

MINIMUM SYMBOL ERROR RATE-BASED CONSTANT ENVELOPE PRECODING FOR MULTIUSER MASSIVE MISO DOWNLINK

Mingjie Shao[†], Qiang Li^{*†}, Wing-Kin Ma[†] and Anthony Man-Cho So[§]

[†] Department of Elec. Eng., The Chinese University of Hong Kong, Hong Kong SAR, China

^{*} School of Info. & Comm. Eng., University of Electronic Science and Technology of China, China

[§] Department of Sys. Eng. & Eng. Mgmt., The Chinese University of Hong Kong, Hong Kong SAR, China

E-mail: [†] {mjshao, wkma}@ee.cuhk.edu.hk, ^{*} lq@uestc.edu.cn, [§] manchoso@se.cuhk.edu.hk

ABSTRACT

This paper considers the problem of constant envelope (CE) precoder designs for multiuser massive MISO downlink channels. The use of CE signals allows one to employ low-cost radio frequency chains and thereby facilitates the implementation of massive MIMO. However, the subsequent CE precoder designs are usually challenging owing to the non-convex CE constraints. The existing CE precoder designs consider minimization of some measures on the distortion levels of the received symbols, and they usually aim at improving the symbol-error rate (SER) performances. In this paper we formulate a minimum SER-based design for CE precoding. The design formulation is non-convex and we propose two low-complexity first-order algorithms using gradient projection. Curiously, our simulation results show that the proposed designs can achieve bit-error rate performance close to that of zero-forcing beamforming without CE signaling restrictions.

Index Terms— massive MIMO, constant envelope, symbol error rate, non-convex projected gradient

1. INTRODUCTION

Massive MIMO, as one of the core physical-layer techniques for the fifth generation mobile system, can provide substantial spectral efficiency gains [1, 2]. In order to fully harness the benefits of massive MIMO, high-performance radio-frequency (RF) chains with large linear dynamic range are desired. However, this would push up the hardware cost tremendously as the number of antennas scales to hundreds or even more. To circumvent this difficulty, a signal processing-based constant envelope (CE) precoding solution was proposed [3], where the amplitudes of the transmit signals are fixed and only the phases are changed from symbol to symbol. Due to the low peak-to-average power ratio (PAPR) of CE signals, one can use very cheap RF chains (with lim-

ited dynamic range) to amplify the signals without incurring much distortion.

While CE precoding can be easily implemented with phase shifters, the design of CE precoder itself is a challenging task, owing to the non-convex nature of the CE constraints. In light of this, a great deal of efforts have been devoted to CE precoder designs under various constellations and user settings. In [3] and [4], a full study for the feasibility of CE precoding in single-user MISO channels was done and an optimal phase recovery algorithm for CE precoding was proposed in [4]. Subsequently, the concept of CE precoding was extended for multiuser MISO [5–9] and point-to-point MIMO [10]. In particular, the works [5, 6] and [7] studied the multiuser interference (MUI) power minimization by CE precoding under frequency-flat channel and frequency-selective channel, respectively. In addition to MUI reduction, the concept of constructive interference using CE precoding is investigated in [8] for PSK modulations; cross-entropy and convex relaxation methods are proposed to handle the CE problem. The joint optimization of CE precoding and receive beamforming is considered for a single-user MIMO system in [10]. Except for the single-user MISO case, the vast majority of the aforementioned works adopt some symbol distortion measures as their guidelines for CE precoder designs. In those studies it is usually anticipated that minimizing those symbol distortion measures should lead to improved symbol-error rate (SER) performance.

In this work we consider a minimum SER-based design for CE precoding in multiuser massive MISO downlink systems. Our minimum SER-based design formulation follows that of our very recent work for one-bit MIMO precoding [11] — which, by nature, is even more difficult than CE precoding and we handled it via a sophisticated alternating minimization algorithm. For self-containedness we will describe the formulation in Section 2. In the present paper, we will propose a simple non-convex projected gradient (PG) algorithm for the CE precoder design problem. To further speed up the convergence of the non-convex PG, we also consider the FISTA-type acceleration [12]. Our simulation results show that the bit-

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2016J011.

error rate (BER) performance of our CE design for 16-QAM and 64-QAM is about 1 – 3dB away from that of zero-forcing beamforming without stringent CE constraints; and the acceleration is helpful in reducing the running times.

