EXTREME-POINT PURSUIT FOR UNIT-MODULUS OPTIMIZATION

Mingjie Shao, Qi Dai and Wing-Kin Ma

Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR of China

ABSTRACT

Unit-modulus constrained optimization is frequently encountered in many engineering problems. In our recent study for massive MIMO precoding, we devised a penalty method for unit-modulus optimization. In this paper, we revisit this penalty method in several other unit-modulus applications in signal processing. Moreover, as a new result, we show that the concept of our penalty method can be generalized to handle a much broader class of problems, such as those with semi-orthogonal matrix constraints. The rationale is to relax the constraint set as its convex hull, and at the same time, we add a penalty function to force the solution to be an extreme point—which lies in the original constraint set. We show conditions under which the penalty formulation is equivalent to the original problem. We test the penalty method on classic and one-bit MIMO detection under M-PSK constraints, and on phase retrieval. Simulation results indicate that the penalty method yields competitive performance on the aforementioned applications.

Index Terms— unit-modulus constraint, penalty method, extreme point, homotopy optimization

1. INTRODUCTION

Consider the following unit-modulus optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{C}^n} f(\boldsymbol{x}), \quad \text{s.t. } \boldsymbol{x}\in\mathcal{V}, \tag{P1}$$

where f is the objective function; $\mathcal{V}$ is either

$$\mathcal{V}=\{\boldsymbol{x}\in\mathbb{C}^n \mid |x_i|=1,\ i=1,\ldots,n\}, \tag{1}$$

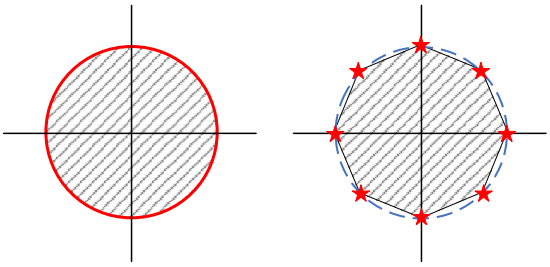
or

$$\mathcal{V}=\mathcal{S}_M^n,\ \mathcal{S}_M=\{s\in\mathbb{C}|\ s=e^{\mathrm{j}\frac{2\pi k}{M}}, k=0,1,\ldots,M-1\}. \tag{2}$$

An illustration of $\mathcal{V}$ with $n=1$ is shown in Fig. 1.

Unit-modulus optimization arises in a variety of applications in signal processing, wireless communication, data science and machine learning. For instance, the continuous unit-modulus constraint (1) can be found in applications such as beamforming/precoding in communication and radar systems [1–3], phase retrieval [4] and phase synchronization [5]; the discrete unit-modulus set (2) is typically used to model phase-shift keying (PSK) modulation in digital communications [6] and digital phase shifters with finite phase states [7]; also, the binary constraint $\mathcal{V}=\{\pm1\}^n$ is a special case of (2) ($M=2$), which has many applications related to combinatorial optimizations, such as community detection [8] and feature selection [9].

This work was supported by a General Research Fund (GRF) of Hong Kong Research Grant Council (RGC), under Project ID CUHK 14208819.



(a) Continuous UM set (1). (b) Discrete UM set (2).

Fig. 1: Unit-modulus (UM) sets. Red: $\mathcal{V}$; shaded: $\mathrm{conv}\mathcal{V}$.

The unit-modulus constraints are non-convex. When f is quadratic, a popular approach to tackle problem (P1) is semidefinite relaxation (SDR) [4, 10, 11]. The elegance of SDR is that it employs a convex optimization problem to approximate (P1), but the application of SDR for large-scale problems is impeded by its complexity—the complexity increases cubically with the problem size. Non-convex optimization for unit-modulus optimization (presumably with a large problem size) has recently drawn attentions. Non-convex projected gradient methods for the continuous unit-modulus constraint (1) have been studied in [1, 5]. There are several methods custom-built for the binary case; see e.g., [12–14]. Our recent works [3, 15] proposed a penalty method that can handle both sets (1) and (2) in MIMO precoding [3] and binary MIMO detection [15], with promising performance reported.

