as a matrix factorization problem with constant modulus constraints. The main contributions of our work are listed as follows: First, we propose sufficient and necessary conditions for hybrid precoding schemes to realize any unconstrained optimal precoders exactly when the number of data streams is equal to the number of radio frequency chains. Second, we show that the power constraint in the hybrid precoding problem can be removed without loss of optimality. Third, we present a trust region Newton method to solve our problem using gradient and Hessian information, and the proposed algorithm converges to a stationary point satisfying the first and second order necessary optimality conditions. Several numerical examples are provided to show that the proposed algorithm outperforms existing hybrid precoding algorithms.

## I. INTRODUCTION

Millimeter wave (mmWave) multiple-input multiple-output (MIMO) communication with linear precoding is a promising technique for future generation wireless communication systems. For conventional MIMO systems, linear precoding techniques are implemented in the digital domain by fully digital precoders. Computing an optimal fully digital precoder is a well understood procedure, and the solution is known to be the unconstrained optimal precoder. However, it is infeasible to employ unconstrained optimal precoders directly at mmWave frequencies due to several hardware constraints [1]. To address this issue, a hybrid precoding scheme has been proposed for mmWave MIMO systems [2]–[7]. This scheme divides the linear precoder into analog and digital precoders, which are implemented in analog and digital domains, respectively. The analog precoder is realized by phase shifters, thus its elements satisfy constant modulus constraints. These nonconvex constant modulus constraints form a major challenge for hybrid precoding design, and several hybrid precoding algorithms have been proposed to tackle this problem.

So far, most existing works on hybrid precoding [2]–[6] assume ideal Gaussian inputs, which are rarely realized in practice. It is well known that practical systems utilize finite-alphabet inputs, such as phase-shift keying (PSK) or

quadrature amplitude modulation (QAM). Furthermore, it has been shown that linear precoding under Gaussian inputs are quite suboptimal for practical systems with finite-alphabet inputs [8]–[12]. Recently, the authors in [7] proposed an iterative gradient descent algorithm for single user mmWave MIMO systems with finite-alphabet inputs. The proposed gradient descent algorithm can achieve up to 0.4 bits/s/Hz gains compared to the Gaussian inputs scenario.

In this paper, we study the hybrid precoding design for mmWave MIMO systems from the matrix factorization perspective. The contributions of this paper are summarized as follows:

- We first provide a sufficient condition under which hybrid precoding schemes can realize any unconstrained optimal precoders exactly. When the sufficient condition does not hold, we also present a necessary condition for hybrid precoding to achieve the performance of unconstrained optimal precoders.
- We prove that the coupled power constraint in the hybrid precoding problem can be removed without loss of local and global optimality. This result greatly simplifies the precoding design, and it provides a theoretical guarantee for previous hybrid precoding algorithms that drop the power constraint directly without any proof.
- We derive closed form expressions for gradient and Hessian of the hybrid precoding problem. Then we utilize these information to design a trust region Newton algorithm. The proposed algorithm converges to a stationary point satisfying the first and second order necessary optimality conditions.

*Notations:* Boldface lowercase letters, boldface uppercase letters, and calligraphic letters are used to denote vectors, matrices and sets, respectively. The real and complex number fields are denoted by  $\mathbb{R}$  and  $\mathbb{C}$ , respectively. The superscripts  $(\cdot)^T$ ,  $(\cdot)^*$  and  $(\cdot)^H$  stand for transpose, conjugate, and conjugate transpose operations, respectively.  $\text{tr}(\cdot)$  is the trace of a matrix;  $\|\cdot\|$  denotes the Euclidean norm of a vector;  $\|\cdot\|_F$  represents the Frobenius norm of a matrix;  $E_x(\cdot)$  represents the statistical expectation with respect to  $x$ ;  $\mathbf{I}$  and  $\mathbf{0}$  denote an identity matrix and a zero matrix, respectively,

with appropriate dimensions;  $\otimes$  and  $\circ$  are Kronecker and Hadamard matrix products, respectively;  $\mathcal{I}(\cdot)$  represents the mutual information;  $\Re$  and  $\Im$  are the real and image parts of a complex value;  $\log(\cdot)$  is used for the base two logarithm.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. System Model

Consider a point-to-point mmWave MIMO system, where a transmitter with  $N_T$  antennas sends  $N_S$  data streams to a receiver with  $N_R$  antennas. The number of RF chains at the transmitter is  $N_{RF}$ , which satisfies  $N_S \leq N_{RF} \leq N_T$ . The received baseband signal  $\mathbf{y} \in \mathbb{C}^{N_R \times 1}$  can then be written as

$$\mathbf{y} = \mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{x} + \mathbf{n} \quad (1)$$

where  $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$  is the mmWave channel matrix;  $\mathbf{F}_{RF} \in \mathbb{C}^{N_T \times N_{RF}}$  is the analog precoder satisfying  $|[\mathbf{F}_{RF}]_{kl}| = \frac{1}{\sqrt{N_T}}, \forall (k, l)$ ;  $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_S}$  is the digital precoder;  $\mathbf{x} \in \mathbb{C}^{N_S \times 1}$  is the input data vector and  $\mathbf{n} \in \mathbb{C}^{N_R \times 1}$  is the independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian noise with zero-mean and covariance  $\sigma^2 \mathbf{I}$ .