2. SYSTEM MODEL AND PROBLEM FORMULATION

2.1. System Model

Consider a multiuser MISO downlink scenario where the base station (BS) equipped with N antennas unicasts K information symbol streams to K single-antenna users. Assuming block fading channel and perfect channel state information (CSI) at the BS, the received signal at user i during the t th time slot of fading block ℓ is given by

$$y_{i,t} = \mathbf{h}_{i,\ell}^T \mathbf{x}_t + n_{i,t}, \quad i = 1, \dots, K, \quad (1)$$

$$t \in ((\ell - 1)T, \ell T]$$

for $\ell = 1, 2, \dots$, where T is the block length; $\mathbf{h}_{i,\ell}^T$ is the downlink channel for user i during fading block ℓ ; $n_{i,t}$ is additive white Gaussian noise with mean 0 and variance σ_n^2 ; $\mathbf{x}_t \in \mathbb{C}^N$ is the transmit signal at the BS. As mentioned before, CE precoding is employed for facilitating massive MIMO implementation. In CE precoding, \mathbf{x}_t conforms to the following constraint:

$$\mathbf{x}_t \in \mathcal{X} \triangleq \{\mathbf{x} \in \mathbb{C}^N \mid |\mathbf{x}_i|^2 = P/N, \quad i = 1, \dots, N\},$$

where P is the total transmission power at the BS; here we assume that each antenna has equal transmit power. In the following derivations, we focus on one fading block and drop the fading block index ℓ for notational simplicity. Similarly, the time slot index t will be suppressed when there is no ambiguity.

The basic idea of CE precoding is to design the transmit signal \mathbf{x} so that the noise-free receive signal at the i th user, given by $\mathbf{h}_i^T \mathbf{x}$, is close to user i 's desired symbol s_i , i.e.,

$$\mathbf{h}_i^T \mathbf{x} \approx d \cdot s_i, \quad \forall i = 1, \dots, K. \quad (2)$$

Herein, $d \geq 0$ is a symbol shaping gain factor, which depends on \mathbf{x} and needs to be optimized; s_i 's are drawn from a QAM constellation \mathcal{S} , viz.

$$\mathcal{S} = \{s_R + js_I \mid s_R, s_I \in \{\pm 1, \pm 3, \dots, \pm(2L - 1)\}\}$$

with L being the order of the QAM constellation. From Eqns. (1) and (2), the symbol detection at users is easily performed as

$$\hat{s}_i = \text{dec}(y_i/d),$$

where $\text{dec}(\cdot)$ is a decision function, mapping y_i/d to the nearest constellation symbol in \mathcal{S} . For simplicity, we assume that all the users know d a priori, say, by estimating d from the training symbols before data block transmission or by broadcasting d from the BS to users via a side channel.

2.2. Problem Statement

Our goal is to minimize the worst SER among all the users. To proceed, let us first characterize user i 's SER as

$$\text{SER}_i = \Pr(\hat{s}_i \neq s_i | s_i).$$

Notice that

$$\text{SER}_i \leq \text{SER}_i^R + \text{SER}_i^I \leq 2 \max\{\text{SER}_i^R, \text{SER}_i^I\}, \quad (3)$$

where $\text{SER}_i^R \triangleq \Pr(\Re\{\hat{s}_i\} \neq \Re\{s_i\} | s_i)$ and $\text{SER}_i^I \triangleq \Pr(\Im\{\hat{s}_i\} \neq \Im\{s_i\} | s_i)$ correspond to the error probabilities of the in-phase and the quadrature components, respectively. Moreover, a straightforward calculation gives

$$\text{SER}_i^R \leq 2Q\left(\frac{d - |\Re\{\mathbf{h}_i^T \mathbf{x}\} - d\Re\{s_i\}|}{\sigma_n/\sqrt{2}}\right) \triangleq M_i^R, \quad (4)$$

$$\text{SER}_i^I \leq 2Q\left(\frac{d - |\Im\{\mathbf{h}_i^T \mathbf{x}\} - d\Im\{s_i\}|}{\sigma_n/\sqrt{2}}\right) \triangleq M_i^I,$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$, $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the real part and the imaginary part, respectively. Using (3) and (4), user i 's SER can be upper bounded by

$$\text{SER}_i \leq 2 \max\{M_i^R, M_i^I\}, \quad \forall i = 1, \dots, K. \quad (5)$$