In this paper, we continue the study of the penalty method in [3, 15] by examining its performance in other unit-modulus applications. Beyond that, we extend the notion of our penalty method to handle a wider scope of non-convex constrained optimization problems, which cover unit-modulus, unit-ellipsoid and semi-orthogonal matrix constraints (also known as the Stiefel manifold). The rationale, simply speaking, is to convex-relax the constraint set, and add a penalty function to encourage the solution to lie in set of extreme points; thus, it is an "extreme-point pursuit". We show that, by choosing a sufficiently large penalty parameter, the penalty formulation is equivalent to the original problem. We test the extreme point pursuit method on the applications of M-PSK MIMO detection (in both classic and one-bit settings) and phase retrieval, and it exhibits satisfactory performance in these applications.

Our notations are standard. x, $\boldsymbol{x}$, and $\boldsymbol{X}$ represent scalar, vector and matrix, respectively; $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{C}$ denote the real domain, real non-negative orthant and complex domain, respectively; $\mathrm{conv}\mathcal{V}$ stands for the convex hull of a set $\mathcal{V}$; $\mathcal{CN}(\boldsymbol{\mu},\boldsymbol{R})$ denotes complex circular Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{R}$. The definitions of Lipschitz continuity, strongly/strictly convex functions are standard and can be found in, e.g., [16].

2. OUR METHOD

2.1. Basic Insight

To provide insight, take the Boolean quadratic program

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}, \quad \mathcal{V} = \{\pm 1\}^n$$

as an example; here, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is symmetric.

Consider the following penalty formulation of (P1)

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F_\lambda(\boldsymbol{x}) := f(\boldsymbol{x}) - \lambda\|\boldsymbol{x}\|_2^2, \quad \text{s.t. } \boldsymbol{x} \in [-1,1]^n, \tag{P2}$$

where $\lambda > 0$ is a given parameter. The idea is that the term $-\lambda\|\boldsymbol{x}\|_2^2$ in (P2) encourages the solution to have large values with x_i^2's, thereby forcing the solution to lie in $\{\pm 1\}^n$. Problem (P2) is a convex-constrained minimization problem, with a non-convex objective, and one may use first-order methods to tackle it. The question of interest is how well (P2) approximates (P1).

We have answered the above question in our prior work [3], and here we want to offer a different perspective to answer the question. Note that $[-1,1]^n = \text{conv}\{\pm 1\}^n$, and F_λ is strictly concave if $\lambda > \|\boldsymbol{A}\|_2$. By listing all the elements of $\{\pm 1\}^n$ as $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$, where $m = 2^n$, we can express any $\boldsymbol{x} \in [-1,1]^n$ as $\boldsymbol{x} = \sum_{i=1}^m \theta_i \boldsymbol{v}_i$, where $\boldsymbol{\theta} \in \Delta^m := \{\boldsymbol{\theta} \in \mathbb{R}_+^n | \boldsymbol{\theta}^\top \mathbf{1} = 1\}$. By Jensen's inequality,

$$F_\lambda(\boldsymbol{x}) \geq \sum_{i=1}^m \theta_i F_\lambda(\boldsymbol{v}_i) \geq \min_{i=1,\ldots,m} F_\lambda(\boldsymbol{v}_i) = \min_{\boldsymbol{x}\in\{\pm 1\}^n} f(\boldsymbol{x}) - n;$$

also, the above equalities are attained if and only if $\boldsymbol{x}$ is one of $\boldsymbol{v}_i$'s and minimizes f, due to the strict concavity of F_λ [16]. Thus, we see that (P2) is an equivalent formulation of (P1) when λ is sufficiently large; note that λ does not need to be infinitely large to obtain this equivalence.