The input data vector  $\mathbf{x}$  is uniformly distributed from a given constellation set with cardinality  $M$ . Suppose that the channel  $\mathbf{H}$  is known at both the transmitter and receiver, then the input-output mutual information is given by [10]

$$\mathcal{I}(\mathbf{x}; \mathbf{y}) = \log K - \frac{1}{K} \sum_{m=1}^K E_{\mathbf{n}} \left\{ \log \sum_{k=1}^K \exp \left( - \frac{\|\mathbf{H}\mathbf{F}\mathbf{e}_{mk} + \mathbf{n}\|^2 - \|\mathbf{n}\|^2}{\sigma^2} \right) \right\} \quad (2)$$

where  $\mathbf{F} = \mathbf{F}_{RF}\mathbf{F}_{BB}$ ,  $K = M^{N_S}$  is a constant, and  $\mathbf{e}_{mk}$  is the difference between  $\mathbf{x}_m$  and  $\mathbf{x}_k$ , with  $\mathbf{x}_m$  and  $\mathbf{x}_k$  being two possible distinct data vectors from  $\mathbf{x}$ .

### B. Channel Model

The mmWave MIMO channel can be characterized by standard multipath models. Suppose the number of physical paths between the transmitter and the receiver is  $L$ . Each path  $l$  is described by three parameters: complex gain  $\alpha_l$ , angle of arrival  $\theta_{R,l}$  and angle of departure  $\theta_{T,l}$ . Under this model, the channel matrix  $\mathbf{H}$  is given by

$$\mathbf{H} = \sqrt{\frac{N_R N_T}{L}} \sum_{l=1}^L \alpha_l \mathbf{a}(\theta_{R,l}) \mathbf{a}(\theta_{T,l})^H \quad (3)$$

where  $\{\theta_{R,l}\}$  and  $\{\theta_{T,l}\}$  are i.i.d. uniformly distributed over  $[0, 2\pi]$ , and  $\{\alpha_l\}$  are i.i.d. complex Gaussian distributed with zero-mean and unit-variance.  $\mathbf{a}(\theta_{T,l})$  and  $\mathbf{a}(\theta_{R,l})$  are array steering vectors of the transmit and receive antenna arrays, respectively. In this paper, the transmitter and receiver adopt uniform linear arrays, and the array steering vector  $\mathbf{a}(\theta)$  is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \left[ 1, e^{-j \frac{2\pi}{\lambda} d \sin \theta}, \dots, e^{-j \frac{2\pi}{\lambda} d(N-1) \sin \theta} \right]^T \quad (4)$$

where  $N$  is the number of antenna element,  $\lambda$  is the wavelength of operation and  $d = \frac{\lambda}{2}$  is the antenna spacing.

### C. Problem Formulation

A fundamental approach to hybrid precoding is to maximize the input-output mutual information under the power and constant modulus constraints. Suppose that the receiver performs ideal decoding, then the hybrid precoding problem is formulated as

$$\begin{aligned} & \underset{\mathbf{F}_{RF} \in \mathcal{U}, \mathbf{F}_{BB}}{\text{maximize}} && \mathcal{I}(\mathbf{x}; \mathbf{y}) \\ & \text{subject to} && \text{tr}(\mathbf{F}_{BB}^H \mathbf{F}_{RF}^H \mathbf{F}_{RF} \mathbf{F}_{BB}) \leq P \end{aligned} \quad (5)$$

where  $\mathcal{I}(\mathbf{x}; \mathbf{y})$  is given in (2),  $P$  is the power budget at the transmitter and  $\mathcal{U}$  is the feasible set of analog precoders, i.e.,  $\mathcal{U} = \{ \mathbf{F}_{RF} : |[\mathbf{F}_{RF}]_{kl}| = \frac{1}{\sqrt{N_T}}, \forall (k, l) \}$ . It is challenging to solve problem (5) directly due to two reasons: First, problem (5) is nonconvex because both  $\mathcal{I}(\mathbf{x}; \mathbf{y})$  and  $\mathcal{U}$  are neither convex nor concave with respect to  $(\mathbf{F}_{RF}, \mathbf{F}_{BB})$ . Second, iterative algorithms for problem (5) have to evaluate the objective function  $\mathcal{I}(\mathbf{x}; \mathbf{y})$  in each iteration, which can be very costly because  $\mathcal{I}(\mathbf{x}; \mathbf{y})$  has no closed form expressions.

To mitigate these difficulties and simplify the precoding design, we adopt the following nonlinear least square formulation [2], where hybrid precoders  $(\mathbf{F}_{RF}, \mathbf{F}_{BB})$  are found by approximating the unconstrained optimal precoder  $\mathbf{F}_{opt}$ , i.e.,

$$\begin{aligned} & \underset{\mathbf{F}_{RF} \in \mathcal{U}, \mathbf{F}_{BB}}{\text{minimize}} && \|\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 \\ & \text{subject to} && \text{tr}(\mathbf{F}_{BB}^H \mathbf{F}_{RF}^H \mathbf{F}_{RF} \mathbf{F}_{BB}) \leq P. \end{aligned} \quad (6)$$

The unconstrained optimal precoder is given by [10], [11]

$$\mathbf{F}_{opt} = \underset{\mathbf{F} \in \mathcal{F}}{\text{argmax}} \mathcal{I}(\mathbf{x}; \mathbf{y}) \quad (7)$$

where  $\mathcal{F} = \{ \mathbf{F} : \text{tr}(\mathbf{F}^H \mathbf{F}) \leq P \}$  is the feasible set.

## III. STRUCTURES OF THE HYBRID PRECODING PROBLEM

In this section, we first present sufficient and necessary conditions, under which hybrid precoding schemes can realize any unconstrained optimal precoder exactly. Then we prove that the power constraint  $\text{tr}(\mathbf{F}_{BB}^H \mathbf{F}_{RF}^H \mathbf{F}_{RF} \mathbf{F}_{BB}) \leq P$  in problem (6) can be removed without loss of local and global optimality.