From (5), we consider the following worst SER-based CE precoding problem:

$$\begin{aligned} \min_{\mathbf{x}, d} \quad & \max_{i=1, \dots, K} \max\{M_i^R, M_i^I\} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X}, \quad d \geq 0, \end{aligned} \quad (6)$$

which can be equivalently written as the following concise form by exploiting the monotonicity of the Q function:

$$\begin{aligned} \min_{\bar{\mathbf{x}}, d} \quad & \max_i |\bar{\mathbf{h}}_i^T \bar{\mathbf{x}} - d\bar{s}_i| - d, \\ \text{s.t.} \quad & \bar{\mathbf{x}} \in \mathcal{X}_{\Re}, \quad d \geq 0, \end{aligned} \quad (7)$$

where

$$\begin{aligned} \mathcal{X}_{\Re} \triangleq \quad & \{\mathbf{x} \in \mathbb{R}^{2N} \mid |x_j|^2 + |x_{j+N}|^2 = P/N, \quad j = 1, \dots, N\}, \\ \bar{\mathbf{x}} = \quad & [\Re\{\mathbf{x}\}^T, \Im\{\mathbf{x}\}^T]^T, \quad \bar{\mathbf{s}} = [\Re\{\mathbf{s}\}^T, \Im\{\mathbf{s}\}^T]^T, \\ \mathbf{H} = \quad & [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K]^T, \\ \bar{\mathbf{H}} = \quad & [\bar{\mathbf{h}}_1, \dots, \bar{\mathbf{h}}_{2K}]^T = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix}. \end{aligned}$$

Problem (7) is derived under a single time slot. The same formulation can be derived for the case of one fading block. Letting $\{\bar{\mathbf{x}}_t\}_{t=1}^T$ be the equivalent real-valued transmit signals within one fading block and following the derivations above, the worst SER-based CE precoding for block transmission can be shown to be

$$\begin{aligned} \min_{\{\bar{\mathbf{x}}_t\}_{t=1}^T, d} \quad & \max_{i=1, \dots, K, t=1, \dots, T} |\bar{\mathbf{h}}_i^T \bar{\mathbf{x}}_t - d\bar{s}_{i,t}| - d \\ \text{s.t.} \quad & \bar{\mathbf{x}}_t \in \mathcal{X}_{\Re}, \quad d \geq 0, \quad t = 1, \dots, T. \end{aligned} \quad (8)$$

Remark 1 Problem (8) is physically sound. Specifically, the objective consists of two parts — One is the distortions of the noise-free received symbols; the other is the symbol shaping gaining factor d . Intuitively, small distortion and large symbol shaping gain factor would make the CE design robust against noise. By jointly optimizing $\{\bar{\mathbf{x}}_t\}_{t=1}^T$ and d , problem (8) can provide a better balance between the two metrics, as compared with most existing CE literatures where d is fixed [3, 5, 6].

The resulting CE precoding problem (8) is a non-smooth non-convex problem. In the following section, we will develop a non-convex projected gradient-based algorithm for problem (8).

3. PROPOSED CE PRECODING ALGORITHM

Our development consists of two steps: First, we tackle the non-smooth objective in (8) by a smooth approximation. Second, a non-convex projected gradient-based method is applied to handle the smoothed problem.

3.1. Smooth Approximation of (8)

We apply the well-known log-sum-exp inequality to smoothen the objective. Specifically, note the following fact [13]:

Fact 1 Given $a_1, \dots, a_K \in \mathbb{R}$, it holds for any $\sigma > 0$ that

$$\max_{i=1, \dots, K} a_i \leq \sigma \log \sum_{i=1}^K e^{\frac{a_i}{\sigma}} \leq \max_{i=1, \dots, K} a_i + \sigma \log K. \quad (9)$$

Moreover, the inequalities become tight as $\sigma \rightarrow 0$.