2.2. New Result

The above example displays a concept of extreme-point pursuit, which can be generalized. Consider

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}), \quad \text{s.t. } \boldsymbol{x} \in \mathcal{V}, \tag{P3}$$

where $\mathcal{D}$ is either $\mathbb{R}^n$ or $\mathbb{C}^n$; $f : \mathcal{D} \to \mathbb{R}$ is the objective function, as before; $\mathcal{V} \subset \mathcal{D}$ now is an extreme point set, i.e., every $\boldsymbol{x} \in \mathcal{V}$ is an extreme point of $\mathcal{V}$. Also, suppose that there exists a strongly convex function $g : \mathcal{D} \to \mathbb{R}$ such that

$$g(\boldsymbol{x}) = C, \ \forall \boldsymbol{x} \in \mathcal{V}, \tag{3}$$

where C is a constant. We therefore consider the following penalty formulation of (P3)

$$\min_{\boldsymbol{x}} F_\lambda(\boldsymbol{x}) := f(\boldsymbol{x}) - \lambda g(\boldsymbol{x}), \quad \text{s.t. } \boldsymbol{x} \in \text{conv}\mathcal{V}. \tag{P4}$$

We can show the following result.

Theorem 1 *Suppose that f has L_f-Lipschitz continuous gradient on $\text{conv}\mathcal{V}$; g is differentiable, β-strongly convex, and satisfies* (3)*. Then, for any $\lambda > L_f/\beta$, we have the following results*

1) *(global optimality) problems* (P3) *and* (P4) *are equivalent in the sense that they share exactly the same optimal solutions.*
2) *(local optimality) any locally optimal solution to problem* (P4) *must be a feasible point of* (P3)*.*
3) *(stationary points) if $\boldsymbol{x}$ is a stationary point of* (P4) *and $\boldsymbol{x} \notin \mathcal{V}$, then $\boldsymbol{x}$ is either a local maximum or a saddle point of* (P4)*.*

Theorem 1 is an adaptation of [3, Theorem 2]. We provide the proof for item *1)* of Theorem 1; the remaining proof is the same as [3, Theorem 2]. For any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \text{conv}\mathcal{V}$, we have

$$\begin{aligned}
&\langle \nabla F_\lambda(\boldsymbol{x}_1) - \nabla F_\lambda(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2\rangle \\
=&\langle \nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2\rangle - \lambda\langle \nabla g(\boldsymbol{x}_1) - \nabla g(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2\rangle \\
\leq&\|\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2)\|_2 \, \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2 - \lambda\beta\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 \\
\leq&(L_f - \lambda\beta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2,
\end{aligned}$$

where the first inequality is due to the Cauchy-Schwarz inequality and the β-strong convexity of g; the second inequality is due to the L_f-Lipschitz continuous gradient of f. By choosing $\lambda > L_f/\beta$, the above inequality implies that $F_\lambda(\boldsymbol{x})$ is strongly concave [17]. By the Krein–Milman theorem, any $\boldsymbol{x} \in \text{conv}\mathcal{V}$ can be represented by $\boldsymbol{x} = \sum_{i=1}^r \theta_i \boldsymbol{v}_i$ where $\boldsymbol{\theta} \in \Delta^r$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ are extreme points of $\text{conv}\mathcal{V}$ for some $r > 0$. We have

$$F_\lambda(\boldsymbol{x}) \geq \sum_{i=1}^r \theta_i F_\lambda(\boldsymbol{v}_i) \geq \min_{\boldsymbol{x}\in\text{conv}\mathcal{V}} F_\lambda(\boldsymbol{x}),$$

and the above equalities hold if and only if $\boldsymbol{x}$ is a solution to (P3).

Let us describe some examples that fall within the scope above.

Example 1 $\mathcal{V} = \{\boldsymbol{x} \in \mathbb{C}^n | |x_i| = 1, \forall i\}$ in (1). We have $\text{conv}\mathcal{V} = \{\boldsymbol{x} \in \mathbb{C}^n | \ |x_i| \leq 1, \forall i\}$. We can choose $g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$.