### A. Optimality of Hybrid Precoding Schemes

The hybrid precoding scheme offers a compromise between performance gain and hardware complexity, and its performance is bounded by the unconstrained optimal precoder in (7). When the hybrid precoding scheme can realize any unconstrained optimal precoder exactly, it is an *optimal* scheme. Then a fundamental question arises:

- Question 1: under what conditions can hybrid precoding schemes realize any unconstrained optimal precoders?

In other words, we want to find necessary and/or sufficient conditions, under which there exist  $(\mathbf{F}_{RF}, \mathbf{F}_{BB})$  such that  $\mathbf{F}_{RF} \in \mathcal{U}$  and  $\mathbf{F}_{opt} = \mathbf{F}_{RF}\mathbf{F}_{BB}$ . The best known result related to this question was shown in [3], [5]. It says that when the number of data streams is restricted to be lower than  $\frac{1}{2}N_{RF}$ , we can construct analog and digital precoders to realize any unconstrained optimal precoder with dimensions  $N_T \times N_S$ . However, this result sacrifices the number of data streams

to satisfy  $\mathbf{F}_{\text{opt}} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ . In order to achieve the maximum mutual information, we should transmit  $N_{\text{RF}}$  data streams rather than  $\frac{1}{2}N_{\text{RF}}$  data streams. This motivates us to reconsider Question 1 under  $N_{\text{S}} = N_{\text{RF}}$ .

First, we transform Question 1 into another existence problem through the following proposition.

*Proposition 1:* Suppose  $\mathbf{F}_{\text{RF}}$  is a full rank matrix, then the following two statements are equivalent:

- 1) There exists  $(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}})$  such that  $\mathbf{F}_{\text{RF}} \in \mathcal{U}$  and  $\mathbf{F}_{\text{opt}} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}$ .
- 2) There exists a full rank square matrix  $\mathbf{S} \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{RF}}}$  such that  $\mathbf{U}_{\text{F}}\mathbf{S} \in \mathcal{U}$ .

Here  $\mathbf{U}_{\text{F}} \in \mathbb{C}^{N_{\text{T}} \times N_{\text{RF}}}$  is a unitary matrix with left singular vectors of  $\mathbf{F}_{\text{opt}}$ .

*Proof:* The proof is omitted due to space limits. ■

The existence problem of  $\mathbf{U}_{\text{F}}\mathbf{S} \in \mathcal{U}$  depends on  $\mathbf{U}_{\text{F}}$  which has a close connection with the channel matrix  $\mathbf{H}$ . Let the singular value decomposition (SVD) of  $\mathbf{H}$  be

$$\mathbf{H} = \mathbf{U}_{\text{H}}\Sigma_{\text{H}}\mathbf{V}_{\text{H}}^H \quad (8)$$

where  $\mathbf{U}_{\text{H}} \in \mathbb{C}^{N_{\text{R}} \times N}$  is a unitary matrix with left singular vectors,  $\Sigma_{\text{H}} \in \mathbb{C}^{N \times N}$  is a diagonal matrix with singular values arranged in decreasing order,  $\mathbf{V}_{\text{H}} \in \mathbb{C}^{N_{\text{T}} \times N}$  is a unitary matrix with right singular vectors, and  $N = \max\{\text{rank}(\mathbf{H}), N_{\text{RF}}\}$ . According to [10, Proposition 2],  $\mathbf{U}_{\text{F}}$  can always be chosen as the first  $N_{\text{RF}}$  columns of  $\mathbf{V}_{\text{H}}$ , i.e.,

$$\mathbf{U}_{\text{F}} = \mathbf{V}_{\text{H}}(:, 1:N_{\text{RF}}). \quad (9)$$

Based on this observation, we obtain a sufficient condition for hybrid precoding schemes to realize the unconstrained optimal precoders.

*Proposition 2:* When the number of paths  $L = N_{\text{RF}} = N_{\text{S}} \leq \min(N_{\text{R}}, N_{\text{T}})$ , there exists a full rank matrix  $\mathbf{S}$  such that  $\mathbf{U}_{\text{F}}\mathbf{S} \in \mathcal{U}$ .

*Proof:* See Appendix A. ■

Combining Propositions 1 and 2, we conclude that hybrid precoding can achieve the performance of unconstrained optimal precoders if the following condition holds:

$$L = N_{\text{RF}} = N_{\text{S}} \leq \min(N_{\text{R}}, N_{\text{T}}). \quad (10)$$

However, condition (10) does not always hold. In the rest of this subsection, we will investigate the existence problem of  $\mathbf{U}_{\text{F}}\mathbf{S} \in \mathcal{U}$  under general scenarios.

First, we rewrite  $\mathbf{U}_{\text{F}}\mathbf{S} \in \mathcal{U}$  as

$$|\mathbf{u}_k^H \mathbf{s}_l| = \frac{1}{\sqrt{N_{\text{T}}}}, \quad k = 1, \dots, N_{\text{T}}, l = 1, \dots, N_{\text{RF}}. \quad (11)$$

where  $\mathbf{u}_k^H$  is the  $k$ -th row of  $\mathbf{U}_{\text{F}}$ , and  $\mathbf{s}_l$  is the  $l$ -th column of  $\mathbf{S}$ . Combining equation (11) and  $\text{rank}(\mathbf{S}) = N_{\text{RF}}$ , our problem is equivalent to finding  $N_{\text{RF}}$  linear independent solutions  $\{\mathbf{s}_l\}$  to the following system of quadratic equations:

$$|\mathbf{u}_k^H \mathbf{s}| = \frac{1}{\sqrt{N_{\text{T}}}}, \quad k = 1, \dots, N_{\text{T}}. \quad (12)$$

Unfortunately, problem (12) is intractable because checking the existence of solutions to a general system of quadratic

equations is nondeterministic polynomial time (NP)-hard [13]. Instead, we investigate necessary conditions for the existence of solutions to problem (12).