Utilizing (9), the CE precoding problem (8) is smoothly approximated as

$$\begin{aligned} \min_{\bar{\mathbf{x}}_t, d} \quad & f(d, \{\bar{\mathbf{x}}_t\}_{t=1}^T) \\ \text{s.t.} \quad & \bar{\mathbf{x}}_t \in \mathcal{X}_{\mathfrak{R}}, \quad d \geq 0, \quad t = 1, \dots, T. \end{aligned} \quad (10)$$

where

$$f(d, \{\bar{\mathbf{x}}_t\}_{t=1}^T) \triangleq \sigma \log \sum_{i,t} \left[e^{\frac{\bar{\mathbf{h}}_i^T \bar{\mathbf{x}}_t - d \bar{s}_{i,t} - d}{\sigma}} + e^{\frac{-\bar{\mathbf{h}}_i^T \bar{\mathbf{x}}_t + d \bar{s}_{i,t} - d}{\sigma}} \right].$$

3.2. Non-convex Gradient Projection Algorithm for (10)

For ease of exposition, we define

$$\mathcal{D} \triangleq \{(d, \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_T) \mid d \geq 0, \bar{\mathbf{x}}_t \in \mathcal{X}_{\mathfrak{R}}, t = 1, \dots, T\}$$

to be the feasible set of problem (10). Let $\mathbf{z} \triangleq (d, \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_T)$. The proposed gradient projection method recursively updates \mathbf{z} according to the following equation

$$\mathbf{z}^{l+1} = \Pi_{\mathcal{D}}(\mathbf{z}^l - \gamma_l \nabla f(\mathbf{z}^l)), \quad (11)$$

where the superscript l denotes the iteration number, and $\gamma_l > 0$ is the stepsize, which can be determined by backtracking line search [12]; $\Pi_{\mathcal{D}}(\mathbf{x})$ denotes the projection of \mathbf{x} onto the set \mathcal{D} . While \mathcal{D} is non-convex, the projection can be easily computed in closed form. Specifically, let

$$\tilde{\mathbf{z}}^l = \left(\tilde{d}^l, \tilde{\mathbf{x}}_1^l, \dots, \tilde{\mathbf{x}}_T^l \right) = \mathbf{z}^l - \gamma_l \nabla f(\mathbf{z}^l).$$

Then,

$$\begin{cases} d^{l+1} = \max\{0, \tilde{d}^l\}, \\ [\bar{\mathbf{x}}_t^{l+1}]_j = \sqrt{\frac{P}{N}} \frac{[\tilde{\mathbf{x}}_t^l]_j}{\sqrt{|[\tilde{\mathbf{x}}_t^l]_j|^2 + |[\tilde{\mathbf{x}}_t^l]_{j+N}|^2}}, \quad j = 1, \dots, N, \forall t, \\ [\bar{\mathbf{x}}_t^{l+1}]_j = \sqrt{\frac{P}{N}} \frac{[\tilde{\mathbf{x}}_t^l]_j}{\sqrt{|[\tilde{\mathbf{x}}_t^l]_{j-N}|^2 + |[\tilde{\mathbf{x}}_t^l]_j|^2}}, \quad j = N+1, \dots, 2N, \forall t, \end{cases} \quad (12)$$

where $[\mathbf{x}]_j$ denotes the j th element of \mathbf{x} . The detailed algorithm is summarized in Algorithm 1.

By adapting the proof in [14], it can be shown that every limit point, denoted as \mathbf{z}^* , generated by Algorithm 1 is a stationary point of problem (10). We omit the proof due to the page limit and will present it in the journal version.

Algorithm 1 : Projected Gradient for CE Precoding

- 1: Initialize $\mathbf{z}^0 = (\bar{\mathbf{x}}^0, d^0)$, σ and $l = 0$
- 2: **repeat**
- 3: Calculate $\nabla f(\mathbf{z}^l)$ as

$$\frac{\partial f(\mathbf{z}^l)}{\partial \bar{\mathbf{x}}_t} = \frac{\sum_i (W_{i,t}^P - W_{i,t}^N) \bar{\mathbf{h}}_i}{\sum_{i,t} (W_{i,t}^P + W_{i,t}^N)}, \quad t = 1, \dots, T,$$