Example 2 $\mathcal{V} = \mathcal{S}_M^n$ in (2). We have $\text{conv}\mathcal{V} = \text{conv}\mathcal{S}_M^n$ with

$$\text{conv}\mathcal{S}_M = \left\{ x \in \mathbb{C} \middle| \ \Re(e^{-\jmath\frac{\pi(2k-1)}{M}} x) \leq \cos(\tfrac{\pi}{M}), k = 1, \ldots, M \right\}.$$

We can choose $g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$. As a remark, the projection onto $\text{conv}\mathcal{S}_M$ can be computed in closed form [3].

Example 3 (An ellipsoidal extension to the unit-modulus constraint) $\mathcal{V} = \{\boldsymbol{x} \in \mathbb{C}^n | \boldsymbol{x}^H \boldsymbol{R}_i \boldsymbol{x} = 1, \ i = 1, \ldots, k\}$, where every $\boldsymbol{R}_i$ is positive semidefinite, and $\boldsymbol{R} = \sum_{i=1}^k \boldsymbol{R}_i$ is positive definite. We have $\text{conv}\mathcal{V} = \{\boldsymbol{x} \in \mathbb{C}^n | \boldsymbol{x}^H \boldsymbol{R}_i \boldsymbol{x} \leq 1, \text{ for } i = 1, \ldots, k\}$. We can choose $g(\boldsymbol{x}) = \boldsymbol{x}^H \boldsymbol{R}\boldsymbol{x}$.

Example 4 (A matrix generalization of the unit-modulus constraint) $\mathcal{V} = \{\boldsymbol{X} \in \mathbb{R}^{m\times n} | \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{I}\}$, with $m \geq n$. We have

$$\text{conv}\mathcal{V} = \{\boldsymbol{X} \in \mathbb{R}^{m\times n} | \ \|\boldsymbol{X}\|_2 \leq 1\}.$$

We can choose $g(\boldsymbol{x}) = \|\boldsymbol{X}\|_F^2$. This example was considered in [18] for orthogonal non-negative matrix factorization.

2.3. Homotopy Optimization

While we showed that (P4) is equivalent to (P3) under some appropriate assumptions, that does not mean that (P4) is easy to solve—if an optimal solution or nearly-optimal solution is desired. Problem (P4) has a convex constraint, and in Examples 1, 2 and 4, we can even leverage on the constraint structures to built efficient first-order methods for tackling (P4). However, the objective function is non-convex, and a first-order method can converge to a bad local minimum. Empirically, we found that the following idea works: start from a small λ; run a first-order method to find a solution (possibly a stationary point) of (P4); increase λ and use the previously obtained

solution as a starting point to run the first-order method; repeat the above step until some stopping rule is met. The above idea is summarized in Algorithm 1. Intuitively, the idea is that if the landscape of F_λ does not change significantly from one λ to another λ, the optimal solutions may not change significantly, too. Also, by considering that the problem should be "easy" for small λ (actually F_λ is convex if $\lambda < -L_f/\beta$), the strategy is to try to track the solution path with respect to λ. In the literature, this solution tracking strategy is widely adopted and is called homotopy optimization [15, 19–22].

An intriguing question is whether, or to what extent, the above intuition can be proven to be valid. This is an open question at present. In the recent advances of non-convex optimization, we have seen an exciting direction where researchers showed that a first-order method may converge to a global minimum when the (non-convex) objective function satisfies some good landscape properties [21]. Such results may require strong local convexity. Unfortunately, our extreme point pursuit appears inapplicable to such assumptions.

Algorithm 1 Homotopy Optimization for Problem (P4)

1: **Input:** Initialization $\boldsymbol{x}^0$ and λ_0.
2: $k = 0$
3: **repeat**
4: Run a descent algorithm for (P4) with $\lambda = \lambda_k$ and initialization $\boldsymbol{x}^k$. Set the output as $\boldsymbol{x}^{k+1}$.
5: Choose λ_{k+1} as an increased version of λ_k.
6: $k = k + 1$
7: **until** some stopping criterion is satisfied.