The main idea is to transform the quadratic system (12) into a linear system by semidefinite programming. Define  $\mathbf{Z} = \mathbf{s}\mathbf{s}^H$  and  $\mathbf{C}_k = \mathbf{u}_k \mathbf{u}_k^H$ , then our problem is equivalent to finding  $N_{\text{RF}}$  linear independent solutions to the following linear system with a nonlinear equality constraint:

$$\text{tr}(\mathbf{C}_k \mathbf{Z}) = \frac{1}{N_{\text{T}}}, \quad k = 1, \dots, N_{\text{T}}, \quad \mathbf{Z} = \mathbf{s}\mathbf{s}^H. \quad (13)$$

To rewrite (13) in standard forms,  $\mathbf{C}_k$  and  $\mathbf{Z}$  are vectorized based on the following results:

$$\text{tr}(\mathbf{C}_k \mathbf{Z}) = \text{vec}(\mathbf{C}_k^T)^T \text{vec}(\mathbf{Z}) \quad (14)$$

$$\text{vec}(\mathbf{s}\mathbf{s}^H) = \mathbf{s} \otimes \mathbf{s}^*. \quad (15)$$

Then problem (13) can be expressed more compactly as

$$\mathbf{C}^T \mathbf{z} = \mathbf{1}, \quad \mathbf{z} = \mathbf{s} \otimes \mathbf{s}^* \quad (16)$$

where  $\mathbf{z} = \text{vec}(\mathbf{Z})$  and  $\mathbf{C}$  is given by

$$\mathbf{C} = N_{\text{T}} \cdot [\text{vec}(\mathbf{C}_1^T), \dots, \text{vec}(\mathbf{C}_{N_{\text{T}}}^T)]. \quad (17)$$

The main barrier for solving problem (16) is the nonlinear constraint  $\mathbf{z} = \mathbf{s} \otimes \mathbf{s}^*$ , which restricts solutions of  $\mathbf{C}^T \mathbf{z} = \mathbf{1}$  with a certain structure. Therefore, we first ignore  $\mathbf{z} = \mathbf{s} \otimes \mathbf{s}^*$  and focus on the linear system  $\mathbf{C}^T \mathbf{z} = \mathbf{1}$ . Clearly, if problem (16) has  $N_{\text{RF}}$  linear independent solutions, then  $\mathbf{C}^T \mathbf{z} = \mathbf{1}$  should have at least  $N_{\text{RF}}$  linear independent solutions. Based on this observation, the following proposition provides a necessary condition for the existence of a full rank  $\mathbf{S}$  such that  $\mathbf{U}_{\text{F}}\mathbf{S} \in \mathcal{U}$ .

*Proposition 3:* If there exist a full rank square matrix  $\mathbf{S}$  satisfying  $\mathbf{U}_{\text{F}}\mathbf{S} \in \mathcal{U}$ , then we have

- 1)  $||\mathbf{s}_l||^2 = 1, \quad l = 1, 2, \dots, N_{\text{RF}}$
- 2)  $\text{rank}(\mathbf{C}) \leq N_{\text{RF}}^2 - N_{\text{RF}} + 1$

where  $\mathbf{C}$  is a matrix function of  $\mathbf{U}_{\text{F}}$ .

*Proof:* The proof is omitted due to space limits. ■

When the number of paths  $L > \min(N_{\text{R}}, N_{\text{T}})$ ,  $\mathbf{C}$  is usually a full rank matrix. In this case, we derive the minimum number of RF chains needed for hybrid precoding to achieve the performance of unconstrained optimal precoders.

*Corollary 1:* When  $\mathbf{C} \in \mathbb{C}^{N_{\text{RF}}^2 \times N_{\text{T}}}$  is a full rank matrix, we require at least  $\sqrt{N_{\text{T}} - \frac{3}{4} + \frac{1}{2}}$  RF chains for hybrid precoding schemes to realize any unconstrained optimal precoder exactly.

*Proof:* The proof is omitted due to space limits. ■

## B. Structures of Problem (6)

The hybrid precoding problem (6) is a nonconvex problem, and theoretical challenges of problem (6) are listed as follows:

- 1) The optimization variables  $\mathbf{F}_{\text{RF}}$  and  $\mathbf{F}_{\text{BB}}$  are coupled through the power constraint. Therefore, we cannot deploy the alternating minimization approach which requires separate variables in constraints. If we jointly optimize  $(\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}})$ , the difficulty also lies in handing the intersection of power and constant modulus constraints.

- 2) The bilinear mapping  $(\mathbf{F}_{RF}, \mathbf{F}_{BB}) \mapsto \mathbf{F}_{RF}\mathbf{F}_{BB}$  is not a one-to-one mapping, thus  $(\mathbf{F}_{RF}, \mathbf{F}_{BB})$  and  $(\mathbf{F}_{RF}\Sigma, \Sigma^{-1}\mathbf{F}_{BB})$  result in the same objective value, where  $\Sigma$  is a diagonal matrix with unit modulus diagonal entries to ensure  $\mathbf{F}_{RF}\Sigma \in \mathcal{U}$ . In other words, we should expect problem (6) to have infinite number of local minima and saddle points.

Existing works [2], [4], [6] address the first issue by simply dropping the power constraint, and then solve the following relaxed problem:

$$\underset{\mathbf{F}_{RF} \in \mathcal{U}, \mathbf{F}_{BB}}{\text{minimize}} \quad \|\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2. \quad (18)$$

Since  $\mathbf{F}_{RF}$  and  $\mathbf{F}_{BB}$  are separate in constraints, problem (18) can be solved by standard optimization techniques, e.g., the alternating minimization approach. After obtaining a high quality solution  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  to problem (18), the normalization

$$\hat{\mathbf{F}}_{BB} := \sqrt{P} \cdot \frac{\hat{\mathbf{F}}_{BB}}{\|\hat{\mathbf{F}}_{RF}\hat{\mathbf{F}}_{BB}\|_F} \quad (19)$$

is utilized to satisfy the power constraint. However, the corresponding solution after normalization may not be a high quality solution to problem (6). To the best of our knowledge, the relationship between problems (6) and (18) has not been fully understood in previous works. In this paper, we fully address this issue by proving the equivalence between problems (6) and (18).