$$\frac{\partial f(\mathbf{z}^l)}{\partial d} = \frac{\sum_{i,t} (-W_{i,t}^P (\bar{s}_{i,t} + 1) + W_{i,t}^N (\bar{s}_{i,t} - 1))}{\sum_{i,t} (W_{i,t}^P + W_{i,t}^N)},$$

where $W_{i,t}^P = \exp\left(\frac{\bar{\mathbf{h}}_i^T \bar{\mathbf{x}}_t^l - d^l \bar{s}_{i,t} - d^l}{\sigma}\right)$ and $W_{i,t}^N = \exp\left(\frac{-\bar{\mathbf{h}}_i^T \bar{\mathbf{x}}_t^l + d^l \bar{s}_{i,t} - d^l}{\sigma}\right)$;

- 4: Update d^{l+1} and $\{\bar{\mathbf{x}}_t^{l+1}\}_{t=1}^T$ according to Eqn. (11) and Eqn. (12);
 - 5: $l = l + 1$;
 - 6: **until** some stopping criterion is satisfied.
-

3.3. Fast Non-convex Gradient Projection for (10)

In this section, we introduce an acceleration scheme to further speed up the convergence of Algorithm 1. The idea is similar to the FISTA algorithm [12] or Nesterov's accelerated gradient method [15], though FISTA was originally developed for convex problems. Specifically, we modify Eqn. (11) as

$$\mathbf{z}^{l+1} = \Pi_{\mathcal{D}}(\mathbf{w}^l - \gamma_l \nabla f(\mathbf{w}^l)),$$

where w^l is an extrapolated point of z^l and z^{l-1} , i.e.,

$$w^l = z^l + \frac{\beta_l - 1}{\beta_{l+1}}(z^l - z^{l-1}),$$

$$\beta_{l+1} = \frac{1 + \sqrt{1 + 4\beta_l^2}}{2},$$

with $z^0 = z^{-1}$ and $\beta_0 = 1$.

It has been shown in [12] that this acceleration can improve the convergence rate of the projected gradient method from $\mathcal{O}(1/l)$ to $\mathcal{O}(1/l^2)$ for convex problems. While our considered problem is non-convex, our numerical experience suggests that this acceleration scheme is still very effective, as illustrated in the ensuing section.

4. SIMULATION RESULTS

In this section, we evaluate the performance of our proposed algorithms by Monte-Carlo simulations. Three schemes are compared, namely, zero-forcing (ZF) without CE constraints, named “ZF”; ZF with naive projection onto the CE set, named “CE ZF”; and the total MUI power minimization algorithm [5], named “MUImin”. We will use “PG” and “FPG” to represent Algorithm 1 and its accelerated version. BER is used as the performance metric.

The simulation setting is as follows: A block Rayleigh fading channel is assumed with transmission block length $T = 10$. The total transmit power is $P = 1$. The elements of channel $\{h_i\}_{i=1}^K$ are i.i.d. generated according to $\mathcal{CN}(0, 1)$. Both 16-QAM and 64-QAM modulations are considered. For both Algorithm 1 and its accelerated variant, the smoothing parameter is set to $\sigma = 0.05$, and the algorithms stop when the improvement of successive iterations is less than 10^{-4} , or the maximum number of iterations 5,000 is reached. All the results were averaged over 10^4 independent channel trials.

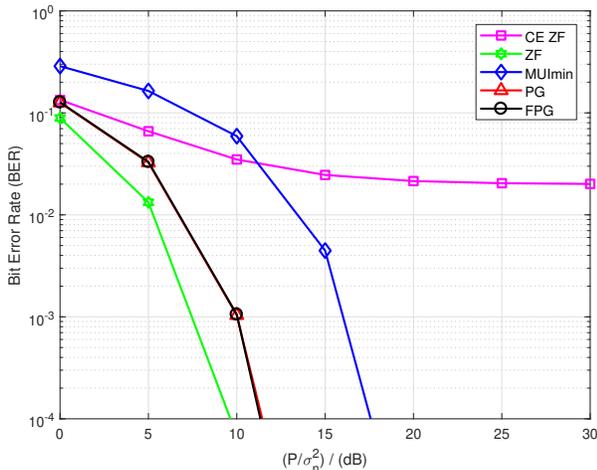


Fig. 1: Average BER performance versus P/σ_n^2 ; 16-QAM.