3. APPLICATIONS AND SIMULATION RESULTS

We apply the above penalty method on some unit-modulus optimization problems in practice. We choose $g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$. The descent algorithm in Line 4 of Algorithm 1 is chosen to be the first-order method in [3], which combines majorization-minimization (MM) method and inexact accelerated projected gradient method; We refer the reader to [3] for details.

3.1. Classic MIMO Detection

First, we consider the MIMO detection problem [23]. The problem is described as follows. The transmitter has n antennas and sends an M-PSK modulated symbol vector $\boldsymbol{x} \in \mathcal{S}_M^n$. After propagating through a wireless channel $\boldsymbol{H} \in \mathbb{C}^{m\times n}$, the m-antenna receiver receives a noisy signal $\boldsymbol{y} \in \mathbb{C}^m$ with $m \geq n$. This can be modeled by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v},$$

where $\boldsymbol{v} \in \mathbb{C}^m \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ is noise.

Provided the channel state information $\boldsymbol{H}$, MIMO detection is to detect the symbol $\boldsymbol{x}$ from the received signal $\boldsymbol{y}$. The maximum-likelihood (ML) detector is considered

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2, \text{ s.t. } \boldsymbol{x} \in \mathcal{S}_M^n. \tag{4}$$

Particularly, we use extreme-point pursuit to deal with (4). We show simulation results. The modulation scheme is 8-PSK or 16-PSK. The channel matrix $\boldsymbol{H}$ is generated based on the following correlated model

$$\boldsymbol{H} = \boldsymbol{R}_r^{1/2}\tilde{\boldsymbol{H}}\boldsymbol{R}_t^{1/2}, \tag{5}$$

where $\tilde{\boldsymbol{H}}$ follows an element-wise i.i.d. Gaussian distribution, i.e., $\tilde{h}_{ij} \sim \mathcal{CN}(0,1)$ for all i, j; $\boldsymbol{R}_r$ and $\boldsymbol{R}_t$ model the antenna correlations at the receiver side and transmitter side, respectively, and are assumed to follow the exponential model with $\rho = 0.2$ [24]. The signal-to-noise ratio (SNR) is defined as $\mathbb{E}[\|\boldsymbol{H}\boldsymbol{x}\|_2^2]/\mathbb{E}[\|\boldsymbol{v}\|_2^2]$. A number of 10,000 Monte-Carlo trials were used to obtain the results.

The compared schemes are: 1) the MMSE detector; 2) semidefinite relaxation (SDR) [10], with 1,000 times Gaussian randomization; 3) LAMA (an approximate message passing method) [25]. The original LAMA in [25] does not work under the correlated channel (5); we applied damping [26], with damping factor 0.7, to improve its performance. We also show a lower bound when no MIMO interference is present.

The numbers of transmit and receive antennas are, respectively, $m = 80$ and $n = 80$. Fig. 2 shows the bit error rate (BER) performance. It is seen that our EXtreme-Point Pursuit (EXPP) method achieves satisfactory performance and is close to the no-interference lower bound. The runtime performance is shown in Table 1. The settings are the same with those in Fig. 2, with SNR = 38 dB and 8-PSK. It is seen that both EXPP and LAMA are computationally efficient; they run much faster than SDR.

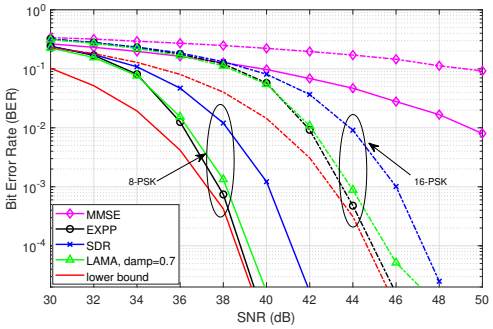


Fig. 2: BER performance of classic MIMO detection; solid: 8-PSK; dashed: 16-PSK.