*Theorem 1:* If  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  is a KKT point (or globally optimal solution) of problem (18), then it is also a KKT point (or globally optimal solution) of problem (6). ■

*Proof:* See Appendix A. ■

According to Theorem 1, any KKT point of problem (18) satisfies  $\text{tr}(\mathbf{F}_{BB}^H \mathbf{F}_{RF}^H \mathbf{F}_{RF} \mathbf{F}_{BB}) \leq P$ , thus the power constraint can be removed without loss of local and global optimality.

Problem (18) is a constant modulus matrix factorization problem where a given matrix  $\mathbf{F}_{opt}$  is factorized into two complex matrices  $(\mathbf{F}_{RF}, \mathbf{F}_{BB})$  under constant modulus constraints on  $\mathbf{F}_{RF}$ . Based on the previous analysis, problem (18) has infinite number of saddle points, and we will design an efficient algorithm that can address this issue in Section IV.

#### IV. CONSTANT MODULUS MATRIX FACTORIZATION

First, we observe that for any given  $\mathbf{F}_{RF}$ , problem (18) is a least square problem

$$\underset{\mathbf{F}_{BB}}{\text{minimize}} \quad \|\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2. \quad (20)$$

Suppose that  $\mathbf{F}_{RF}$  has full column rank, then the optimal solution to problem (20) is

$$\mathbf{F}_{BB} = \mathbf{F}_{RF}^+ \mathbf{F}_{opt} \quad (21)$$

where  $\mathbf{F}_{RF}^+ = (\mathbf{F}_{RF}^H \mathbf{F}_{RF})^{-1} \mathbf{F}_{RF}^H$  is the Moore-Penrose pseudoinverse of  $\mathbf{F}_{RF}$ . Inserting (21) into problem (18),  $\mathbf{F}_{BB}$  is eliminated and we obtain the modified problem:

$$\underset{\mathbf{F}_{RF} \in \mathcal{U}}{\text{minimize}} \quad f(\mathbf{F}_{RF}) = \|\mathbf{F}_{opt} - \mathbf{F}_{RF}\mathbf{F}_{RF}^+ \mathbf{F}_{opt}\|_F^2 \quad (22)$$

This elimination technique is called the variable projection method, which was proposed in [14] to solve unconstrained problems in real field. The following theorem guarantees that problems (18) and (22) are equivalent.

*Theorem 2:* If  $\hat{\mathbf{F}}_{RF}$  is a KKT point of problem (22) and  $\hat{\mathbf{F}}_{BB} = \hat{\mathbf{F}}_{RF}^+ \mathbf{F}_{opt}$ , then  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  is a KKT point of problem (18). Furthermore,  $\hat{\mathbf{F}}_{RF}$  is a globally optimal solution of problem (22) if and only if  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  is a globally optimal solution of problem (18).

*Proof:* The proof is omitted due to space limits. ■

The variable projection technique reduces the search space from  $(\mathbf{F}_{RF}, \mathbf{F}_{BB})$  to  $\mathbf{F}_{RF}$ , thus improving efficiency and possibly reducing the number of suboptimal KKT points. As a result, problem (22) is more likely to converge to a globally optimal solution than problem (18).

Problem (22) can be further transformed into a unconstrained problem via change of optimization variables. Note that nonconvex constant modulus constraints imply that we can only change the phase of the analog precoder  $\mathbf{F}_{RF}$ . Therefore, instead of using  $\mathbf{F}_{RF}$  as the optimization variable, it is more convenient to optimize the phase of  $\mathbf{F}_{RF}$  directly. Let the phase of  $\mathbf{F}_{RF}$  be  $\Phi$ , i.e.,  $[\mathbf{F}_{RF}]_{kl} = \frac{1}{\sqrt{N_T}} e^{j[\Phi]_{kl}}$ . Using  $\Phi$  as the optimization variable and rewriting  $\mathbf{F}_{RF}$  as  $\mathbf{F}_{RF}(\Phi)$ , we can drop the constant modulus constraints  $\mathcal{U}$  and reformulate (22) as the following unconstrained minimization problem:

$$\underset{\Phi}{\text{minimize}} \quad \varphi(\Phi) = \|\mathbf{F}_{opt} - \mathbf{F}_{RF}(\Phi)\mathbf{F}_{RF}^+(\Phi)\mathbf{F}_{opt}\|_F^2. \quad (23)$$

The gradient and Hessian matrices of  $\varphi(\Phi)$  are given in the following theorem.

*Theorem 3:* The gradient and Hessian matrices of  $\varphi(\Phi)$  are given respectively as

$$\nabla \varphi(\Phi) = 2\Im[\mathbf{Z}_1 \mathbf{F}_{opt} \mathbf{Z}_2^H \circ \mathbf{F}_{RF}^*] \quad (24)$$

$$\nabla^2 \varphi(\Phi) = 2\Re[\text{diag}(\text{vec}[\mathbf{Z}_1 \mathbf{F}_{opt} \mathbf{Z}_2^H \circ \mathbf{F}_{RF}^*]) - \mathbf{M}] \quad (25)$$

where  $\mathbf{Z}_1 = \mathbf{F}_{RF}\mathbf{F}_{RF}^+ - \mathbf{I}$ ,  $\mathbf{Z}_2 = \mathbf{F}_{RF}^+ \mathbf{F}_{opt}$ , and  $\mathbf{M}$  is given by

$$\begin{aligned} \mathbf{M} = & \mathbf{D}^* \left[ (\mathbf{Z}_2 \mathbf{Z}_2^H)^T \otimes \mathbf{Z}_1 + (\mathbf{F}_{RF}^H \mathbf{F}_{RF})^{-T} \otimes \mathbf{Z}_1 \mathbf{F}_{opt} \mathbf{F}_{opt}^H \mathbf{Z}_1^H \right] \mathbf{D} \\ & - \mathbf{D}^* [\mathbf{E} + \mathbf{E}^T] \mathbf{D}^*. \end{aligned} \quad (26)$$