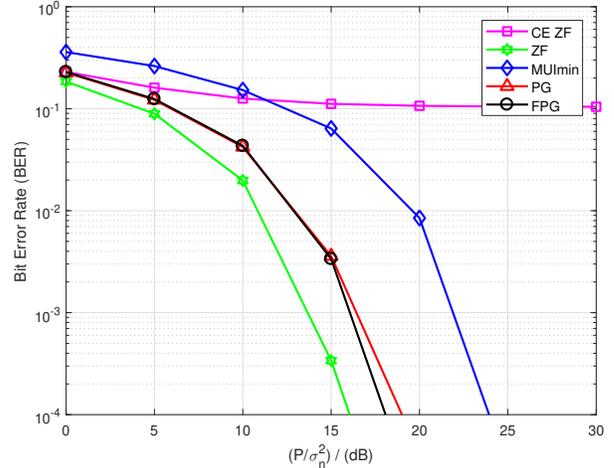


Fig. 2: Average BER performance versus P/σ_n^2 ; 64-QAM.

In Fig. 1, we compare the BER performance under 16-QAM modulation scheme; the number of transmit antennas is $N = 128$ and the number of users is $K = 16$. It is seen that both PG and FPG achieve much better BER performance than CE ZF and MUImin. Notice that the performance gap between ZF (without CE constraints) and FPG is only about 2.5dB at $\text{BER} = 10^{-3}$. Fig. 2 shows the simulation result under the same setting as Fig. 1, except that 64-QAM is used. The figure shows similar performance behaviors.

While PG and FPG have similar BER performances, FPG is much more computationally efficient than PG. To verify this, we compared the runtime of PG and FPG under different number of antennas with fixed number of users and modulation scheme. The results are shown in Table 1. As seen, FPG can achieve significant runtime reduction, especially for large problem sizes.

Table 1: Average runtime (in Sec.) for each transmission block ($K = 16$, 64-QAM)

N	50	100	150	200
PG	0.531	0.701	1.053	1.29
FPG	0.415	0.436	0.545	0.607

5. CONCLUSION

In this paper, we have considered the CE precoder design for multiuser massive MISO downlink channels. A worst SER-based CE formulation is employed. By exploiting the problem structure of the CE precoding, a simple and efficient non-convex projected gradient algorithm and its accelerated variant were derived. Simulation results showed that our proposed algorithms can achieve better BER performance than the existing CE precoder designs.

6. REFERENCES

- [1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.
- [2] 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.
- [3] S. K. Mohammed and E. G. Larsson, "Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3992–4005, 2012.
- [4] J. Pan and W.-K. Ma, "Constant envelope precoding for single-user large-scale MISO channels: Efficient precoding and optimal designs," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 982–995, 2014.
- [5] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, 2013.
- [6] J.-C. Chen, C.-K. Wen, and K.-K. Wong, "Improved constant envelope multiuser precoding for massive MIMO systems," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1311–1314, 2014.
- [7] S. K. Mohammed and E. G. Larsson, "Constant-envelope multi-user precoding for frequency-selective massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 547–550, 2013.
- [8] P. V. Amadori and C. Masouros, "Constant envelope precoding by interference exploitation in phase shift keying-modulated multiuser transmission," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 538–550, 2017.
- [9] F. Liu, C. Masouros, P. V. Amadori, and H. Sun, "An efficient manifold algorithm for constructive interference based constant envelope precoding," *IEEE Signal Process. Lett.*, vol. 24, no. 10, pp. 1542–1546, 2017.
- [10] S. Zhang, R. Zhang, and T. J. Lim, "Constant envelope precoding for MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 149–162, Jan 2018.
- [11] M. Shao, Q. Li, and W.-K. Ma, "One-bit massive MIMO precoding via a minimum symbol-error probability design," in to appear in *Proc. IEEE Int. Conf. Acous., Speech, Signal Process. (ICASSP)*, 2018.
- [12] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [14] J. Tranter, N. D. Sidiropoulos, X. Fu, and A. Swami, "Fast unit-modulus least squares with applications in beamforming," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2875–2887, 2017.
- [15] Y. Nesterov, *Introductory Lectures on Convex Optimization*. Springer Science & Business Media, 2004, vol. 87.