Table 1: Average runtime performance of different algorithms.

Alg.	SDR	EXPP	LAMA
time (in sec.)	1.017	0.026	0.021

3.2. One-Bit MIMO Detection

Second, we consider the one-bit MIMO detection problem, which has attracted interest recently in massive MIMO. This problem considers a quantized received signal

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{H}\boldsymbol{x} + \boldsymbol{v}),$$

where $\mathcal{Q}(x) := \text{sgn}(\Re(x)) + \mathrm{j}\,\text{sgn}(\Im(x))$ is the one-bit quantizer applied on the real and imaginary parts. The corresponding ML detector is:

$$\min_{\boldsymbol{x}\in\mathcal{S}_M^n} -\sum_{i=1}^m \Big(\log\Big(\frac{\Re(y_i)\Re(\boldsymbol{h}_i^\top\boldsymbol{x})}{\sigma/\sqrt{2}}\Big) + \log\Big(\frac{\Im(y_i)\Im(\boldsymbol{h}_i^\top\boldsymbol{x})}{\sigma/\sqrt{2}}\Big)\Big); \tag{6}$$

see [27]. Here, $\boldsymbol{h}_i^\top$ denotes the ith row of $\boldsymbol{H}$; the modulation scheme of the symbol $\boldsymbol{x}$ is again assumed to be M-PSK.

We show the simulation results. The compared methods are: 1) the zero-forcing (ZF) detector; 2) nML [27], which applies a spherical relaxation $\|\boldsymbol{x}\|_2^2 \leq n$ for problem (6); and 3) nML-two-stage, which involves an additional local search procedure on the nML solution [27].

Fig. 3 shows the BER performance. The number of transmit antennas and the number of receive antennas are $n = 32$ and $m = 256$, respectively. The channel matrix $\boldsymbol{H}$ follows (5) with $\boldsymbol{R}_r = \boldsymbol{I}$ and $\boldsymbol{R}_t = \boldsymbol{I}$. 8-PSK is used. It is seen that EXPP outperms the other algorithms. Table 2 shows the runtimes of the considered algorithms at the SNR=25 dB. EXPP is slightly slower than nML, but is faster than nML-two-stage.

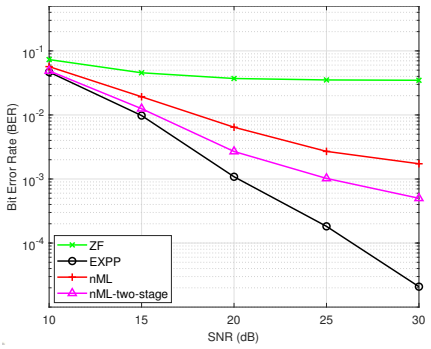


Fig. 3: BER performance of one-bit MIMO detection; 8-PSK.

Table 2: Average runtime performance of different algorithms.

Alg.	nML	nML-two-stage	EXPP
time (in sec.)	0.015	0.028	0.019

3.3. Phase Retrieval

Phase retrieval is an important problem in physical sciences and engineering [28]. The problem stems from the fact that the detectors can only record the intensity of the signal but not its phase [29]. To be specific, the recorded data, or samples, $\boldsymbol{y} \in \mathbb{R}^m$ takes the form

$$y_i = |\boldsymbol{a}_i^H \boldsymbol{u}|, \; i = 1, \ldots, m, \tag{7}$$

where $\boldsymbol{a}_i \in \mathbb{C}^n$ are known sampling vectors, $\boldsymbol{u} \in \mathbb{C}^n$ is the signal to be identified. The problem of phase retrieval is to recover $\boldsymbol{u}$ from $\boldsymbol{y}$. By introducing an auxiliary phase variable $\boldsymbol{x} \in \mathbb{C}^m$ with $|x_i| = 1$ for all i, the phase retrieval problem can be posed as [4]