Here  $\mathbf{E} = (\mathbf{Z}_1 \mathbf{F}_{opt} \mathbf{Z}_2^H)^T \otimes (\mathbf{F}_{RF}^+)^H \mathbf{K}_{N_T, N_{RF}}$  with  $\mathbf{K}_{N_T, N_{RF}}$  being the commutation matrix, and  $\mathbf{D} = \text{diag}[\text{vec}(\mathbf{F}_{RF})]$ .

*Proof:* The proof is omitted due to space limits. ■

With the gradient and Hessian matrices at hand, we propose a trust region Newton method to solve problem (23). At each iteration of the trust region Newton method for minimizing  $\varphi(\Phi)$ , we have an iterate  $\Phi_k$ , a radius  $\Delta_k$  of the trust region, and a quadratic model

$$q(\mathbf{S} | \Phi_k) = \mathbf{g}_k^T \text{vec}(\mathbf{S}) + \frac{1}{2} \text{vec}(\mathbf{S})^T [\nabla^2 \varphi(\Phi_k)] \text{vec}(\mathbf{S}) \quad (27)$$

where  $\mathbf{g}_k = \text{vec}[\nabla \varphi(\Phi_k)]$ , and  $q(\mathbf{S} | \Phi_k)$  serves as the approximation of the value  $\varphi(\Phi_k + \mathbf{S}) - \varphi(\Phi_k)$ . Next, we determine the step  $\mathbf{S}_k$  by solving the following trust region subproblem:

$$\mathbf{S}_k = \underset{\mathbf{S} \in \mathcal{S}}{\text{argmin}} \quad q(\mathbf{S} | \Phi_k) \quad (28)$$

where  $\mathcal{S} = \{\mathbf{S}: \|\mathbf{S}\|_F \leq \Delta_k\}$ . Then we update  $\Phi_k$  and  $\Delta_k$  by checking

$$\rho_k = \frac{\varphi(\Phi_k + \mathbf{S}_k) - \varphi(\Phi_k)}{q(\mathbf{S}_k | \Phi_k)} \quad (29)$$

where  $\rho_k$  is the ratio of the actual reduction in the function to the predicted reduction in the quadratic model. Updating rules for  $\Phi_k$  and  $\Delta_k$  are given as follows [15]:

$$\Phi_{k+1} = \begin{cases} \Phi_k + \mathbf{S}_k & \text{if } \rho_k > \eta_1 \\ \Phi_k & \text{if } \rho_k \leq \eta_1 \end{cases} \quad (30)$$

$$\Delta_{k+1} = \begin{cases} \gamma_1 \|\mathbf{S}_k\|_F & \text{if } \rho_k \leq \eta_1 \\ \Delta_k & \text{if } \eta_1 < \rho_k < \eta_2 \\ \max(\gamma_2 \|\mathbf{S}_k\|_F, \Delta_k) & \text{if } \rho_k \geq \eta_2 \end{cases} \quad (31)$$

where  $\eta_1 < \eta_2 < 1$  and  $0 < \gamma_1 < 1 < \gamma_2$  are pre-specified positive values. The details of our trust region algorithm are given in Algorithm 1.

#### Algorithm 1 Trust region algorithm for problem (23)

1. Given  $\mathbf{F}_{\text{opt}}$ . Set initial  $\Phi_0$  and  $\Delta_0$ .
2. For  $k = 0, 1, \dots$  (outer iterations)
  - If  $\|\nabla \varphi(\Phi_k)\|_F < \epsilon$ , stop.
  - Solve the trust region subproblem (28) to obtain  $\mathbf{S}_k$ .
  - Compute  $\rho_k$  via (29).
  - Update  $\Phi_k$  to  $\Phi_{k+1}$  according to (30).
  - Obtain  $\Delta_{k+1}$  according to (31).
3. Return  $\mathbf{F}_{\text{RF}} = \frac{1}{\sqrt{N_T}} \exp(j\Phi_{\text{opt}})$  and  $\mathbf{F}_{\text{BB}} = \mathbf{F}_{\text{RF}}^+ \mathbf{F}_{\text{opt}}$ .

In each iteration of Algorithm 1, a solution to the trust region subproblem (28) can be computed in polynomial time by the safeguarded root finding method [16]. Moreover, the convergence of Algorithm 1 is guaranteed by the following proposition.

*Proposition 4:* The solution obtained by Algorithm 1 satisfies the first and second order necessary optimality conditions, i.e.,  $\nabla \varphi(\Phi_{\text{opt}}) = \mathbf{0}$  and  $\nabla^2 \varphi(\Phi_{\text{opt}}) \succeq \mathbf{0}$ .

*Proof:* The proof follows directly from Theorem 4.7 of [16]. ■

## V. NUMERICAL RESULTS

In this section, we provide a numerical example to evaluate the performance of our proposed algorithm. The parameters in Algorithm 1 are set as

$$\eta_1 = 0.05, \eta_2 = 0.9, \gamma_1 = 0.25, \gamma_2 = 2.5. \quad (32)$$

Consider a transmitter with  $N_T = 32$  antennas and  $N_{\text{RF}} = 2$  RF chains sends  $N_S = 2$  data streams to a receiver with  $N_R = 8$  antennas. The number of paths  $L$  is set as 2. The input signal is drawn from QPSK modulation, and the signal-to-noise ratio (SNR) is defined as  $\text{SNR} = \frac{P}{\sigma^2}$ . Finally, the average mutual information is plotted versus SNR over 500 channel realizations.

Fig. 1 shows that our proposed algorithm outperforms the gradient descent (GD) algorithm with finite-alphabet inputs

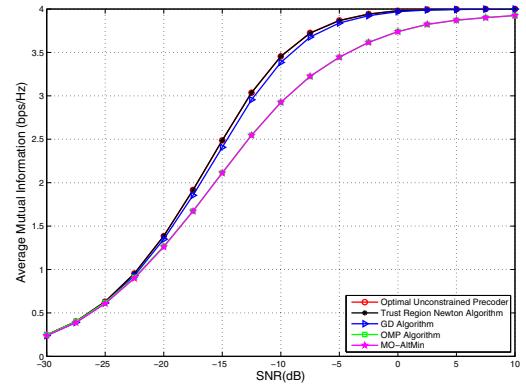


Fig. 1: Average mutual information versus SNR for different methods in a  $8 \times 32$  system with  $N_{\text{RF}} = N_S = 2$ .

[7], the orthogonal matching pursuit (OMP) algorithm with Gaussian inputs [2] and the MO-AltMin algorithm with Gaussian inputs in whole SNR regimes. Note that our proposed algorithm has about 5 dB gain as compared to the OMP and MO-AltMin algorithm, and it has about 0.1 bps/Hz improvement compared with the GD algorithm. Since mmWave provide very large bandwidths, a gain of 0.1 bps/Hz would translate to a large increase in the effective data rate. Moreover, our proposed algorithm achieves nearly the same performance as the unconstrained optimal precoder, and this result is in accord with Proposition 2.

## VI. CONCLUSION

This paper considers the hybrid precoding design for mmWave MIMO systems with finite-alphabet inputs. The precoding problem is formulated as a matrix factorization problem with constant modulus constraints. We first propose sufficient and necessary conditions for hybrid precoding to achieve the performance of unconstrained optimal precoders. Then we show that the power constraint can be removed without loss of optimality. Finally, we propose a trust region Newton method to solve the precoding problem. A numerical example is provided to show the performance gain of our proposed algorithm over existing algorithms.

## APPENDIX A PROOFS OF PROPOSITIONS 2 AND THEOREM 1

*Proof of Proposition 2:* We first rewrite  $\mathbf{H}$  in (3) in a more compact form as

$$\mathbf{H} = \mathbf{A}_{\text{R}} \text{diag}(\boldsymbol{\alpha}) \mathbf{A}_{\text{T}}^H \quad (33)$$

where  $\boldsymbol{\alpha} = \sqrt{\frac{N_{\text{R}} N_{\text{T}}}{L}} [\alpha_1, \dots, \alpha_L]^T$ ,  $\mathbf{A}_{\text{R}} \in \mathbb{C}^{N_{\text{R}} \times L}$  and  $\mathbf{A}_{\text{T}} \in \mathbb{C}^{N_{\text{T}} \times L}$  are array steering matrices with constant modulus entries, given by

$$\mathbf{A}_{\text{R}} = [\mathbf{a}(\theta_{\text{R},1}), \dots, \mathbf{a}(\theta_{\text{R},L})] \quad (34)$$

$$\mathbf{A}_{\text{T}} = [\mathbf{a}(\theta_{\text{T},1}), \dots, \mathbf{a}(\theta_{\text{T},L})]. \quad (35)$$

According to [2], when  $L = N_{\text{RF}} = N_S \leq \min(N_{\text{R}}, N_{\text{T}})$ ,  $\text{rank}(\mathbf{H}) = L$ . Therefore, the rows of  $\mathbf{A}_{\text{T}}^H$  form an orthogonal

basis of  $\mathcal{R}(\mathbf{H})$ , where  $\mathcal{R}(\cdot)$  represents the space spanned by rows of a matrix. Moreover, since  $\mathbf{H}$  can be expressed as  $\mathbf{H} = \mathbf{U}_\mathbf{H} \boldsymbol{\Sigma}_\mathbf{H} \mathbf{V}_\mathbf{H}^H$ , the rows of  $\mathbf{V}_\mathbf{H}^H$  also form a basis of  $\mathcal{R}(\mathbf{H})$ . Therefore, there exists a full rank square matrix  $\mathbf{S}$  such that

$$\mathbf{A}_T = \mathbf{V}_\mathbf{H} \mathbf{S} = \mathbf{U}_F \mathbf{S} \in \mathcal{U}. \quad (36)$$

where the second equality holds due to equation (9). This completes the proof. ■

*Proof of Theorem 1:* If  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  is a KKT point of problem (18), then it satisfies KKT conditions:

$$(\mathbf{F}_{opt} - \hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{BB}) \hat{\mathbf{F}}_{BB}^H = \boldsymbol{\Upsilon} \circ \hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB} = \hat{\mathbf{F}}_{RF}^+ \mathbf{F}_{opt}, \hat{\mathbf{F}}_{RF} \in \mathcal{U} \quad (37)$$

where  $\boldsymbol{\Upsilon}_{kl}$  is the lagrangian multiplier associated with constant modulus constraints. Inserting  $\hat{\mathbf{F}}_{BB} = \hat{\mathbf{F}}_{RF}^+ \mathbf{F}_{opt}$  into  $\text{tr}(\hat{\mathbf{F}}_{BB}^H \hat{\mathbf{F}}_{RF}^H \hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{BB})$ , we obtain