$$\min_{\boldsymbol{x},\boldsymbol{u}} \|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{y} \odot \boldsymbol{x}\|_2^2, \text{ s.t. } |x_i| = 1, \; i = 1, \ldots, m, \tag{8}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^H$, and $\odot$ is the Hadamard product. By eliminating the variable $\boldsymbol{u}$ (a least-square problem given any $\boldsymbol{x}$), we can recast the problem as

$$\min_{\boldsymbol{x}} \boldsymbol{x}^H \boldsymbol{W} \boldsymbol{x}, \text{ s.t. } |x_i| = 1, \; i = 1, \ldots, m, \tag{9}$$

where $\boldsymbol{W} = \text{Diag}(\boldsymbol{y})(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^\dagger)\text{Diag}(\boldsymbol{y})$.

Let us show the simulation results. The elements of $\boldsymbol{u} \in \mathbb{C}^n$ are i.i.d. generated from $\mathcal{CN}(0, 1)$ with $n = 64$. We test the recovery performance under two commonly used models of $\boldsymbol{a}_i$'s: 1) $\boldsymbol{A}$ takes the first n columns of the m-point discrete Fourier transform (DFT) matrix; 2) the elements of $\boldsymbol{a}_i$'s are i.i.d. generated from $\mathcal{CN}(0, 1)$. The number of measurements is set as $m = 256$. The simulation data are generated by the signal model (7) with noise, i.e.,

$$y_i = |\boldsymbol{a}_i^T \boldsymbol{u} + v_i|,$$

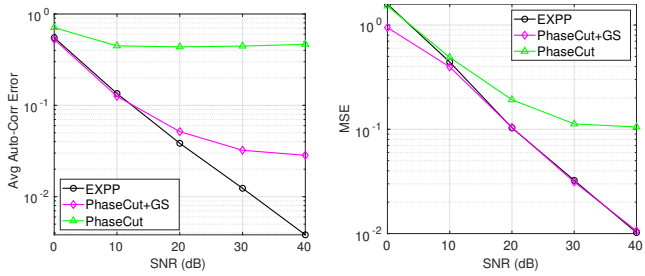
where $v_i \sim \mathcal{CN}(0, \sigma^2)$. The SNR is defined as $\mathbb{E}[\|\boldsymbol{A}\boldsymbol{u}\|_2^2]/\mathbb{E}[\|\boldsymbol{v}\|_2^2]$. Under the Gaussian setting of $\boldsymbol{a}_i$'s, we adopt mean square error (MSE)

$$\min_{\phi \in \mathbb{C}, |\phi|=1} \|\phi \hat{\boldsymbol{u}} - \boldsymbol{u}\|_2^2 / \|\boldsymbol{u}\|_2^2$$

as the performance metric, where $\hat{\boldsymbol{u}}$ is the estimated result obtained by the phase retrieval algorithm, and ϕ accounts for the global phase ambiguity. Under the DFT setting of $\boldsymbol{a}_i$'s, there exist ambiguities on the recovered signal [30], and so we adopt the autocorrelation function as the performance metric [31].

The benchmark algorithm is PhaseCut [4], which tackles the phase retrieval problem (9) by SDR. We also take the suggestion in [4] to improve the performance of PhaseCut by a second stage refinement, which performs 100 step Gerchberg-Saxton (GS) algorithm [32]; we name it PhaseCut+GS. We ran 100 Monte-Carlo trials.

We test the performance of the considered algorithms under different SNR levels. The results are shown in Fig. 4. First, it is seen that EXPP performs better than PhaseCut; second, the performance gap between EXPP and PhaseCut can be narrowed by the GS refinement on PhaseCut solution; third, EXPP is computationally attractive, as indicated by the runtime performance in Table 3.



(a) DFT $\boldsymbol{a}_i$'s (b) Gaussian $\boldsymbol{a}_i$'s

Fig. 4: Error performance of different phase retrieval algorithms.

Table 3: Average runtime performance (in sec.), SNR = 40 dB.