$$\text{tr}(\hat{\mathbf{F}}_{BB}^H \hat{\mathbf{F}}_{RF}^H \hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{BB}) = \text{tr}(\hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{RF}^+ \mathbf{F}_{opt} \mathbf{F}_{opt}^H) \quad (38)$$

$$\leq \sum_{i=1}^{N_T} \lambda_i(\hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{RF}^+) \lambda_i(\mathbf{F}_{opt} \mathbf{F}_{opt}^H) \quad (39)$$

where  $\lambda_i(\cdot)$  represents the eigenvalue of a Hermitian matrix in decreasing order. The inequality in (39) follows from [17, Lemma II.1]:

$$\sum_{i=1}^n \lambda_i(\mathbf{A}) \lambda_{n-i+1}(\mathbf{B}) \leq \text{tr}(\mathbf{AB}) \leq \sum_{i=1}^n \lambda_i(\mathbf{A}) \lambda_i(\mathbf{B}) \quad (40)$$

where  $\mathbf{A} \in \mathbb{C}^{n \times n}$  and  $\mathbf{B} \in \mathbb{C}^{n \times n}$  are Hermitian matrices. Let the singular value decomposition (SVD) of  $\hat{\mathbf{F}}_{RF}$  to be

$$\hat{\mathbf{F}}_{RF} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^H \quad (41)$$

where  $\mathbf{U} \in \mathbb{C}^{N_T \times N_{RF}}$  is an unitary matrix,  $\boldsymbol{\Sigma} \in \mathbb{R}^{N_{RF} \times N_{RF}}$  is a diagonal matrix of singular values arranged in decreasing order, and  $\mathbf{V} \in \mathbb{C}^{N_{RF} \times N_{RF}}$  is an unitary matrix. Using the SVD of  $\hat{\mathbf{F}}_{RF}$  and standard mathematical manipulation, the eigendecomposition of  $\hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{RF}^+$  can be expressed as

$$\hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{RF}^+ = \mathbf{U} \mathbf{U}^H. \quad (42)$$

Therefore, the eigenvalues of  $\hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{RF}^+$  are

$$\lambda_i(\hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{RF}^+) = \begin{cases} 1, & i = 1, 2, \dots, N_{RF} \\ 0, & \text{otherwise} \end{cases} \quad (43)$$

and  $\text{tr}(\hat{\mathbf{F}}_{BB}^H \hat{\mathbf{F}}_{RF}^H \hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{BB})$  can be further upper bounded by

$$\text{tr}(\hat{\mathbf{F}}_{BB}^H \hat{\mathbf{F}}_{RF}^H \hat{\mathbf{F}}_{RF} \hat{\mathbf{F}}_{BB}) \leq \sum_{i=1}^{N_{RF}} \lambda_i(\mathbf{F}_{opt} \mathbf{F}_{opt}^H) \leq \text{tr}(\mathbf{F}_{opt} \mathbf{F}_{opt}^H) \leq P.$$

Since any KKT point of problem (18) satisfies the power constraint automatically, it is also a KKT point of problem (6).

If  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  is a globally optimal solution of problem (18), then  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  is also a KKT point. Based on the previous analysis,  $(\hat{\mathbf{F}}_{RF}, \hat{\mathbf{F}}_{BB})$  satisfies the power constraint, and it is a globally optimal solution of problem (6). This completes the proof. ■

## ACKNOWLEDGMENT

The work of Y.R. Zheng and C. Xiao was supported in part by US National Science Foundation under grants ECCS-1231848, ECCS-1408316 and ECCS-1539316. The work of W. Chen was supported by NSFC 61671294, by Shanghai S&T Key Project 16JC1402900, by Guangxi NSFC 2015GXNSFDA139037, by National Major Project 2017ZX03001002-005, and by National 863 project 2015AA01A710. This work has been carried out while Mr. Juening Jin is visiting Missouri University of Science and Technology.

## REFERENCES

- [1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [2] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [3] E. Zhang and C. Huang, "On achieving optimal rate of digital precoder by RF-baseband codesign for MIMO systems," in *Proc. 80th IEEE Veh. Technol. Conf. (VTC Fall)*, Vancouver, BC, Sept. 2014, pp. 1–5.
- [4] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.
- [5] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [6] C. Rusu, R. M  ndez-Rial, N. Gonz  lez-Prelcic, and R. W. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.
- [7] R. Rajashekhar and L. Hanzo, "Hybrid beamforming in mm-wave MIMO systems having a finite input alphabet," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3337–3349, Aug. 2016.
- [8] C. Xiao and Y. R. Zheng, "On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs," in *Proc. IEEE Globecom*, 2008, pp. 1–5.
- [9] F. P  rez-Cruz, M. R. Rodrigues, and S. Verd  , "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.
- [10] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Jul. 2011.
- [11] W. Zeng, C. Xiao, and J. Lu, "A low-complexity design of linear precoding for mimo channels with finite-alphabet inputs," *IEEE Wireless Commun. Lett.*, vol. 1, no. 1, pp. 38–41, Feb. 2012.
- [12] J. Jin, C. Xiao, M. Tao, and W. Chen, "Linear precoding for fading cognitive multiple access wiretap channel with finite-alphabet inputs," *IEEE Trans. Veh. Technol.*, accepted for publication.
- [13] Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 739–747.
- [14] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, Apr. 1973.
- [15] A. R. Conn, N. I. Gould, and P. L. Toint, *Trust region methods*. Philadelphia, PA, U.S.: SIAM, 2000.
- [16] J. J. Mor   and D. C. Sorensen, "Computing a trust region step," *SIAM J. Sci. Stat. Comp.*, vol. 4, no. 3, pp. 553–572, Sept. 1983.
- [17] J. B. Lasserre, "A trace inequality for matrix product," *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1500–1501, Aug. 1995.