Alg.	EXPP	PhaseCut	PhaseCut+GS
DFT $\boldsymbol{a}_i$'s	0.4598	0.7644	0.7702
Gaussian $\boldsymbol{a}_i$'s	0.5505	2.4484	2.4532

4. CONCLUSION

In this paper, we revisited the penalty method for unit-modulus constraint in [3, 15], and described a generalization of its concept as an act of extreme-point pursuit. We applied the method to two previously untested applications, and the results are promising.

5. REFERENCES

[1] 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.

[2] 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.

[3] M. Shao, Q. Li, W.-K. Ma, and A. M.-C. So, "A framework for one-bit and constant-envelope precoding over multiuser massive MISO channels," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5309–5324, 2019.

[4] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Math. Program.*, vol. 149, no. 1, pp. 47–81, 2015.

[5] N. Boumal, "Nonconvex phase synchronization," *SIAM J. Optim.*, vol. 26, no. 4, pp. 2355–2377, 2016.

[6] S. Verdu, *Multiuser Detection*. Cambridge University Press, 1998.

[7] Microwaves101, "Phase Shifters," https://www.microwaves101.com/encyclopedias/phase-shifters.

[8] M. E. Newman, "Modularity and community structure in networks," *Proc. National Academy Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006.

[9] I. Rodriguez-Lujan, C. Elkan, C. Santa Cruz, and R. Huerta, "Quadratic programming feature selection," *J. Mach. Learn. Research*, 2010.

[10] A. M.-C. So, J. Zhang, and Y. Ye, "On approximating complex quadratic optimization problems via semidefinite programming relaxations," *Math. Program.*, vol. 110, no. 1, pp. 93–110, 2007.

[11] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[12] G. Yuan and B. Ghanem, "Binary optimization via mathematical programming with equilibrium constraints," *arXiv preprint arXiv:1608.04425*, 2016.

[13] B. Wu and B. Ghanem, "ℓ_p-box ADMM: A versatile framework for integer programming," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 41, no. 7, pp. 1695–1708, 2018.

[14] H. Liu, M.-C. Yue, A. M.-C. So, and W.-K. Ma, "A discrete first-order method for large-scale MIMO detection with provable guarantees," in *Proc. Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*. IEEE, 2017.

[15] M. Shao and W.-K. Ma, "Binary MIMO detection via homotopy optimization and its deep adaptation," *IEEE Trans. Signal Process.*, vol. 69, pp. 781–796, 2021.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[17] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003, vol. 1.

[18] Y. Liu, M. Shao, and W.-K. Ma, "A homotopy optimization method for orthogonal non-negative matrix factorization," in *Proc. European Signal Process. Conf. (EUSIPCO)*, 2021, pp. 1085–1089.

[19] Z. Wu, "The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation," *SIAM J. Opt.*, vol. 6, no. 3, pp. 748–768, 1996.

[20] D. M. Dunlavy and D. P. O'Leary, "Homotopy optimization methods for global optimization," *Report SAND2005-7495, Sandia National Laboratories*, 2005.

[21] E. Hazan, K. Y. Levy, and S. Shalev-Shwartz, "On graduated optimization for stochastic non-convex problems," in *Prof. Int. Conf. Machine Learn.* PMLR, 2016, pp. 1833–1841.

[22] M. Simões, A. Themelis, and P. Patrinos, "Lasry-Lions envelopes and nonconvex optimization: A homotopy approach," *arXiv preprint arXiv:2103.08533*, 2021.

[23] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys & Tut.*, vol. 17, no. 4, pp. 1941–1988, 2015.

[24] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.

[25] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimal data detection in large MIMO," *arXiv preprint arXiv:1811.01917*, 2018.

[26] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5339–5351, 2019.

[27] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, 2016.

[28] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[29] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[30] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, 2015.

[31] T. Qiu, P. Babu, and D. P. Palomar, "PRIME: Phase retrieval via majorization-minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 19, p. 5174–5186, Oct 2016.

[32] R. W. Gerchberg and O. W. Saxton, "A practical algorithm for the determination ofthe phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 237, 1972.